Published online 31 July 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 17* **10293–10305**
doi: 10.1093/nar/gkx685

The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition

Elsa Franco-Echevarría[1], Noelia González-Polo[2], Silvia Zorrilla[3], Santiago Martínez-Lumbreras[4,5], Clara M. Santiveri[6], Ramón Campos-Olivas[6], Mar Sánchez[2], Olga Calvo[2], Beatriz González[1,*] and José Manuel Pérez-Cañadillas[5,*]

[1]Departament of Crystallography and Structural Biology, Institute of Physical-Chemistry "Rocasolano", CSIC, C/ Serrano 119, 28006 Madrid, Spain, [2]Instituto de Biología Funcional y Genómica, IBFG-CSIC, Universidad de Salamanca, [3]Department of Cellular and Molecular Biology, Biological Research Center, CSIC, [4]Department of Chemistry, King's College London, [5]Department of Biological Physical Chemistry, Institute of Physical-Chemistry "Rocasolano", CSIC, C/ Serrano 119, 28006 Madrid, Spain and [6]Spectroscopy and Nuclear Magnetic Resonance Unit, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre

Received June 09, 2017; Revised July 18, 2017; Editorial Decision July 23, 2017; Accepted July 25, 2017

ABSTRACT

Transcription termination of non-coding RNAs is regulated in yeast by a complex of three RNA binding proteins: Nrd1, Nab3 and Sen1. Nrd1 is central in this process by interacting with Rbp1 of RNA polymerase II, Trf4 of TRAMP and GUAA/G terminator sequences. We lack structural data for the last of these binding events. We determined the structures of Nrd1 RNA binding domain and its complexes with three GUAA-containing RNAs, characterized RNA binding energetics and tested rationally designed mutants *in vivo*. The Nrd1 structure shows an RRM domain fused with a second α/β domain that we name split domain (SD), because it is formed by two non-consecutive segments at each side of the RRM. The GUAA interacts with both domains and with a pocket of water molecules, trapped between the two stacking adenines and the SD. Comprehensive binding studies demonstrate for the first time that Nrd1 has a slight preference for GUAA over GUAG and genetic and functional studies suggest that Nrd1 RNA binding domain might play further roles in non-coding RNAs transcription termination.

INTRODUCTION

Genetic information is transcribed by three RNA polymerases in eukaryotes, specialised in different types of transcripts. The RNA pol II transcribes the bulk of protein coding RNAs (mRNAs), some 'classical' non-coding genes of well-known function (snoRNAs and snRNAs), and more recently described non-coding transcripts of less known function (i.e. CUTs, SUTs, etc.) (1). The biosynthesis of all these transcripts is a tightly regulated process that is coordinated with other events of RNA metabolism like nuclear export and quality control/degradation. Transcription termination represents the final step and is performed through two different pathways in budding yeast, dependent on the nature of the transcript (for recent reviews, see (2–6)): the cleavage and polyadenylation pathway, performed by multiprotein factor CPF, processes the mRNAs by cleaving the transcript in the 3′-UTR and adding a poly(A) tail; and the Nrd1–Nab3–Sen1 pathway (NNS) deals with the termination of functional non-coding RNAs (snoRNAs and snRNAs) (7) and promotes the degradation of products arising from Pol II pervasive transcription (e.g. CUTs) (8). This second pathway appears to be less conserved along the eukaryote kingdom than the polyadenylation one and involves three RNA-binding proteins (Nrd1, Nab3 and Sen1) that couple RNA processing and degradation. Nab3 (9) and Nrd1 (10) have been extensively studied and contain RRM domains with sequence specificity for UCUUG (7,11–13) and GUAA/G (10,12,13) respectively. These elements are frequent in snoRNA terminator sequences and their simultaneous recognition is surely boosted by Nrd1-Nab3 heterodimerization (14). In addition, Nrd1 N-terminal CTD interaction domain (CID) interacts specifically with phospho-Ser5 (pS5) repeats of C-terminal domain (CTD) of RNA Pol II Rpb1 subunit (15,16) and with Trf4 from TRAMP complex (17,18) using mutually exclusive interfaces. To further increase this sophisticated landscape of interactions, Nab3 contains low complexity regions and potential oligomerization domains for autoassociation (19).

Currently, there are several structural studies that shed light into this complex network of biomolecular interac-

*To whom correspondence should be addressed. Tel: +34 91 561 94 00; Fax: +34 91 564 24 31; Email: jmperez@iqfr.csic.es
Correspondence may also be addressed to Beatriz González. Tel: +34 91 561 94 00; Fax: +34 91 564 24 31; Email: xbeatriz@iqfr.csic.es

tions in the NNS complex. The NMR and X-ray structures of Nrd1 CID in complex with pS5 (16) and with Trf4 peptides (18) provide key insights into Nrd1 recruitment to early elongation complexes and the incompatibility between this binding event and TRAMP recognition (thought to occur later). The structure of Nrd1 CID-CTD (pS5) complements those of Pcf11 CID-CTD(pS2) (20) and Rtt103 CID-CTD(pS2) (21) and gives mechanistic clues about how these budding yeast Pol II CTD code readers are recruited to the elongation complex at different stages. Conversely, the protein-RNA recognition in NNS has been less studied at high-resolution. Complexes between Nab3 RRM and UCUU have been solved by X-ray crystallography (22) and NMR (23), but the recognition details of the higher affinity sites (UCUUG) (11) remain unknown. The structure of Nrd1 RNA binding domain has been studied by NMR though only the fold for the RRM counterpart has been determined. (24). In the same work, authors proposed that Nrd1–RNA recognition is semi-specific.

Here, we report the X-ray and NMR structures of Nrd1 RBD. The structures show an unusual fold in which the RRM-flanking sequences define a second well-folded α/β domain (in contrast with previous data (24)). More importantly, we present the X-ray structures of Nrd1 RBD in complex with several RNA sequences containing the GUAA motif, which reveal a unique binding mode that involves both domains. These data, together with our ITC and fluorescence anisotropy binding affinity studies, provide the first explanation at atomic level for Nrd1 specific recognition of GUAA and GUAG previously seen both *in vitro* (10,12,13) and *in vivo* (25,26). We use this detailed structural and biophysical information to rationally design a battery of mutants and study their changes in binding affinity, growth defect phenotypes and snRN13 transcription termination defects *in vivo*. Our work provides high-resolution key structure-function knowledge to progress in our understanding of the mechanism of transcription termination through the NNS pathway.

MATERIALS AND METHODS

Cloning, protein expression and purification

Plasmids used in this work are summarised in Supplementary Table S1. Nrd1 sequences were amplified (KOD polymerase; Novagen) from *Saccharomyces cerevisiae* genomic DNA and cloned in a pET28-modified vector containing an N-terminal fusion cassette (thioredoxin A+6xHis+TEV sequence) ((27) for details). Mutants were obtained with a QuikChange Lightning Kit (Agilent genomics), both in the pET28 and pRS415 plasmid backgrounds using specific DNA oligos (IDT and Macrogen).

Samples of Nrd1 wild type and mutant proteins used in structural and biophysical studies were produced by overexpression in *Escherichia coli* BL21(DE3) cells (Novagen). Cultures were grown at 37°C in LB with 30 μg/l of kanamycin (Sigma-Aldrich). Upon reaching OD_{600nm} = 0.6–0.8, cells were transferred to 12°C and induced with 0.5 μM IPTG (Sigma-Aldrich) during 72–96 h. We noticed that expression at higher temperatures result in production of misfolded protein. Uniformly (^{15}N and/or ^{13}C) or selectively unlabeled samples (see specific details in Supplementary Figure S1) were produced in a similar way in chemically defined media ((27) for details). Selenomethionine substituted $Nrd1_{301–489}$ was expressed similarly, just using minima media supplemented with a mix of inhibition aminoacids and Selenomethionine.

For protein purification, cell pellets were resuspended in buffer A (25 mM Potassium Phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol and 1 tablet/50 ml of protease inhibitors (Roche)), lysed by sonication and cleared by ultracentrifugation. The supernatant was filtered (0.22 μm) and loaded in a HisTrap™ HP 5 ml column (GE Healthcare), washed with buffer B (as buffer A but with 500 mM NaCl and 30 mM Imidazole) and eluted with buffer C (as buffer A but with 300 mM imidazole). The samples were exchanged to buffer A and cleaved overnight at 16°C with homemade TEV protease (100 μg/ml). After complete cleavage, the sample was reloaded on the HisTrap column, further washed with buffer A and selectively eluted with buffer B (TEV protease, cleaved fusion and traces of uncleaved product remain bound to the column). Nrd1 samples were finally polished by gel filtration (Superdex 200 preparative, GE Healthcare) equilibrated in 20 mM Tris–HCl pH 8, 200 mM NaCl and 0.1 mM DTT. Purified Nrd1 samples were concentrated and/or buffer exchanged according to their posterior use.

Protein crystallization

Crystallization experiments with $Nrd1_{301–489}$ and $Nrd1_{290–468}$ constructs were performed at 291K using the sitting-drop vapor-diffusion method and Hampton Research and Quiagen commercial screens using 96-well plates (Innovaplate SD-2 microplates, Innovadyne Technologies Inc). The crystallization conditions were scaled up and refined in 24-well/48-well plates (Hampton Research). Initially, $Nrd1_{301–489}$ (23 mg ml^{-1}) crystallized in 15–18% (v/v) PEG 8000 and 0.1 M Bicine pH 9.3. Selenomethionine (SeMet) modified $Nrd1_{301–489}$ crystals were obtained in very similar conditions consisting of 12% (v/v) PEG 8000 and 0.1 M Bicine pH 9.0. Best condition obtained for a shorter construct, $Nrd1_{290–468}$ (30 mg ml^{-1}), is 0.2 M potassium thiocyanate, 23% PEG 3350 (v/v). A microseeding technique (28) was necessary to improve the quality of these crystals. Apart from these conditions, crystals were obtained in a variety of conditions including a broad range of pHs and using a protein concentration ranging from 20 to 30 mg ml^{-1}. Nrd1 crystals appeared from one to seven days after setting up the crystallization trials.

We tried to prepare crystal complexes between Nrd1 and RNAs with different lengths (4, 5, 6, 10 and 36 bases) by co-crystallization or soaking using all variety of crystals obtained. We had success only with soaking experiments using crystals of $Nrd1_{290–468}$ (24 mg ml^{-1}) grown in 1 M sodium potassium phosphate pH 7.4. The RNA samples were dissolved in the precipitant solution to a final concentration of 4 mM. $Nrd1_{290–468}$ complexes with GUAA, CGUAAA and UUAGUAAUCC RNAs (IBA and IDT) were obtained by slowly adding of 1 μl of the RNA solution to the crystals drops followed by overnight (GUAA) or three hours (CGUAAA and UUAGUAAUCC) incuba-

tion. We only could obtain a single crystal in complex with the longer RNA, since crystals were quickly crashed after its addition. For complex with CGUAAA, Nrd1 $_{290–468}$ was dialyzed in 20 mM sodium potassium phosphate pH 8, 200 mM NaCl, 1 mM DTT buffer to remove the tris molecule found in the structure.

Data collection and processing

Prior to data collection, all crystals were transferred for a few seconds to the crystallization solution plus 20–25% (v/v) ethylene glycol or 25%(v/v) glucose and then flash cooled in liquid nitrogen. Diffraction data for Nrd1$_{301–489}$ crystals were collected at the European Synchrotron Radiation Facility (ESRF) (beamline ID23-1) (Grenoble, France), while data sets for the SeMet Nrd1$_{301–489}$ and Nrd1 $_{290–468}$ (apo and soaked with RNA) were collected in ALBA (beamline BL13-XALOC) synchrotron facilities at −173°C (Supplementary Table S2). Data processing was performed with XDS package (29) and merging with Aimless (30) from CCP4 suite (Collaborative Computational Project, Number 4, 1994). All constructs used crystallized in the tetragonal space group $P4_32_12$, whereas the shorter construct Nrd1$_{290–468}$ also crystallized in the hexagonal group $P6_5$, in all cases with one molecule in the asymmetric unit (Supplementary Table S2). The structure of Nrd1$_{301–489}$ was solved by single-wavelength anomalous dispersion (SAD) technique from the SeMet-Nrd1$_{301–489}$ dataset collected at the Se fluorescence peak wavelength. Two of three expected selenium positions were located using SHELX (31). Buccaneer was used for preliminary model building (32). Subsequently the structure of non-substituted Nrd1$_{301–489}$ was obtained at 2.3 Å resolution by difference Fourier synthesis using the experimental model above. The shorter construct Nrd1$_{290–468}$ hexagonal crystals diffracted to 1.6 Å resolution and its structure was obtained by molecular replacement with MOLREP (33) using Nrd1 $_{301–489}$ refined model as a template. Finally, the Nrd1$_{290–468}$ tetragonal crystals allowed us to get three different RNA-Nrd1 complexes (Supplementary Table S2), which diffracted to 2.45 Å maximum resolution in the best case (GUAA complex). The structure was solved by molecular replacement using the coordinates of Nrd1$_{290–468}$ as a search model. The RNAs were manually built into the electron density maps using COOT (34). Model refinement in all cases was performed by alternating cycles of automatic refinement with REFMAC (35) and manual building with COOT. Statistics for all data processing and refinement are summarized in Supplementary Table S2. The electron density maps allowed building of all chain except for some N- and C-terminal residues (Supplementary Table S2). The stereochemistry has been checked with PROCHECK (36). The figures of the models were generated with PYMOL (37).

NMR

Nrd1 samples (100–800 μM) were prepared in NMR buffer (25 mM Potassium Phosphate pH 6.5, 25 mM NaCl, 1 mM DTT and 10% D_2O) and experiments acquired at 25°C on cryoprobe-equipped Bruker AV800 MHz spectrometer. Resonance assignments (^{1}H, ^{15}N and ^{13}C) were obtained with a battery of triple and double resonance 3D experiments (HNCA, HNCO, CBCA(CO) NH, HCCH-TOCSY) (38,39) that were processed with NMRPipe (40) and analysed with CcpNmr Analysis (41).

The Nrd1$_{290–468}$ NMR structure was calculated from NOE-derived distance restraints (Supplementary Tables S3, S4 and Supplementary Figure S2) obtained from 2D NOESY and 3D ^{13}C/^{15}N-HSQC-NOESY experiments and from a set of 2D ^{13}C–^{15}N (F1-filtered) NOESY and 2D ^{13}C–^{15}N (F1/F2-doublefiltered) (42) NOESY experiments recorded on samples with amino acid selective reverse unlabeling (Supplementary Figures S1 and S2C). Backbone dihedral angle restrains (ϕ and ψ) were obtained with TALOS+ (43) from Cα, Cβ, CO and NH chemical shifts. Structures were calculated with CYANA 2.1 (44), starting from 50 randomly-generated conformers, using a restrained simulated annealing protocol. The final 20 conformers with lower target function and having no distance (>0.2 Å) and angle (> 5°) violations were selected and subjected to energy minimization with AMBER

Nrd1$_{290–468}$-GUAA complex formation was followed by ^{1}H–^{15}N-HSQCs by progressively adding the RNA (IBA), in the same NMR buffer as free protein, until reaching slight excess of (1:1.2). A series of 3D HNCO, 3D HNCA and 3D CBCA(CO)HN were taken to assign Nrd1 bound state. The chemical shift perturbations (CPS) were calculated for each NH peak with the equation: $\Delta\partial^{av} = (1/2\bullet((\Delta\partial_H)^2+(0.2\bullet\Delta\partial_N)^2))^{1/2}$.

Fluorescence anisotropy

Measurements were performed in a BMG Polarstar Galaxy plate reader essentially as described in (27). The temperature was 26°C, the concentration of the fluorescein labeled oligonucleotides (IDT) was 40 nM, and the buffer was 20 mM Tris–HCl, 150 mM NaCl, 1 mM DTT pH 8.0. A 1:1 binding model compatible with the experimental data was fitted to the isotherms using BIOEQS software as previously described (27). Errors in the fitting parameters were obtained by confidence limit testing, using the same software, at the 67% confidence level. The intensity of the emission of the fluorescein dye remained essentially unchanged in the presence of the protein.

Isothermal titration calorimetry

ITC experiments were performed at 25°C using a MicroCal iTC200 (Malvern Instruments, UK) calorimeter. Protein and RNA samples were prepared in 20 mM Potassium Phosphate buffer pH 7.0, 150 mM NaCl and 2 mM β-mercaptoethanol. CCGUAACC (230 μM) and CCGUAGCC (470 μM) RNAs (IDT) were titrated into Nrd1$_{290–468}$ at 24–25 and 46–48 μM concentration, respectively, placed in the 200 μl sample cell. The reference cell was filled with distilled water. Titration experiments consisted of 19 injections of 2 μl (with a first injection of 0.4 μl) separated by 150 s to allow thermal power to return to baseline. For homogeneous mixing in the cell, the stirring speed was 1000 rpm. Data were analysed with Origin 7.0 (OriginLab) using a one-site binding model. The experiments were carried out in duplicate. Control experiments of dilution of the

RNAs into buffer were performed and the resulting heats were subtracted from the protein-RNA titration data.

Yeast strains

Plasmids and strains used are listed in Supplementary Table S1. Strains construction and other genetic manipulations were performed following standard procedures (45). Briefly, we used the EJS101-9d strain (10) to construct all *nrd1* mutants analysed in this study. In this strain, *NRD1* has been replaced by the *HIS3* marker, and a genomic wild-type copy of *NRD1* is expressed from a centromeric *URA* plasmid (pRS316). We transformed EJS101–9d strain with the centromeric *LEU* plasmid pRS415, containing either wild-type (pRS415-*NRD1*) (15) or *nrd1* mutant genes. Then, the transformant were selected in –URA-LEU selective media, and thereafter grown in 5-FOA containing media, forcing pRS316-*NRD1* loss, and expression of *NRD1 wt* and *nrd1* mutant versions (obtained as described above) from the *LEU* plasmid.

In the case of *nrd1* mutations that caused lethality, we carried out another approach to confirm it. We used then the DLY883 strain (8), where *NRD1* is expressed under the control of a GAL promoter. Thus, in glucose containing media, *NRD1* expression is abolished. We transformed DLY883 with an empty *LEU* plasmid (pRS415) or with the above mentioned plasmids, bearing *NRD1 wt* or *nrd1* gene mutants. Yeast transformants were selected in 2% GAL–LEU media and then transferred to a 4% glucose–LEU media to turn off Nrd1 expression.

The EJS101-9d and DLY883 strains, as well as the pRS415-*NRD1* plasmid used as template to generate all *nrd1* mutated genes, were kindly provided by S. Buratowski.

Northern blot analysis

Total RNA was isolated using Qiagen Purification Kit from cells grown at 28°C and after a shift to 37°C for 1 h. Approximately 25 μg of total RNA was loaded onto a 1.5% MOPS-formaldehyde agarose gel, and then RNA was transferred onto nylon membranes. Prehybridization and hybridization were performed in a solution containing 1% SDS, 1 M NaCl and 10% dextran sulphate at 60°C. In the case of the prehybridization, it was carried out for at least 2 h, and in the case of the hybridization, the solution contained in addition the radiolabeled probe (*SNR13* or *ADH1*) and it was carried out overnight. Thereafter, the membrane was twice washed with 2× SSC, 1% SDS at 60°C for 5 and 30 min, respectively and once with 0.1× SSC for 1 h at room temperature. After washes, the membranes were exposed to X-ray film.

RESULTS

Nrd1 RBD forms a novel structure with two domains

The sequence alignment along Nrd1 orthologues (Supplementary Figure S3) shows high conservation in regions flanking the canonical RRM domain (339–410). The ^{1}H–^{15}N HSQC spectra of Nrd1$_{301–489}$ and Nrd1$_{290–468}$ constructs are very similar, with differences in the N- and C-terminal regions (Supplementary Figure S4). The ^{13}C conformational shifts (Supplementary Figure S5) revealed regular secondary structure elements (β-sheets and α-helices) outside the RRM. In the Nrd1$_{301–489}$ the C-terminal segment 468–489 is disordered (random coil ^{13}C chemical shifts and sharp linewidths; Supplementary Figures S4 and S5). Given these observations we decided to setup crystallization experiments with constructs including and lacking this fragment.

Nrd1$_{301–489}$ and Nrd1$_{290–468}$ crystallized under different conditions and two different space groups, $P4_32_12$ and $P6_5$ (Supplementary Table S2). The structures of the two constructs are nearly identical with remarkable differences at the N-terminus (before Asp308) and the C-terminus (after Ile463). The Nrd1 RBD adopts a dumbbell-like structure with two α/β domains occupying the positions of the weights (Figure 1A). The RRM domain (residues 339 to 407) shows the classic α/β topology (Figure 1B). The second domain is built up from two discontinuous segments (304–338, 408–464) that flank the RRM (Figure 1A&B). Topologically, its structure is accurately described as an insertion of the RRM into one of the loops of this second domain. For simplicity, we will refer to this second domain as 'split domain' (SD). The SD architecture is different to that of the RRM. It contains a mixed 5-stranded β-sheet with the two central strands in parallel orientation, and an α-helix that runs nearly perpendicular to the β-sheet (Figure 1B). Among the RRM structures having N- or C-terminal extensions, the structure of the Nrd1 RBD represent an utterly new configuration. While we were preparing this manuscript the structure of Seb1 RBD (Nrd1 homolog in *S. pombe*) was published (46) showing an equivalent fold to our Nrd1 structure.

The crystal structure of the Nrd1$_{301–489}$ construct displays a longer C-terminus (464–471) than the structures of the Nrd1$_{290–468}$ construct (Supplementary Figure S6). These residues fold back and loosely interact with the SD, in agreement with their high temperature factors and sharp NMR signals. The two crystal structures have different conformation at the N-terminus. In the Nrd1$_{290–468}$ construct Leu305 interacts with Trp353 and the polypeptide chain can be traced up to Asp302 (Supplementary Figure S6A) whereas for the construct Nrd1$_{301–489}$ the electron density is lost before His304 and Trp353 shows a flipped out conformation (Supplementary Figure S6B).

We also obtained the NMR structure of Nrd1$_{290–468}$ in solution (Figure 1C and Supplementary Table S3) that is essentially equivalent to the crystal one (Supplementary Figure S7A) but reveals great differences from the previously published NMR structure (24), which shows an RRM with very similar fold and an SD with no tertiary fold (Supplementary Figure S7B). We think that this discrepancy probably responds to differences in constructs used or in protein expression protocols. The conformation of the N-terminus (Supplementary Figure S6C) shows the Leu305/Trp353 interaction seen in the crystal structure of the equivalent construct (Supplementary Figure S6A), and an additional contact with Phe298 (disordered in the crystal). The Trp353 Nε1–Hε1 signal in the ^{1}H–^{15}N HSQC spectrum of Nrd1$_{290–468}$ is duplicated, a minor peak of 20% population, evidencing conformational heterogeneity. The position of this minor crosspeak coincides with the Trp353 Nε1-Hε1 peak in the Nrd1$_{301–489}$ spectrum (Supplementary Figure S4). Since this construct lacks Phe298, we interpret

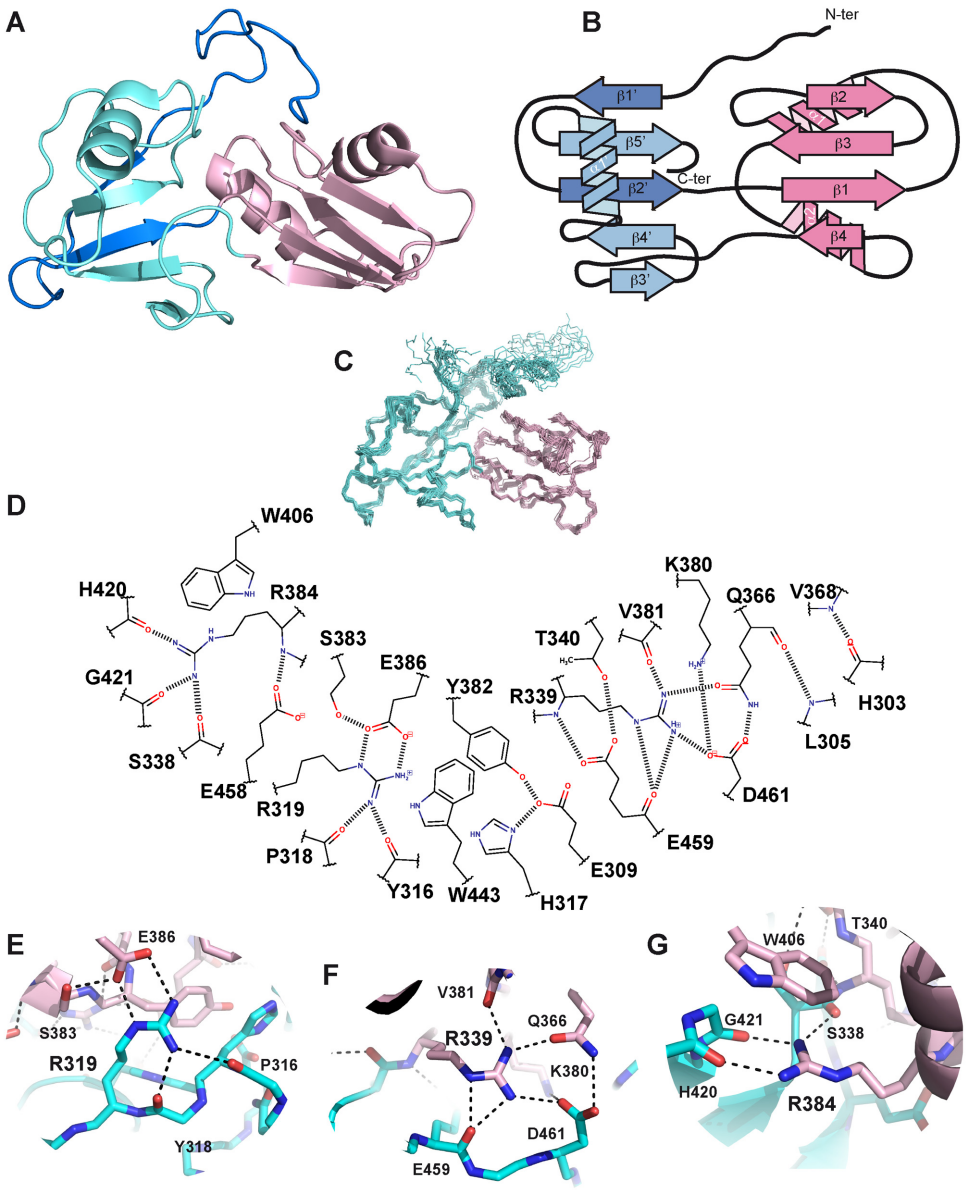

Figure 1. Structures of Nrd1 RNA binding domain. (**A**) Cartoon representation showing the RRM domain in pink and the SD in blue (different blue color is used for fragments non consecutive in sequence). (**B**) Schematic representation of the protein topology showing the elements of secondary structure. (**C**) Superposition of the NMR ensemble of the $Nrd1_{290–468}$. (**D**) Scheme of interactions formed at the RRM–SD interface. Hydrogen bonds are shown in dashed lines. (**E**) Structural detail around the Arg319 cluster. Residues are shown as sticks and colored in the same code as in A. Hydrogen bonds are shown in dashed lines. (**F**) Structural detail around the Arg339 cluster. (**G**) Structural detail around the Arg384 cluster.

that the heterogeneity in the longer construct might arise from slow exchange equilibrium of Phe298 coming in and out of the Leu305/Trp353 site. This hypothesis is compatible with the NOE data (Supplementary Figure S6D).

Altogether the X-ray and NMR data for Nrd1 RBD show the coexistence of a well-folded two-domain core with N-terminal and C-terminal tails that can interact loosely or in multiple ways with it.

Nrd1 RRM and SD domains interact through a conserved polar interface

The SD domain contains long loops with non-regular structure elements that interact with the upper loops of RRM defining a large convex interface (Supplementary Figure S8A). This intramolecular interface is predominantly polar and maintained by an intricate network around three arginine residues (Arg319, Arg339 and Arg384; Figure 1D–G). On one side, Arg319 (partially exposed) forms a salt-bridge with Glu386 and hydrogen bonds with backbone carbonyls that freeze the orientation of the guanidium group (Figure 1E). At the most buried region of the interface, Arg339 coordinates an even more complex network of hydrogen bonds/salt bridges with Glu459, Asp461 and Gln366 (Figure 1F). A second basic residue (Lys380) takes part on this network and is probably important to keep this buried cluster neutral. A layer of aromatics (Trp443, Tyr382, Tyr306

and Trp353) separates the Arg339 and Arg319 clusters (Supplementary Figure S8B). Finally, Arg384 (partially exposed) closes the opposite end of the interface by a similar hydrogen bond network with backbone carbonyls of domain SD (Figure 1G). Despite its polar character, the SD–RRM interface is dry; only two water molecules (with low solvent accessibility and temperature factors) take part on it (Supplementary Figure S8C).

Nrd1 recognizes GUAA specifically

RRM domains typically recognize RNA using exposed residues in the β-sheet (47–49), a potential interface that is accessible in the structure of the Nrd1 RBD. Classical studies (10,12,13) and recent genome-wide CLIP maps (25,26) proposed Nrd1 selectivity for the sequences GUAA and GUAG. We have determined the binding affinities by ITC and fluorescence anisotropy showing that Nrd1$_{290–468}$ recognizes GUAA with slightly higher affinity than GUAG and both of them with significantly larger affinity than the AUAA, AUAU and AUUA probes (Figure 2 and Table 1). The relatively high affinity of the Nrd1:GUAA interaction suggests that it might involve further residues than the canonical ones in RRMs, because single RRMs typically bind weaker than 2 μM and without sequence specificity. To unravel this issue we attempted to obtain the structures of protein-RNA complexes by soaking and co-crystallization with short RNAs. Only the soaking method, and using tetragonal crystals from the Nrd1$_{290–468}$ construct, was successful and we obtained three complex structures of Nrd1 with GUAA, CGUAAA and UUAGUAAUCC RNAs (Figure 3). The three sequences contain the GUAA motif, found in the majority of transcripts processed through the NNS pathway (12,25,26), and the last one derives from the *SNR13* terminator, one of the best characterized NNS targets (13). The core GUAA shows the same conformation in the three Nrd1:RNA complexes and there was enough electron density to build extra bases for the longer RNAs (Figure 3C). However these additional bases have higher temperature factors than the GUAA core and no evident contacts with the protein, thus we reasoned that they are loosely stabilized by intra RNA contacts. The Nrd1$_{290–468}$:GUAA crystal structure is consistent with NMR chemical shift perturbation data in solution, which map the RNA binding pocket to the depression between RRM and SD domains (Figure 3B). In conclusion, the structures suggest that Nrd1 RBD mainly recognises the core GUAA.

The Nrd1$_{290–468}$:GUAA structure, with the highest resolution among the three complexes (Supplementary Table S2), shows that both RRM and SD domains contribute to GUAA recognition (Figure 3). The RNA interacts with the RRM domain with its backbone running in the classical 5′ to 3′ orientation in all the RRM-RNA complexes reported to date. A series of hydrogen bonds (G1 O2′-His376 Nδ2, U2 O2′-A3 OP1 and A3 O2′-A4 O5′) explain ribose specificity; in the case of U2, an unusual 2′-endo configuration for the sugar puckering favors this ribose-specific recognition (Figure 3D,F). The RRM domain interacts with the first three nucleotides: U2 (Figure 3F) and A3 (Figure 3G) form archetypal planar stacking interactions with Phe342 and Phe378 respectively, whereas G1 stacks to the edge of Phe342 ring (Figure 3E). SD Residues Ile369 and Tyr418 make additional stacking interactions with A3 and U2, respectively. The fourth base (A4) forms a planar stacking with A3 (Figure 3H), which together with Phe378 define a three-layer aromatic array (Figure 3D&H). Residues from SD (His303, Ile462 and Val464) make further contacts with A4 (Figure 3H).

RNA base specificity is achieved by direct and water-mediated hydrogen bonds to the protein. Positions 3 and 4 are specified by direct hydrogen bonds between the N6 positions of the adenines and the backbone carbonyls of Val408 and Gly409 (Figure 3G and H). A4 N7 position (purine over pyrimidine selectivity) is specifically recognized by Arg413 and its mutation lowers affinity ∼10-fold (Table 1), reinforcing the structural role of this conserved residue (Supplementary Figure S3) in purine recognition. A3 N7 position is recognized by a water molecule that contacts simultaneously to U2 O2. The remaining acceptors/donors of A3 and A4 are satisfied by interactions with four water molecules that are trapped between the RNA and the protein (Figure 3G and H). These waters are well structured, according to their temperature factors, but are not completely buried by the interaction; a narrow channel on the back of this water pocket communicates it with the bulk solvent (Supplementary Figure S9). U2 position is specified by the U2 N3-Trp406 O and by the U2 O2-water hydrogen bonds (Figure 3F). The conserved Tyr418 interacts with U2 and its mutation decreases RNA affinity (Table 1). Finally, two hydrogen bonds (G1 N2-U2 OP1 and G1 N7-Gly345 NH) specify guanine at position 1 (Figure 3E). There is a third base-specific contact with a Tris molecule (from the buffer) that bridges G1 with the Arg403/Arg405 pair. The Nrd1$_{290–468}$ crystal structure in complex with CGUAAA, which was solved in phosphate buffer (no Tris present), showed that only Arg403 makes base-specific contacts (Arg403 Nϵ-G1 O6) (Supplementary Figure S10). This result is consistent with mutational data on the equivalent residue in Seb1 which lowers RNA affinity dramatically (46). To enquire about the role of Arg405 in RNA recognition we measured energetics of binding of the R405A mutant finding slightly higher affinity than for *wt* protein (Table 1). Therefore, we confirm Arg403 as the one involved in G1 O6 recognition. To further confirm the guanine requirement at position 1, we tested substitution by other purines. Substantial affinity losses were observed upon inosine (∼15-fold: loss of N3 group) and adenine (>40-fold: full remodeling of the Watson–Crick face) substitutions (Table 1 and Figure 2C).

Structural integrity of the RRM–SD tandem is key for RNA recognition

RNA-binding causes little structural changes on Nrd1 (Supplementary Figure S11), suggesting that the RNA binding site is structurally preconfigured. The X-ray structures of the different protein-RNA complexes reveal the key elements for the RNA recognition. We made a thorough mutagenesis analysis to gain insight into different contributions to RNA binding energetics.

R374A, R413G and Y418A show a rather homogeneous decrease in affinity (15-fold K_D increase) (Table 1),

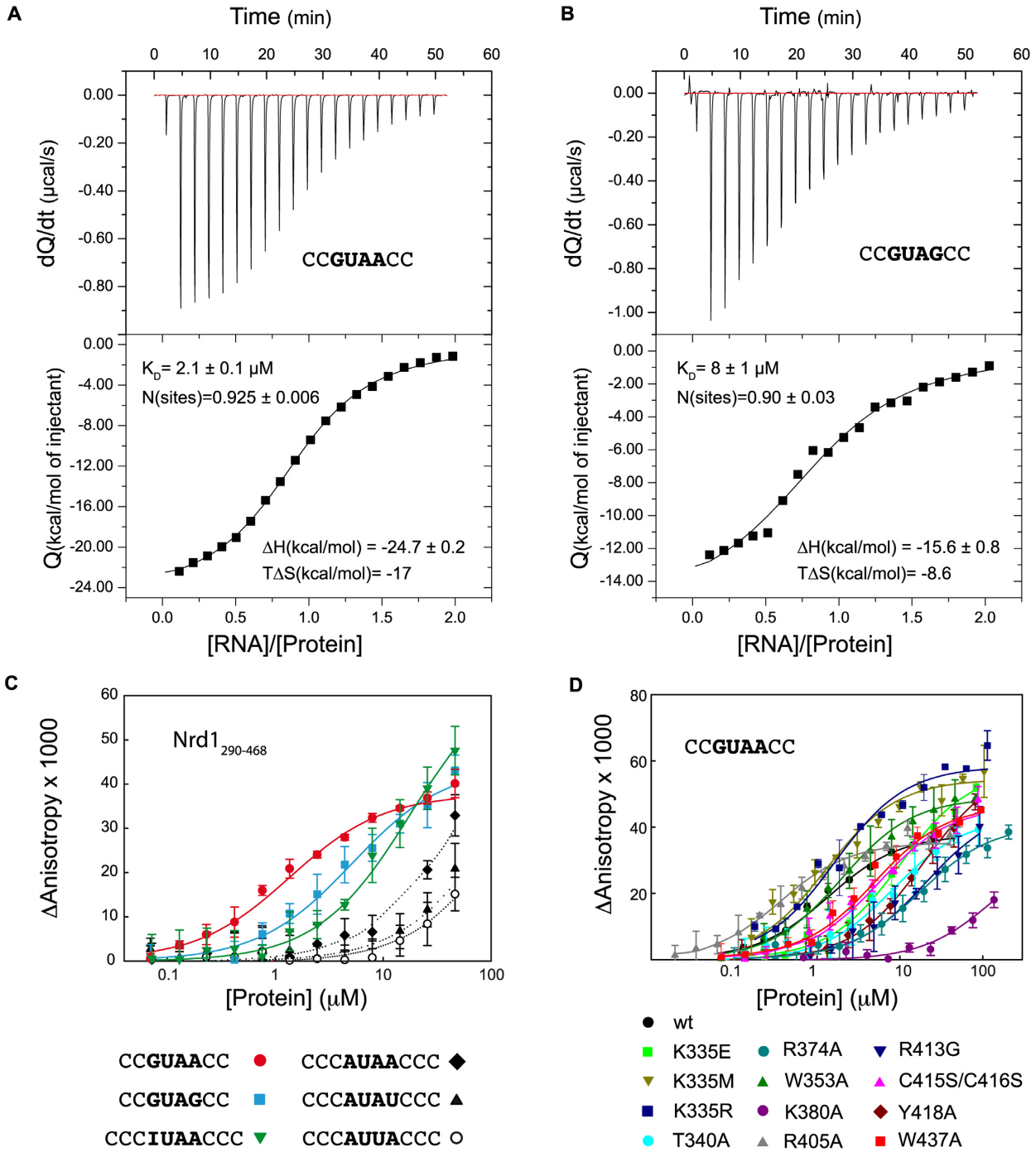

Figure 2. Biophysical analysis of RNA recognition by Nrd1 RBD. (**A** and **B**) Isothermal titration calorimetry thermograms (upper panels) and binding isotherms (lower panels) obtained for titrations of CCGUAACC and CCGUAGCC RNAs over $Nrd1_{290–468}$. Enthalpic and entropic contributions to binding, K_D ($1/K_B$) and stoichiometry (N) are indicated on each curve. (**C**) Fluorescence anisotropy $Nrd1_{290–468}$ titration curves over different RNAs (I = ionosine). (**D**) Fluorescence anisotropy CCGUAACC titration curves over various $Nrd1_{290–468}$ mutants. Fitted values are summarized in Table 1.

presumably due to loss of protein–RNA contacts (hydrogen bonds: R413G, hydrophobic: Y418A and columbic: R374A). We also performed mutations in the second coordination sphere (stabilizes residues in direct contact with RNA): T340A, W437A and C415S/C416S, exhibiting more modest effects (Table 1). Next we altered the RRM–SD interface with different results: W353A mutant, that should affect the interaction of the N-terminus with the RRM, has a very little impact on GUAA recognition. In stark contrast, K380A, at the heart of the RRM–SD interface, lowers RNA affinity dramatically, likely due to a disruption of the interdomain arrangement. Finally, we mutated the conserved residue Lys335 (Supplementary Figure S3), placed at the centre of the SD β-sheet and far away from

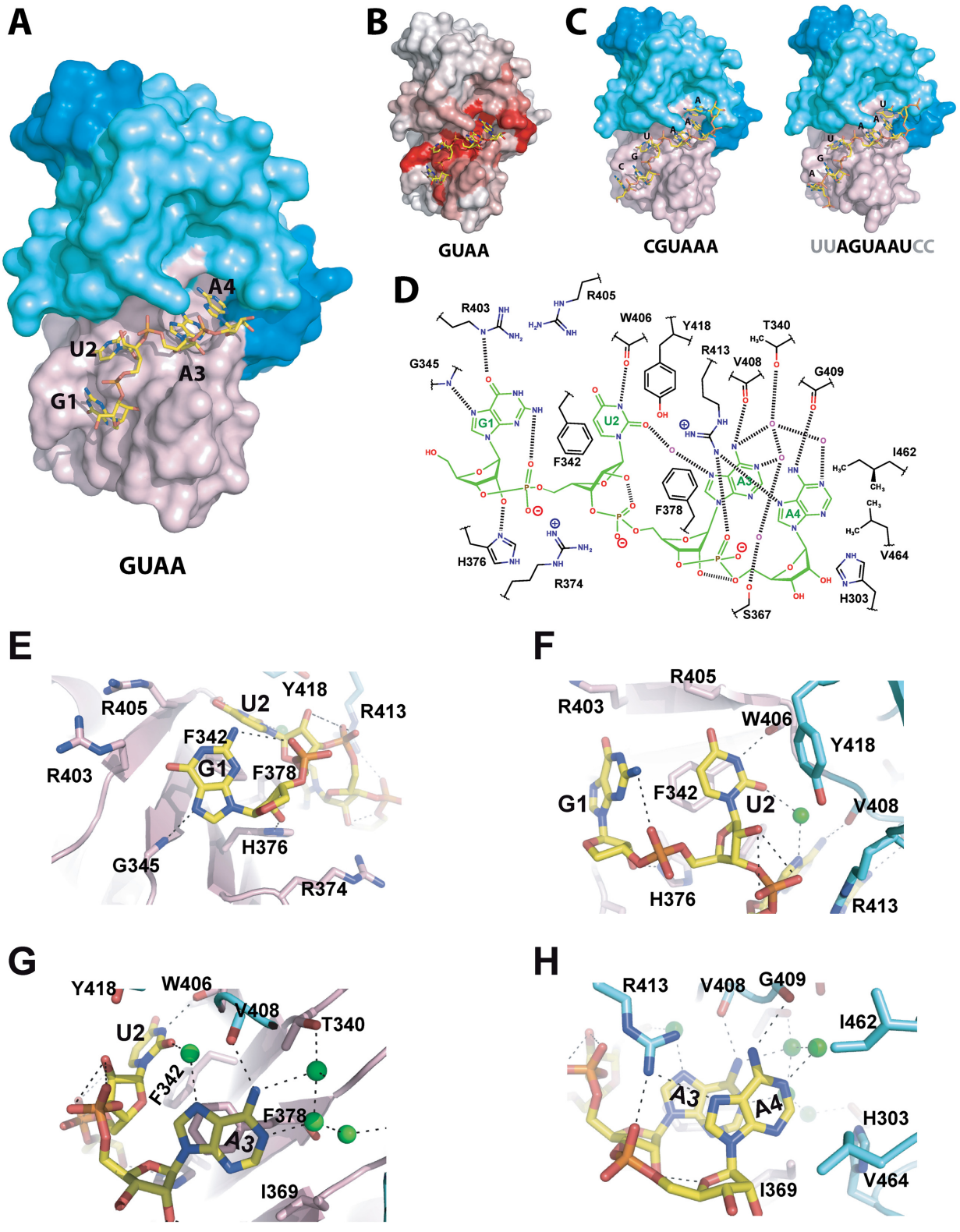

Figure 3. X-ray structures of the Nrd1$_{290–468}$:RNA complexes. (**A**) Nrd1$_{290–468}$:GUAA complex represented as surface and colored by domains as in Figure 1. The RNA is shown as sticks (C: yellow, O: red, N: blue and P: orange). (**B**) Nrd1$_{290–468}$:GUAA complex in the same orientation as A and with the surface colored by chemical shift mapping obtained for the same interaction by NMR (shades of red). (**C**) Nrd1$_{290–468}$:CGUAAA (left) and Nrd1$_{290–468}$:UUAGUAAUCC (right) complexes. (**D**) Schematic representation of the Nrd1:GUAA interface. Hydrogen bonds are represented in dashed lines. (**E**) Structural detail around G1. (**F**) Structural detail around U2. (**G**) Structural detail around A3. (**H**) Structural detail around A4. Water molecules are shown as green spheres. Side-chains of residues interacting thought their backbone carbonyls or amide groups have been omitted for clarity.

Table 1. Thermodynamic parameters of *wt* and mutants $Nrd1_{290–468}$/RNA interactions determined by isothermal titration calorimetry (ITC) and fluorescence anisotropy (FA)

Exp.	Protein	Mutant	RNA[a]	ΔG (kcal/mol)[b]	K_D (μM)[c]
ITC	Nrd1	wt	CC**GUAA**CC	7.74 ± 0.05	2.1 ± 0.1
ITC	Nrd1	wt	CC**GUAG**CC	6.95 ± 0.08	8 ± 1
FA	txA-Nrd1	wt	CC**GUAA**CC	8.0 ± 0.1	1.4 ± 0.2
FA	txA-Nrd1	wt	CC**GUAG**CC	7.2 ± 0.2	5 ± 1
FA	txA-Nrd1	wt	CCC**AUAA**CCC	n.d.	∼50[d]
FA	txA-Nrd1	wt	CCC**IUAA**CCC	6.5 ±. 0.2	18 ± 5
FA	txA-Nrd1	wt	CCC**AUUA**CCC	n.d.	∼160[d]
FA	txA-Nrd1	wt	CCC**AUAU**CCC	n.d.	∼110[d]
FA	txA-Nrd1	K335E	CC**GUAA**CC	6.8 ± 0.1	11 ±1
FA	txA-Nrd1	K335M	CC**GUAA**CC	8.1 ± 0.1	1.2 ±0.2
FA	txA-Nrd1	K335R	CC**GUAA**CC	7.9 ± 0.2	1.7 ±0.4
FA	txA-Nrd1	T340A	CC**GUAA**CC	6.9 ± 0.2	9 ± 2
FA	txA-Nrd1	W353A	CC**GUAA**CC	7.8 ± 0.1	2.0 ± 0.3
FA	txA-Nrd1	R374A	CC**GUAA**CC	6.5 ± 0.2	18 ± 5
FA	txA-Nrd1	K380A	CC**GUAA**CC	n.d.	∼135[d]
FA	txA-Nrd1	R405A	CC**GUAA**CC	8.8 ± 0.1	0.37±0.05
FA	txA-Nrd1	R413G	CC**GUAA**CC	6.4 ± 0.2	21 ± 6
FA	txA-Nrd1	C415S/C416S	CC**GUAA**CC	7.2 ± 0.2	5 ± 1
FA	txA-Nrd1	Y418A	CC**GUAA**CC	6.4 ± 0.1	21 ± 3
FA	txA-Nrd1	W437A	CC**GUAA**CC	7.3 ± 0.2	5 ± 1

[a]Sequences 5′- to 3′. All RNAs used for fluorescence anisotropy were fluorescein-labeled at 5′. I = inosine.
[b]Dissociation ΔG obtained in FA from non-linear least squares fitting of the data. ITC values and uncertainties were calculated from $\Delta G = RT \bullet \ln(K_{B,obs})$ and error propagation.
[c]$K_D = (K_{B,obs})^{-1}$ in ITC. Apparent K_D values and uncertainties in FA were calculated from $K_D = e^{(-\Delta G/RT)}$ and error propagation.
[d]Estimations assuming similar fluorescence anisotropy changes as those obtained in the curves in which saturation was reached.
n.d.: not determined due to insufficient saturation.

the RNA-binding site (Supplementary Figure S11). This is not involved in RNA recognition, however the K335E exhibits a small GUAA affinity loss (Table 1). Lys335 forms a conserved salt bridge with Asp326, thus we argue that the K335E mutant might introduce a destabilizing repulsion effect. To prove this, we measured RNA binding affinity of more structurally compatible K335R and K335M mutants, which rendered values indistinguishable from *wt*.

Altogether, we conclude that the SD plays a scaffolding role allowing the structural preconfiguration of the GUAA binding site in Nrd1.

The split domain of Nrd1 is essential for viability

Genomic removal of Nrd1 compromises cell viability (10). We aimed to find out if the Nrd1 SD (Figure 4A) affects cell survival by introducing LEU plasmids containing wild-type *NRD1* (*wt*), complete SD deletion (Δ301–336/Δ412–463) or partial deletions (ΔSD_A = Δ301–336 and ΔSD_B = Δ412–463) of this domain in a yeast strain harbouring a centromeric URA plasmid expressing a genomic copy of *NRD1* (10). Loss of the URA wild-type plasmid upon 5-FOA treatment and plasmid shuffling results in cell lethality in all the SD deletion: only the cells bearing *wt NRD1* remain viable (Figure 4B left panel). We further tested synthetic lethality of these SD deletion mutants using another method: controlling genomic *wt NRD1* expression under the GAL promoter. As shown, cells harboring SD deletion mutations are unable to grow in glucose containing media (Figure 4B, right panel). Therefore, the SD is a key element for Nrd1 activity due to either its role in specific recognition of RNA terminators (as described in our Nrd1:RNA structures) or to other unknown functions. We extended our mutagenesis study to delve deeper into these important questions.

Nrd1 has a conserved sequence at the C-terminus (residues 551–575) (Supplementary Figure S12), which includes a polyglutamine tail. We introduced stop codon mutations at several positions after the RBD (L558stop, Q567stop and K490stop) that do not cause cell growth defects (Figure 4C), suggesting that the conserved Nrd1 C-terminus is not critical for its function. In contrast, disrupting Nrd1 from the Val468 residue (V468stop) provokes significant growth defects at all tested temperatures (Figure 4C, upper panel). The *nrd1Δ301–306* mutant, at the N-terminus of the RBD, shows less marked growth defects (Figure 4C, bottom panel).

Next we explored how changes in RNA affinity affect Nrd1 function *in vivo*. In general, mutations causing minor affinity losses (W353A, C415S/C416S, W437A) do not display noticeable growth defects. Although in the R374A, R413G and Y418A mutants, affecting direct Nrd1–RNA contacts and displaying similar K_D increases (Table 1), only Y418A cells present slow growth at 37°C (Figure 4C, bottom panel). Conversely, K335E, with less impact in RNA recognition, has a strong growth defect, which is comparable with the previously reported defects on T340A and K380A (24) (Figure 4C). More importantly, Y418A/K335E double mutation (with both residues belonging to the SD) causes an extraordinary slow growth at 28°C and lethality at 37°C (Figure 4C). Therefore, this double mutation is almost as aggressive as the SD deletion mutants (Figure 4B). The growth defect caused by the K335E is not totally rescued in the case of K335M and K335R (Figure 4C). These two mutants show RNA binding affinities indistinguishable from that of *wt* (Table 1), therefore we

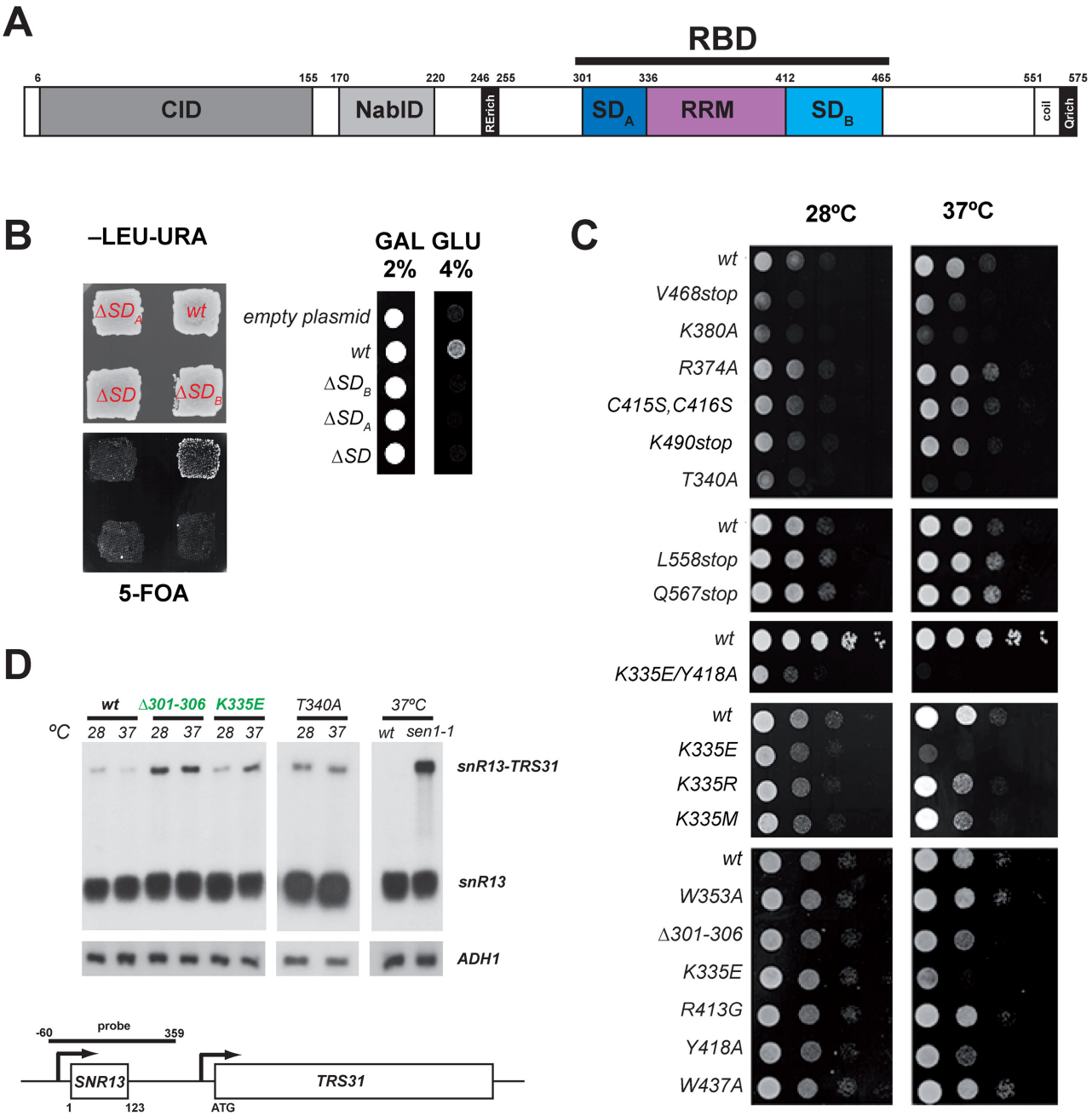

Figure 4. Nrd1 RBD functional analysis. (**A**) Schematic representation of Nrd1 domain organization. (**B**) Partial or total deletion of the Nrd1-SD causes cell lethality. Left panel, the indicated strains were grown in -LEU-URA selective media and then replicated onto 5-FOA containing media and grown for 2–3 days at 28°C. Right panel, strains with the indicated genotypes were grown in galactose (GAL) or glucose (GLU) containing media and grown for 2–3 days. (**C**) Analysis of *nrd1* mutants growth phenotypes. Serial dilutions (1:10) of *wt* and *nrd1* mutant strains were spotted on selective SC media and grown for 2–3 days at the indicated temperatures. (**D**) Northern blot analyses of the *SNR13* and *ADH1* genes. Total RNA was purified from the indicated strains grown at 28°C and 37°C. In addition, *wt* and *sen1–1* isogenic strains were used as positive control for transcription termination defects (50). Full-length snR13 and readthrough transcripts are indicated by arrows. A schematic representation of *SNR13* gene and 3′ end flanking region is also shown, where the size and localization of the probe is represented by a black bar above *SNR*13 coding region.

propose that they alter some functions of Nrd1 RBD not related to their RNA binding; probably unknown protein-protein interactions with other components of the NNS pathway.

Finally we decided to test whether the mutants exhibiting reduced cell growth phenotypes also showed transcription termination defects at both 28°C and 37°C. For that purpose, we analysed the well-known *SNR13* transcript by northern blot assay as described in (50). We used as a positive control of snoRNA transcription termination defects the *sen1–1* mutant grown at 37°C (50). *SNR13* displayed strong transcription termination defects at 28°C and 37°C in *nrd1Δ301–306* mutant, and to a lesser extent, but significantly, in *nrd1-K335E* cells at 37°C, consistent with the extraordinary reduced growth of that mutant at this temperature. In the case of *nrd1-T340A* cells, whose mutation lies into the RRM domain, *SNR13* transcription termination defects are also observed at both temperatures (Figure 4D). Therefore, these data suggest that both RRM and SD have a role in snoRNAs transcription termination. The fact that K335E and T340A mutants do not show a very marked RNA affinity lost (Table 1) lead us to think the biological role of Nrd1 RBD in snoRNAs transcription termination might not be solely dependent on RNA binding.

DISCUSSION

Widely found in eukaryotes, the RRMs have a typical architecture ($\beta_1-\alpha_1-\beta_2-\beta_3-\alpha_2-\beta_4$) in which the RNA recognition residues (RNP1 and RNP2 motifs) are on the exposed face of β2 and β3 (47–49,51,52). From this basic fold, there are a number of RRMs with extended features, typically built from additional regions on the C- and/or N-termini: extra β-strands (53), α-helices (54–56) and β-sheets (27). The structure of Nrd1 RBD is a new class of RRM; it contains additional N- and C-terminal extensions which, rather than adding extra secondary structural elements, form together a domain by itself (SD). The SD is fused to the RRM through an intricate polar interface, defining a rigid body structure. The structures of Nrd1 and Seb1 (46) RBDs have similar folds, differing in that Nrd1 RBD includes long N-terminal extensions that make further contacts with the RRM domain (Supplementary Figure S13 A&B). The RRM–SD interface is mostly conserved between Nrd1 and Seb1 RBDs (Supplementary Figure S13C–E). Only the Arg319 cluster is not present in Seb1 and the rest of RRM–SD interfacial residues: Arg339, Lys380, Arg384, Gln366, Trp406, Pro460 and Asp461, are totally conserved in fungi (Supplementary Figure S3), suggesting a conservation of the Nrd1/Seb1 RBD structure at least in yeast and moulds.

Seb1 RBD binds AUUAGUAAAA with 1.8 μM affinity (46), a remarkably close value to the 1.5–2.1 μM affinity that we measured for Nrd1 RBD binding to CCGUAACC. Both Seb1 and Nrd1 RNA targets contain the GUAA motif. These similarities suggest that Seb1 might have a similar RNA binding mode than Nrd1, hence we constructed a model of the Seb1:RNA complex by placing the RNA on the Seb1 structure in the same conformation as in the $Nrd1_{290–468}$:GUAA complex (Supplementary Figure S13F). This simple model is likely realistic because, as it happens for Nrd1 (Supplementary Figure S11), we expect little conformational changes in Seb1 upon binding. In support of this model, the mutagenesis data (46) show that the biggest changes in RNA binding affinity affect interfacial residues, while other mutants, far from the hypothetical binding site, cause negligible effects. In conclusion, we predict that Seb1 will bind RNAs with the GUAA/G motif by using a similar mode as Nrd1 RBD, which is probably conserved across the family with minor variations.

Both GUAA and GUAG sequences have been identified at NNS terminator regions (10,12,13). Nrd1 binds GUAG terminators with four times lower affinity than for GUAA ones. In both cases binding is enthalpically driven (Figure 2) in agreement with the formation of multiple polar interactions: hydrogen bonds and salt bridges. However, the magnitude of ΔH is lower for the GUAG complex (Figure 2) suggesting that there are less polar contacts in this case. In the Nrd1:GUAA complex, A4 acts as a sort of molecular gate trapping four water molecules between the RNA and the SD (Figure 3G). We hypothesize that a suboptimal binding of G4 would facilitate the release of these ligand-binding waters providing an explanation for the loss in ΔH contribution in GUAG complex formation with respect to that of GUAA one.

The uniqueness of Nrd1-like RBD structure opens the question of whether this domain might play further roles beyond RNA binding. We demonstrated that the SD accompanying the canonical RRM is important for cell viability (Figure 4B) and its role in RNA binding is demonstrated by structural and mutagenesis data. SD mutants *nrd1-K335E* and *nrd1Δ(301–306)* exhibit both cell growth and *SNR13* termination defects (Figure 4C and D). In the case of Lys335, a conserved residue in the opposite face of the RNA binding pocket, its mutation might slightly perturb the SD stability/architecture but not sufficiently to cause a dramatic change in RNA binding affinity (Table 1). Instead the *nrd1-K335E* phenotype is likely due to the disruption of other interactions within the NNS pathway. Currently, we have sufficient structural/biochemical information to designate potential targets for Nrd1 RBD protein–protein interactions. The two known Nrd1 anchoring points to the RNA pol II are the nascent RNA and the pS5 Rpb1 CTD repeats, both facing toward the same side of the enzyme (Supplementary Figure S14). Nrd1 interacts with Nab3 through its heterodimerization domain (14) and both co-purify with the RNA helicase Sen1 (57). However, Sen1 recruitment to the NNS complex seems to be mediated by Nab3, rather than by Nrd1 (58). On the other hand, the RNA Pol II stalk, formed by the Rpb4/Rbp7 heterodimer, participates in the recruitment of RNA processing factors and CTD modifying enzymes (59–61). Interestingly, nascent RNA exits making contacts with Rpb7 (62) and Nrd1 interacts physically with Rpb7 through its C-terminus (including half of the SD) (63). The temperature sensitive phenotype of *nrd1*-V468stop might be due to a partial or total disruption of this interaction. If so, the Nrd1-Rpb7 interaction would locate the Nrd1 RBD on the direct exit route of the newly synthetized RNA, allowing an early recognition of GUAA/G sequences, which perhaps triggers transcription termination.

ACCESSION NUMBERS

Atomic coordinates of the protein structures and their complexes have been deposited in the Protein Data Bank (PDB) with the following accession codes: 5O1W (crystal structure of $Nrd1_{301–489}$); 5O1X (crystal structure of $Nrd1_{290–468}$); 5O1Y (crystal structure of $Nrd1_{290–468}$ in complex with GUAA RNA); 5O1Z (crystal structure of $Nrd1_{290–468}$ in complex with CGUAAA RNA); 5O20 (crystal structure of $Nrd1_{290–468}$ in complex with UUAGUAAUCC RNA (corresponds to SNR13 148–157)) and 5O1T ($Nrd1_{290–468}$ NMR structure). The chemical shifts for $Nrd1_{290–468}$ construct has been deposited on the Biological Magnetic Resonance Bank (BMRB) with the accession code 34140.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr S. Buratowski for gratefully providing yeast strains and plasmids. Data collection was performed at ESRF, ID23-1 beamline (Grenoble, France) and at ALBA Synchrotron, XALOC-BL13 beamline, with the collaboration of ESRF and ALBA staff.

FUNDING

Spanish Ministry of Economy and Competitiveness (MINECO) [BFU2014-53762-P to B.G., BFU2015-71978-REDT and BFU2013-48374-P to O.C., CTQ2011-26665 and CTQ2014-52633 to J.M.P.C.]; E.F.-E. was supported by BFU2014–53762-P grant. Funding for open access charge: MINECO.
Conflict of interest statement. None declared.

REFERENCES

1. Tudek,A., Candelli,T. and Libri,D. (2015) Non-coding transcription by RNA polymerase II in yeast: Hasard or necessite? *Biochimie*, **117**, 28–36.
2. Porrua,O., Boudvillain,M. and Libri,D. (2016) Transcription Termination: Variations on Common Themes. *Trends Genet.*, **32**, 508–522.
3. Loya,T.J. and Reines,D. (2016) Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research*, **5**, F1000.
4. Porrua,O. and Libri,D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.*, **16**, 190–202.
5. Fasken,M.B., Laribee,R.N. and Corbett,A.H. (2015) Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS Genet*., **11**, e1005044.
6. Arndt,K.M. and Reines,D. (2015) Termination of transcription of short noncoding RNAs by RNA polymerase II. *Annu. Rev. Biochem.*, **84**, 381–404.
7. 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.
8. 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.
9. Wilson,S.M., Datar,K.V., Paddy,M.R., Swedlow,J.R. and Swanson,M.S. (1994) Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae. *J. Cell Biol.*, **127**, 1173–1184.
10. 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.
11. Porrua,O., Hobor,F., Boulay,J., Kubicek,K., D'Aubenton-Carafa,Y., Gudipati,R.K., Stefl,R. and Libri,D. (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1–Nab3–Sen1-dependent transcription termination. *EMBO J.*, **31**, 3935–3948.
12. 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.
13. 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.
14. 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.
15. 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.
16. Kubicek,K., Cerna,H., Holub,P., Pasulka,J., Hrossova,D., Loehr,F., Hofr,C., Vanacova,S. and Stefl,R. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.*, **26**, 1891–1896.
17. Grzechnik,P. and Kufel,J. (2008) Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell*, **32**, 247–258.
18. Tudek,A., Porrua,O., Kabzinski,T., Lidschreiber,M., Kubicek,K., Fortova,A., Lacroute,F., Vanacova,S., Cramer,P., Stefl,R. *et al.* (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell*, **55**, 467–481.
19. Loya,T.J., O'Rourke,T.W. and Reines,D. (2012) Yeast Nab3 Contains a Self-Assembly Domain found in Human hnRNP-C that is Necessary for Transcription Termination. *J. Biol. Chem.*, **288**, 2111–2117.
20. Meinhart,A. and Cramer,P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature*, **430**, 223–226.
21. Lunde,B.M., Reichow,S.L., Kim,M., Suh,H., Leeper,T.C., Yang,F., Mutschler,H., Buratowski,S., Meinhart,A. and Varani,G. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, **17**, 1195–1201.
22. Lunde,B.M., Horner,M. and Meinhart,A. (2011) Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.*, **39**, 337–346.
23. Hobor,F., Pergoli,R., Kubicek,K., Hrossova,D., Bacikova,V., Zimmermann,M., Pasulka,J., Hofr,C., Vanacova,S. and Stefl,R. (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.*, **286**, 3645–3657.
24. Bacikova,V., Pasulka,J., Kubicek,K. and Stefl,R. (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.*, **42**, 8024–8038.
25. Jamonnak,N., Creamer,T.J., Darby,M.M., Schaughency,P., Wheelan,S.J. and Corden,J.L. (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA*, **17**, 2011–2025.
26. Schulz,D., Schwalb,B., Kiesel,A., Baejen,C., Torkler,P., Gagneur,J., Soeding,J. and Cramer,P. (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, **155**, 1075–1087.
27. Martínez-Lumbreras,S., Taverniti,V., Zorrilla,S., Séraphin,B. and Pérez-Cañadillas,J.M. (2016) Gbp2 interacts with THO/TREX through a novel type of RRM domain. *Nucleic Acids Res.*, **44**, 437–448.

28. Bergfors,T. (2003) Seeds to crystals. *J. Struct. Biol.*, **142**, 66–76.
29. Kabsch,W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
30. Evans,P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D, Biol. Crystallogr.*, **67**, 282–292.
31. Sheldrick,G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 479–485.
32. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D, Biol. Crystallogr.*, **62**, 1002–1011.
33. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 22–25.
34. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.
35. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D, Biol. Crystallogr.*, **53**, 240–255.
36. 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.
37. DeLano,W.L. (2002). *DeLano Scientific*, San Carlos, CA.
38. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93–158.
39. Kay,L.E., Xu,G.Y., Singer,A.U., Muhandiram,D.R. and Forman-Kay,J.D. (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. *J. Magn. Reson. Ser. B*, **101**, 333–337.
40. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
41. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
42. Breeze,A.L. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Mag. Res. Sp.*, **36**, 323–372.
43. Shen,Y., Delaglio,F., Cornilescu,G. and Bax,A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
44. Guntert,P., Mumenthaler,C. and Wuthrich,K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
45. Bruke,D., Dawson,D. and Stearns,T. (2000) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, NY.
46. Wittmann,S., Renner,M., Watts,B.R., Adams,O., Huseyin,M., Baejen,C., El Omari,K., Kilchert,C., Heo,D.H., Kecman,T. *et al.* (2017) The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat. Commun.*, **8**, 14861.
47. Daubner,G.M., Clery,A. and Allain,F.H. (2013) RRM-RNA recognition: NMR or crystallography…and new findings. *Curr. Opin. Struct. Biol.*, **23**, 100–108.
48. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
49. 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.
50. Kim,M., Vasiljeva,L., Rando,O.J., Zhelkovsky,A., Moore,C.L. and Buratowski,S. (2006) Distinct pathways for snoRNA and mRNA termination. *Mol. Cell*, **24**, 723–734.
51. Pérez-Cañadillas,J.M. and Varani,G. (2001) Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol.*, **11**, 53–58.
52. Messias,A.C. and Sattler,M. (2004) Structural basis of single-stranded RNA recognition. *Acc. Chem. Res.*, **37**, 279–287.
53. 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.
54. Jacks,A., Babon,J., Kelly,G., Manolaridis,I., Cary,P.D., Curry,S. and Conte,M.R. (2003) Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure*, **11**, 833–843.
55. Pérez-Cañadillas,J.M. and Varani,G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.*, **22**, 2821–2830.
56. Santiveri,C.M., Mirassou,Y., Rico-Lastres,P., Martínez-Lumbreras,S. and Pérez-Cañadillas,J.M. (2011) Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. *PLoS One*, **6**, e24481.
57. Vasiljeva,L. and Buratowski,S. (2006) Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239–248.
58. Nedea,E., Nalbant,D., Xia,D., Theoharis,N.T., Suter,B., Richardson,C.J., Tatchell,K., Kislinger,T., Greenblatt,J.F. and Nagy,P.L. (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol. Cell*, **29**, 577–587.
59. Allepuz-Fuster,P., Martinez-Fernandez,V., Garrido-Godino,A.I., Alonso-Aguado,S., Hanes,S.D., Navarro,F. and Calvo,O. (2014) Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation. *Nucleic Acids Res.*, **42**, 13674–13688.
60. Garavís,M., González-Polo,N., Allepuz-Fuster,P., Louro,J.A., Fernández-Tornero,C. and Calvo,O. (2017) Sub1 contacts the RNA polymerase II stalk to modulate mRNA synthesis. *Nucleic Acids Res.*, **45**, 2458–2471.
61. Runner,V.M., Podolny,V. and Buratowski,S. (2008) The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3′ processing factors. *Mol. Cell. Biol.*, **28**, 1883–1891.
62. Chen,C.Y., Chang,C.C., Yen,C.F., Chiu,M.T. and Chang,W.H. (2009) Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 127–132.
63. Mitsuzawa,H., Kanda,E. and Ishihama,A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.*, **31**, 4696–4701.