Published online 30 August 2010

Nucleic Acids Research, 2011, Vol. 39, No. 1 ***337–346***
doi:10.1093/nar/gkq751

Structural insights into *cis* element recognition of non-polyadenylated RNAs by the Nab3-RRM

Bradley M. Lunde, Maximilian Hörner and Anton Meinhart*

Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahn Strasse 29, 69120 Heidelberg, Germany

Received July 1, 2010; Revised August 6, 2010; Accepted August 7, 2010

ABSTRACT

Transcription termination of non-polyadenylated RNAs in *Saccharomyces cerevisiae* occurs through the action of the Nrd1–Nab3–Sen1 complex. Part of the decision to terminate via this pathway occurs via direct recognition of sequences within the nascent transcript by RNA recognition motifs (RRMs) within Nrd1 and Nab3. Here we present the 1.6 Å structure of Nab3-RRM bound to its UCUU recognition sequence. The crystal structure reveals clear density for a UCU trinucleotide and a fourth putative U binding site. Nab3-RRM establishes a clear preference for the central cytidine of the UCUU motif, which forms pseudo-base pairing interactions primarily through hydrogen bonds to main chain atoms and one serine hydroxyl group. Specificity for the flanking uridines is less defined; however, binding experiments confirm that these residues are also important for high affinity binding. Comparison of the Nab3-RRM to other structures of RRMs bound to polypyrimidine RNAs showed that this mode of recognition is similar to what is observed for the polypyrimidine-tract binding RRMs, and that the serine residue involved in pseudo-base pairing is only found in RRMs that bind to polypyrimidine RNAs that contain a cytosine base, suggesting a possible mechanism for discriminating between cytosine and uracil bases in RRMs that bind to polypyrimidine-containing RNA.

INTRODUCTION

In *Saccharomyces cervisiae* at least two pathways for 3′-end formation have been described (1), and the decision which pathway to use depends on both *cis*-acting sequence elements within the RNA and *trans*-acting factors. For messenger RNAs, 3′-end formation involves recognition of several conserved sequence elements near the processing site, as well as recognition of the phosphorylation status of the RNA polymerase II (Pol II) C-terminal domain (CTD) by two multi-protein processing complexes, which then direct the assembly and function of the cleavage and polyadenylation machinery (2,3). For short Pol II transcripts, such as the small nuclear and nucleolar RNAs (sn/snoRNAs), a distinct 3′-end processing pathway termed the Nrd1–Nab3–Sen1 dependent pathway exists, which also depends on the recognition of signals from the RNA (4–8) as well as the phosphorylation state of the CTD (9). Both of these complexes appear to travel with the polymerase and the decision which termination pathway to choose is driven by this combination of signals from the polymerase and the nascent transcript (10). One element propelling the decision to terminate via the Nrd1–Nab3–Sen1 dependent pathway is direct contact to the CTD of Pol II by both Nrd1 and Sen1 (9,11). In particular, Nrd1 contains an N-terminal CTD interacting domain (CID) that binds preferentially to the Ser5 phosphorylated CTD (9). Recognition of the Ser5 phosphorylated CTD facilitates processing of short RNAs, as this modification predominates early in transcription (12). Further enhancing elements for the decision to terminate via the Nrd1–Nab3–Sen1 pathway are binding sites within the transcript for both Nrd1 and Nab3. These binding sites are recognized by single RNA recognition motifs (RRMs) found within Nrd1 and Nab3, where the Nrd1-RRM preferentially binds to GUA[A/G] and Nab3-RRM to UCUU sequence motifs (6). These sequences are often found in many of the RNAs terminated via the Nrd1–Nab3–Sen1 pathway (4,6,13,14); however, the frequency of occurrence, the distances separating them, and their order of appearance are not conserved (6).

Nrd1 and Nab3 form a heterodimer and this seems to facilitate high affinity binding to RNAs containing UCUU and GUA[A/G] sequences (15). Heterodimerization of Nrd1 and Nab3 may serve as a mechanism to increase the proteins' ability to recognize non-poly(A) RNAs (16). Often RRMs are found in multiple copies within genes, and this modularity can have dramatic influences

*To whom correspondence should be addressed. Tel: +49 6221 486 505; Fax: +49 6221 486 585; Email: anton.meinhart@mpimf-heidelberg.mpg.de

on the affinity and specificity for particular RNA targets (17). Since Nrd1 and Nab3 each contain single RRMs, heterodimerization may facilitate the same sort of mechanism to increase their affinity and specificity for particular RNAs as is seen in proteins containing multiple RRMs. In addition, it has been suggested that the presence of multiple copies of the Nrd1 and Nab3 binding sequences found in non-poly(A) terminators may lead to cooperative interactions which could exert significant additional control over when and where these proteins act (6). However, there is no strict requirement for non-coding RNAs to contain both Nrd1 and Nab3 binding sites and in some cases there are multiple Nab3 sites but no Nrd1 sites observed or *vice versa* (13). Thus, it remains unclear how cooperative interactions may occur with varied RNA targets.

In order to understand the role of RNA recognition in the Nrd1–Nab3–Sen1 dependent pathway, we have explored the structural details of Nab3-RRM binding to an RNA containing the UCUU recognition sequence derived from the 3′-region of the snR47 gene. The structure of the Nab3-RRM:RNA complex demonstrates that Nab3-RRM binds to a UCU trinucleotide and contains a highly specific interaction with the central cytidine nucleotide. Comparison of the Nab3-RRM:RNA complex to previously solved RRM:RNA complexes with polypyrimidine-containing RNAs reveals that Nab3-RRM binds similarly to the polypyrimidine-tract binding protein (PTB) and suggests that a serine residue within the unstructured C-terminus of the Nab3 and PTB RRM can facilitate specific recognition of a cytidine nucleotide. Binding experiments exploring the specificity of the interaction revealed that the central cytidine as well as the flanking uridine nucleotides are important for a high-affinity binding. We also assessed the effect of neighboring Nab3 binding sites within a single RNA, but could find no evidence for cooperative binding by Nab3-RRM. Taken together, our results demonstrate the structural basis for the recognition of UCUU elements within non-poly(A) RNAs.

MATERIALS AND METHODS

Cloning and expression

The coding sequences for the Nab3-RRM(404) (encoding for residues 329–404) and Nab3-RRM(419) (residues 329–419) were amplified by polymerase chain reaction (PCR) using the primer Nab3_329_forward 5′-GCATCA TATGAAGTCAAGATTATTCATTGG-3′ and Nab3_ 404_reverse 5′-GCGCGGCCGCTTAACGAGCATTCG AGC-3′ or Nab3_419_reverse 5′-GCGCGGCCGCTTA AGTAGAACTACTGTTTGTACC-3′ from genomic *S. cerevisiae* DNA. After restriction digestion using NdeI and NotI (New England Biolabs), PCR products were ligated into pET28b expression vector DNA (Novagen) resulting in an N-terminal fusion with a hexahistidine tag. All expression constructs were verified by DNA sequencing.

Escherichia coli BL-21(DE3) RIL cells (Stratagene) transformed with the pET28b Nab3-RRM(419) or Nab3-RRM(419) constructs were grown at 310 K in LB medium containing kanamycin and chloramphenicol until reaching an OD_{600}~0.6. The temperature was reduced to 293 K and protein expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside for 16 h. Cells were harvested by centrifugation and resuspended in buffer A (50 mM Tris–HCl pH 7.3, 300 mM NaCl and 5 mM 2-mercaptoethanol). Cell walls were broken by sonication and the cell debris was clarified by centrifugation at 20 000*g*. The supernatant was loaded onto Ni–NTA beads (Qiagen) equilibrated with buffer A. Following a wash with buffer B (50 mM Tris–HCl pH 7.3, 1 M NaCl and 5 mM 2-mercaptoethanol), the bound proteins were eluted in buffer C (50 mM Tris–HCl pH 7.3, 300 mM NaCl, 250 mM imidazole and 5 mM 2-mercaptoethanol). The fused hexahistidine tag was removed by thrombin digestion in buffer A supplemented with 2 mM $CaCl_2$ over night at 277 K and the cleaved protein was separated from uncleaved protein by re-chromatography over Ni–NTA beads. The flow-through was dialyzed against buffer D (50 mM Tris–HCl, pH 7.3, 50 mM NaCl and 2 mM dithioerythritol) and loaded onto a HiTrap Heparin column (GE Healthcare) equilibrated with buffer D. The bound protein eluted in a linear gradient of 20 column volumes to buffer E (50 mM Tris–HCl pH 7.3, 1 M NaCl and 2 mM dithioerythritol). Nab3-RRM proteins were further purified by size exclusion chromatography on a Superdex 75 column (GE Healthcare) equilibrated with buffer F (50 mM Hepes–NaOH pH 7.3, 200 mM NaCl and 2 mM dithioerythritol).

Crystallization, crystal harvesting and data collection

Crystals of the free Nab3-RRM(404) grew at 293 K in a hanging-drop vapor diffusion setup using a reservoir solution containing 100 mM sodium cacodylate buffer pH 6.5, 200 mM sodium acetate and PEG 8000 at concentrations between 20 and 30% (w/v) within 24 h. Single crystals were harvested into a cryoprotectant solution containing the mother liquor and additional 10 % (v/v) glycerol and subsequently flash cooled in liquid nitrogen.

For crystallization of the Nab3-RRM(404):RNA complex, synthetic wt snR47 (Table 1) was purchased from Eurofins MWG. The RNA was dialyzed against buffer F using a 1 ml Spectra/Por Float-A-Lyzer 1000 MWCO (Spectrum Laboratories) at 277 K overnight. Nab3–RRM(404) protein and the RNA were mixed at a ~1:1.2 molar ratio. Protein/RNA crystals grew within a week in a sitting-drop vapor diffusion setup using a reservoir solution containing 85 mM sodium acetate buffer

Table 1. List of RNAs used in this study

RNA	Sequence
wt snR47	5′-UUCUUAUUCUUA-3′
FAM-wtsnR47	5′-FAM-UUCUUAUUCUUA-3′
FAM-snR47one	5′-FAM-UUCUUAUUUUUA-3′
FAM-snR47U	5′-FAM-UUUUUAUUUUUA-3′
FAM-snR47C	5′-FAM-UCCCUAUCCCUA-3′

pH 4.6, 170 mM ammonium acetate, 25.5% (w/v) PEG 4000 and 15 % (v/v) glycerol. This crystal condition contained sufficient cryoprotectant and the crystals were immediately flash-frozen in liquid nitrogen. Synchrotron diffraction data were collected at the beamline X10SA at the Swiss Light Source (Switzerland) at 100 K and processed with XDS (18).

Phasing and structure refinement

Initial phases for free Nab3-RRM were obtained by molecular replacement using PHASER (19) with the hnRNP C structure (PDB ID: 1WF2) as a search model. For the diffraction data of Nab3-RRM:RNA complex crystals, phases were obtained by molecular replacement with the free Nab3-RRM as a search model. Model building was performed in cycles of manual building with COOT (20) and automatic refinement using the simulated annealing protocol from CNS (21) and restrained refinement with REFMAC (22). During the final cycles of refinement, phases were improved using TLS mode (23) with the protein and RNA determined as individual groups by the TLSMD server (24). Model quality was evaluated with PROCHECK (25) and MolProbity (26). Figures were prepared using PYMOL (27) and LIGPLOT (28).

Fluorescence anisotropy titration experiments

5′ 5,6-Carboxyfluorescein (FAM) labeled RNAs (Table 1) were purchased from Eurofins MWG, resuspended in an RNase-free 10 mM sodium phosphate buffer pH 7 and dialyzed as described earlier. For the fluorescence anisotropy titration experiments, the FAM-labeled RNA was diluted to a final concentration of 500 nM in a buffer containing 10 mM Hepes–NaOH pH 7.3, 150 mM NaCl and 1 mM dithioerythritol into which the Nab3-RRM protein was titrated. All of the fluorescence anisotropy titrations were carried out in a fluorescence spectrometer in T-configuration (Model FL322, Jobin Yvon) at 25°C. The fluorophore was excited at 477 nm using vertically polarized light, and both vertical and horizontal emissions were recorded at 525 nm. The data were modeled using non-linear least squares curve-fitting routines as implemented in the program Dynafit (29) and the binding model selection was performed based on the Akaike Information Criterion.

RESULTS AND DISCUSSION

Nab3-RRM folds into a canonical RRM

Crystals of Nab3-RRM(404) diffracted to 1.3 Å resolution and thus enabled us to determine the 3D structure at atomic resolution. In addition to residues 329–403 from the Nab3 polypeptide chain, four N-terminally fused residues from the residual artificial tag could be modeled unambiguously into the electron density. Only the electron density for the loop region connecting β1 to α1 (residues 339–341) was poorly defined, indicating that this part of the polypeptide chain is disordered in the crystal and thus may be flexible (Supplementary Figure S1a). Much to our surprise, electron density for the last arginine residue (Arg404) at the C-terminus was missing. Instead, we observed density consistent with a C-terminal carboxylate group at Ala403. By mass spectrometry, we could verify the lack of the C-terminal last residue, although it was present in the coding sequence (data not shown). Thus, Arg404 is proteolytically removed during purification. A second unexpected finding was residual electron density interpreted as an s-dimethylarsinoyl-cysteine residue, which most probably arose from a covalent modification of Cys383 by cacodylate in the crystallization buffer (Supplementary Figure S1b). Overall, the model has excellent stereochemical quality with 97.2% of the residues lying in the most favored regions and 2.8 % in the generally allowed regions of the Ramachandran plot. Details of the diffraction data and refinement statistics are listed in Table 2.

The Nab3-RRM adopts the canonical β1–α1–β2–β3–α2–β4 topology and folds into a four-stranded, antiparallel β-sheet packed against two α-helices common to previously characterized RRM structures (Figure 1). However, among the RRM structures determined to date the Nab3-RRM is exceptional, since unlike most other RRM domains, the loop region connecting the β2

Table 2. Data collection and refinement statistics

	Nab3-RRM(404)	Nab3-RRM(404) + $U_1C_2U_3$
Data collection		
Space group	$P2_12_12_1$	$P4_1$
Cell dimensions		
a, *b*, *c* (Å)	26.95, 42.07, 55.67	30.68, 30.68, 82.55
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	1.1271	0.978243
Resolution (Å)[a]	50–1.3 (1.4–1.3)*	50–1.6 (1.7–1.6)
R_{meas}[a,b]	5.7 (20.7)	8.6 (42.9)
$<I/\sigma I>$[a]	19.8 (7.54)	14.91 (5.26)
Completeness (%)[a]	98.3% (95.0)	98.9% (96.3)
Redundancy[a]	6.0 (5.5)	6.3 (6.3)
Refinement		
Resolution (Å)[a]	33.6–1.3 (1.33–1.3)	30.6–1.6 (1.642–1.6)
No. of unique reflections[a]	15 337 (808)	9510 (501)
R_{work}[c]/R_{free}[d]	0.17/0.19	0.16/0.18
No. of atoms		
Protein	647	605
Acetate ion	8	8
Water	81	47
RNA	—	65
B-factors		
Protein	12.1	17.5
RNA	NA	36.2
Cacodylate	16.1	NA
Acetate ion	24.2	47
Water	17.8	21
r.m.s.d.		
Bond lengths (Å)[e]	0.010	0.009
Bond angles (°)	1.335	1.25

[a]Values in parentheses are for highest-resolution shell.
[b]$R_{meas} = \sum_h [n_h/(n_h-1)]^{1/2} \sum_i |I_{hi}-I_h| / \sum_h \sum_i I_{h,i}$, where I_h is the mean intensity of symmetry-equivalent reflections and n_h is the redundancy.
[c]$R_{work} = \sum ||F_o|-|F_c|| / \sum |F_o|$.
[d]R_{free} is the same as R_{work}, but calculated based on 5% of the data excluded from refinement.
[e]r.m.s.d. from target geometries.

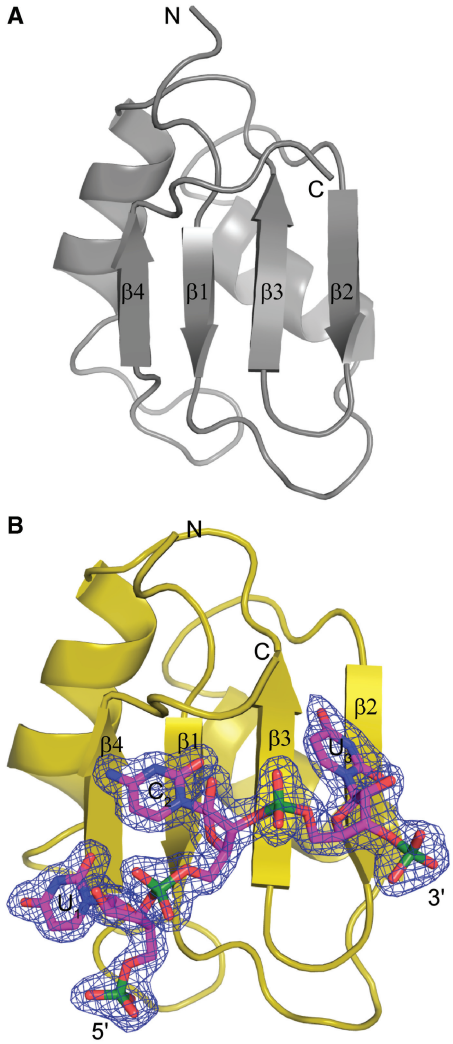


Figure 1. The Nab3-RRM of *S. cerevisiae*. (**A**) Nab3-RRM(404) (gray), (**B**) Nab3-RRM(404) (gold) bound to wt snR47 RNA (magenta). The Nab3-RRM is shown as ribbon diagram, and the bound $U_1C_2U_3$ RNA nucleotides as stick model. Electron density from a $2F_o$–F_c map contoured at 1σ for the RNA is illustrated as a blue mesh.

to β3 strand is very short and the first two residues of the ribonucleoprotein 1 (RNP 1) motif are lacking in the amino acid sequence. Position 1 of RNP 1 is often involved in RNA binding with either a lysine or an arginine residue occupying this position in other RRMs (in ~70% of RRMs), while position 2 is generally a glycine residue (30,31). In other RRMs, the positively charged position 1 residue binds to the RNA phosphodiester backbone between nucleotides bound at RNP1 and RNP2 and neutralizes the negative charge (30,31), a feature apparently missing in the Nab3-RRM. Like most RRM domains, the 3D Nab3-RRM protein structure alone does not allow one to draw any conclusions concerning RNA binding and specificity. Thus, in order to better understand the RNA recognition properties of Nab3-RRM, we aimed at obtaining the 3D structure of Nab3-RRM in complex with its cognate RNA.

Nab3-RRM specifically recognizes UCU trinucleotides

Crystals of the Nab3-RRM(404):wt snR47 RNA complex diffracted to 1.6 Å resolution and residues 329–402 of the Nab3-RRM polypeptide chain as well as an additional amino acid from the N-terminal tag could be modeled into the electron density. Overall, the model has excellent stereochemical quality with 97.3% of the residues lying in the most favored regions and 2.7% in the generally allowed regions of the Ramachandran plot. In addition to the electron density for the protein moiety, detailed residual electron density for an RNA trinucleotide was observed (Figure 1B). Based on previous reports in which the Nab3-RRM was shown to bind specifically to a UCUU sequence element (5,6) we decided to use the snR47 RNA Nab3 recognition motif (Table 1) for co-crystallization, in which only 2 nts are inserted between two consecutive UCUU sequence elements. Although the RNA used for co-crystallization contained 12 nts, only a $U_1C_2U_3$ trinucleotide was unambiguously defined in the electron density. Additionally, electron density for a bridging phosphate to the fourth uracil nucleotide was observed. Notably, further residual but poorly defined electron density protruding into the solvent was found adjacent to the 5′- and to the 3′-ends of the trinucleotide. Most likely this electron density can be attributed to disordered nucleotides from one snR47 RNA molecule that traverse the solvent channels between symmetry related Nab3-RRM(404) molecules in the crystal. However, PISA (32) did not detect any other pronounced contact interface than the RRM–RNA interaction zone, and thus two separate Nab3-RRM(404) molecules related by the crystal symmetry bind independently to the two Nab3 recognition motifs on the snR47 RNA without significant protein–protein contacts.

Comparison of the free and bound structures demonstrates that there are only subtle changes in the overall structure of the Nab3-RRM upon RNA binding (r.m.s.d. for all atoms is 1.75 Å) and most of these are confined to the loop regions surrounding the binding site, particularly β1–α1 which becomes ordered upon RNA binding, β2–β3 and α2–β4 Supplementary Figure S2). Furthermore, the C-terminal residues (400–402) are displaced by ~2.2 Å in order to accommodate the RNA. The majority of changes are in side chain conformations once bound to the RNA with the most pronounced being the RNP 1 aromatic chains of Phe366 and Phe368, which flip to accommodate binding of the nucleotide bases (Supplementary Figure S2).

The bound $U_1C_2U_3$ trinucleotide extends over the entire β-sheet of the Nab3-RRM(404) (Figure 1B) and the individual nucleotides are bound by side chains of individual β-strands (for instance U_1 is bound by side chains of the β4 strand, C_2 by β1 and U_3 by β3). Binding of the RNA is primarily driven by interactions with the bases and only a single hydrogen bond to the phosphate backbone (Ser401 side chain hydroxyl group with the C_2–U_3 bridging phosphate) is established, suggesting that the Nab3-RRM specifically recognizes the RNA sequence.

The central C_2 nucleotide of the $U_1C_2U_3$ motif is anchored in a groove lined by the side chains of Phe333

and Ser400 between which the cytosine base is sandwiched. Phe333 forms a hydrophobic stacking interaction against the cytosine base and the side chain of Ser400 closes over the base (Figure 2). Additional van der Waals interactions are formed between the aliphatic atoms of the side chain of Glu397, and the ribose is packed against the aromatic side chains of Phe366 and Phe368 from the neighboring β3-strand. By forming this binding groove for C_2, the Nab3-RRM establishes a steric constraint which allows for discrimination for pyrimidine nucleotides against purines. Moreover, Nab3-RRM forms an exquisite Watson–Crick-like hydrogen bonding interaction network guaranteeing specificity for a cytidine at this position (Figure 2). This protein mimicry of a complementary base is performed by hydrogen bonds from the side chain hydroxyl group of Ser399 to the cytosine carbonyl atom, and most importantly, by a hydrogen bond of the backbone amide NH group of Ser400 to the cytosine N3 nitrogen atom and a hydrogen bond of the cytosine amine to the backbone carbonyl group of Val398. The later interactions ensure that Nab3-RRM specifically binds a cytosine base into this groove rather than a uracil (Figure 2).

In contrast to the highly specific C_2 recognition site, binding of the 5′-adjacent U_1 nucleotide is ensured through a hydrogen bond between the NH group at position N3 of the base and the carboxylate side chain of Glu397. However, the latter is also within acceptor distance for a hydrogen bond to the O4 carbonyl function of the base. Thus, it seems plausible that Nab3-RRM(404) could also interact with a cytidine nucleotide through an interaction with the amine at position 4 of the base. Additionally, van der Waals contacts with the hydrophobic side chain of Ile395 are formed with the ring system of U_1 and its ribose forms a hydrophobic stacking interaction against the phenylic side chain of Phe333 found in the neighboring β1 strand (Figure 2). We note that this nucleotide is further contacted by a side chain amine of Lys340 from a symmetry related molecule which forms a hydrogen bond to the O4 carbonyl function of the base.

Similar to U_1, the U_3 base does not establish a full Watson–Crick like hydrogen bonding network. The base could potentially form hydrogen bonds with the side chain carbonyl of Asn361 to its N3 and with a water molecule through its O4 function (Figure 2). This water molecule is also coordinated by the carbonyl function of Asn361 and the backbone carbonyl group of Phe368. The guanidinium group of Arg331 is also in close proximity and could potentially form a hydrogen bond to the O4 carbonyl function of U_3 and thereby contribute to specificity (Figure 2). However, this side chain exists in multiple conformations in the crystal structure and the electron density for the guanidinium group is poorly defined. The U_3 base also forms a planar stacking arrangement against the aromatic side chain of Phe368. Additional interactions with the nucleotide are through van der Waals contacts from the aliphatic region of the Lys363 side chain which packs against the ribose (Figure 2) and the bridging phosphate to U_4 forms a salt bridge with Lys363 side chain (Figure 2).

A putative U_4 recognition site on the Nab3-RRM

Previous studies clearly established a $U_1C_2U_3U_4$ tetranucleotide sequence to be the recognition site for Nab3 on the nascent RNA (6) and mutation at position U_4 interfered with Nab3 binding to the RNA (6,16). We note that we observe a hydrophobic pocket formed at the base of the short β2–β3 loop with walls formed by the aromatic side chain of Phe366 and the aliphatic side chain of Lys363 in the vicinity of the bridging phosphate between U_3 and U_4. Indeed, in this region, we observe residual but ambiguous electron density which could be caused by a partially occupied or flexible U_4 base (Supplementary Figure S3). However, no significant electron density for the ribose was observed and modeling of a fourth nucleotide was not performed. Additional support for the hypothesis of weaker binding of nucleotides 3′ to the central C_2 nucleotide comes from comparison of the temperature factors of the individual bases which showed that already the U_3 nucleotide revealed increased values when compared with those obtained for U_1 or C_2. Finally, nucleotide binding by the β2–β3 loop region has been observed in other RRM–RNA complexes (33,34) as well; however, those RRMs have a longer β2–β3 loop insertion when compared to Nab3-RRM. For instance, NMR studies of PTB RRMs 3 and 4 bound to RNA showed that the base is flipped back into a hydrophobic pocket that is formed near the β2–β3 loop (35). The NMR ensembles for the flipped back bases in the PTB structure, however, clearly show that this residue is more loosely bound than the other bases (35). Thus it is likely that the electron density observed is caused by a weakly bound U_4 nucleotide. Whether the β2–β3 loop region indeed forms a weak binding site or whether additional C-terminal regions, which we lack in the Nab3-RRM(404) construct, enhance binding, can not be derived from the crystal structure.

Comparison of Nab3-RRM to other polypyrimidine binding RRMs

There are a number of RRM:RNA structures in the PDB that contain RRMs bound to polypyrimidine-containing RNAs. Thus, we compared the Nab3-RRM with other polypyrimidine binding RRMs in order to determine if there are any general features that may be important for recognition of these RNAs. An amino acid sequence alignment of Nab3-RRM with other polypyrimidine bound RRM structures (Figure 3) revealed that the PTB RRMs and Nab3 all contain a serine, Ser399 in Nab3-RRM(404), at their C-terminus, while the other RRMs do not (35–37). This serine is important for the recognition of the central cytidine nucleotide by the Nab3-RRM, and examination of all four of the PTB RRMs showed that they also engage in a similar interaction (Figure 4). Interestingly, RRMs lacking a serine at this position bind primarily to polyU sequences. Thus, this raises the possibility that a serine at this position indicates a preference for a cytidine at this position. However, it should be noted that this residue is found at the C-terminus of the protein which varies significantly from RRM to RRM (30), and the RNA

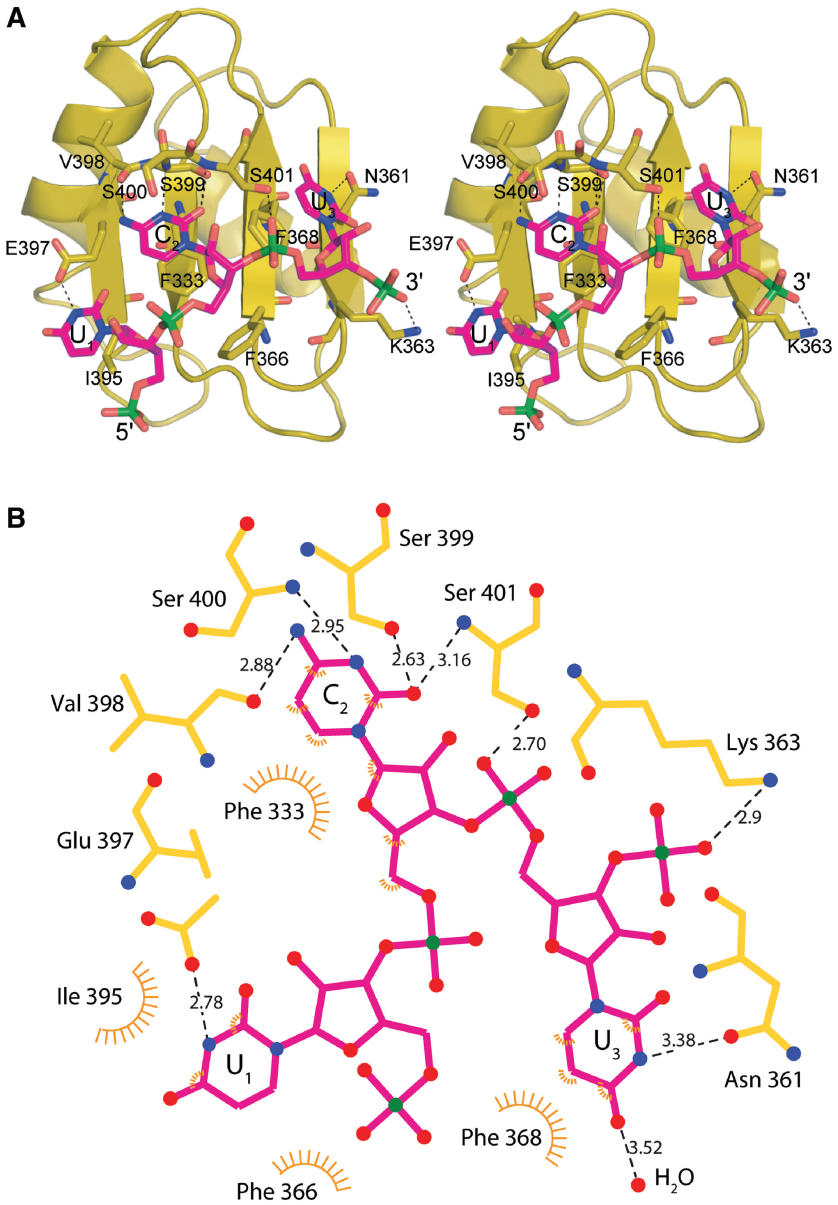


Figure 2. Details of $U_1C_2U_3$ recognition by Nab3-RRM. (**A**) Stereo view of the structure of the Nab3-RRM(404):wt snR47 RNA complex with residues important for recognition of the RNA highlighted as a stick model. Hydrogen bonding interactions are indicated as black dashed lines. (**B**) Schematic drawing of the Nab3-RRM(404) $U_1C_2U_3$ RNA interaction. Coloring according to (A) except hydrophobic interactions which are illustrated as orange semi-circles. Distances are given in Å. Note that Ser399 and Ser401 both form hydrogen bonds with their side chain hydroxyl function and not their carbonyl main chain groups.

recognition mode between RRMs can be quite different due to the plasticity of the RNA as well as differences in functional residues on the RRM surface. Thus, while it seems quite plausible that the presence of a serine at this position suggests a preference for cytidine at this position, the overall specificity will also depend on the context of the surrounding residues.

The Nab3-RRM binds RNA with low-affinity

We next asked to which extent the individual bases of a $U_1C_2U_3$ trinucleotide are important for specificity and/or binding. Thus, we performed fluorescence anisotropy titration experiments in which Nab3-RRM(404) was titrated with a series of fluorescently labeled RNAs (Table 1). Equilibrium dissociation constants (K_d) were determined using non-linear least squares fitting assuming that two Nab3-RRM(404) molecules bind independently to the two UCUU motifs of the snRNA74 RNA. Binding curves generated for these data and tabulated affinities are shown in Figure 5.

Whereas a Nab3-RRM(404) binds a wt snR47 RNA with moderate affinity with an apparent K_d of 110 μM, mutation of C_2 to U_2 (snR47U) lead to a significantly

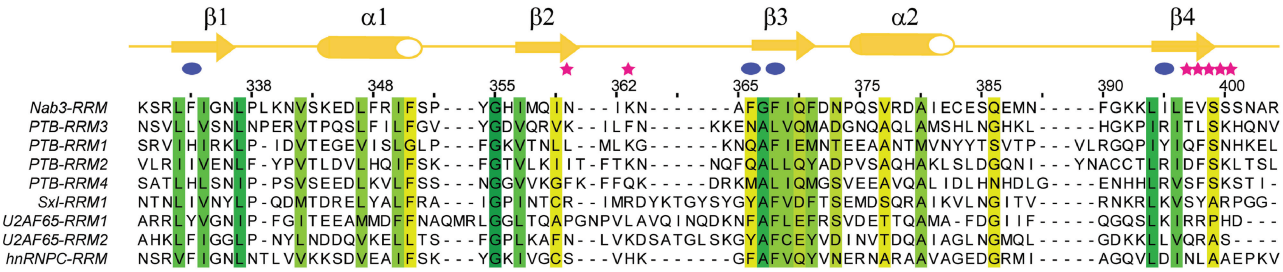


Figure 3. Comparison of Nab3-RRM to polypyrimidine binding RRMs. Amino acid sequence alignment of RRMs bound to polypyrimidine containing RNAs were aligned using ClustalW (39) and illustrated using Jalview (40). Identical residues are colored dark green and declining sequence similarity is shown in going from light green to yellow. Residues important for RNA binding in the Nab3-RRM:wt snR47 RNA structure are highlighted. Pink stars indicate residues involved in hydrogen bonds to the RNA. Blue circles indicate residues that make hydrophobic interactions.

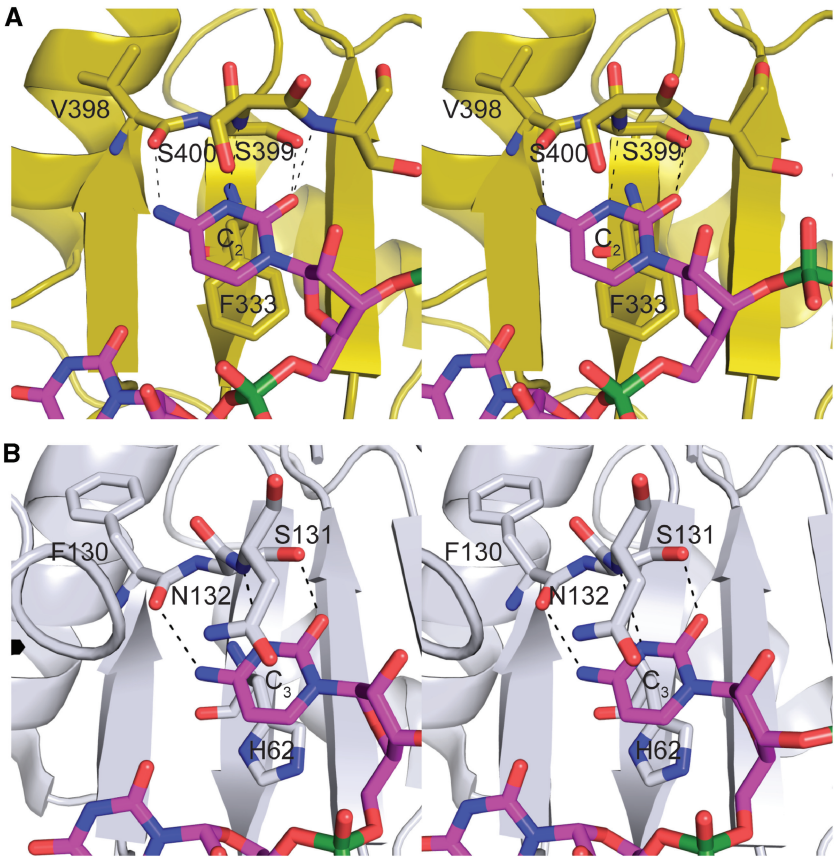


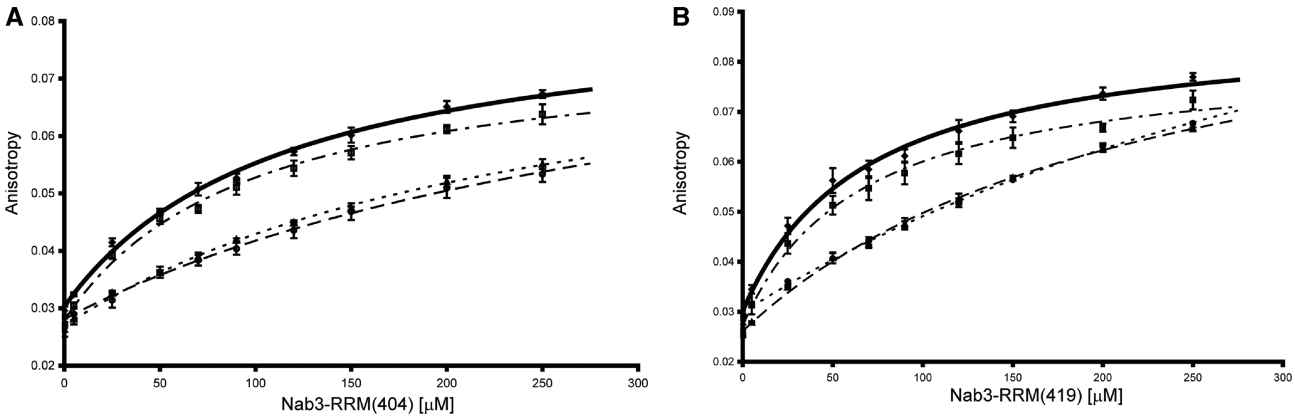
Figure 4. A common mechanism for base recognition between Nab3 and PTB RRMs. (**A**) Stereo close-up view of the cytosine base recognition of Nab3-RRM illustrated and colored similarly as in Figure 2. (**B**) Illustrated similar as in (A) but shown for the PTB-RRM1.

decreased affinity (K_d of 350 μM) (Figure 5C). When the $U_1C_2U_3$ trinucleotides were replaced by $C_1C_2C_3$ motifs (snR47C) we saw a less pronounced decrease in affinity (K_d of 250 μM), demonstrating that in addition to the mimicry of a complementary base for C_2 recognition, binding of the flanking nucleotides is important for affinity and specificity as well.

In order to verify whether Nab3-RRM(404) alone can cooperatively bind and align on an RNA containing several UCUU motifs or whether two Nab3-RRM(404) bind to snR47 RNA independently, similarly as observed in the crystal structure, we mutated one of the two UCUU motifs in the snR47 RNA (snR47one). As expected, mutation of the second UCUU motif to a weaker binding UUUU motif did not interfere significantly with binding and the apparent K_d was similar to that obtained for wt snR47 RNA (Figure 5C). In general, the non-cooperative binding we observe is in accordance with previous observations, however, the affinities we obtained deviate significantly when compared to those reported earlier (6).

Nab3-RRM core is sufficient for binding

It has been demonstrated for several RRMs that C-terminal extensions to the canonical RRM facilitate



C

RNA	Nab3-RRM (404) Affinity [μM]	Nab3-RRM (419) Affinity [μM]
FAM-wtsnR47	110 (+20/-20)	70 (+20/-10)
FAM-snR47one	100 (+20/-20)	70 (+20/-10)
FAM-snR47U	350 (+140/-90)	470 (+130/-90)
FAM-snR47C	250 (+60/-40)	230 (+40/-30)

Figure 5. Fluorescence anisotropy titration experiments for Nab3-RRM(404) (**A**) or Nab3-RRM(419) (**B**). Filled diamonds represent titration with FAM-wtsnR47, open squares with FAM-snR47one, filled triangles with FAM-snR47C and open circles with FAM-snR47U, respectively. (**C**) Best fit to the data using a non-linear least squares algorithm and equilibrium dissociation constants with confidence intervals were calculated. Numbers in parentheses denote a 95% confidence interval.

binding (30). Since the affinity observed for a Nab3-RRM(404) construct deviates significantly from previously reported results, in which longer Nab3 constructs have been used (6), we wondered whether a C-terminally extended Nab3-RRM(419) version binds tighter to RNA in our assays. However, extending the polypeptide chain of Nab3 only slightly increased affinity for a wt snR47 RNA (Figure 5C). Further, when we tested this longer Nab3-RRM(419) construct for binding to RNAs in which the UCUU motifs have been changed, the change in the K_ds is similar to those obtained for the shorter Nab3-RRM(404) version (Figure 5C). Thus, specificity and most probably affinity to the RNA is determined by the core of the RRM. This finding is further supported by the recently reported chemical shift assignment of the Nab3-RRM (38), showing that residues C-terminal to the core mainly exist in random coil conformation and a C-terminal helix as observed for other RRMs contributing to RNA binding (30) is not present. Thus, the crystal structure reported establishes the atomic details for specificity and binding of Nab3 to the RNA. Furthermore, the apparent K_d calculated from experiments when Nab3-RRM(419) was titrated into wt snR47 RNA and snR47one RNA in which the second UCUU motif was mutated are similar (Figure 5C) and thus cooperative binding facilitated by the C-terminal extension is also unlikely.

CONCLUSION

The structure of the Nab3-RRM(404):wt snR47 RNA demonstrates that specific recognition of the UCUU motif is through direct recognition of the $U_1C_2U_3$ trinucleotide (Figure 2). As observed in other RRM:RNA complexes (30), the RNA is bound across the β-sheet of the RRM with aromatic residues in the RNP1 and RNP2 consensus motifs providing a platform onto which the bases stack. While the structure provides an unambiguous explanation for binding of the central cytidine nucleotide through recognition of its Watson–Crick face (Figure 4A), specificity for the surrounding nucleotides is less defined. U_1 is recognized by Glu397 at the N3 position (Figure 2); however a cytidine at this position could perhaps also form a similar interaction with the amine of the base. U_3 recognition could potentially be achieved through an interaction of Arg331 with the O4 position of the base; however, we do not observe electron density for the guanidinium group. This suggests that this particular side chain is disordered or flexible and, if at all, this interaction is transient. Specificity for U_3 could also be achieved through interactions with a water molecule as well as the side chain carbonyl of Asn361, but similar interactions are also possible for binding a cytidine nucleotide. Finally, we observe some residual electron density suggesting the possibility of a loosely bound U_4 in a pocket formed by the β2–β3 loop, which would also be specific for a U-base (Supplementary Figure S3).

Comparison of the Nab3-RRM(404):RNA structure with other structures of polypyrimidine binding RRMs revealed PTB-RRMs to have a similar interaction with the central cytidine nucleotide as observed for Nab3-RRM (Figure 4A and B). Intriguingly, this interaction is dependent on the presence of a serine residue

at the C-terminus of the protein. Its serine hydroxyl group establishes a hydrogen bond to the O2 of the base and restricts the adjacent residues to adopt a conformation that is beneficial for a cytosine-specific pseudo-base pairing interaction (Figure 4). Our hypothesis is further supported by comparison to other polypyrimidine binding RRMs (Figure 3) demonstrating that RRMs lacking an equivalent serine residue instead bind to a uracil base at this position.

From the crystal structure we would have predicted that replacement of the central cytosine base by a uracil should be reflected in a dramatic reduction in binding affinity, at least more pronounced than observed in our binding experiments. Apparently, Nab3-RRM seems to bind to polypyrimidine containing RNAs with moderate affinity, but has a preference for a UCUU containing RNA. Our results clearly demonstrate that residues C-terminally adjacent to the RRM are not significantly involved in binding the UCUU motif and the core RRM of Nab3 is sufficient for binding and specificity. In contrast to previous studies suggesting cooperative binding of Nab3 to RNA sequences with multiple repeats (6), our results show that binding to multiple copies of UCUU motifs is not enhanced by any cooperativity; neither in the case of Nab3-RRM(404) nor in a C-terminally extended Nab3-RRM(419) version.

ACCESSION NUMBERS

The coordinates and structure factors have been deposited in the Protein Data Bank for the Nab3-RRM(404) and the Nab3-RRM(404):RNA complex with the accession numbers 2XNQ and 2XNR, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank M. Mueller for performing the MALDI-TOF experiments and H. Mutschler, J. Reinstein for advice on the fluorescence anisotropy titration experiments. We are grateful to A. Steinmetz and T. Barends for carefully reading the article and I. Schlichting for continuous encouragement and support. We thank the scientific staff at the beamline X10SA, Swiss Light Source, Paul Scherrer Institute (Villigen, Switzerland) for help with data collection and I. Vetter for support of the crystallographic software.

FUNDING

German Research Foundation (grant ME3135/1-2 to A.M.). Funding for open access charge: Max Planck Society.

Conflict of interest statement. None declared.

REFERENCES

1. Richard,P. and Manley,J.L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.*, **23**, 1247–1269.
2. Mandel,C.R., Bai,Y. and Tong,L. (2008) Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.*, **65**, 1099–1122.
3. Zhao,J., Hyman,L. and Moore,C. (1999) Formation of mRNA 3′ ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405–445.
4. Steinmetz,E.J., Conrad,N.K., Brow,D.A. and Corden,J.L. (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature*, **413**, 327–331.
5. Morlando,M., Greco,P., Dichtl,B., Fatica,A., Keller,W. and Bozzoni,I. (2002) Functional analysis of yeast snoRNA and snRNA 3′-end formation mediated by uncoupling of cleavage and polyadenylation. *Mol. Cell. Biol.*, **22**, 1379–1389.
6. Carroll,K.L., Pradhan,D.A., Granek,J.A., Clarke,N.D. and Corden,J.L. (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol. Cell. Biol.*, **24**, 6241–6252.
7. Steinmetz,E.J. and Brow,D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.*, **16**, 6993–7003.
8. Steinmetz,E.J. and Brow,D.A. (1998) Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc. Natl Acad. Sci. USA*, **95**, 6699–6704.
9. Vasiljeva,L., Kim,M., Mutschler,H., Buratowski,S. and Meinhart,A. (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, **15**, 795–804.
10. Kim,M., Vasiljeva,L., Rando,O.J., Zhelkovsky,A., Moore,C. and Buratowski,S. (2006) Distinct pathways for snoRNA and mRNA termination. *Mol. Cell*, **24**, 723–734.
11. Ursic,D., Chinchilla,K., Finkel,J.S. and Culbertson,M.R. (2004) Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res.*, **32**, 2441– 2452.
12. Gudipati,R.K., Villa,T., Boulay,J. and Libri,D. (2008) Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat. Struct. Mol. Biol.*, **15**, 786–794.
13. Arigo,J.T., Eyler,D.E., Carroll,K.L. and Corden,J.L. (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-Binding proteins Nrd1 and Nab3. *Mol. Cell*, **23**, 841–851.
14. Thiebaut,M., Kisseleva-Romanova,E., Rougemaille,M., Boulay,J. and Libri,D. (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the Nrd1-Nab3 pathway in genome surveillance. *Mol. Cell*, **23**, 853–864.
15. Conrad,N.K., Wilson,S.M., Steinmetz,E.J., Patturajan,M., Brow,D.A., Swanson,M.S. and Corden,J.L. (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics*, **154**, 557–571.
16. Carroll,K.L., Ghirlando,R., Ames,J.M. and Corden,J.L. (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA*, **13**, 361–373.
17. Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell. Bio.*, **8**, 479–490.
18. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, **26**, 795–800.
19. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

20. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
21. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
22. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
23. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.
24. Painter,J. and Merritt,E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.*, **39**, 109–111.
25. Laskowski,R.A., Macarthur,M.W., Moss,D.S. and Thornton,J.M. (1993) Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
26. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. 3rd, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
27. DeLano,W.L. (2008) *The PyMOL Molecular Graphics System*. DeLano Scientific LLC, Palo Alto, CA, USA.
28. Wallace,A.C., Laskowski,R.A. and Thornton,J.M. (1995) Ligplot-a program to generate schematic diagrams of protein ligand interactions. *Protein Eng.*, **8**, 127–134.
29. Kuzmic,P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**, 260–273.
30. Maris,C., Dominguez,C. and Allain,F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
31. Birney,E., Kumar,S. and Krainer,A.R. (1993) Analysis of the Rna-recognition motif and Rs and Rgg domains - conservation in metazoan pre-messenger-Rna splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.
32. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
33. Deo,R.C., Bonanno,J.B., Sonenberg,N. and Burley,S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, **98**, 835–845.
34. Hargous,Y., Hautbergue,G.M., Tintaru,A.M., Skrisovska,L., Golovanov,A.P., Stevenin,J., Lian,L.Y., Wilson,S.A. and Allain,F.H.T. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.*, **25**, 5126–5137.
35. Oberstrass,F.C., Auweter,S.D., Erat,M., Hargous,Y., Henning,A., Wenter,P., Reymond,L., Amir-Ahmady,B., Pitsch,S., Black,D.L. *et al.* (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054– 2057.
36. Sickmier,E.A., Frato,K.E., Shen,H.H., Paranawithana,S.R., Green,M.R. and Kielkopf,C.L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell*, **23**, 49–59.
37. Handa,N., Nureki,O., Kurimoto,K., Kim,I., Sakamoto,H., Shimura,Y., Muto,Y. and Yokoyama,S. (1999) Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature*, **398**, 579–585.
38. Pergoli,R., Kubicek,K., Hobor,F., Pasulka,J. and Stefl,R. (2010) H-1, C-13, and N-15 chemical shift assignments for the RNA recognition motif of Nab3. *Biomol. NMR Assign.*, **4**, 119–121.
39. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) Clustal-W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
40. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.