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

Elsa Franco-Echevarría1, Noelia González-Polo2, Silvia Zorrilla3, Santiago Martínez-Lumbreras4,5, Clara M. Santiveri6, Ramón Campos-Olivas6, Mar Sánchez2, Olga Calvo2, Beatriz González1,* and José Manuel Pérez-Cañadillas5,*

1Department of Crystallography and Structural Biology, Institute of Physical-Chemistry “Rocasolano”, CSIC, C/ Serrano 119, 28006 Madrid, Spain, 2Instituto de Biología Funcional y Genómica, IBFG-CSIC, Universidad de Salamanca, 3Department of Cellular and Molecular Biology, Biological Research Center, CSIC, 4Department of Chemistry, King’s College London, 5Department of Biological Physical Chemistry, Institute of Physical-Chemistry “Rocasolano”, CSIC, C/ Serrano 119, 28006 Madrid, Spain and 6Spectroscopy 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 (snRNAs 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 (snRNAs 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 (10,12,13) respectively. These elements are frequent in snRNA 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-
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 Rtu103 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 GAUG 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 snR13 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 QuickChange Lightning Kit (Agilent genomics), both in sequence) ((27) for details). Mutants were obtained with a QuickChange Lightning Kit (Agilent genomics), both in

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 Qiagen commercial screens using 96-well plates (Innovaplate SD-2 microplates, Innovadyne Technologies Inc). The crystallization conditions were scaled 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 UUGAUAUCC RNAs (IBA and IDT) were obtained by slow addition of 1 μl of the RNA solution to the crystals drops followed by overnight (GUAA) or three hours (CGUAAA and UUGAUAUCC) incubation.
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 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 datasets 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 P43212, whereas the shorter construct Nrd1 290–468 also crystallized in the hexagonal group P65, 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 and refinement are summarized in Supplementary Table S2. The stereochemistry has been checked with CcpNmr Analysis (41).

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% D2O) and experiments acquired at 25°C on cryoprobe-equipped Bruker AV800 MHz spectrometer. Resonance assignments (1H, 15N and 13C) were obtained with a battery of triple and double resonance 3D experiments (HNCA, HNCO, CBCA(CO)NH, HCH-TOCSY) (38,39) that were processed with NMRPipe (40) and analyzed 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 [13C–15N-HSQC-NOESY experiments and from a set of 2D 1C–15N (F1-filtered) NOESY and 2D 15C–13N (F1/F2-doublefiltered) (42) NOESY experiments recorded on samples with amino acid selective reverse labeling (Supplementary Figures S1 and S2C). Backbone dihedral angle restraints (φ and ψ) were obtained with TLS+ (43) from CO, CB, 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.

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 analyzed with Origin 7.0 (OriginLab) using a one-site binding model. The experiments were carried out in duplicate. Control experiments of dilution of the...
**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 ($\beta$-sheets and $\alpha$-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, $P4_12_2$ 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 $\alpha$/$\beta$ domains occupying the positions of the weights (Figure 1A). The RRM domain (residues 339 to 407) shows the classic $\alpha$/$\beta$ 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 $\beta$-sheet with the two central strands in parallel orientation, and an $\alpha$-helix that runs nearly perpendicular to the $\beta$-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$_{300-468}$ 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 structure. The 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 Ne1–He1 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 Ne1–He1 peak in the Nrd1$_{301-489}$ spectrum (Supplementary Figure S4). Since this construct lacks Phe298, we interpret
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 carboxyls 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...
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 proposed Nrd1 selectivity for the sequences GUAA and ure3C). However, these additional bases have higher temperature factors than the GUAA core and no evident conformation among the three complexes (Supplementary Table 1). 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 S2), shows that both RRM and SD domains contribute to the depression between RRM and SD domains 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 recognizes the core GUAA.

The 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: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 N82, U2 O2′-A3 OPI 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 OPI 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:GUAA crystal structure in complex with CGUAAA, which was solved in phosphate buffer (no Tris present), showed that only Arg403 makes base-specific contacts (Arg403 Ne-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),
presumably due to loss of protein–RNA contacts (hydrogen bonds: R413G, hydrophobic: Y418A and cumbic: 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 Nrd1290–468:RNA complexes. (A) Nrd1290–468: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) Nrd1290–468: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) Nrd1290–468:CGUAAA (left) and Nrd1290–468: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 through their backbone carbonyls or amide groups have been omitted for clarity.
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 (ΔSDA = Δ301–336 and ΔSDB = 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 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 nrd1 Δ301–306 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 KD 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

Table 1. Thermodynamic parameters of wt and mutants Nrd1290–468/RNA interactions determined by isothermal titration calorimetry (ITC) and fluorescence anisotropy (FA)

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Protein</th>
<th>Mutant</th>
<th>RNAa</th>
<th>ΔG (kcal/mol)b</th>
<th>KD (µM)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITC</td>
<td>Nrd1</td>
<td>wt</td>
<td>CCGUAACC</td>
<td>7.4 ± 0.05</td>
<td>2.1 ± 0.1</td>
</tr>
<tr>
<td>ITC</td>
<td>Nrd1</td>
<td>wt</td>
<td>CCGUGGCC</td>
<td>6.95 ± 0.08</td>
<td>8 ± 1</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>CCGUAACC</td>
<td>8.0 ± 0.1</td>
<td>1.4 ± 0.2</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>CCGUGGCC</td>
<td>7.2 ± 0.2</td>
<td>5 ± 1</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>CCCUAACC</td>
<td>n.d.</td>
<td>~50d</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>CCCUAACC</td>
<td>6.5 ± 0.2</td>
<td>18 ± 5</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>CCCUAACC</td>
<td>n.d.</td>
<td>~160d</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>K335E</td>
<td>6.8 ± 0.1</td>
<td>11 ± 1</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>K335M</td>
<td>8.1 ± 0.1</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>K335R</td>
<td>7.9 ± 0.2</td>
<td>1.7 ± 0.4</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>T340A</td>
<td>6.9 ± 0.2</td>
<td>9 ± 2</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>W353A</td>
<td>7.8 ± 0.1</td>
<td>2.0 ± 0.3</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>R374A</td>
<td>6.5 ± 0.2</td>
<td>18 ± 5</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>K380A</td>
<td>n.d.</td>
<td>~135d</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>R405A</td>
<td>8.8 ± 0.1</td>
<td>0.37±0.05</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>R413G</td>
<td>6.4 ± 0.2</td>
<td>21 ± 6</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>C415S/C416S</td>
<td>7.2 ± 0.2</td>
<td>5 ± 1</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>Y418A</td>
<td>6.4 ± 0.1</td>
<td>21 ± 3</td>
</tr>
<tr>
<td>FA</td>
<td>txA-Nrd1</td>
<td>wt</td>
<td>W437A</td>
<td>7.3 ± 0.2</td>
<td>5 ± 1</td>
</tr>
</tbody>
</table>

aSequences 5′- to 3′. All RNAs used for fluorescence anisotropy were fluorescein-labeled at 5′. I = inosine.
bDissociation ΔG obtained in FA from non-linear least squares fitting of the data. ITC values and uncertainties were calculated from ΔG = RTln(KR,obs) and error propagation.
cKD = (KR,obs)−1 in ITC. Apparent KD values and uncertainties in FA were calculated from KD = e(−ΔG/RT) and error propagation.
dEstimations assuming similar fluorescence anisotropy changes as those obtained in the curves in which saturation was reached. n.d.: not determined due to insufficient saturation.
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 SNR13 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 senf1–1 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 nrd1ΔK335E cells at 37°C, consistent with the extraordinary reduced growth of that mutant at this temperature. In the case of nrd1ΔT340A 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 loss (Table 1) leads us to think that 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 (β1–α1–β2–β3–α2–β4) 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 AUUAGUAAA 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 Nrd1290–468: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 nrd1ΔK335E and nrd1Δ301–306 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 nrd1ΔK335E 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 Rbp1 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 nrd1Δ–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 synthesized 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 Nrd1d301–489); 5O1X (crystal structure of Nrd1d290–468); 5O1Y (crystal structure of Nrd1d290–468 in complex with GUAA RNA); 5O1Z (crystal structure of Nrd1d290–468 in complex with CGUAA RNA); 5O20 (crystal structure of Nrd1d290–468 in complex with UUAGUAUCCRNA (corresponds to SNR13 148–157)) and 5O1T (Nrd1d290–468 NMR structure). The chemical shifts for Nrd1d290–468 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


