

Research Article

STRUCTURAL CHARACTERIZATION OF THE RNA BINDING DOMAIN OF HUMAN STEM LOOP BINDING PROTEIN

Maruthi Kashyap and Neel Sarovar Bhavesh*

Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology
Aruna Asaf Ali Marg, New Delhi 110067, India

Abstract: A gene encoding the RNA binding domain (RBD) of human stem loop binding protein (SLBP) was cloned in pET 28a vector and over-expressed in *E. coli* codon plus cells. The over-expressed SLBP-RBD carried no tag and aggregated as inclusion bodies in the cell lysate. Inclusion bodies were semi-purified to >85% purity by establishing a method involving detergent washing and subsequently denatured in 8 M urea. Refolding of the denatured RBD was carried out by step dialysis in decreasing concentrations of urea and L-arginine. Refolded SLBP-RBD was analyzed using size exclusion chromatography that revealed its monomeric nature and folded state. Uniformly ^{15}N and $^{15}\text{N},^{13}\text{C}$ labeled SLBP-RBD was prepared at concentrations for solution NMR studies. Approximately, 60% of the sequence specific backbone resonance assignments have been achieved through standard triple resonance NMR experiments. Analyses of secondary chemical shifts reveal presence of a small helical secondary structural elements and large intrinsically disordered regions.

Keywords: Human SLBP refolding; Refolded RBD (r-RBD); Solution NMR; Backbone assignment

Introduction

Eukaryotic mRNAs are characterized by the presence of a poly-A tail at their 3' ends which are generated from their respective pre-mRNAs in a two-step mechanism involving the polyadenylation machinery and the presence of two conserved sequence elements in the pre-mRNAs namely a hexanucleotide sequence element AAUAAA and a downstream GU rich sequence element. The first step is characterized by a cleavage of the pre-mRNAs between these sequence elements followed by an addition of a poly-A tail at the 3' end which constitutes the second step (Wahle and Ruegsegger, 1999). On the contrary, mRNAs that code for the histone proteins in eukaryotes are not polyadenylated

instead they end in a highly conserved 26-nucleotide stem loop at the 3' end (Marzluff, 2005). This stem loop is formed at the 3' end as a result of a cleavage between the stem loop and a downstream purine rich element, called the histone downstream element (HDE), in the pre-mRNAs (Dominski and Marzluff, 1999). HDE interacts with U7 snRNP (Mowry and Steitz, 1987) while the stem loop binding protein (SLBP) interacts with the 26 nucleotide hairpin and is necessary for efficient 3' processing of the histone mRNA (Wang *et al.*, 1996; Dominski *et al.*, 1999). It has also been reported that SLBP associates with the stem loop RNA even after processing the histone pre-mRNA into mRNA and plays an important role in its stability, export and translation (Sanchez and Marzluff, 2002; Sullivan *et al.*, 2009). A novel RNA binding domain in SLBP mediates interaction between SLBP and the stem loop (Martin *et al.*, 1997) and there are no other proteins with a similar sequence to the SLBP RBD in the human, *Drosophila* or *Caenorhabditis* genomes.

Corresponding Author: Neel Sarovar Bhavesh
E-mail: neelsb@icgeb.res.in

Received: November 9, 2013

Accepted: December 5, 2013

Published: December 14, 2013

Possibly it is a unique RNA binding protein that evolved as a part of regulatory mechanism for coordinating and precisely regulating the histone mRNA levels during cell cycle.

Architecturally, SLBP contains a nuclear localization signal sequences (NLS-1 and NLS-2) and an RNA binding domain of ~85 residues (M121-E202). The N-terminal region of *Drosophila* SLBP is intrinsically disordered (Thapar *et al.*, 2004) and previous studies have shown the importance of TPNK sequence within the RBD, where the phosphorylation of threonine increases the affinity for the stem loop RNA (Borchers *et al.*, 2006). Sequence alignment of four related metazoan RBDs reveals the highly conserved residues (Figure 1). In this study we have successfully established a refolding method for the RBD of human SLBP and solution NMR studies reveal the presence of helical secondary structural elements and disordered regions.

Materials and Methods

Cloning of RBD in expression vector - The following primers were designed for sub cloning of the RBD (M121-E202) of SLBP in pET 28a expression vector with Nco1 and Xho1 restriction sites.

5' CGCGCTCCATGG GATCTACTGTGCCGGC 3' forward and

5' GCGCGCCTCGAGTTATTCCGCTGGAGGATC 3' reverse primers.

RBD gene was PCR amplified and gel purified using a gel-extraction kit (Qiagen, Germany). Both PCR product and pET28a expression vector (Novagen) were sequentially digested overnight with Nco1 and Xho1 and gel purified. The purified PCR product was ligated with the double digested and purified pET28a at 6:1 molar ratio and was incubated overnight at 16°C. The ligation

mixture was transformed in *E. coli* Top 10 cells and positive colonies were screened by colony PCR. The identity of the clone was confirmed by sequencing (Macrogen, South Korea). The introduction of a stop codon in the reverse primer results in the absence of a C-terminal hexahistidine tag in the over expressed protein.

Over-expression and preparation of inclusion bodies - The recombinant plasmid was transformed into *Escherichia coli* strain BL21 (DE3) Codon plus for protein expression. A single colony was picked up and inoculated in Luria-Bertani (LB) medium containing 50 µg ml⁻¹ kanamycin and incubated in a shaker incubator at 37°C, 175 rev min⁻¹. At sufficient cell density, 1% of the primary inoculum was inoculated in 1 l of LB medium and grown at 37°C, 175 rev min⁻¹. At OD₆₀₀ ~1.0, the cells were induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 37°C for 6 h, 175 rev min⁻¹. The cells were harvested and re-suspended in sonication buffer (20 mM Tris, pH 8.0, 1M NaCl) and lysed by sonication. The lysate was centrifuged at 4300g for 60 min and the pellet constituting the inclusion bodies was thoroughly washed with Triton buffer (2% Triton-X 100, 20 mM Tris, pH 8.0) and incubated at 37°C, 175 rev min⁻¹ in a shaker incubator for 1h. The inclusion bodies were recovered by centrifugation at 4300g for 30 min followed by extensive washing of the pellet with sonication buffer to remove any traces of the detergent. The inclusion bodies were again recovered by centrifugation and denatured in 8M urea followed by incubation for 4 h in a shaker incubator at 37°C, 175 rev min⁻¹. In the final step, the denatured RBD was centrifuged at 25000g for 45 min to remove any traces of solid impurity. The supernatant which contains the denatured RBD is >80% pure as analyzed by SDS-PAGE.



Figure 1: Multiple sequence analysis of human, *Xenopus*, *Drosophila* and *C. elegans* RBDs. Fully conserved residues are indicated by an asterisk including the crucial TPNK sequence. The alignment was generated using ClustalX (Larkin *et al.*, 2007)

Refolding of RBD - The solution of denatured protein (in 8 M Urea, 20 mM Tris, pH 8.0) was diluted to 0.5 mg.ml⁻¹ and dialyzed against 2L of dialysis buffers containing decremented concentrations of urea (4 M, 2 M, 1 M, 0.5 M and 0 M) in 20 mM phosphate buffer, pH 6.0 using a 3 kDa pre-treated dialysis membrane. Since the protein had a propensity to aggregate at <1M urea concentration, 500 mM L-arginine was introduced in the dialysis buffer in the penultimate step. In the last step of dialysis (0 M step) the concentration of L-arginine was reduced to 200 mM. Every refolding step was carried out at 4 °C for 6 h with constant gentle stirring. The final solution contained refolded SLBP-RBD in 200 mM L-arginine, 20 mM sodium phosphate buffer, pH 6.0.

Size exclusion chromatography - Refolded SLBP-RBD was passed through a HiLoad 16/60 Superdex-75 gel permeation column (GE Healthcare) pre-equilibrated with 200 mM L-arginine, 20 mM phosphate buffer, pH 6.0. The area under the symmetrical curve in the chromatogram corresponding to the purified refolded SLBP-RBD fraction was collected and concentrated to 0.4 mM using a 3 kDa molecular-weight cutoff centrifugal filter (Millipore).

Preparation of labeled proteins - Uniform isotope enrichment by ¹³C and ¹⁵N of protein is essential for multi-dimensional NMR spectroscopy. For this, uniformly isotopically labeled proteins (¹⁵N and ¹⁵N,¹³C) were prepared based on the protocol standardized (as discussed previously) in minimal media using ¹⁵NH₄Cl and ¹³C Glucose (Isotec) as the nitrogen and carbon source respectively.

Solution NMR spectroscopy - The final NMR sample contained 5% (v/v) D₂O, 0.4 mM labeled SLBP-RBD, 200 mM L-arginine and 20 mM sodium phosphate buffer, pH 6.0. Sequence-specific resonance backbone assignment was performed using standard triple resonance experiments namely HNCA, HNCO, HN(CA)CO, HNCACB and CBCA(CO)NH spectra (Kay *et al.*, 2011). All NMR spectra were recorded at 298 K on a Bruker Avance III spectrometers equipped with cryogenic triple resonance probes, operating at field strengths of 500 and 700 MHz. All of the spectral data was processed with Topspin 2.1 (Bruker AG) and analyzed with Cara (Keller, 2004).

Results and Discussion

RBD expression and refolding

The RBD of SLBP was over-expressed in *E. coli* as inclusion bodies (Figure 2a). A construct expressing a maltose binding protein (MBP) tag along with the RBD was successful in solubilizing the recombinant fusion protein in the cell lysate. However, purification of the fusion protein followed by TEV protease treatment (to remove MBP) resulted in immediate precipitation of RBD (Figure 2b). This indicates that MBP acts as a solubility enhancer and RBD as an independent entity fails to fold properly in the heterologous host and may require modifications post translation and/or RNA binding to adopt a native conformation. Taking this into consideration the RBD was cloned in pET28a vector and over-expressed at 37°C as inclusion bodies without any tag. This strategy has been successful in maximizing the yield of inclusion bodies of SLBP-RBD.

To enhance the purity of inclusion bodies a protocol was standardized as elaborated in materials and methods. This protocol was

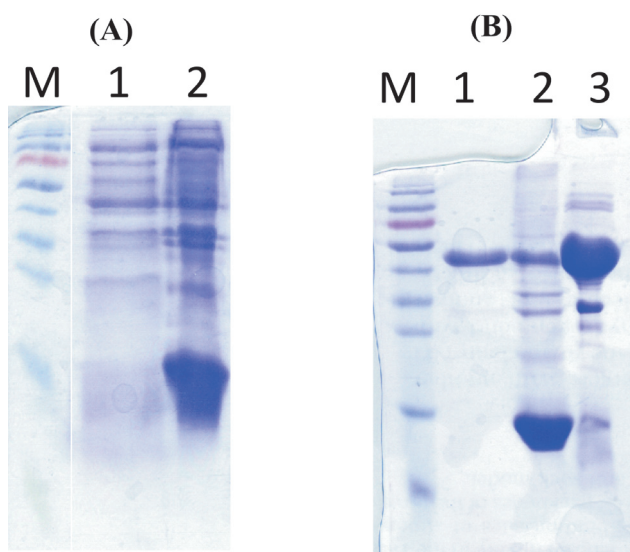


Figure 2: 15% SDS-PAGE migration profile showing the (A) expression analysis of SLBP-RBD in pET 28a vector. Lane 1 corresponds to the supernatant portion after cell lysis and lane 2 corresponds to the pellet fraction, which contains inclusion bodies of SLBP-RBD. (B) Purified MBP-RBD fusion protein following TEV cleavage. Lane 1 corresponds to MBP control (~45 kDa). Lane 2 corresponds to the pellet fraction following TEV cleavage, which resulted in precipitation of SLBP-RBD. Lane 3 corresponds to the supernatant fraction following TEV cleavage, which corresponds to the MBP.

efficient in the preparation of relatively pure inclusion bodies, which were later denatured with 8 M urea solution. Since the denatured RBD was >85% pure, no further purification was required and refolding attempts were carried out by fast dilution and step dialysis methods. Fast dilution method resulted formation of soluble aggregates of the protein and hence refolding was carried out by the conventional step dialysis method at decreasing concentrations of urea (Table 1). Step dialysis method was successful and the introduction of 0.5 M and 0.2 M of L-Arginine in the last two steps (at 1 M and 0.5 M urea concentration) was crucial for the stabilization of the protein. In the absence of L-arginine the protein readily precipitated at urea concentrations of 1M and below. The overall results of refolding are indicated in Table 1.

Table 1
Refolding of RBD by step dialysis

Concentration of urea used for denaturation	8M
Concentration gradient of urea in steps	8M-4M-2M-1M-0.5M-0M
Concentration of L-arginine	At 1M to 0.5M dialysis step- 0.5M At 0.5M to 0M dialysis step- 0.2M
Starting yield/liter of culture	20mg
Yield after refolding/liter of culture	8mg

Refolded SLBP-RBD was concentrated and analyzed on a size exclusion column. Refolded SLBP-RBD eluted at 80 ml V_e in the gel permeation column which corresponds to a molecular weight of ~ 10 kDa indicating its monomeric nature (Figure 3).

SLBP-RBD is a flexible protein

Prior to proceeding with the triple resonance and other NMR experiments the temperature was optimized and a series of 2D [^{15}N , ^1H] HSQC spectra was measured at 293K, 298K, 303K and 308K. Spectra at 298K and above had better line widths relative to the spectra recorded at 293K and 298K was chosen as the optimal temperature keeping in view of the stability of refolded SLBP-RBD. All experiments were measured at a protein concentration of 0.4 mM as SLBP-RBD had a propensity to precipitate at a higher concentration. The 2D [^{15}N , ^1H] HSQC spectrum measured after standardizing the initial conditions revealed less dispersed and crowded backbone resonances indicating the presence of helical and disordered regions in the SLBP-RBD. A total of 50 backbone NH resonances against 74 resonances (68% of peaks observed) and all the amino side chain resonances were observed. The absence of 24 backbone resonances can be attributed to conformational exchange happening at micro-to-milli second time scales that causes line broadening of resonances.

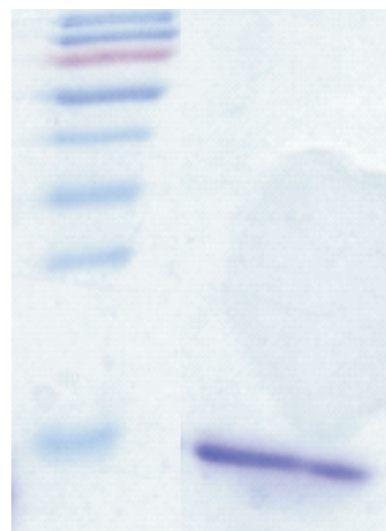
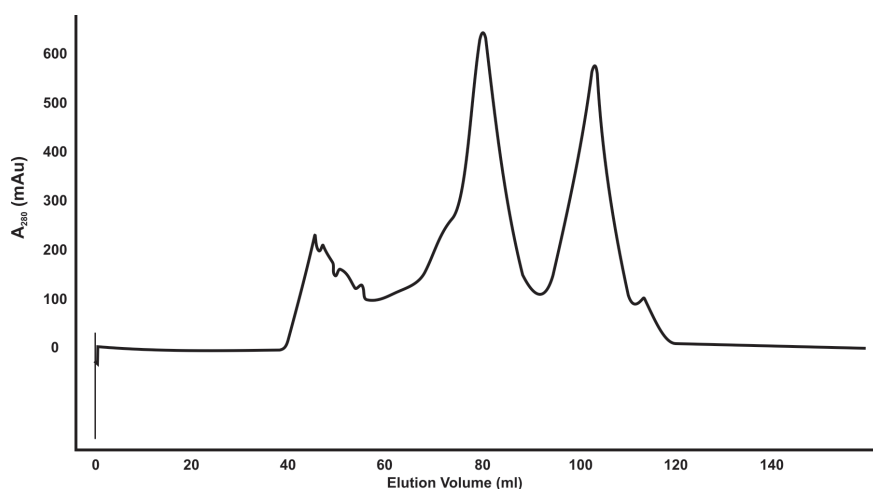


Figure 3: S-75 Gel permeation column chromatography elution profile of refolded SLBP-RBD. Area under the major peak corresponding to V_e 80ml shows the elution of refolded SLBP-RBD. The protein was analyzed on a 15% SDS-PAGE.

Out of the 50 observable backbone resonances only 40 of them could be unambiguously assigned using a series of standard triple resonance experiments. Partially assigned 2D [^{15}N , ^1H] HSQC spectrum is shown in Figure 4. Analysis of C^α and C^β resonances of these 40 assigned residues based on chemical shift differences with respect to random coils reveal the presence of a helical region comprising D131-R138 and the rest being disordered (Figure 5). The absence of backbone resonances due to intermediate conformational chemical exchange or exchange with water due to the inherent disordered property of the molecule and the crowding of peaks due to chemical shift degeneracies has been a major obstacle to obtain complete backbone resonance assignments.

From the NMR data it is clear that except for a small helical stretch SLBP-RBD protein is natively flexible and undergoing slow conformational averaging. The flexible nature of SLBP-RBD is similar to the *Drosophila* SLBP (dSLBP). NMR studies of RNA processing domain (RPD) of *Drosophila* SLBP have shown it to be intrinsically unstructured in absence of bound RNA. Phosphorylation of serines and binding of its cognate RNA is necessary for proper folding dSLBP-RPD protein (Thapar *et al.*, 2004a,b). SLBP-RBD also might be acquiring a folded structure only upon binding to stem loop RNA. A further optimization of conditions to obtain RNA bound SLBP-RBD is required to achieve complete structural characterization of the protein and explain the structure-function relationship.

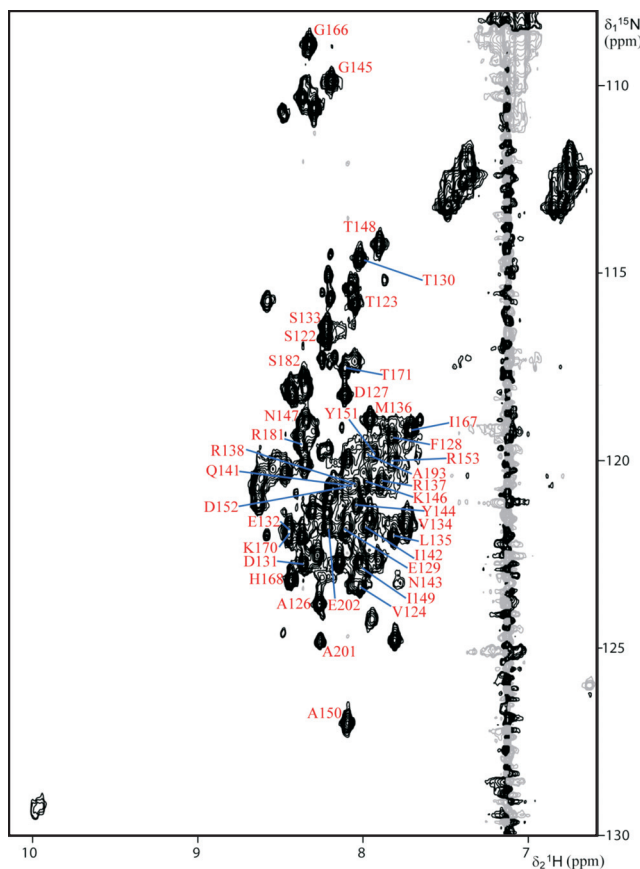


Figure 4: 2D [^{15}N , ^1H]-HSQC spectrum of SLBP-RBD recorded at 298 K at pH 6.5 on a Bruker Avance III 500 MHz spectrometer equipped with a cryogenic triple-resonance 5 mm TCI probe-head. The spectrum was acquired on uniformly ^{13}C , ^{15}N -labeled protein (ca. 0.4 mM) with 8 scans per t_1 increment in ^{15}N dimension. Acquisition times for ^{15}N and ^1H dimension were 98.6 ms ($t_{1\text{max}}$) and 170.4 ms ($t_{2\text{max}}$) respectively. Spectral-widths for ^{15}N and ^1H dimensions were 1621.8 and 6009.6 Hz respectively. The -NH resonances at 7.1 ppm corresponds to L-Arginine, which is present at high concentration in the final NMR sample.

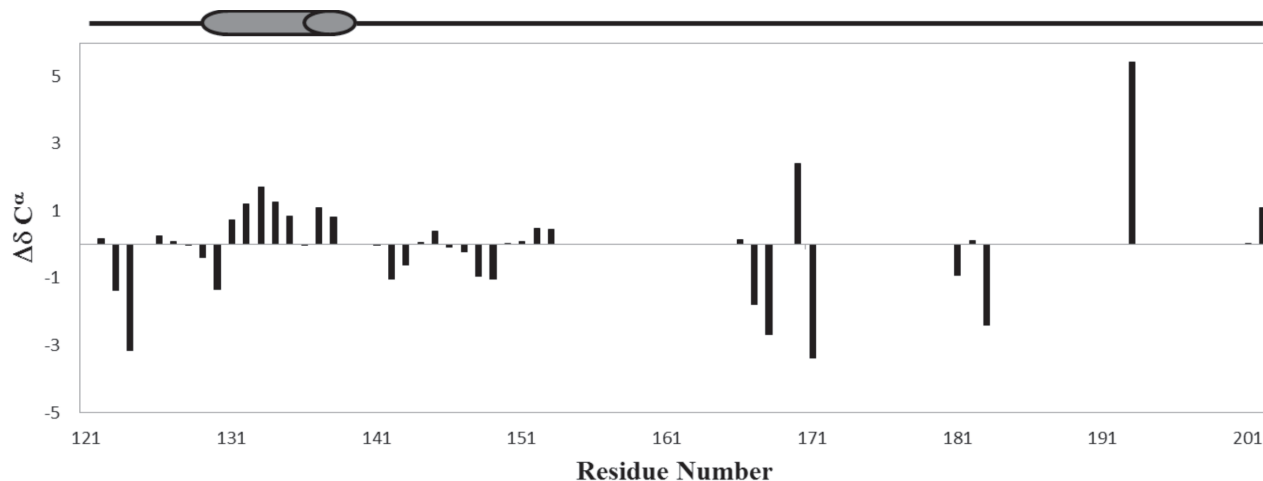


Figure 5: Chemical shift deviations of C^α atoms for SLBP-RBD. Secondary structure elements are denoted on the top, helical segment is denoted by a cylinder and the disordered region by a line segment.

Acknowledgements

The authors wish to thank DBT for providing financial support for the High Field NMR spectrometers at the ICGB, New Delhi and NII, New Delhi. MK is a recipient of Indian Council of Medical Research-Senior Research Fellowship (ICMR-SRF).

References

- Borchers, C. H., Thapar, R., Petrotchenko, E. V., Torres, M. P., Speir, J. P., Easterling, M., Dominski, Z. and Marzluff, W. F. (2006). Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity RNA binding. *Proc. Natl. Acad. Sci. USA* **103**, 3094-3099.
- Dominski, Z. and Marzluff, W. F. (1999). Formation of the 3' end of histone mRNA. *Gene* **239**, 1-14.
- Dominski, Z., Zheng, L. X., Sanchez, R. and Marzluff, W. F. (1999). Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol. Cell. Biol.* **19**, 3561-3570.
- Kay, L. E., Ikura, M., Tschudin, R. and Bax, A. (2011). Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J Magn Reson* **213**, 423-441.
- Keller, R. (2004). The computer aided resonance assignment tutorial. Cantina, Goldau.
- Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. and Lopez, R. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- Martin, F., Schaller, A., Eglite, S., Schumperli, D. and Muller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J.* **16**, 769-778.
- Marzluff, W. F. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr. Opin. Cell Biol.* **17**, 274-280.
- Mowry, K. L. and Steitz, J. A. (1987). Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's. *Science* **238**, 1682-1687.
- Sanchez, R. and Marzluff, W. F. (2002). The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Mol. Cell. Biol.* **22**, 7093-7104.
- Sullivan, K. D., Mullen, T. E., Marzluff, W. F. and Wagner, E. J. (2009). Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA* **15**, 459-472.
- Thapar, R., Mueller, G. A. and Marzluff, W. F. (2004a). The N-terminal domain of the *Drosophila* histone mRNA binding protein, SLBP, is intrinsically disordered with nascent helical structure. *Biochemistry* **43**, 9390-9400.
- Thapar, R., Mueller, G. A. and Marzluff, W. F. (2004b). Electrostatic Contribution of Serine Phosphorylation to the *Drosophila* SLBP-Histone mRNA Complex. *Biochemistry* **43**, 9401-9412.
- Wahle, E. and Ruegsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* **23**, 277-295.
- Wang, Z. F., Whitfield, M. L., Ingledue, T. C., 3rd, Dominski, Z. and Marzluff, W. F. (1996). The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev.* **10**, 3028-3040.