Published in final edited form as:

J Mol Biol. 2010 August 20; 401(3): 334–349. doi:10.1016/j.jmb.2010.06.032.

Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1

Thomas C. Leeper^{1,4}, Xiangping Qu², Connie Lu³, Claire Moore², and Gabriele Varani^{1,3}

¹University of Washington, Department of Chemistry, Seattle, WA 98195-1700

²Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, USA

³University of Washington, Department of Biochemistry, Seattle, WA 98195-1700

Abstract

Precise 3'-end processing of mRNA is essential for correct gene expression, yet in yeast, 3'-processing signals consist of multiple ambiguous sequence elements. Two neighboring elements upstream of the cleavage site are particularly important for the accuracy (positioning element) and efficiency (efficiency element) of 3'-processing and are recognized by the RNA-binding proteins Rna15 and Hrp1, respectively. *In vivo*, these interactions are strengthened by the scaffolding protein Rna14 that stabilizes their association. The NMR structure of the 34 kDa ternary complex of the RRM domains of Hrp1 and Rna15 bound to this pair of RNA elements was determined using Residual Dipolar Coupling and Paramagnetic relaxation experiments. It reveals how each of the proteins binds to RNA, and introduces a novel class of protein-protein contact in regions of previously unknown function. These interdomain contacts had previously been overlooked in other multi-RRM structures, although a careful analysis suggests that they may be frequently present. Mutations in the regions of these contacts disrupt 3'-end processing, suggesting that they may structurally organize the ribonucleoprotein (RNP) complexes responsible for RNA processing.

Keywords

Rna15; Hrp1; mRNA 3'-end processing; protein-RNA; NMR

Introduction

Correct 3'-end processing is essential for mRNA biogenesis and transcription termination. Execution of the cleavage and polyadenylation reactions preserves RNAs from degradation,¹ promotes export,² enhances translation,³ ensures correct termination of transcription⁴ and promotes transcription activation^{5,6}. Because the poly(A) polymerase enzyme reacts with any free 3'-hydroxyl, recognition of RNA processing sequences by the cleavage and

⁴Present address: University of Akron, Department of Chemistry, Akron, OH 44325-3601

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

polyadenylation complex is critical in all eukaryotes for the identification of both canonical and alternative 3'-end processing sites. This step of mRNA synthesis is of even greater significance in budding yeast, compared to metazoans. In this organism, efficient transcription termination is more critical because intergenic regions are shorter, and efficient mRNA export is more dependent on polyadenylation instead of splicing because so few genes contain introns⁷⁸.

Given the importance of polyadenylation, it is surprising that individual RNA sequence elements required for 3'-end processing are relatively poorly defined in yeast,^{9,10} although similarities to the metazoan polyadenylation element AAUAAA and to the pyrimidine-adenine (PyA) nucleotides at the cleavage site are clearly identifiable (Figure 1A). The efficiency element (EE), typically UAUUAU,^{9,10} stimulates the efficiency of polyadenylation and is within 50 nucleotides upstream of the yeast cleavage site. The A-rich positioning element (PE), usually AAUAAN,^{9,10} is just downstream of the EE, about 10 to 30 nucleotides upstream of the cleavage site. The PE is essential for the precise execution of the cleavage reaction. Flanking the cleavage site are two degenerate U-rich sequences.^{9,10,11} The combination of these four sequences lead to accurate and efficient cleavage via recruitment of the cleavage and polyadenylation machinery.¹²

In all eukaryotes, recognition of the 3' end processing signals generally occurs by rather weak interactions of factors with these elements, which nevertheless leads to a ribonucleoprotein competent for processing.¹² Cross-factor interactions must stabilize this complex on the collection of signals comprising an authentic poly(A) site and thus prevent premature processing in regions which are lacking one or more of the signals. However, the nature of such interactions is poorly understood. In yeast, the positioning element sequence is recognized by the Rna15 subunit of Cleavage Factor Ia (CFIa), through its N-terminal RNA recognition motif (RRM, Figure 1). Mutations in the putative RNA-binding face of Rna15, or in the A-rich positioning sequence, lead to imprecise cleavage and functional defects.^{13,14,15,16} The enhancer element is recognized by Hrp1 through its two RRM (Figure 1).¹⁷ The interaction of Rna15 and Hrp1 with the scaffolding protein Rna14 increases the efficiency of with which these sequences are recognized, yet RNA association is solely mediated by the RNA-binding domains of Hrp1 and Rna15.^{13,18} By acting in concert, these interactions may anchor the cleavage and polyadenylation machinery relative to the cleavage site. Accordingly, we will refer to the combined efficiency and positioning elements sequence as the "anchor RNA" sequence throughout this report.

In order to further understand how 3'-end processing sites are identified and a stable processing complex assembled, we determined the structure of the ternary "anchoring" complex formed by the RRM of Hrp1 and Rna15 on the anchor RNA sequence from the *GAL7* mRNA, whose 3' end processing has been well characterized *in vivo* and *in vitro*. We found robust and simultaneous protein-RNA recognition by these two proteins under NMR conditions, even though Rna14 is not part of the complex. Employing paramagnetic relaxation enhancements (PREs) and residual dipolar couplings (RDCs) defined the overall envelope of structures to allow us to observe confidently new protein-protein interactions, and permitted calculation of the structure of this large complex in spite of the absence of assignable protein-protein intermolecular NOEs. In the structure of this ternary RNP, we

observe canonical RRM-RNA 5 interactions, as well as novel RRM/RRM interactions. Mutations of these protein-protein contacts perturb cleavage and polyadenylation, suggesting that Hrp1 and Rna15 may be structurally organized by these interactions to promote accurate recognition of the 3'-end processing signals. Although the interaction interface is small, the conserved nature of similarly-positioned amino acid side chains indicates that this may be a common feature of other RRMs.

Results

The interaction of Hrp1 with the enhancer element has been studied by NMR,¹⁷ as well as that of CstF 64, the vertebrate counterpart of Rna15, with its cognate GU-rich RNA target.^{19,20} However, to fully understand how RNA recognition leads to 3'-end processing and transcription termination, structural information is needed for the functional multi-protein complex. So far, it has been impossible to crystallize a complex of Rna14 and Rna15, the proteins that act in concert with Hrp1 to recognize the yeast anchor RNA region. Thus, we focused first our investigation on the ternary complex formed when the RRMs of Hrp1 and Rna15 bind to the anchoring RNA.

Formation of the ternary complex

We sought to identify sequences that gave good NMR spectra of the Rna15 RRM domain (referred to just as Rna15 henceforth) bound to RNA, indicative of well-defined complex. However, for most of the RNAs tested, we observed exchange-broadened spectra suggestive of promiscuous binding. We attribute this observation to the 'slippage' of this poorly specific protein onto the RNA.^{21; 22; 23} Only two RNAs, the positioning element (PE) and the complete anchor RNA, gave good NMR spectra for the Rna15- RNA complex. Although both RNAs bind to Rna15 in the fast exchange regime (Figure 2A), indicative of relatively weak binding (K^d much weaker than μM), these interactions are well defined: mapping of the chemical shift perturbations resulted in a clear patch showing RNA recognition by the RRM β -sheet surface (Figure 2C). To the pre-formed anchor RNA/Rna15 complex, we added a protein construct consisting of the two RRMs of Hrp1 and observed binding of Hrp1 to the 5'-portion of the anchoring RNA, even in the presence of Rna15, as determined by both 1D (Supplementary Figure 1B) and 2D NMR (Figure 2B and 2D). Using these proteins and RNA constructs, we were able to determine the structure of the ternary complex of Hrp1 and Rna15 simultaneously bound to the anchoring RNA signal.

NMR data collection and structure determination

Different preparations of proteins and RNAs (Supplementary Table 1) were combined to generate NMR samples of these complexes (Supplementary Table 2). The most difficult task was to establish the relative orientation of the three protein domains because protein-protein interfaces were small (as judged from chemical shift perturbations). While a satisfactory number of intermolecular RNA to protein NOEs were observed to characterize the Rna15 and Hrp1 interfaces with RNA (Figure 3A and Table 1), no unambiguously assignable inter-protein NOEs were found between Rna15 and Hrp1. In retrospect, this is not unexpected since the molecular interface is composed of long amino acid side chains and the interactions primarily electrostatic. However, the correlation time of the complex was ~ 13

ns, as monitored from the T_1/T_2 relaxation time ratios for both Rna15 and Hrp1, suggesting that the three RRM domains tumble together at a fixed orientation and do not form a ‘pearl-on-a-string’ complex.

To circumvent the lack of NOEs between Rna15 and Hrp1, we used residual dipolar couplings (RDCs) to determine relative domain orientations and long-range paramagnetic relaxation enhancements (PREs) to provide the translational restriction lacking in complexes refined only using RDCs.^{24,25} By labeling the anchor RNAs in these complexes with nitroxide spin-labels (Figure 3B), long-range distance constraints (e.g. 12 to 30 Å) were measured between the RNA and the protein domains. Comparison of PRE broadening between ^{15}N Rna15 with spin labeled anchoring RNA (black, Figure 3B) and a ternary complex of ^{15}N Rna15, unlabeled Hrp1, and spin labeled anchoring RNA (blue, Figure 3B) shows that in the absence of the Hrp1 protein, the paramagnetic center uniformly broadens the Rna15 signals, suggesting an absence of defined position relative to its non-cognate sequence. This is consistent with initial observations that Rna15 slides along the RNA and suggests that Hrp1 is required to position Rna15 precisely and confine it to a unique spatial relationship with respect to the spin-label. This result agrees with relaxation measurements and with the biochemical properties of Rna15 in the complete 3'-end processing complex.¹³ The location of some PRE pairs are illustrated in Supplementary Figure 2.

Initial structure calculations in CYANA used 5,349 NOE restraints including 201 intermolecular restraints between proteins and RNA and 143 distances from PREs. NOE assignments of intramolecular protein NOEs utilized the CANDID module of *cyana*, while all intermolecular NOEs were manually assigned. The calculation was migrated into XPLOR-NIH for refinement against RDCs (Table 1). Alignment tensors were approximated from powder pattern distributions of the RDC values and refined by structure calculations using a grid search.²⁶ Structures of the individual RRM domains converged well and have topologies and binding modes consistent with other RRM domain structures.¹⁷ Ramachandran statistics are satisfactory and adherence to the NOE and dihedral data is similar to other large RNP structures.²⁵ The RMS deviation for the RDCs is similar to the experimental error (± 2.5 Hz).

Structure of the complex

The overall architecture of the complex resembles a “horseshoe” (Figure 4A and B) with the β -sheet surfaces of each RRM domain rotated $\sim 90^\circ$ with regard to each adjacent domain (Figure 4A). The exact orientation of these domains is slightly variable between structures, as shown by higher RMSDs for the full complex (Table 1 and Supplementary Figure 3). However, the juxtaposition of Rna15 loop 3 and Hrp1 loop 5 is well maintained and established by PRE contacts (Supplementary Figure 2). The two helical regions of each RRM domain are pointing away from one another, as would be expected considering their acidic nature (Figure 4A). Surface rendering indicates that contacts between the RRM domains of Rna15 and Hrp1 (Figure 4B) are limited to loop 3 of Hrp1 RRM 1 and loop 5 of Rna15 (Figure 4A, 4F, and Supplementary Figure 4). While these contacts account for only a small buried surface area (Table 1), a similar loop 3/loop 5 contact is seen within Hrp1

between RRM domains 2 and 1. As discussed below, mutations within this region have functional consequences on mRNA processing.

The structure of Hrp1 in this ternary complex is very close to the previously determined structure of the Hrp1 complex (Supplementary figure 5A and B).¹⁷ RRM1 recognizes the four 3'-terminal AUAU nucleotides of the enhancer element, and the most well conserved residues within the RNP1 and RNP2 motifs make direct contact with A6 and U7 (Figure 4C). As is characteristic of the Hrp1 architecture, the *syn*-oriented base of A4 is shared between the two RRM domains and stacks over W168 from RRM1. In addition, the carbonyls of Hrp1 Gly165 and Gly201 share a hydrogen bond donated by the imino of U5. Consistent with this contact, the NH signal from the UA-repeat sequence of the anchoring RNA shifts slightly and diminishes in intensity upon Hrp1 binding, but never fully disappears (Supplementary Figure 1B). Thus, the residual NH signal probably reflects protection of the U5 imino NH from solvent exchange via hydrogen bonding. An analogous Uracil to Gly/Gly contact is seen in RRM2 (U1 to Gly249/Gly285). The last nucleotide of the efficiency element at the boundary with the positioning element in the anchor RNA is Uracyl 7. The base of this residue stacks over M191 in Hrp1, as it did in the previously published Hrp1 binary complex,¹⁷ suggesting that the two RNA binding sites are somewhat independent of each other with no requirement for linker-nucleotides.

In the structure of Rna15 (Figure 4E), the RNP1 and RNP2 motifs interact with the RNA in a fairly “canonical” manner: three aromatic side-chains (Y21, Y61, F63, Figure 4E) interact specifically with two RNA bases (A9 and U10) and one of their ribose groups (A9). In addition, the methyl group of M50 stacks under the next base in the PE (A11), a contact frequently seen in other RRM-RNA interactions.¹⁷ However, a number of features make this structure unique. As required by the large number of intermolecular NOEs involving H2 atoms (Figure 3A), the bases of A8 and A9 stack with the ring of Y21 sandwiched between them (Figure 4E). The recognition surface of Rna15 is atypically extended well into β 4 (Figure 4E), due to contacts between K90 and the A8 phosphate. Although residues A12 and U13 are poorly defined in the structure, internucleotide A11/A12 NOEs suggest that A12 is partially stacked upon A11. Proximity of A12 and U13 to the amino group of K48 in some converged structures suggests an electrostatic interaction. Surprisingly, R58 does not form consistent interactions with RNA, suggesting that this residue from loop 3 is not used for RNA recognition, unlike comparable residues from other RRM domains that bind to RNA with high affinity and specificity, such as U1A.²⁷ Altogether, these direct contacts to four RNA bases with a potential indirect interaction via the A11/A12 stacking highlights that the “AAUAA” sequence is specifically recognized by Rna15, as suggested by sequence conservation¹⁶ and direct biochemical evidence.¹³ The propensity for burial of the hydrophobic ‘minor groove’ faces of adjacent adenines explains the favorable association with A-rich sequences. Lastly, the overall orientation of the RNA is noteworthy, with the Watson-Crick faces of all four of the directly recognized bases (A8, A9, U10, A11) pointing directly into or “at” the β -sheet surface.

Loop3-loop5 interactions

Examining the interdomain contacts between RRM1s, we noticed that loops 3 or one RRM domain contacts loop 5 of another, both between *rna15* and *hrp1* RRM1 (Figure 4A and 4F) and between RRM1s within *hrp1* (Figure 4A and 4G). This contact involves a salt bridge between a basic residue located in loop 5 and a strongly conserved aspartic acid in loop 3 of the juxtaposed domain (Figures 4F, 4G, and Supplementary Figure 4). In particular, with regard to *rna15*, R87 points directly at *hrp1* RRM1 D193 and may even form hydrogen bonds in addition to the salt bridge. Within *hrp1*, K226 of RRM1 bridges with D279 of RRM2 in a similar fashion.

Functional analysis of loop3-loop5 interactions

In order to test whether the interaction between the Hrp1 loop 3 and the loop 5 of Rna15 contribute to the function of these proteins, we generated yeast strains in which the wild type genes were replaced with either the D193R mutation in Hrp1 or the R87D mutation in Rna15. The mutant forms were stably expressed at levels comparable to the respective wild-type proteins (Supplementary Figure 6A). Both mutant strains showed wild type growth on rich media at 25°C and did not display cold or heat sensitivity (data not shown). However, the *rna15-R87D* strain grew more slowly than wild type in the presence of 1.5% formamide (Supplementary Figure 6B), a reagent known to enhance defects in protein/protein interactions in yeast mutants.²⁸

Extracts were then prepared from the wild type and mutant strains and tested for mRNA 3' end processing activity using *in vitro* cleavage and polyadenylation assays. The D193R mutation within Hrp1 loop 3 noticeably reduced the efficiency of the coupled *in vitro* cleavage/polyadenylation reaction in comparison to that of wild-type extracts, and the separate steps of the reaction were similarly affected (Figure 5A, B, and C). The R87D mutation in Rna15 loop 5 had an even more dramatic effect: extract containing this mutant is almost completely non-functional for the coupled reaction, or when 3'-end cleavage was assayed in the absence of polyadenylation (Figure 5A and B). The Rna15-R87D extract was also defective for processing a precursor containing the CYC1 poly(A) site (Figure 6A). However, it was as efficient as wild-type extract for the poly(A) addition step (Figure 5C), suggesting that the CPF factor, the Hrp1 protein, and the Pcf11, Clp1 and Rna14 subunits of CF IA, which are all needed for both cleavage and polyadenylation, are normal in the Rna15-R87D extract.

Based on the structural analysis, the Rna15-R87D mutation should weaken the interaction with Hrp1 and prevent effective anchor RNA recognition. In order to test this hypothesis, we immunoprecipitated the products of *in vitro* processing reactions with an Rna15 antibody. Examination of the RNAs in the input samples confirmed that processing is efficient in wild-type extract but defective in the mutant extract (Figure 6B, lanes marked "Input"). With wild-type extract, a significant amount of unprocessed RNA was precipitated with the anti-Rna15 antibody compared to aliquots treated with preimmune serum (Figure 6B, Pellet). Interestingly, the Rna15 antibody enriched the cleaved, unadenylated species. However, only a small amount of polyadenylated RNA was immunoprecipitated, suggesting that the processing complex had been released from most of the final product. With the *rna15-R87D*

extract, much less unprocessed RNA was associated with Rna15 (Figure 6B), confirming that interaction of the mutant protein with the polyadenylation signal sequence is less stable.

In order to fully establish the structural rationale for the processing defects, various mutants of the Rna15 and Hrp1 constructs were individually evaluated for protein stability and their interaction with RNA using NMR. The D193R mutation in Hrp1 unfolded RRM1, as determined by 2D NMR, thus providing a ready explanation for the defects observed. However, a variant D193N mutation resulted in a well-folded protein that was still capable of binding to the anchoring RNA element. The R87D mutation in loop 3 of Rna15 resulted instead in a protein incapable of binding to the anchoring RNA (Supplementary Figure 7A). These results provide a molecular basis for the strong *in vitro* defects in cleavage; as noted above, in this structure this residue does not make any direct contacts to the RNA but is located near the protein-protein interface. Based on this result, we generated a milder R87A mutant and screened it for binding *via* NMR to determine if this basic residue is truly necessary for RNA recognition. In contrast to R87D, the R87A mutation could still bind to RNA (Supplementary Figure 7B), although the chemical shift changes were somewhat smaller than for the wild type protein. Interestingly, wildtype Hrp1 protein could compete the anchoring RNA away from Rna15 R87A (Supplementary Figure 7B), but not from wild-type Rna15 (Figure 2A). These observations confirm that favorable protein-protein interactions permit simultaneous loading of Hrp1 and Rna15 onto the anchoring RNA, and that their mutation disrupts RNA recognition and function in pre-mRNA cleavage.

Discussion

We have determined the structure of the ternary complex of the RNA-binding domains of Hrp1 and Rna15 bound to the anchoring RNA sequence containing the efficiency and positioning elements that identify 3'-end processing sites in yeast. In this structure refined using RDC's and paramagnetic relaxation effects (PREs), we observe protein-protein contacts between Rna15 and Hrp1. Although the surface area of these contacts is small (Figure 4F and 4G), it involves conserved residues that may help position the two proteins relative to each other in a horseshoe structure (Figure 4A and 4B). This structure creates a contiguous binding surface for recognizing adjacent RNA signals and maximizes the distance between the strongly negative backside helices of each of the RRM domains (Figure 7). As discussed below, this ternary RNP structure helps explain the spatial juxtaposition of the yeast enhancer and positioning elements that has been revealed by sequence analysis^{9; 10} and shows how the poor RNA specificity of the essential Rna15 protein is functionally tolerated.

In the structure, the RNA is well encapsulated by the RRM domains, so that insertions between enhancer and positioning elements would require a slight parting of the two domains at the N-terminal end of Hrp1 β 2 and C-terminal end of Rna15 β 4. A gap could be accommodated and still maintain the protein/protein contacts, yet it would also result in close phosphate contacts that are unfavorable in the absence of stabilizing RNA structures. Thus, an optimal linker between the two elements is likely to be relatively short. Consistent with this reasoning, the two RNA sequences are frequently found next to one another, with generally only very short intervening sequences^{9; 10}. The accessory domains of Hrp1 and

Rna15 (Figure 1B) and their essential partner Rna14, could be accommodated onto the convex peripheral acidic surface of this structure (Figure 7B).

The RNA specificity of Rna15

The optimal binding site of Hrp1 matches the consensus that has been identified for the yeast enhancer element sequence. However, the specificity of Rna15 is still debated. Rna15 favors G/U-rich motifs that differ from the positioning element in selection experiments²⁰, but the binding affinity is only about 20 uM and selectivity is modest, only 4-5 fold²⁹. Consistent with the poor selectivity of this protein, when we tried to bind Rna15 to different sequences early in the project, we found that Rna15 is in fast exchange between alternate binding sites. The vertebrate orthologue of Rna15, CstF-64, seems to have a clearer specificity for the GU-rich elements that help identify 3'-end processing sites in vertebrates.^{20; 30} A recent crystal structure of Rna15 bound to a GU element supported the assertion that GU sequences are the preferred binding site for Rna15²⁹. However, it has been convincingly shown that Rna15 interacts with the A-rich positioning element during 3'-end processing.¹³ In the *GAL73*'-end processing signal, mutations in this element that reduce RNA binding efficiency for Rna15 also cause cleavage to no longer be localized to a specific site.³¹ Thus, the A-rich positioning element studied here is a functional recognition site of Rna15.¹³

Comparison of the Rna15-GU crystal structure with our results (Supplementary Figure 5c) reveals a very similar overall topology with some limited differences in loop 3.²⁹ With regards to GU-recognition, the crystal structure identified two binding sites for G or GU sequences²⁹: site-II, which is seen in only one crystal form, is the binding site for a single guanine to the canonical RNP1 and RNP2 motifs in the β -sheet surface. That interaction is typical of other RRMs, for example with the stacking of a guanine base over RNP2 residue Y21. Site-I is at the bottom of the domain and is formed at the junction between β 1 and α 1. This second site is unusual, although others have observed binding of RNA to the "bottom" of RRM domains³². Our solution chemical shift perturbation experiments using the anchor RNA sequence (Figure 2c) demonstrate that this site is not occupied, since we observe no change there. This is not unexpected, since only G or U can bind to this second site in the way illustrated in the structure. The functional significance of this site is unclear.

Comparison of our Rna15 structure with the previous model of CstF-64 RNA recognition suggests several possible reasons for the enhanced specificity of CstF-64.³⁰ A first significant difference is the presence of several arginine residues in the surface of the CstF-64 β -sheet, while Rna15 tends to have lysine residues in comparable positions (Supplementary Figure 3). For example, R46 of CstF-64 was proposed to hydrogen bond to O2 of the 5' uracil of the GU recognition element. The Lys in this position in Rna15 would be a poorer hydrogen bond donor. The RNA binding surface of CstF-64 is also enriched in neutral hydrogen bonding donors (Supplementary Figure 3), while Rna15 has fewer S, T, N, or Q residues there. In particular, S17 of CstF-64 was proposed to make direct hydrogen bonds to the O4 carbonyl of the 5' uracil.³⁰ Residue S17 of CstF-64 is replaced with a valine in Rna15 (V19) that builds a hydrophobic pocket along with Y21 (corresponding to F19 in CstF-64) for burial of the bases of A8 and A9. Lastly, the geometry of the backbone

for N22 and G58 of CstF-64 is more reminiscent of the G165/G201 and G249/G285 regions or Hrp1 than of the analogous position (S24/G60) in Rna15. In this structure, a hydrogen bond interaction between the S58 hydroxyl and the gamma oxygen of S24 pinches off that region and prevents access of G60 to the RNA. In Hrp1, uracil residues hydrogen bond to the aforementioned juxtaposed glycines in both RRM domains (165 and 201 in RRM1 and 249 and 285 in RRM2). This region is freely available for uracil or guanine imino hydrogen-bond donor access in Cstf-64, but is blocked from access in Rna15.

Rna15 interacts with an A-rich sequence, like the Poly-A-binding protein (PABP), so we compared our structure with PABP (1cvj)³³ (Supplementary figure 8), though this is not straightforward since PABP has two RRMs and RRM only 1. Rna15 resembles more closely RRM1 of PABP, with a C α RMSD difference of only 1.5 Å (Supplementary figure 8a) arising mostly in the structure of loop 3 and in the longer loop-5 of PABP. While domain 1 of PABP recognizes AAAA, Rna15 binds to AAUA. Interestingly, the RNA takes a more circuitous path over the β -sheet surface for PABP whereas a more direct path is taken for Rna15 resulting in a change in register. Although in both cases all three aromatic are involved in the typical contacts observed in most RRM structures, the specific contacts to individual nucleotides differ between the two proteins.

Protein-protein interactions modulate the specificity of Rna15

In vivo, the complex of Rna15 and Hrp1 on the pre-mRNA is undoubtedly stabilized by Rna14.^{34; 35} However, in our structure we observed unexpected direct contacts between Rna15 and Hrp1, mediated by loop 3 of Hrp1 and loop 5 of Rna15 (Figure 4F and 4G). These contacts are inferred first of all by the relative positions of the domains as defined by RNA/protein intermolecular NOEs against independent binding elements displayed on a topological connected anchor RNA, but they are also constrained from the PRE and RDC data. These interactions appear to be functionally relevant, because mutations in the corresponding loops perturb cleavage *in vitro* and induce slow growth *in vivo* in the presence of additives that destabilize protein-protein interactions. These mutations are nonetheless not lethal, probably because additional interactions, such as the bridging of Rna15 and Hrp1 by Rna14, can compensate for their loss.

The poor specificity of Rna15 for RNA discussed in the previous section may explain the need for multiple protein-protein interactions to allow a functional interaction with the positioning element RNA. The contact with Hrp1 bound at the adjacent enhancer element, positions Rna15 precisely; Rna14 further locks these proteins in place by increasing the efficiency of the Hrp1/Rna15 interaction. In this way, specificity is achieved by multiple weak protein-RNA and protein-protein interactions, rather than by recognition of well-defined sequence elements. The weak character of these interactions may allow other proteins to compete with Rna15 for the same binding site. For example, Npl3 and Rna15 compete for the positioning element³⁶ and Npl3 has been suggested to directly interact with Hrp1.³⁶ In this case, the presence or absence of the anti-termination factor Npl3, and the competition for the A-rich sequence, defines whether a site would be processed or not.

To investigate if RRM-RRM interactions mediated by loop 3 and loop 5, as we observe for Hrp1 and Rna15, are a general property of other RRMs, we performed a survey of structures

of RNP's containing multiple RRM domains. In these structures, loops 3 and 5 are frequently used to make interdomain contacts in adjacent asymmetric units or between domains in solution (Supplementary Figure 3, right column, red check marks). For example, in the second RRM domain of both HUD (1fxl) and PABP (1ovj), loop 3 contacts loop 5 from an adjacent asymmetric unit. Loop 3 of RRM 1 of both of these proteins contacts the interdomain helix, and generates crystal contacts as well. In hnRNPA1, both loop 3 of RRM1 (1up1) and RRM2 (1ha1) bridge over to loop 5's in an adjacent unit. For nucleolin (1fje), the contacts are less clear due to larger distances between domains, but loop 3 of RRM2 is poised to interact with loop 5 of RRM1 from an adjacent unit cell. Interestingly, recapitulation of the lattice by generation of symmetry-related Rna15 molecules reveals that Rna15 inter-domain contacts exist between loops 3 of one molecule and loop 5 of another as well. Of the remaining RRM domain structures, loop 3's are either used for intramolecular contacts (1h2t) or protein/protein interactions (1rk8 and 1uw4, annotated "ptn" in Supplementary Figure 3, right) that do not involve loop 5.

We also analyzed the primary sequence of structurally characterized RRMs to investigate what features of loop 3 and loop 5 may mediate these inter-RRM interactions (Supplementary Figure 3). One class of RRM domains lack the Asp seen in loop 3 of Hrp1 and Rna15, and have more variable loop 3 regions with a tendency toward basic residues, such as the classic U1A protein.^{37, 38} Aligning these sequences by secondary structure boundaries (Supplementary Figure 3) revealed a second, more frequent class which exhibited conservation of residues in extended loops that do not have readily apparent function in RNA binding or protein stability. Conserved features within loop 3's and loop 5's involved in RRM-RRM interactions tend to be an Asp (e.g. D193 in RRM1 of Hrp1) within a longer loop 3 of 8-10 residues and one or more basic residues in loop 5 at the junction between loop 5 and β 4 (e.g. R87 in Rna15). The Asp is near the Nterminal end of loop 3, just as the structure turns to form an extended hairpin loop (Supplementary Figure 3). None of the residues of loop 3 penetrate into the protein hydrophobic core, so it is unlikely that loop3 dominates the thermodynamics of domain stability. However, the aspartic acid (D52 in Rna15) may provide some slight structural stabilization by accepting a hydrogen bond from a backbone amide from the opposite strand of the β -hairpin turn, and its mutation in Hrp1 to D193N has little effect on protein structure, as established by NMR. Instead, the strong conservation of loop 3 and juxtaposition against loop 5 in this and other structures suggests that loop 5 is a potential docking site for the conserved Asp in loop 3. Thus, to our knowledge, our study explains for the first time, the apparent conservation of loop 3 sequences and places the phylogenetic conservation of an apparent inter-protein salt bridge within the context of multi-RRM domain RNA associations.

Conclusions

In summary, we have determined NMR structure of the ternary RNP complex of Rna15, Hrp1 and the RNA element essential for efficient cleavage and polyadenylation and accurate site selection in yeast 3' mRNA processing. This complex contains a *bona fide* yeast 3' processing signal ("the anchor RNA") with both the efficiency and positioning elements required for efficient and accurate mRNA 3'-processing. The structure shows how the RRM domains from these proteins recognize RNA and identifies protein-protein interactions

involving conserved loop sequences in the RRM of Hrp1 and Rna15. We suggest that the juxtaposition of loop 3 and loop 5 of these two different RRMs plays a role in organizing this and perhaps other ribonucleoproteins onto specific regions of mRNA by a combination of weak RNA sequence preference and protein-protein interactions (Figure 7A). The underlying low binding efficiency for individual RRM domains would allow facile RNP remodeling during RNA biogenesis as the location and status of the RNA changes,³⁹ while the enhancement afforded by protein-protein contacts generates signal fidelity.

Materials and Methods

Protein Expression and purification

The Rna15p RRM, residues 6 to 102, was cloned into pET28. while the Hrp1 expression construct with RRM1 and RRM2 was obtained from Dr. Perez-Canadillas¹⁷ (Supplementary Table 1). These proteins were expressed and purified using standard nickel affinity and size exclusion chromatography, followed by removal of the N-terminal tags with appropriate proteases. The NMR buffer was 20 mM Tris-Cl pH 7, 150 mM KCl, 2 mM DTT. Labeled proteins were made by substituting ¹⁵N NH₄Cl and/or ¹³C glucose and all mutants were generated by Quick-change site-directed mutagenesis.

RNA synthesis

Several RNAs were generated, as shown in Supplementary Table 1 and purified by Polyacrylamide gel electrophoresis (PAGE). The complete anchoring RNA (GGAUAUAUAUAUAUAU) was made by T7 *in vitro* transcription.⁴⁰ RNAs for the binary complexes, AUAUAUAU and AAUAUAU, were chemically synthesized. Nitroxide spin-labeled RNAs (UAUAUAUAUAUAU) were prepared from singly substituted 2'-amino RNA oligonucleotides using iso-cyanato-TEMPO.⁴¹ These RNAs were re-purified to near purity post-coupling by PAGE and checked with electron paramagnetic resonance (EPR).

NMR sample preparation

NMR samples were 0.3 to 0.9 mM in 0.3 ml. Complexes were formed by titrating unlabeled RNA into labeled protein samples while monitoring ¹⁵N HSQC chemical shift changes, until a previously established "saturated" spectrum was obtained. For Hrp1, binding was also monitored by the disappearance of a single RNA imino signal upon complex formation. For sample concentrations >0.3 mM, 1:1:1 ratios of all wild-type protein and RNA components resulted in complete ternary complex formation.

NMR data collection

Multiple NMR experiments were collected on single proteins and on binary and ternary complexes (Supplementary tables 1 and 2) to facilitate spectral assignments and structure determination. ¹⁵N HSQC, ¹³C HSQC, HNCO, HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HCCH-TOCSY, ¹⁵N-TOCSY-HSQC, ¹⁵N and ¹³C NOESYHSQC, and 2D D₂O-NOESY.⁴² TROSY versions of ¹⁵N HSQC, ¹³C HSQC, ¹⁵N NOESY-HSQC, and ¹³C NOESY-HSQC, were used with ¹⁵N/¹³C labeled Hrp1 in ternary complex. Intermolecular RNA/protein NOEs were measured by 2D NOESY spectra with and without decoupling on labeled protein/ unlabeled RNA complexes (Figure 3A). Resonance assignments and T₁ and

T_2 measurements on ternary complexes were recorded and analyzed via standard methods.^{42; 43}

Triple resonance and carbon-edited spectra were collected on 500 and 600 Mz Bruker spectrometers equipped with cryo-probes. Two-dimensional NOESY spectra and the 3D ^{15}N NOESY were collected at 750 MHz with a conventional triple resonance probe. NMR acquisition parameters were as follows: ^1H and ^{15}N sweep widths were 12 and 36 ppm. ^{13}C sweep widths were 16 ppm for the HNCO, 32 ppm for the HNCA and HN(CO)CA, and 80 ppm for the remaining carbon spectra. Echo-anti-echo sensitivity enhancement,^{44; 45} was employed for amide detected spectra, and TPPI or STATES-TPPI quadrature modes⁴⁶ were used for the other indirectly detected dimensions. Acquisition times for indirectly detected ^1H , ^{15}N and ^{13}C were 18 msec, 16.4 msec, and 3.5 to 8 msec respectively, except for 2D experiments where these times were approximately doubled. NOE mixing times were 100 msec and $^1\text{H}/^1\text{H}$ TOCSY mixing times were 30 msec.

Spectral Assignments

For protein sequential assignments, triple resonance spectra were compared (HNCA, HN(CO)CA, CBCA(CO)NH and HNCACB) to assign backbone and CB carbon resonances. Side chains were assigned by correlating HA to HN using ^{15}N edited TOCSY and then the HA, CA, and CB shifts were used as starting points to analyze the HCCH-TOCSY spectrum. Aromatic side chain protons were assigned using 2D NOESYs in D_2O . Because no large chemical shift changes occurred in the relevant protein backbone resonances, ternary complex chemical shifts could be inferred from the binary complex spectra. Resonance assignments for Hrp1, free and bound to a short oligonucleotide of sequence UAUAUAU, were available from previous studies¹⁷ and had only small deviations when bound to the complete anchoring RNA element studied here. Rna15 was assigned *de novo* from triple resonance data collected on the free protein and protein bound to AAUAAU as well as to the full anchoring RNA. NOE cross-peaks were assigned by a combination of manual and automated NOE assignment with the *candid* module of *cyana* (see below). Inter molecular NOEs between the proteins and RNA were assigned manually by comparison of coupled and decoupled NOESY spectra where the pattern of cross-peak splitting indicated whether a given peak was intramolecular protein (quadruplet), intramolecular RNA (singlet) or intermolecular protein/RNA (doublet) (Figure 3A).

Residual dipolar couplings

RDC's were measured for both proteins simultaneously to eliminate deviations in the alignment tensors between samples that would complicate the structure refinement. Two samples were generated: ^{13}C Rna15 - ^{15}N Hrp1 - anchor RNA and ^{15}N Rna15 - ^{13}C Hrp1 - anchor RNA. One bond couplings were measured using IPAP ^{15}N and ^{13}C HSQC⁴⁷ before and after alignment with Pf1 phage (Asla Biotech).⁴⁸ Because of the labeling scheme (Supplementary tables 1 and 2), RDCs for each protein could be measured simultaneously. The magnitude (D_a) and asymmetry (R) of the alignment tensors for these datasets were separately evaluated by obtaining an initial estimate from RDC distributions⁴⁹ and refined by structure calculations.⁵⁰

Paramagnetic Relaxation Enhancements

PREs were obtained using anchoring RNA⁵¹ spin-labeled at each of three uracils. A total of 143 PRE restraints were measured from two sites (U3 and U5). RNA labeled at U3 and U5 was capable of binding both Rna15 and Hrp1 simultaneously and spectra on binary and ternary complexes labeled with ¹⁵N-labeled Rna15 were recorded. RNA spin-labeled at position U7 was able to bind the RRM's individually, but could not form a ternary complex, suggesting that the spin-label was not permitted at this position in the complex. Resonances broadened to < 30% intensity were conservatively restrained to be separated < 20 Å from the relevant 2'-hydroxyl position. Those broadened to between 30% and 80% intensity were restrained to < 30 Å and those that were unaffected or broadened only slightly (> 80% intensity) were restrained to > 12 Å. Since a fixed and inflexible distance constraint was employed, these boundaries are intentionally generous.

Structure calculations

Structure calculations began on binary complexes to test completeness and accuracy of the NOE data with *cyana*.⁵² Once final *cyana* target function values for binary complexes were satisfactorily under 2Å, we initiated structure calculations on the ternary complexes with binary complex restraint sets plus PREs. Structure calculations then were migrated into xplor-NIH to permit RDC refinement. The following annealing protocol was used for RNP refinement in Xplor-NIH: 2000 steps of unrestrained Powell minimization using the *cyana* refined structure as a starting model, restrained dynamics over 238 steps from 3000 to 25K, and 500 steps of unconstrained *powell* minimization. During the cooling, the *Sani* force constant was ramped up to prevent conflicts between degenerate bond vector orientations and distance constraints during the higher temperatures of the annealing. A total of 100 structures were calculated and the top 15 lowest energy structures were selected as the final structure ensemble.

Yeast strains, media and plasmids

The yeast MHY942/*HRP1* shuffle strain (*MATa hrp1::HIS3, his3, trp1, ura3, leu2, ade2, ade8, lys1*, plus *pRS316-HRP1* containing *HRP1* and *URA3*) was a gift from Michael Henry⁵³ and the LM31 strain (*MATa rna15::TRP1 ura3 trp1 ade2 leu2 his3*, plus *pLM13* containing *RNA15* and *URA3*) was a gift from F. Lacroute.⁵⁴ Yeasts were grown and maintained on YPD (1% yeast extract, 2% peptone, 2% glucose) or on selective media as needed. For temperature sensitivity assays, yeasts were grown at 16, 25, 30, and 37°C.

Amino acid substitutions were generated by the QuickChange method (according to QuickChange Mutagenesis Kit manual, Stratagene). Briefly, PCR amplification with PfuTurbo DNA polymerase (Stratagene) to construct plasmid MHY50-D193R from plasmid MHY50 containing *MYCHIS6-HRP1* and *LEU2*⁵³ and to construct plasmid Yeplac-rna15-R87D plasmid from Yeplac-RNA15H6⁵⁵ DNA sequencing was used to confirm that the intended mutation was the only one present in the target gene.

To construct yeast mutant strains *hrp1*-D193R, the yeast strain MHY942 was transformed with plasmid MHY50-D193R and for the *rna15*-R87D strain, LM31 was transformed with plasmid Yeplac-rna-15-R87D. The transformed yeast cells were plated on synthetic complete

media lacking leucine and loss of *URA3* plasmids from the cells was accomplished by plating on solid medium containing 5-fluoroorotic acid. To construct yeast strain D193R/R87D, the *hrp1*-D193R strain was mated with the *rna15*-R87D strain. Following sporulation and tetrad dissection, the desired D193R/R87D spores were identified by growing on solid synthetic complete medium lacking both tryptophan and histidine.

In vitro cleavage and polyadenylation assay and RNA immunoprecipitations

In vitro 3' end cleavage and polyadenylation assays using yeast extract followed previously described protocols.^{34; 56; 57; 58} Briefly, precursor RNAs included *GAL7*-1, which contains the *GAL7* poly(A) site and flanking sequences; *GAL7*-9, a substrate lacking sequences downstream of the *GAL7* site; and a with the *CYC1* poly(A) site and flanking sequences (transcribed from plasmid pT3-CYC1, a gift of Mohamed Ghazy). ATP was used in the coupled cleavage and polyadenylation reactions with full-length precursor RNAs and in the polyadenylation assays with pre-cleaved precursor RNA, while dATP replaced ATP in the cleavage assays. Immunoprecipitation of RNAs from *in vitro* processing reactions was performed as described⁵⁹ with slight modifications. Reactions containing 20 μ l extract in a 200 μ l total reaction volume were stopped after for 20 minutes at 30°C by adding 800 μ l IP Buffer (10 mM Tris-HCl, 50 mM KCl, 1 mM EDTA, pH 7.5), followed by incubation with protein A – agarose beads for 40 minutes at 4°C to clear the samples of proteins binding nonspecifically. Supernatants were then incubated with protein A – agarose beads absorbed with either anti-Rna15 antiserum or pre-immune serum as a control for one hour at 4°C, and the beads then washed 4 times with 1 ml of IP buffer at 25°C. RNAs were released by incubating the beads in 200 μ l of Elution Buffer (50 mM Tris-HCl, 5 mM EDTA, 10 mM DTT, 1% SDS, pH 7.5) for 30 min at 65°C. The supernatant was collected and combined with 150 μ l of TE buffer (10 mM Tris, 1 mM EDTA, pH. 7.5) that was added to the beads after the Elution Buffer. After purification by phenol-chloroform-isoamyl alcohol extraction and concentration by ethanol precipitation, the immunoprecipitated RNAs were separated on a denaturing 5% polyacrylamide gel containing 8.3 M urea and visualized with a phosphorimager. Western blotting was performed as previously described.⁵⁷

PDB sequence and structural analysis

Structural features of RRM domains present in the PDB 17; 25; 29; 30; 33; 60; 61; 62; 63; 64; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 77; 78; 79 were correlated to sequence elements by visual inspection of each set of coordinates. Particular attention was paid to the conserved sequences present in loops that were not apparently involved in direct RNA contacts or in stabilization of the hydrophobic core of the domains. The tabulated results were aligned based upon these structural motifs as anchor points. The sequences were then clustered into two groups based upon the presence of canonical loop 3 sequences. Omitted from the analysis were Tap (1fo1)⁸⁰ and U2AF35 (1jmt).⁸¹ In the case of the former protein, missing density and an absence of any clear RNP1 or RNP2 motifs suggested that this domain is not an RRM. In the case of U2AF35, the extra-long helical extensions and insertions of several residues throughout the β -sheet precluded reliable alignment.

Accession numbers

Coordinates have been deposited in the Protein Data Bank with accession number 2km8.
NMR data and chemical shifts are deposited with BMRB accession code 16425.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Brad Lunde, Yu Chen, and Andrew Bohm for helpful discussions and suggestions. A portion of the research was performed using EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. This work was supported in part by NIH R01 grants to GV and CM.

References

- Schmid M, Jensen TH. Quality control of mRNP in the nucleus. *Chromosoma*. 2008; 117:419–29. [PubMed: 18563427]
- Iglesias N, Stutz F. Regulation of mRNP dynamics along the export pathway. *FEBS Lett*. 2008; 582:1987–96. [PubMed: 18394429]
- Mangus DA, Evans MC, Jacobson A. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol*. 2003; 4:223. [PubMed: 12844354]
- Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes Dev*. 2009; 23:1247–69. [PubMed: 19487567]
- Lainé JP, Singh BN, Krishnamurthy S, Hampsey M. A physiological role for gene loops in yeast. *Genes & Development*. 2009; 23:2604–2609. [PubMed: 19933150]
- Tan-Wong SM, Wijayatilake HD, Proudfoot NJ. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes & Development*. 2009; 23:2610–2624. [PubMed: 19933151]
- Guisbert KSK, Li H, Guthrie C. Alternative 3' pre-mRNA processing in *Saccharomyces cerevisiae* is modulated by Nab4/Hrp1 in vivo. *Plos Biology*. 2007; 5:15–22.
- Rondon AG, Proudfoot NJ. Nuclear roadblocks for mRNA export. *Cell*. 2008; 135:207–8. [PubMed: 18957194]
- Graber JH, McAllister GD, Smith TF. Probabilistic prediction of *Saccharomyces cerevisiae* mRNA 3'-processing sites. *Nucleic Acids Res*. 2002; 30:1851–8. [PubMed: 11937640]
- Graber JH, Cantor CR, Mohr SC, Smith TF. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res*. 1999; 27:888–94. [PubMed: 9889288]
- Dichtl B, Keller W. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J*. 2001; 20:3197–209. [PubMed: 11406596]
- Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci*. 2008; 65:1099–122. [PubMed: 18158581]
- Gross S, Moore CL. Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. *Mol Cell Biol*. 2001; 21:8045–55. [PubMed: 11689695]
- Irniger S, Braus GH. Saturation Mutagenesis of a Polyadenylation Signal Reveals a Hexanucleotide Element Essential for Messenger-Rna 3' End Formation in *Saccharomyces Cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*. 1994; 91:257–261. [PubMed: 8278376]
- Duvel K, Egli CM, Braus GH. A single point mutation in the yeast TRP4 gene affects efficiency of mRNA 3' end processing and alters selection of the poly(A) site. *Nucleic Acids Res*. 1999; 27:1289–95. [PubMed: 9973616]
- Guo Z, Sherman F. 3'-end-forming signals of yeast mRNA. *Mol Cell Biol*. 1995; 15:5983–90. [PubMed: 7565751]

17. Perez-Canadillas JM. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J.* 2006; 25:3167–78. [PubMed: 16794580]
18. Noble CG, Walker PA, Calder LJ, Taylor IA. Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic Acids Res.* 2004; 32:3364–75. [PubMed: 15215336]
19. Takagaki Y, Manley JL, Macdonald CC, Wilusz J, Shenk T. A Multisubunit Factor, Cstf, Is Required for Polyadenylation of Mammalian Pre-Messenger-Rnas. *Genes & Development.* 1990; 4:2112–2120. [PubMed: 1980119]
20. Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol.* 1997; 17:3907–14. [PubMed: 9199325]
21. Lutz-Freyermuth C, Query CC, Keene JD. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc Natl Acad Sci U S A.* 1990; 87:6393–7. [PubMed: 1696729]
22. Perez I, McAfee JG, Patton JG. Multiple RRM contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry.* 1997; 36:11881–90. [PubMed: 9305981]
23. Banerjee H, Rahn A, Davis W, Singh R. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA.* 2003; 9:88–99. [PubMed: 12554879]
24. Leulliot N, Quevillon-Cheruel S, Graille M, van Tilbeurgh H, Leeper TC, Godin KS, Edwards TE, Sigurdsson ST, Rozenkrants N, Nagel RJ, Ares M, Varani G. A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *EMBO J.* 2004; 23:2468–77. [PubMed: 15192703]
25. Varani L, Gunderson SI, Mattaj IW, Kay LE, Neuhaus D, Varani G. The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat Struct Biol.* 2000; 7:329–35. [PubMed: 10742179]
26. Leeper TC, Athanassiou Z, Dias RLA, Robinson JA, Varani G. TAR RNA recognition by a cyclic peptidomimetic of Tat protein. *Biochemistry.* 2005; 44:12362–12372. [PubMed: 16156649]
27. Allain FH, Howe PW, Neuhaus D, Varani G. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* 1997; 16:5764–72. [PubMed: 9312034]
28. Hampsey M. A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast.* 1997; 13:1099–133. [PubMed: 9301019]
29. Pancevac C, Goldstone DC, Ramos A, Taylor IA. Structure of the Rna15 RRMRNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res.* 38:3119–32.
30. Perez Canadillas JM, Varani G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* 2003; 22:2821–30. [PubMed: 12773396]
31. Guo Z, Sherman F. 3'-end-forming signals of yeast mRNA. *Trends Biochem Sci.* 1996; 21:477–81. [PubMed: 9009831]
32. Auweter SD, Fasan R, Raymond L, Underwood JG, Black DL, Pitsch S, Allain FH. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* 2006; 25:163–73. [PubMed: 16362037]
33. Deo RC, Bonanno JB, Sonenberg N, Burley SK. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell.* 1999; 98:835–45. [PubMed: 10499800]
34. Gross S, Moore C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci U S A.* 2001; 98:6080–5. [PubMed: 11344258]
35. Kessler MM, Zhao J, Moore CL. Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends. *J Biol Chem.* 1996; 271:27167–75. [PubMed: 8900210]
36. Bucheli ME, He X, Kaplan CD, Moore CL, Buratowski S. Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. *RNA.* 2007; 13:1756–64. [PubMed: 17684230]

37. Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, Varani G. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature*. 1996; 380:646–50. [PubMed: 8602269]
38. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*. 1994; 372:432–8. [PubMed: 7984237]
39. Singh R, Valcarcel J. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol*. 2005; 12:645–53. [PubMed: 16077728]
40. Milligan JF, Uhlenbeck OC. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol*. 1989; 180:51–62. [PubMed: 2482430]
41. Edwards TE, Sigurdsson ST. Site-specific incorporation of nitroxide spinlabels into 2'-positions of nucleic acids. *Nat Protoc*. 2007; 2:1954–62. [PubMed: 17703207]
42. Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 1999; 34:93–158.
43. Kay LE, Torchia DA, Bax A. Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear Nmr-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry*. 1989; 28:8972–8979. [PubMed: 2690953]
44. Palmer AG, Cavanagh J, Wright PE, Rance M. Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation Nmr-Spectroscopy. *Journal of Magnetic Resonance*. 1991; 93:151–170.
45. Kay LE, Keifer P, Saarinen T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *Journal of the American Chemical Society*. 1992; 114:10663–10665.
46. States DJ, Haberkorn RA, Ruben DJ. A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *Journal of Magnetic Resonance*. 1982; 48:286–292.
47. Andersson P, Weigelt J, Otting G. Spin-state selection filters for the measurement of heteronuclear one-bond coupling constants. *J Biomol NMR*. 1998; 12:435–41. [PubMed: 9835050]
48. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*. 1998; 5:1065–74. [PubMed: 9846877]
49. Clore GM, Gronenborn AM, Bax A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson*. 1998; 133:216–21. [PubMed: 9654491]
50. Clore GM, Gronenborn AM, Tjandra N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson*. 1998; 131:159–62. [PubMed: 9533920]
51. Ramos A, Bayer P, Varani G. Determination of the structure of the RNA complex of a double-stranded RNA-binding domain from *Drosophila* Staufen protein. *Biopolymers*. 1999; 52:181–96. [PubMed: 11295750]
52. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol*. 2004; 278:353–78. [PubMed: 15318003]
53. Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev*. 1997; 11:2545–56. [PubMed: 9334319]
54. Minvielle-Sebastia L, Winsor B, Bonneaud N, Lacroute F. Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein. *Mol Cell Biol*. 1991; 11:3075–87. [PubMed: 1674817]
55. Qu X, Perez-Canadillas JM, Agrawal S, De Baecke J, Cheng H, Varani G, Moore C. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. *J Biol Chem*. 2007; 282:2101–15. [PubMed: 17116658]
56. Chen J, Moore C. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. *Mol Cell Biol*. 1992; 12:3470–81. [PubMed: 1352851]

57. Zhao J, Kessler M, Helmling S, O'Connor JP, Moore C. Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol Cell Biol.* 1999; 19:7733–40. [PubMed: 10523662]
58. Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH. Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell.* 2008; 31:91–103. [PubMed: 18614048]
59. Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH, Moore C. Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation. *Mol Cell Biol.* 2009; 29:5327–38. [PubMed: 19635808]
60. Wang X, Tanaka Hall TM. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol.* 2001; 8:141–5. [PubMed: 11175903]
61. Allain FH, Bouvet P, Dieckmann T, Feigon J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* 2000; 19:6870–81. [PubMed: 11118222]
62. Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR. Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure.* 1997; 5:559–70. [PubMed: 9115444]
63. Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol.* 1997; 4:215–22. [PubMed: 9164463]
64. Volpon L, D'Orso I, Young CR, Frasch AC, Gehring K. NMR structural study of TcUBP1, a single RRM domain protein from *Trypanosoma cruzi*: contribution of a beta hairpin to RNA binding. *Biochemistry.* 2005; 44:3708–17. [PubMed: 15751947]
65. Bono F, Ebert J, Unterholzner L, Guttler T, Izaurralde E, Conti E. Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep.* 2004; 5:304–10. [PubMed: 14968132]
66. Kadlec J, Izaurralde E, Cusack S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol.* 2004; 11:330–7. [PubMed: 15004547]
67. Perez-Alvarado GC, Martinez-Yamout M, Allen MM, Grosschedl R, Dyson HJ, Wright PE. Structure of the nuclear factor ALY: insights into postranscriptional regulatory and mRNA nuclear export processes. *Biochemistry.* 2003; 42:7348–57. [PubMed: 12809490]
68. Inoue M, Muto Y, Sakamoto H, Kigawa T, Takio K, Shimura Y, Yokoyama S. A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino-terminal RNA-binding domain of *Drosophila* sex-lethal. *J Mol Biol.* 1997; 272:82–94. [PubMed: 9299339]
69. Lee AL, Kanaar R, Rio DC, Wemmer DE. Resonance assignments and solution structure of the second RNA-binding domain of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry.* 1994; 33:13775–86. [PubMed: 7524663]
70. Ito T, Muto Y, Green MR, Yokoyama S. Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)). *EMBO J.* 1999; 18:4523–34. [PubMed: 10449418]
71. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell.* 2003; 11:965–76. [PubMed: 12718882]
72. Mazza C, Segref A, Mattaj IW, Cusack S. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* 2002; 21:5548–57. [PubMed: 12374755]
73. Enokizono Y, Konishi Y, Nagata K, Ouhashi K, Uesugi S, Ishikawa F, Katahira M. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J Biol Chem.* 2005; 280:18862–70. [PubMed: 15734733]
74. Conte MR, Grune T, Ghuman J, Kelly G, Ladas A, Matthews S, Curry S. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* 2000; 19:3132–41. [PubMed: 10856256]

75. Deka P, Bucheli ME, Moore C, Buratowski S, Varani G. Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals. *J Mol Biol.* 2008; 375:136–50. [PubMed: 18022637]
76. Price SR, Evans PR, Nagai K. Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature.* 1998; 394:645–50. [PubMed: 9716128]
77. Jacks A, Babon J, Kelly G, Manolaridis I, Cary PD, Curry S, Conte MR. Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure.* 2003; 11:833–43. [PubMed: 12842046]
78. Simpson PJ, Monie TP, Szendroi A, Davydova N, Tyzack JK, Conte MR, Read CM, Cary PD, Svergun DI, Konarev PV, Curry S, Matthews S. Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure.* 2004; 12:1631–43. [PubMed: 15341728]
79. Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science.* 2005; 309:2054–7. [PubMed: 16179478]
80. Liker E, Fernandez E, Izaurralde E, Conti E. The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. *EMBO J.* 2000; 19:5587–98. [PubMed: 11060011]
81. Kielkopf CL, Rodionova NA, Green MR, Burley SK. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell.* 2001; 106:595–605. [PubMed: 11551507]

Abbreviations

RRM	RNA recognition motif
PE	positioning element
EE	enhancer element
DUE	downstream u-rich element
UUE	upstream u-rich element
RDC	residual dipolar couplings
PRE	paramagnetic relaxation enhancements
NOE	nuclear Overhauser enhancements

Figure 1. mRNA 3'-end processing components in yeast. **A)** Schematic representation of the 3'-end processing complex on the *GAL7* pre-mRNA, with RNA elements important for correct processing highlighted in boxes. For the Hrp1 binding site, only the RNA element identified with a solid blue box was studied in this work, although the sequence is duplicated (blue dashed box). Below the RNA sequence, the cartoon identifies additional protein factors. Not shown are the other components of the CPF complex that are responsible for the chemical steps of the cleavage and polyadenylation reactions. **B)** The domain organization of Hrp1 and Rna15 proteins, with the RRM domains studied in this work highlighted. The accessory N- and C-terminal domains that mitigate additional protein-protein contacts, e.g. between Rna15 C-terminus and Rna14, were not examined.

Figure 2.

Chemical shift perturbation of the NMR signals of Hrp1 and Rna15 upon complex formation. **A)** ^{15}N HSQC spectra of Rna15 with (red) and without (black) one equivalent of the anchoring RNA element. **B)** ^{15}N HSQC spectra of Hrp1 with (red) and without (black) one equivalent of the complete anchoring-RNA/Rna15 complex. **C)** Chemical shift perturbations for Rna15 ($\text{ppm} = [(\Delta\text{H})^2 + (\Delta\text{N} \cdot 0.2)^2]^{1/2}$) plotted as a function of sequence and displayed on the structure in the inset; residues with CSP > 0.075 ppm are indicated in green, while residues with CSP > 0.15 are in dark green. **D)** Surface renderings for Hrp1 RRM1 (top) and RRM2 (bottom) showing sites of RNA contacts, with residues with chemical shift perturbation > 0.2 and 0.4 (supplementary figure 1A) painted in green and dark green, respectively.

Figure 3.

Experimental constraints for structure determination. **A)** Intermolecular NOEs between Rna15 and RNA monitored by comparison of coupled (black) and decoupled (red) NOESY spectra in D₂O on ¹⁵N and ¹³C labeled Rna15 bound to the unlabeled anchor RNA. Some representative intermolecular NOEs are highlighted by the blue boxes. **B)** Representative PRE data: U5 nitroxide spin labeled anchoring RNA bound to ¹⁵N-labeled Rna15 (black) or ¹⁵N-labeled Rna15 and unlabeled Hrp1 (blue). The reference spectrum of the later sample collected after the spin label was reduced is also shown (red).

Figure 4.

Structure of the Rna15/Hrp1/anchoring RNA ternary complex. **A)** Cartoon and stick representation of the ternary complex. Rna15 is colored in cyan and yellow, the anchor RNA is green, and the Hrp1 domains are in magenta and yellow. Loop 3 regions of RRM domains are colored in red and loop 5 regions are colored in bright blue. The functionally important Rna15-Hrp1 loop3-loop5 contacts mutated and analyzed *in vivo* are enclosed by the red dotted rectangle. **B)** Surface rendering of the ternary complex with a similar pose and domain coloring as in panel A. Atoms of the conserved aspartic acid to basic residue

contacts between domains are indicated by the blue and red spheres, respectively. **C)** Ensemble superposition for Hrp1 RRM1; RNA residues 4 to 7 (green) contact Hrp1 residues (yellow) F162, W168, M191, F202, F204. **D)** Ensemble superposition for Hrp1 RRM2; RNA residues 1 to 4 (green) and Hrp1 residues F246, M275, F286, and F288 are explicitly shown (yellow). **E)** Ensemble superposition of the Rna15 RRM with RNA residues 8 to 11 (green) with Rna15 residues Y21, M50, Y61, F63 from the lowest energy structure shown (yellow). **F)** Intermolecular interactions between Hrp1 RRM1 loop3 involving D193 (red) and loop5 in Rna15 involving R87 (blue). Side chains for the relevant side chains are shown as spheres. **G)** Intramolecular interactions between loop 3 in Hrp1 RRM2 involving D279 (red) and loop 5 in RRM 1 involving K226 (blue). **H)** Close up view of the interaction between RNP1 and RNP2 in Rna15 and the RNA. Residues 8 to 11 of the anchor RNA are shown from dark blue to cyan. The Rna15 side chains that make typical RRM interactions are Y21 in yellow, M50 in orange, Y61 in red-orange and F63 in red.

Figure 5.

Analysis of the effect of the *rna15-R87D* and *hrp1-D193R* mutations on mRNA cleavage and polyadenylation in vitro. **A)** Coupled cleavage and polyadenylation assays. Extracts were prepared from the indicated mutants or isogenic wild-type cells grown in liquid YPD at 25°C, and processing reactions performed by incubation of 1 or 2 µl of extract with ATP and a full-length radioactively labeled precursor RNA (*GAL7-1*) containing the *GAL7* poly(A) site (lane marked “Pre”) at 30°C for 20 min. **B)** Cleavage assays. Processing reactions performed by incubation of the extracts described in panel A with dATP and full-length

GAL7 precursor at 30°C for 20 min. C) Polyadenylation assays. Processing reactions performed by incubation of 1 μ l of each of the extracts described in panel A with ATP and pre-cleaved *GAL7* precursor (GAL7-9) at 30°C for 20 min.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6.

Deficiencies in *rna15-R87D* are not limited to the *GAL7* transcript and are due to an inability to form a stable RNP poly-protein processing complex *in vivo*. **A)** Coupled cleavage/polyadenylation assay using the same conditions as Figure 5A but with precursor containing the *CYC1* poly(A) site. **B)** The *rna15-R87D* mutation weakens the association of Rna15 with *GAL7-1* precursor incubated with extract. Processing reactions were incubated at 30°C for 30 min, and RNAs immunoprecipitated using Rna15p antibody (+) or preimmune serum (-). “Input” represents an aliquot of the reaction taken before immunoprecipitation, and “Pellet” indicates RNAs that are pulled down with the antibodies.

Figure 7.

Model for anchor RNA recognition by multiple RRMs. **A)** Topological arrangement of the RRM domains and RNA in the Hrp1/Rna15/anchoring RNA ternary complex. Covalent connections between the RNA segments are shown by dotted lines. The helical segment between RRM1 and RRM2 in Hrp1 is drawn as a purple spiral. Basic surface patches are indicated in blue, while acidic regions are in red. The multicolored arrows highlight the potential charge/charge interactions and shape complementarities between loop 3 and loop 5. Arrows to question marks suggest potential contact sites to heretofore-unidentified RRM domains. **B)** Similar to panel A, but the surface rendering and topological arrangement of the

RRM domains is flipped 180° with Rna15 at the left and Hrp1 RRM2 at the right; from this perspective, the strongly negative feature of the RRM α -helical faces can be appreciated.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

NMR structural statistics for the 10 lowest energy conformers of Rna15 and Hrp1 simultaneously bound to the anchor RNA sequence. Dihedral (phi/psi) restraints were generated with TALOS. Ramachandran statistics were evaluated with PROCHECK-NMR. Buried surface area was calculated in MolMol.

Restraints:	
Distance Restraints (≈ 17 per residue)	5349
Intermolecular (Hrp1 to RNA)	131
Intermolecular (Rna15 to RNA)	70
Spin-label	143
Dihedral	172
RDCs	201
RMSD (restraints):	
Total Explor Energy	-2854.84
Dihedrals	0.64°
NOEs	0.02 Å
RDCs	2.9 Hz
RMSD (geometry):	
Bond lengths	0.003 Å
Bond angles	0.59°
Improper	0.49°
Ramachandran plot regions:	
Most favored	69 %
Additionally allowed	27.2 %
Generously allowed	3.8 %
Disallowed	0.0 %
RMSDs (ensemble backbone atoms to the mean structure):	
All ordered residues (protein and RNA)	2.0 Å
Hrp1 RRM domain 1 + anchor RNA residues 1 to 4	0.6 Å
Hrp1 RRM domain 2 + anchor RNA residues 4 to 7	0.5 Å
Rna15 RRM domain + anchor RNA residues 8 to 11	0.5 Å
Buried surface area:	
RNA	1792 Å ²
Rna15	673.6 Å ²
Rna15/Hrp1 interface	113 Å ²