*Published online 23 November 2015* *Nucleic Acids Research, 2016, Vol. 44, No. 1* ***437–448***
*doi: 10.1093/nar/gkv1303*

# Gbp2 interacts with THO/TREX through a novel type of RRM domain

**Santiago Martínez-Lumbreras[1], Valerio Taverniti[2], Silvia Zorrilla[3], Bertrand Séraphin[2,*] and José Manuel Pérez-Cañadillas[1,*]**

[1]Department of Biological Physical Chemistry, Instituto de Química-Física 'Rocasolano', CSIC, Serrano-119, 28006 Madrid, Spain, [2]Equipe Labellisée La Ligue, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGMBC), Centre National de Recherche Scientifique (CNRS) UMR 7104/Institut National de Santé et de Recherche Médicale (INSERM) U964/Université de Strasbourg, 67404 Illkirch, France and [3]Department of Cellular and Molecular Biology, Centro de Investigaciones Biológicas, CSIC, Madrid, Spain

Received February 25, 2015; Revised October 26, 2015; Accepted November 07, 2015

## ABSTRACT

**Metazoan SR and SR-like proteins are important regulatory factors in RNA splicing, export, translation and RNA decay. We determined the NMR structures and nucleic acid interaction modes of Gbp2 and Hrb1, two paralogous budding yeast proteins with similarities to mammalian SR proteins. Gbp2 RRM1 and RRM2 recognise preferentially RNAs containing the core motif GGUG. Sequence selectivity resides in a non-canonical interface in RRM2 that is highly related to the SRSF1 pseudoRRM. The atypical Gbp2/Hrb1 C-terminal RRM domains (RRM3) do not interact with RNA/DNA, likely because of their novel N-terminal extensions that block the canonical RNA binding interface. Instead, we discovered that RRM3 is crucial for interaction with the THO/TREX complex and identified key residues essential for this interaction. Moreover, Gbp2 interacts genetically with Tho2 as the double deletion shows a synthetic phenotype and preventing Gbp2 interaction with the THO/TREX complex partly supresses gene expression defect associated with inactivation of the latter complex. These findings provide structural and functional insights into the contribution of SR-like proteins in the post-transcriptional control of gene expression.**

## INTRODUCTION

Classic SR proteins belong to a family of metazoan proteins interacting with RNA. These factors have been identified as key players in the regulation of constitutive and alternative splicing in metazoans (1–5). They have since been shown to contribute to many steps of the post-transcriptional regulation of gene expression including miRNA production, mRNA transport, translational control and mRNA decay regulation (6). SR proteins display a prototypical domain organization that includes RNA binding domains (RRM), sometimes associated with pseudoRRM modules, and serine/arginine rich regions whose phosphorylated forms appear to be recognized by a specific antibody (7). The yeast *Saccharomyces cerevisiae* is devoid of this class of proteins but has three SR-like proteins (3): Npl3, an important factor for RNA metabolism including pre-mRNA splicing (8) that reminds of the role of its metazoan cousins; Gbp2 and Hrb1. These last two contain three RRM domains (Npl3 has two) and probably arose from an ancient genome duplication event (9). Unlike for Npl3, deletion of Gbp2 or Hrb1 (or both) does not compromise the overall performance of mRNA splicing (8). Instead the two proteins were proposed to work as quality control factors for spliced mRNA (10) and may possibly be involved in other processes as well. Gbp2 and Hrb1 interact with Mtr4, a component of the RNA degradation machinery TRAMP, but also with Mex67, a key adaptor in the mRNA export pathway (10). The latter interaction only occurs upon efficient splicing; else Gbp2/Hrb1 remain associated to TRAMP and the transcript is degraded in the nucleus (10). Unlike Npl3, Gbp2 and Hrb1 interact with the THO/TREX complex on nascent RNAs (11,12). The three yeast SR-like proteins are exported to the cytoplasm as part of the messenger RiboNucleoProtein particle (mRNP) and are present in the polysomes (13). Npl3 has been recently shown to be important for translation initiation (14), whereas the role of Gbp2 and Hrb1 in translation and/or post-transcriptional control is still unclear.

We have performed a structural, biophysical, and functional characterization of Gbp2 and Hrb1 proteins. We determined the NMR structures of several RRM domains

*To whom correspondence should be addressed. Tel: +34 915619400; Fax: +34 915642431; Email: jmperez@iqfr.csic.es
Correspondence may also be addressed to Bertrand Séraphin. Tel: +33 3 88 65 32 16; Fax: +33 3 88 65 32 03; Email: seraphin@igbmc.fr
Present address: Santiago Martínez-Lumbreras, Department of Chemistry, School of Natural & Mathematical Sciences, King's College London, London, UK.

and analysed their DNA/RNA binding by NMR, fluorescence anisotropy, CD and EMSA. Gbp2 binds preferentially RNA via the RRM1-RRM2 tandem (hereafter RRM12), with RRM2 playing a leading role in recognition of sequences containing the core signal (GGUG). Recognition surface involves a non-canonical interface clustered around helix α1, equivalent to the recently reported pseudoRRM in SR proteins (15). Additionally, Gbp2 binds with high affinity G-strand telomeric DNAs preventing G-quadruplex formation *in vitro*. Gbp2/Hrb1 RRM3 domains do not bind nucleic acids due to the presence of an unusual structure that blocks the access to the β-sheet recognition sites. Here, we show that this new class of RRM interacts physically with the THO/TREX complex by using a new protein–protein interface likely conserved in some mammalian SR proteins. Finally, we found that Gbp2 and Tho2 (a component of the THO complex) interact genetically and functionally.

## MATERIALS AND METHODS

### DNA cloning

Gbp2 and Hrb1 coding sequences were cloned from *Saccharomyces cerevisiae* genomic DNA by PCR using Hifi KOD DNA polymerase (Novagen) and a library of DNA oligonucleotides (Sigma, IDT). Amplified products were purified, digested with corresponding restriction enzymes and ligated into a home-modified pET28 (Novagen) vector that contains: a thioredoxin A N-terminal tag to enhance expression, a 6xHis tag for purification, and a TEV site for fusion removal. Bacterial plasmids used in this work are summarized in Supplementary Figure S1.

### Protein expression and purification

Gbp2 (Hrb1) plasmids were transformed in *E. coli* BL21 (DE3) (Novagen) chemically competent cells and the cultures were grown either in LB medium for non-labelled samples or in K-MOPS minimal medium (16) with $^{15}NH_4Cl$ (1 g/l) and/or $^{13}$C-glucose (4 g/l) for isotopic labelling, both containing 30 μg/l of kanamycin (Sigma–Aldrich). In general, cultures were grown at 37°C until reaching $OD_{600nm}$ = 0.6–0.8, equilibrated to 20°C and induced overnight (>18 h) with 0.5 mM IPTG (Sigma–Aldrich). Centrifuged cell pellets (15 min at 3000 g) were resuspended in lysis buffer (20 mM potassium phosphate pH 8.0, 300 mM NaCl and 1 tablet/50 ml of protease inhibitors (Roche)), lysed by sonication and pelleted at high speed (30 min at 15 000 g). Recombinant proteins were purified from clear lysate by metal affinity chromatography (HisTrap™ HP 5 ml, GE Healthcare) and eluted with 300 mM imidazole-containing buffer. Samples were dialysed against cleavage buffer (20 mM potassium phosphate pH 8.0) and simultaneously digested with homemade TEV protease (100 μg/ml) at 4°C overnight (for difficult-to-cleave constructs proteolysis was performed at 16°C using double TEV concentration). After complete cleavage (checked by PAGE-SDS) a second metal affinity chromatography step was done to remove digested tags, undigested fusion proteins and TEV, recovering in the flow through the purified Gbp2 (Hrb1) proteins. Finally, all Gbp2 and Hrb1 proteins were purified and concentrated by anion exchange chromatography (MonoQ 5 ml, GE Healthcare) with a linear salt gradient elution (from 100 to 1000 mM NaCl) in a 20 mM Tris-HCl pH 8.0 buffer. Sample purity and homogeneity was asserted by PAGE-SDS, mass spectrometry and NMR.

### NMR

Samples of Gbp2 and Hrb1 constructs were prepared at concentrations of 200–800 μM in 25 mM potassium phosphate (pH 6.5), 25 mM NaCl, 0.1 mM DTT and 10% $D_2O$ buffer (also containing 10 μM of DSS for proton chemical shift referencing). NMR experiments were acquired at 25°C, on Bruker AV600 and AV800 spectrometers with cryoprobes. 3D triple resonance experiments (HNCA, HNCO, CBCA(CO)NH) (17) were recorded for backbone assignment of Gbp2 and Hrb1 single-domain constructs and 3D-HCCH TOCSY experiments (18) for side chain assignments. $^{15}$N relaxation data $T_1$, $T_2$ and $^{15}N\{^1H\}$ heteronuclear NOE were recorded on a Bruker AV600 spectrometer for all Gbp2/Hrb1 constructs under the same experimental conditions using standard pulse sequences (19) (see Supplementary Figure S6, S7 and S8 for more details). NMR spectra were processed/analysed with NMRPipe (20), Topspin (Bruker), and CcpNmr Analysis (21).

Distance restrains were obtained from 2D NOESY experiments (50–80 ms mixing time) of Hrb1 RRM1, RRM2, RRM3 and Gbp2 RRM3 and backbone angle restrains (ϕ and ψ) $^{13}C/^1H$ chemical shifts using TALOS+ (22). Structures were calculated with CYANA 2.1 (23) starting from 50 random conformers followed by restrained simulated annealing protocol. The 20 lowest target function conformers without distance (>0.2 Å) and angle (>5°) violations were selected and subjected to a final molecular dynamics simulation refinement at 0 K with AMBER 9.0 (24) using implicit water solvent model (ibg = 5). The final ensemble of structures for each protein were analysed and represented using the molecular graphic packages MOLMOL (25) and PyMOL v0.98 (DeLano Scientific LLC, Palo Alto, CA, USA). Gbp2 RRM1 and RRM2 structural models were created by homology with Hrb1 structures using PyMOL.

### NMR titrations

DNA and RNA probes used for titration experiments were chemically synthetized by IDT (Supplementary Table S3). Protein samples were prepared in NMR buffer (see above) and titrated with the nucleic acid probes reaching 1:1 and 1:2 excess of DNA/RNA (100–200 μM). $^1H$–$^{15}N$ HSQC spectra were acquired for each titration point and the chemical shift perturbation (CSP) was calculated for each amide signal using the following formula:

$$\Delta\delta^{av} = \sqrt{\left((\Delta\delta_{1H})^2 + \left(\Delta\delta_{15N}/5\right)^2\right) \cdot 0.5}$$

CSP results were mapped into the calculated structures/models using PyMOL v0.98 software.

### Fluorescence anisotropy binding titrations

Measurements were performed in a BMG Polarstar Galaxy plate reader using 96-well black plates (Corning) regulated at 26°C, with 485 and 520 nm excitation and emission filters, respectively. The concentration of the fluorescein labelled oligonucleotides (IBA GmbH) was 40 nM and the buffer was 20 mM Tris–HCl, 150 mM NaCl, pH 8.0. A 1:1 binding model compatible with the experimental data was fitted to the isotherms using BIOEQS software (26), to obtain the free energy of formation of complexes from their individual elements employing a numerical solver engine. Errors in the fitting parameters were obtained by confidence limit testing, using the same software, at the 67% confidence level. No changes in the fluorescence emitted by the fluorescein dye were observed upon binding of any of the tested proteins.

Experimental details of other biophysical experiments (fEMSA and CD) are further explained on Supplementary Figures S12 and S16.

### *cerevisiae* genetic manipulations and TAP purifications

All yeast strains used in this work were derived from BMA64 (MATα *ade2-1; his3-11,15; leu2-3,112; trp1Δ; ura3-1*) (27). Tandem affinity purification (TAP) tags were integrated into C-terminal positions by homologous recombination (28). Gbp2 mutants were obtained by PCR using plasmid pBS4473 carrying the Gbp2 coding sequence as template. After fusion to the TAP tag cassette, mutations were introduced in the genome by homologous recombination. Knockout strains were constructed by standard gene disruption while double mutant strains were generated by crossing and dissection. Strains used in this study are listed in Supplementary Table S2.

Tandem Affinity Purifications were performed as described (29). Eluates were fractionated on 4–20% gradient PAGE-SDS and bands of interest were identified by mass spectrometry after trypsin digestion.

### β-Galactosidase activity assays

The pLG-SD5 plasmid (30) was introduced in yeast by transformation (31). ß-Galactosidase activities were measured as indicated previously (32).

## RESULTS

### Gbp2 and Hrb1 RRM3 have novel N-terminal extensions

We obtained several Gbp2 and Hrb1 protein constructs containing one to three of the putative RRM domains of each protein (Supplementary Figure S1) and assigned their NMR spectra. The $^{13}$C conformational shifts from single-domain constructs confirm the predicted RRM with the exception of RRM3 domains that appeared to have an extra N-terminal helix (α0) (Supplementary Figure S2). Gbp2 RRM1 coexists in slow exchange equilibrium between folded and unfolded forms, the latter populated ∼20% at 25°C (Supplementary Figure S3). The unfolded population rises to ∼40% at 35°C, value close to the $T_{1/2}$ estimated in the CD-monitored thermal denaturation experiments (Supplementary Table S1). Gbp2 RRM1 equilibrium remains on RRM12 and RRM123 constructs (data not shown).

We solved the NMR structures of Hrb1 RRM1, RRM2, RRM3 and Gbp2 RRM3 domains (Figure 1, Supplementary Figures S4, S5 and Table 1). The conformational equilibrium in Gbp2 RRM1 and the tendency to aggregate for Gbp2 RRM2 impeded to obtain their structures but we could model them from the Hrb1 RRMs (49–71% sequence identity). Hrb1 RRM1 and RRM2 have the typical RRM-fold βαββαβ (Supplementary Figure S5A and B), the first one contains a disordered histidine/arginine-rich loop (sequence RGHHRG) that could not be assigned perhaps due to complex histidine acid/base equilibrium exchange (pH 6.5). Hrb1 RRM2 has a conserved Trp residue (also present in Gbp2 and homologs) in helix α2 that plays a chief role in nucleic acid recognition (see below).

Gbp2 and Hrb1 RRM3 include an additional structural element (Figure 1 and Supplementary Figure S4): the N-terminus of both constructs interacts with the C-terminus of the domain (that is also the C-terminus of the proteins) to form a singular network of interactions not observed previously in RRM domains (Figure 1 and Supplementary Figure S4). The domains start with a short helix (α0; 1 and a half turn) followed by a short extended sequence that forms a small two-stranded β-sheet with the C-terminus of the protein; these two elements pack against the central β-sheet blocking the canonical RNA binding interface (Figure 1B and Supplementary Figure S4). The 'mini-sheet' is defined by up to three interstrand hydrogen bonds and generates the characteristic Hα–Hα NOE peaks in the 2D NOESY spectra (Figure 1B). Finally, the N-terminal extension shows an unusual glycine-rich tight turn that fastens around the C-terminal of the protein. This so-called 'closed-loop' is stabilized by a conserved network of interactions involving the guanidinium group of Arg 374 in the turn (Figure 1B and Supplementary Figure S4), which forms two hydrogen bonds: one with Tyr 450 Oη via Hϵ1 (Arg 347 Hϵ1-Tyr 423 Oη in Gbp2) and another through Hη21 protons with Val 426 O (Arg 347 Hη21- Leu 399 CO in Gbp2). The first hydrogen bond probably contributes to slow down the Tyr ring flipping, resulting in the detection of the four ring resonances instead of two (Figure 1B). On the other hand, the interaction involving Hη21 protons reduces the exchange rates enough to allow the observation of the Nη21–Hη21 correlations in the $^{1}$H–$^{15}$N HSQC spectra (Figure 1B).

### Gbp2 and Hrb1 multidomain constructs have different dynamics

We used $^{15}$N relaxation data to investigate the dynamics of the RRMs in different molecular contexts (Supplementary Figures S6 and S7). Gbp2/Hrb1 RRM3 and Hrb1 RRM2 have correlation times ($\tau_c$) of ∼5.5 ns, typical for their molecular weight (∼10 KDa) (Supplementary Figure S8). Conversely, Gbp2 RRM1 has the largest $\tau_c$ of a single-domain construct probably due to the conformational equilibrium described above. Hrb1 RRM1 and Gbp2 RRM2 also have $\tau_c$ larger than expected likely due to the contribution of internal disordered loops or protein aggregation.

Correlation times of the RRMs domains increase in Gbp2/Hrb1 RRM12 constructs (with the exception of

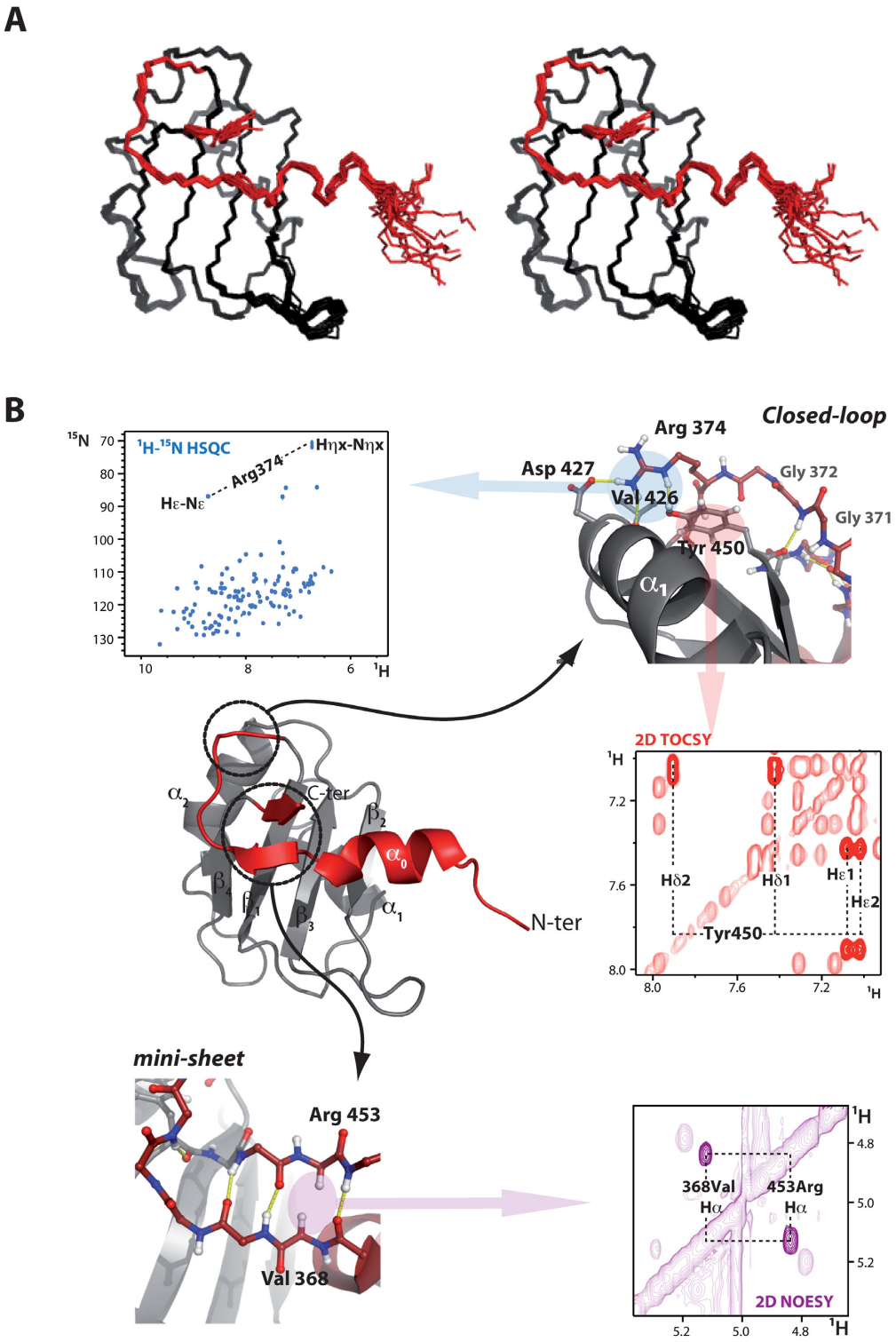


**Figure 1.** (**A**) Stereoview of the NMR structure (20 conformers) of Hrb1 RRM3 (PDB: 2MZT). The C-terminal domain shows a N-terminal extension (in red) also present in the structure of Gbp2 RRM3 (PDB code 2MZQ) (Supplementary Figure S4). (**B**) Structural close-up of the details of the novel element in Hrb1 RRM3. The N-terminal helix packs against the β-sheet. The structure of the 'closed loop' is maintained by a hydrogen bond network involving Arg374, Asp427 and Tyr450. The resonances of some guaninidium group NHs are downfield shifted (Hϵ) or exchange with the solvent slower than usual (Hηx). The observation of the four ring resonances of Tyr450 indicates a slow down on the ring flipping dynamics. Finally the C-terminal part of Hrb1 forms a small two-stranded antiparallel β-sheet spectroscopically characterized by the Hα-Hα interstrand NOE.

**Table 1.** Summary of NMR restrains and structural calculation statistics for different domain structures of Gbp2p and Hrb1p

| | Hrb1 RRM1 | Hrb1 RRM2 | Hrb1 RRM3 | Gbp2 RRM3 |
|---|---|---|---|---|
| **NMR experimental restrains** | | | | |
| *NOE-derived* | | | | |
| Intraresidue | 274 | 190 | 589 | 533 |
| Sequential | 163 | 137 | 540 | 521 |
| Medium-range ($1 < i\text{-}j < 4$) | 164 | 131 | 476 | 476 |
| Long-range ($i\text{-}j > 4$) | 375 | 283 | 1312 | 1370 |
| Total per residue | 13.0 | 10.0 | 30.1 | 30.2 |
| *TALOS+ obtained restrains* | | | | |
| $\phi$ angle restrains | 49 | 58 | 70 | 71 |
| $\psi$ angle restrains | 51 | 56 | 69 | 68 |
| **Structure statistics** | | | | |
| *Mean AMBER energies (kcal/mol ± SD)* | | | | |
| Total | −3772 ± 23 | −2537 ± 12 | −3300 ± 11 | −2925 ± 12 |
| Van der Waals | −643 ± 15 | −599 ± 10 | −764 ± 6 | −751 ± 7 |
| Restrains (distance + angle) | 4 ± 1 | 6 ± 1 | 27 ± 1 | 19 ± 1 |
| *Violations* | | | | |
| Distance[a] | 0.2 ± 0.4 | 1.1 ± 0.2 | 1.5 ± 1.0 | 1.4 ± 0.9 |
| Maximum distance violation (Å) | 0.15 | 0.20 | 0.26 | 0.17 |
| Angle[b] | 0.5 ± 07 | 0.5 ± 0.7 | 0.8 ± 0.9 | 0.2 ± 0.4 |
| Maximum angle violation (°) | 7.6 | 7.4 | 6.6 | 3.6 |
| *Ramachandran Plot analysis (%)*[c] | | | | |
| most favoured regions | 79.2 | 80.2 | 84.3 | 85.1 |
| additionally allowed regions | 17.6 | 16.6 | 13.6 | 12.7 |
| generously allowed regions | 2.0 | 2.4 | 1.7 | 2.1 |
| disallowed regions | 1.3 | 0.8 | 0.4 | 0.2 |
| *RMSD from ideal geometry* | | | | |
| Bond lengths (Å) | 0.010 | 0.010 | 0.009 | 0.009 |
| Bond angles (°) | 1.98 | 2.22 | 1.85 | 2.05 |
| *Averages RMSD to mean structure (range)* | *(162–236)* | *(262–336)* | *(357–454)* | *(330–427)* |
| N, CO, Cα (Å) (± SD) | 1.39 ± 0.40 | 0.72 ± 0.19 | 0.33 ± 0.09 | 0.43 ± 0.08 |
| All heavy (Å) (± SD) | 4.47 ± 2.18 | 3.87 ± 1.63 | 0.72 ± 0.09 | 0.85 ± 0.09 |

[a]Averaged value per structure of distance violations > 0.15 Å ± SD.
[b]Averaged value per structure of total angle violations ± SD.
[c] Obtained from PROCHECK-NMR.

Gbp2 RRM1 affected by the fold/unfold equilibrium) and are in the expected range for a molecule of the size of these tandems (Supplementary Figure S8). This is consistent with a model in which RRM1 and RRM2 contact in the tandem and couple their dynamics. The higher thermal stability of the tandems in comparison to the single-domain constructs (Supplementary Table S1) supports this model.

The correlation times of the domains further increase in Hrb1 RRM123 and nicely correlate to the expected for a spherical particle in the Stokes model (Supplementary Figure S8). In contrast, in Gbp2 RRM123 the rotational dynamics of the RRMs are faster suggesting that RRM3 and RRM12 are largely independent. Indeed the comparison among the $^{1}H$–$^{15}N$ HSQC spectra of Gbp2 RRM123, RRM12 and RRM3 shows fewer differences than the equivalent with Hrb1 constructs (Supplementary Figure S9), suggesting that Hrb1 RRM3 makes contacts with the tandem, which affects both the chemical shifts and the $\tau_c$.

## Gbp2 RRM2 recognises the GGUG core sequence

Gpb2 RNA binding specificity has been investigated *in vitro* (33) and *in vivo* (34) and the protein was early identified as a telomeric DNA binding protein (35). We studied the interaction of various Gbp2 constructs with the RNA probe 5′-UUGGUGUU-3′ which derives from the SELEX consensus sequence (HGGUGW; H = A/C/U W = A/U) (33) and contains the core tetranucleotide signal (GGUG) recently identified *in vivo* by PAR-CLIP (35). Fluorescence anisotropy titrations showed that this RNA interacts with Gbp2 RRM2 (Figure 2A and Table 2) ($K_d \sim 50\ \mu M$). Binding to Gbp2 RRM1 is at least 10 times weaker and Gbp2 RRM3 does not interact in the concentration range tested. RRM12 and RRM123 bind with similar apparent affinity but ∼10 times stronger than RRM2 (Figure 2A and Table 2). The results show that RRM1 cooperates with RRM2 in RNA recognition, a common scenario in RRM tandems, and confirm the neutral role of RRM3. None of the Gbp2 protein constructs bind significantly to the $U_{12}$ and $A_{12}$ RNA probes suggesting that the interaction is sequence-specific (Supplementary Figure S10A). As Gbp2 has been described as a DNA-binding protein we also explored Gbp2 DNA/RNA selectivity. Affinities for the d(TTGGTGTT) probe is ∼5-fold lower than for the cognate RNA for all Gbp2 constructs demonstrating that the protein prefers RNA (Figure 2A and Table 2).

In line with these results, Gbp2 RRM2 $^{1}H$–$^{15}N$ HSQC spectrum shows larger perturbations than Gbp2 RRM1 one upon titration with 5′-UUGGUGUU-3′ (Figure 2B and Supplementary Figure S11). Gbp2 RRM3 NMR spectrum remains unperturbed with this probe and others ($U_{12}$, $A_{12}$, $(UA)_7$ and $(CU)_6$ (data not shown)). The chemical shift perturbation and line broadening effects map to the β-sheet (canonical interface) and to helix α1 (Figure 2C) in Gbp2 RRM2 model. This helix contains a WQxLKD motif that

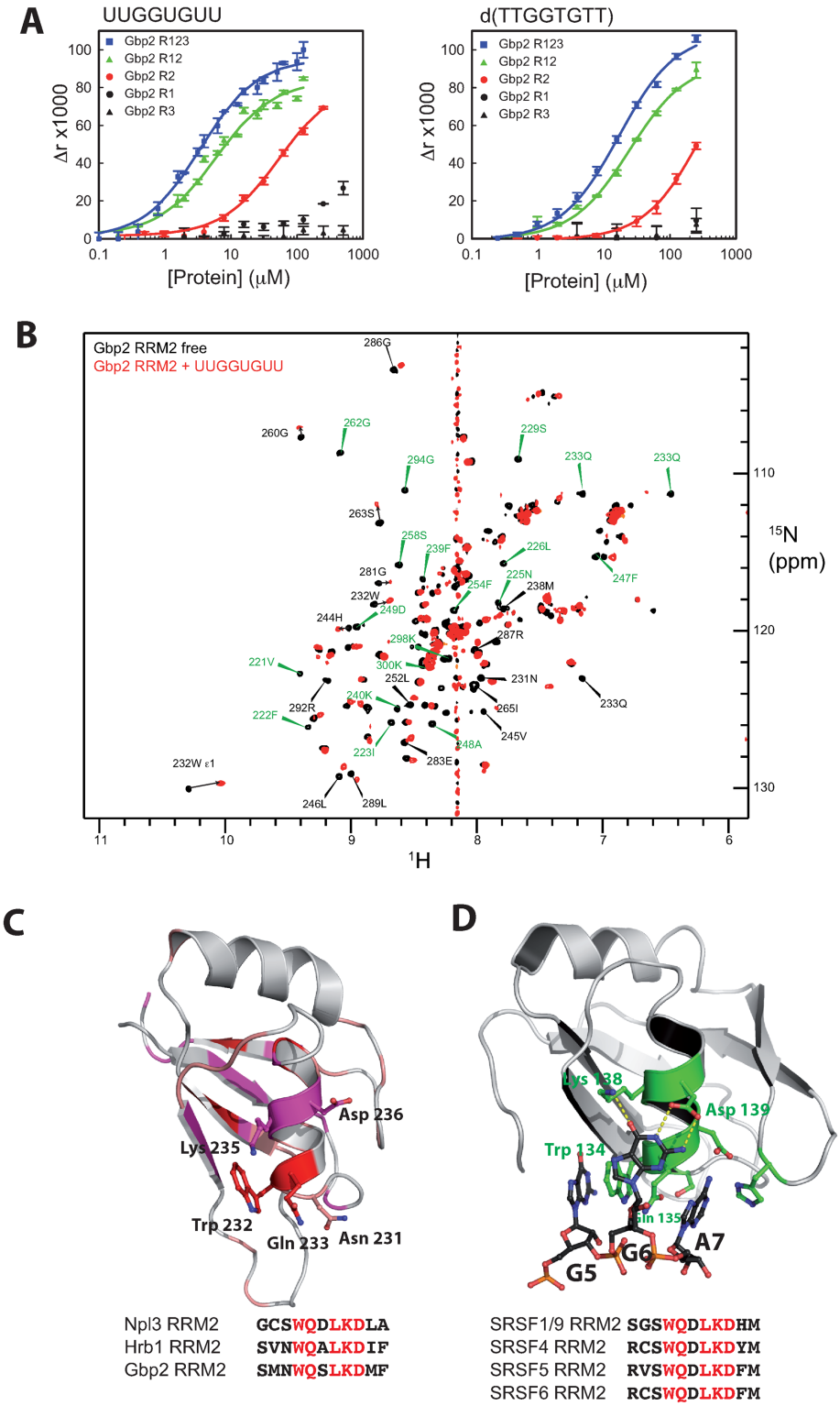


**Figure 2.** Nucleic acids binding mode of Gbp2 RRM2. (**A**) Fluorescence anisotropy binding isotherms obtained upon titration of fluorescein labelled RNA and DNA with various Gbp2 protein constructs (fitted values on Table 2). (**B**) Superposition of full $^{1}H$–$^{15}N$ HSQC spectra of Gbp2 RRM2 on its free (black) and 5-'UUGGUGUU-3′ RNA bound form (red). Black labels identify some crosspeaks experience shift that could be assigned and green labels crosspeaks that disappear upon titration. Chemical shift perturbations are plotted on Supplementary Figure S10. (**C**) Mapping of the CSP (Supplementary Figure S10) on the structure of Gbp2 RRM2 (modelled from Hrb1 RRM2). Residues with broadened HSQC peaks are depicted in magenta, those with CSP > 0.1 ppm in red and those with 0.1 > CSP > 0.05 ppm in light red. The changes span through the β-sheet (canonical RNA binding interface) and through helix α1. (**D**) NMR structure of SRSF1 pseudo-RRM2 in complex with 5′-UGUGGAC-3′ (15) showing side-chains involved in protein RNA-recognition. Hydrogen bonds responsible of G6 specificity are shown in yellow. Equivalent residues are found in Gbp2 RRM2 (**C**) and define a conserved motif in helix α1, which is found in Npl3 and in some SRSF human proteins (**D**).

**Table 2.** Energetic parameters of the interactions between different Gbp2 constructs/mutants and RNA/DNA probes determined by fluorescence anisotropy

| Gbp2 construct | RNA/DNA[a] | $\Delta G$ (kcal.mol$^{-1}$) | $K_D$ (μM) |
|---|---|---|---|
| RRM2 | UUGGUGUU | 5.84 ± 0.05 | 54 ± 4 |
| RRM12 | | 7.2 ± 0.1 | 5.5 ± 0.7 |
| RRM123 | | 7.5 ± 0.2 | 3.3 ± 0.8 |
| RRM1 | | n.d. | > 1000[d] |
| RRM3 | | n.i. | n.i. |
| RRM2 | d(TTGGTGTT) | 4.8 ± 0.2 | 310 ± 90 |
| RRM12 | | 6.31± 0.05 | 24 ± 2 |
| RRM123 | | 6.5 ± 0.1 | 17 ± 2 |
| RRM1 | | n.d. | > 1000 [d] |
| RRM3 | | n.i. | n.i. |
| RRM2 W232A | d(TTGGTGTT) | n.d | > 1000 [d] |
| RRM12 W232A | | 5.9±0.1 | 49±9 |
| RRM2 W232A | UUGGUGUU | n.d | > 1000 [d] |
| RRM12 W232A | | 6.6±0.2 | 15±4 |
| RRM2 | $U_{12}$ | n.d. | > 1000 [d] |
| RRM12 | | n.d. | > 1000 [d] |
| RRM2 | $A_{12}$ | n.d. | > 1000 [d] |
| RRM123 | | n.d. | > 1000 [d] |

[a]Sequences 5′- to 3′ and fluorescein-labeled (at 5′-phosphates).
[b]Free energy values correspond to the dissociation reaction.
[c]Apparent $K_D$ values obtained from the free energies determined as indicated in the main text.
[d]Lower limit estimation obtained by simulation assuming similar fluorescence anisotropy changes.
n.d.: not determined due to insufficient saturation.
n.i.: no interaction was detected. The changes on anisotropy are within the error along the entire concentration range.

is conserved in yeast SR-like and in several members of the SRSF protein family and it is directly involved in RNA recognition in SRSF1 pseudoRRM2 (15) (Figure 2C and D). The mutation of the key residue Trp232 in Gbp2 RRM2 severely impairs RNA/DNA recognition (Table 2 and Supplementary Figure S10B and C). A similar result was reported for SRSF1 (15) strengthening the idea that Gbp2 RRM2 adopts the same binding mechanism. However, this mutation in the Gbp2 RRM12 construct causes a more modest impact showing that the tandem can exploit additional interfaces (probably involving RRM1 and RRM2 β-sheets) that are not accessible to single-domain constructs. It also should be noticed that Gbp2 RRM12 W232A retains sequence specificity (affinity is still tighter than wild-type for $U_{12}$ and $A_{12}$), either because the tandem interface has on itself GU-rich specificity or because it plays a supportive role that alleviates the impact of the mutation on the pseudoRRM interface. Indeed, key RNA recognition residues are still present in the mutant (Figure 2C and D).

### Gbp2 RRM2 binds telomeric DNA

Next we studied Gbp2 binding of telomeric DNA (35). Fluorescence-EMSA show that Gbp2 RRM123 causes a quantitative band shift at equimolecular concentrations and that RRM2 is specifically involved in the interaction (Supplementary Figure S12A). In contrast, Hrb1 binds TG-43 less efficiently and its RRM2 does not exhibit the typical bandshift observed for Gbp2 (Supplementary Figure S12B). Gbp2 RRM123 recognises TG-43 better than Hrb1, and with sub-μM affinity ($K_d$ ∼59 nM) and once again RRM3 seems to be dispensable for binding (Supplementary Figura S13A). The changes in the $^{1}H$–$^{15}N$ HSQC spectrum upon TG-43 binding are located to RRM12 region, whose signals disappear (Supplementary Figure S13B). This global depletion of the NMR signals is also observed for Gbp2 RRM2 (Supplementary Figure S13C) and is compatible with an scenario in which the tandem RRM12 forms large protein/DNA oligomers that increase $\tau_c$ (shorter $T_2$). The unperturbed RRM3 domain does not seem to interact with the DNA and retains substantial mobility, probably because it protrudes out of the oligomer core.

TG-43 DNA forms different types of DNA G-quadruplexes (Supplementary Figure S14) in potassium-containing media with the characteristic circular dichroism band at 260 nm (Supplementary Figure S15 left panel) (36). Gbp2 binding to linear TG-43 (in potassium ion-free solutions) blocks its ability to form G-quadruplexes (Supplementary Figure S15 middle) but at the same time Gbp2 binding to TG-43 already in the quadruplex form cannot melt them (Supplementary Figure S15 right). These data suggest that Gbp2 binding sites overlap (at least partially) with the GGG motifs.

### Gbp2 interacts with THO via its RRM3 domain

As RMM3 of Gbp2/Hrb1 does not appear to bind nucleic acids, we reasoned that it could be involved in protein-protein interactions, as shown for other RRM variants (see (37) and references therein). To investigate this hypothesis, we purified Gbp2 from a Gbp2-TAP strain following the standard tandem affinity purification protocol (29). Consistent with previous investigations (12,38), we detected 2 major bands of high molecular weight that were confirmed by mass spectrometry as THO/TREX subunits Tho2 and Hpr1 (Figure 3A). Next we constructed the Gbp2ΔRRM3-TAP strain, replacing RRM3 by the TAP cassette, and found no evidences of Tho2 and Hrp1 co-purification de-

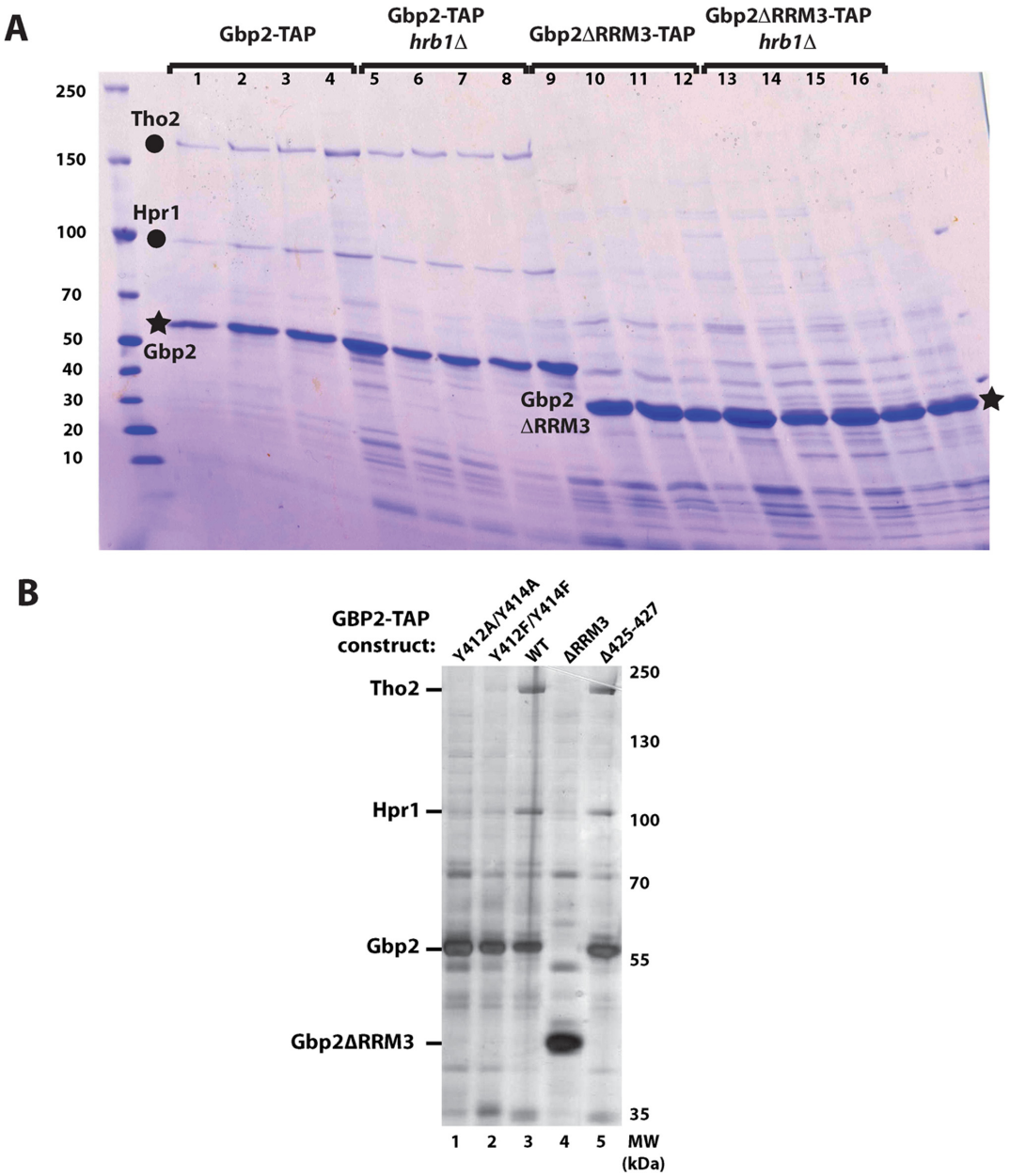


**Figure 3.** (**A**) Gradient (5–20%) SDS-PAGE of the eluted fractions of each TAP purification: lines 1–4 and 5–8 correspond to the GBP2-TAP obtained in the Hrb1 wild type and deletion backgrounds; lines 9–12 and 13–16 correspond to the GBP2ΔRRM3-TAP obtained in the Hrb1 wild type and deletion backgrounds. Stars mark the positions of TAP-tagged proteins and black dots mark label identified (by mass spectrometry) components of THO. Subunits of the THO complex smaller than Gbp2 are poorly visible due to degradation fragments thereof. (**B**) Silver stained gel of the eluted fractions from TAP of various Gbp2 mutants. Similar levels of Tho2 and Hpr1 components of THO complex are observed for wild type and closed-loop mutant Δ425–427; the levels of Tho2 and Hpr1 are much reduced in the Y412F/Y414F double point mutant and not detectable in Y412A/Y414A and ΔRRM3.

spite recovery of similar amount of Gbp2 (Figure 3A). This absence of THO/TREX is not due to competition with Hrb1 as the results are similar in a ΔHrb1 genetic background (Figure 3A), indicating that Gbp2 RRM3 is required for interaction with the THO/TREX complex.

Deletion experiments indicated that the Gbp2 nucleic acid binding tandem is not required for THO/TREX binding as the Tho2 subunit co-purifies with the Gbp2ΔRRM12-TAP (Supplementary Figure S16A). To definitively demonstrate that Gbp2 RRM3 is mediating THO/TREX recruitment we fused it to Npl3, a non-THO/TREX binding SR-like protein (12), and compared the TAP-purified protein profiles of Npl3-TAP and Npl3-Gbp2RRM3-TAP strains. As expected the Tho2 was absent from the Npl3-TAP purification (12) but associated with the Npl3-Gbp2 chimera (confirmed by mass spectrometry analysis) (Supplementary Figure S16B). These important results demonstrate that RRM3 is necessary and sufficient for THO/TREX recognition

and show that this property can be transferred to other molecular contexts.

Given the robustness of the TAP assay, we used it to delineate the interface of Gbp2 RRM3 interacting with THO. Gbp2/Hrb1 contain two distinctive features: the 'closed-loop' revealed by our structural analyses (see above) and a surface motif (YxYGG) in loop 5 (Supplementary Figure S17 A) with the two tyrosines exposed (Supplementary Figure S17 B) highlighted by sequence conservation in fungal orthologs. To investigate which of these two features are relevant for THO recognition, three new Gbp2-TAP mutants were obtained: Gbp2 (Y412F/Y414F) and Gbp2 (Y412A/Y414A) target the conserved tyrosines in the loop 5 motif whereas Gbp2 (Δ425–427) was designed to disrupt the 'closed-loop' by removing the final mini-β-sheet. Substitution of Y412 and Y414 by alanine abolished Tho2/Hrp1 co-purification while their mutation to structurally related phenylalanine residues reduced the presence of Tho2 and Hpr1 in TAP eluates to just above background levels (Figure 3B). The Δ425–427 mutation does not affect THO recruitment, thus demonstrating a specific effect of the tyrosine substitutions. The NMR spectra of both Y412/414 mutants show that their overall structure is the same as wild type (Supplementary Figures S18 and S19) therefore the loss-of-function is not attributable to a destabilization effect. These results strongly suggest that the conserved motif in loop 5 is involved in THO recognition.

### Gbp2 interacts functionally with Tho2

Unlike their metazoan counterparts, only the Npl3 yeast SR-like protein has been implicated in splicing (8). Recently, however, a study described Gbp2 and Hrb1 as quality control/surveillance factors delivering incorrectly processed splicing transcripts to the nuclear exosome (10). Consistent with previous results, we observed that deletion of *GBP2* and *HRB1*, alone or in combination, did not affect constitutive splicing using a variety of reporters (data not shown). Moreover, single and double deletion mutant strains grew on different media and temperature conditions without showing noticeable phenotypes (Supplementary Figure S20) apart from a minor (but reproducible) increase on cell density for late cultures in the Δhrb1Δgbp2 double mutant (Supplementary Figure S20). This result suggests that these proteins have a subtle and long-term negative effect on cell growth consistent with results recently obtained for the Gbp2/Hrb1 homolog in *Aspergillus nidulans* (39).

As Gbp2 and Hrb1 interact with the THO/TREX complex, we next tested for genetic interactions between these genes and *THO2*. In particular, the C-terminal domain of Tho2 was reported to bind RNA/DNA (40) and we reasoned interaction of THO/TREX with Gbp2 and/or Hrb1 could become functionally more important in strains carrying Tho2 mutants with impaired RNA/DNA binding. Thus we crossed *Δtho2*, *tho2Δ*$_{1271–1597}$ or *tho2Δ*$_{1408–1597}$ (40) with a *Δgbp2 Δhrb1* strain. Tetrad dissection revealed that spores carrying the *gbp2* deletion associated with any of the *tho2* allele grew very poorly (Figure 4A and data not shown). This was particularly striking in the case of the *tho2Δ*$_{1408–1597}$ that by itself display only a very weak growth

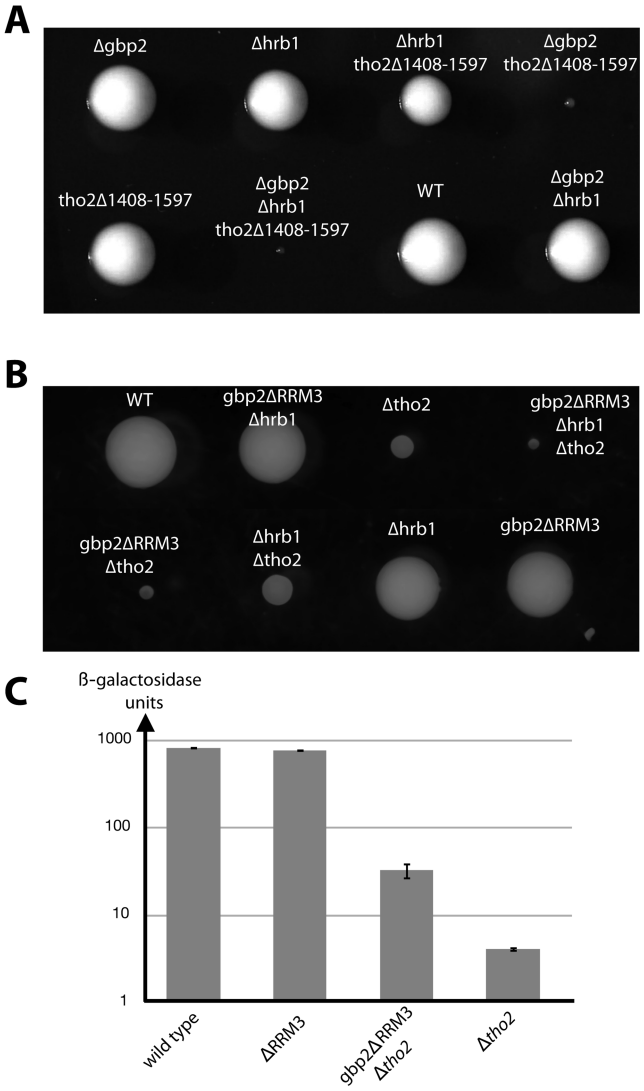


**Figure 4.** (**A**) Growth of spores from tetrad dissection of *Δgbp2 Δhrb1* with *tho2Δ*$_{1408–1597}$. Spores were incubated for 72 h at 30°C. (**B**) Growth of spores from tetrad dissection of *gbp2*ΔRRM3 *Δhrb1* with *Δtho2*. (**C**) Expression of a lacZ reporter in *tho2* and/or *gbp2*ΔRRM3 mutants was monitored by assaying ß-galactosidase activity. Note that the data are plotted on a logarithmic scale. Two biological samples assayed each in duplicate were used to monitor variations.

phenotype ((40), and Figure 4A). In contrast, Δ*hrb1*Δ*tho2*, Δ*hrb1tho2Δ*$_{1271–1597}$ and Δ*hrb1tho2Δ*$_{1408–1597}$ spores grew like single *tho2* mutants and the Δ*hrb1* mutation has little additional impact on the growth of *gbp2*/*tho2* double mutants (Figure 4A and data not shown). This synthetic genetic interaction between *GBP2* and *THO2* reveals a functional link between Gbp2 and the THO/TREX complex and represents a marked difference between *GBP2* and *HRB1*.

To test whether Gbp2 RRM3 is required for Gbp2 function, we crossed a *gbp2*ΔRRM3 Δ*hrb1* strain with a *Δtho2* strain and monitored the growth of the spores after dissection. We observed a synthetic growth phenotype between *gbp2*ΔRRM3 and Δ*tho2* as spores grew very poorly (Fig-

ure 4B). This result indicates that Gbp2 RRM3 is required for the functional interaction of *GBP2* and *THO2*.

Because both Gbp2 and Tho2 are subunits of mRNPs, we tested the impact of Δ*tho2* and *gbp2*ΔRRM3 mutation, alone or in combination, on the expression of the lacZ reporter that is known to be sensitive to Tho2 function (40). The pLS-SD5 reporter was introduced in otherwise wild type, single mutant or double mutant strains and ß-galactosidase expression was assayed (Figure 4C). Consistent with the lack of phenotype of Δgbp2 strains, inactivation of Gbp2 by deletion of RRM3 had almost no effect on lacZ expression. In contrast, deletion of *tho2* reduced lacZ expression by >2 orders of magnitude as previously reported (40). Surprisingly, deleting Gbp2 RRM3 in the Δ*tho2* background partly suppressed the reduction of lacZ expression: 6–8-fold more ß-galactosidase was detected in the Δ*tho2 gbp2*ΔRRM3 double mutant as compared to the Δ*tho2* strain. This observation indicates that besides interacting physically and genetically, the THO/TREX complex and Gbp2 interact functionally contributing to mRNP expression.

## DISCUSSION

The SR protein superfamily comprises a large number of eukaryotic RNA binding proteins (5 families and 11 subgroups (41)) involved in different aspects of mRNA metabolism and characterized by a common architecture composed by RNA binding domains (principally RRM-type) and serine/arginine rich domains. *S. cerevisiae* lacks canonical SR proteins; instead it has three SR-like proteins (Gbp2, Hrb1 and Npl3). The splicing factor Npl3 (8) has two RRMs and is more similar to mammalian SR proteins. Gbp2 and Hrb1 have an additional RRM domain and have been recently linked to pre-mRNA splicing quality control (10). However, the absence of a clear deletion phenotype ((42) and this work) challenges the search of other biological roles of Gbp2/Hrb1.

### RRM domains of Gbp2/Hrb1 are specialized

In our work, we provide an extensive biophysical characterization of Gbp2/Hrb1 RRM domains, finding that the tandem RRM1–RRM2 binds nucleic acids, while RRM3 is a protein-protein interaction platform. Gbp2 binds selectively RNA sequences containing the GGUG element through an interface built by helix α1 and strand β2 of RRM2. This binding mode is equivalent to that of SRSF1 pseudoRRM2 (15), but Gbp2 RRM2 binds weakly (50 μM versus 0.7 μM). The most likely reason is that the SRSF1/RNA interface is larger and involves additional contacts with the α2–β4 loop that are structurally precluded in Gbp2 (because this loop is not long enough). Despite this low affinity Gbp2 has some selectivity as it can discriminate GU-rich from U-rich and A-rich sequences. Because A/U-rich elements are frequently found in yeast 3′-UTR (43) we would expect Gbp2 to be preferentially distributed along the ORFs of mRNAs, which is in fact the case (33,34,44).

Gbp2/Hrb1 RRM2 are probably involved in the recruitment of Sky1 kinase using a binding mode similar to that proposed for SRSF1 RRM2 (15). The protein–protein interface would involve Trp 232 and partially overlaps with the RNA interface. Sky1 phosphorylation occurs in the cytoplasm and is necessary for Gbp2, Hrb1 and Npl3 reimport to the nucleus. Binding to Sky1 'activation loop' by these SR-like proteins might function in two ways: to recruit Sky1 and to disassemble the RNA-binding proteins from the mRNPs facilitating their remodelling in a translation-ready form.

We show that Gbp2 RRM3 is necessary and sufficient for recruitment to the THO/TREX. We confirm that this interaction is RNA/DNA independent (12) because the construct Gbp2ΔRRM12-TAP, lacking the RNA-binding region, co-purifies with THO/TREX. RRM3 contains two new features: the 'closed-loop' and the loop 5 motif (YxYGG). The first one appears to have evolved to suppress the intrinsic RNA binding ability of RRMs by forming a well-defined blockage over the canonical β-sheet interface. There are other examples of non-canonical structural elements interacting with the canonical β-sheet, but to our knowledge none of them disrupt RNA recognition to the level observed in RRM3. Mutational analysis indicates however that this feature is not required to maintain a RRM3 structure competent to interact with the THO complex. The second feature is critical for THO recruitment, in particular the two conserved tyrosine residues. Gbp2/THO interaction is not probably just driven by shape complementarity recognition. Our data reveals that the tyrosine hydroxyl groups are not structurally important for Gbp2 RRM3 fold (Supplementary Figure S18A), but are crucial for THO recognition (Tyr to Phe mutation reduces binding dramatically) very likely by participating to a complex network of hydrogen bond interactions.

The human proteome has around ∼1600 RRM annotated. We found that only SRSF1 RRM1 and SRSF9 RRM1 have a YxYGG motif in loop 5. Moreover, the structure of SRSF1 RRM1 shows remarkable similarities with Gbp2/Hrb1 RRM3 in the region around this loop (Supplementary Figure S17B), which leads us to predict that SRSF1 and SRSF9 might be recruited to the human THO in a similar manner than Gbp2. To our understanding this is a novel hypothesis and might serve to discover new functional links in the SRSF field.

### Gbp2 and Hrb1 participate in different biological processes

We find out that Gbp2 function becomes crucial in spores carrying simultaneously a *tho2* C-terminal truncation that abolished the DNA/RNA binding of THO without compromising its structural integrity. Hrb1 does not appear to contribute to this phenotype, possibly because it is not expressed in these conditions or because it affects transcripts that are not essential for this step. This indicates that Gbp2 and Hrb1 are not functionally redundant at least during spore germination. The hypothesis of a differential specialization of Gbp2 and Hrb1 has been discussed in the past. However only our data and the recent high-resolution transcriptome-wide mapping of Gbp2/Hrb1 binding sites (34) provide direct evidences in this way. In this study, the PAR-CLIP analysis shows a strong correlation between Gbp2/Hrb1 binding sites with those of THO/TREX. Additionally and only for Hrb1, there is also a strong correlation

with other splicing factors, leading the authors to suggest an additional role of Hrb1 in intron removal.

We propose a model in which Gbp2 would function as an adaptor, interacting with RNA during transcription of the coding part of the genes. Giving the ability of Gbp2 to prevent G-quadruplex formation *in vitro*, it may act co-transcriptional to prevent the formation of such stable structures. Gbp2 and Tho2 contain nucleic acid binding domains and it is tempting to speculate that Gbp2 and the THO/TREX complex collaborate in recognizing nascent RNAs (or ssDNA at the transcription bubble). In such context, the observation that a Gbp2 C-terminal truncation partially suppresses the reduced ß-galactosidase expression observed in *Δtho2* strain is unexpected. However, Gbp2 has a putative second RNA binding region - the N-terminal SR-rich region-, besides its tandem RRM12, and previous studies have noticed the remarkable similarities between these regions and the C-terminal of Tho2 (1410–1597) (45). In the absence of Tho2, the Gbp2 SR-rich region might usurp the role of Tho2 C-terminal domain and impair gene expression. A second working hypothesis would be that the partial suppression of the *Δtho2* phenotype in the Gbp2 mutant is due to the lack of recruitment of THO/TREX. A damaged THO/TREX complex could block downstream steps of mRNP maturation and/or expression. Alternatively Gbp2 may efficiently recruit factors negatively impacting mRNA expression that are usually outcompeted in the presence of the THO/TREX complex. Future studies will reveal which of these mechanisms (or others) better describes the relationship between Gbp2/THO at molecular level.

## ACCESSION NUMBERS

Atomic coordinates for the NMR structure ensembles have been deposited on the PDB: Hrb1 RRM1 (2MZR), Hrb1 RRM2 (2MZS), Hrb1 RRM3 (2MZT) and Gbp2 RRM3 (2MZQ). NMR data are available from the Biological Magnetic Resonance Data Bank (accession numbers: 25497 for Hrb1 RRM1, 25498 for Hrb1 RRM2, 25499 for Hrb1 RRM3 and 25496 for Gbp2 RRM3).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the members of our groups for discussion and advice, C. Faux and D. Velázquez for technical assistance and IGBMC services for support.

## FUNDING

Spanish Ministerio de Economía y Competitividad (MINECO) [CTQ2011-26665 and CTQ2014-52633-P to J.M.P.C.]; Comunidad Autónoma de Madrid [CPI/0265/2008 to S.M.L.]; Federation of European Biochemical Societies (FEBS) [FEBS Summer Fellowship to S.M.L.]; Ligue Contre le Cancer (Equipe Labellisée 2014); Centre National pour la Recherche Scientifique; CERBM-IGBMC; Investissements d'Avenir ANR-10-IDEX0002-02 [ANR-10LABX-0030-INRT]; French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; INSTRUCT as part of the European Strategy Forum on Research Infrastructures. The open access publication charge for this paper has been waived by Oxford University Press - *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Long,J.C. and Cáceres,J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
2. Shepard,P.J. and Hertel,K.J. (2009) The SR protein family. *Genome Biol.*, **10**, 242.
3. Busch,A. and Hertel,K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley interdiscipl. Rev. RNA*, **3**, 1–12.
4. Lin,S. and Fu,X.D. (2007) SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.*, **623**, 107–122.
5. Zhou,Z. and Fu,X.D. (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, **122**, 191–207.
6. Twyffels,L., Gueydan,C. and Kruys,V. (2011) Shuttling SR proteins: more than splicing factors. *FEBS J.*, **278**, 3246–3255.
7. Roth,M.B., Zahler,A.M. and Stolk,J.A. (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J. Cell Biol.*, **115**, 587–596.
8. Kress,T.L., Krogan,N.J. and Guthrie,C. (2008) A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol. Cell*, **32**, 727–734.
9. Dujon,B. (2010) Yeast evolutionary genomics. *Nat. Rev. Genet.*, **11**, 512–524.
10. Hackmann,A., Wu,H., Schneider,U.M., Meyer,K., Jung,K. and Krebber,H. (2014) Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat. Commun.*, **5**, 3123.
11. Häcker,S. and Krebber,H. (2004) Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J. Biol. Chem.*, **279**, 5049–5052.
12. Hurt,E., Luo,M.J., Röther,S., Reed,R. and Strässer,K. (2004) Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 1858–1862.
13. Windgassen,M., Sturm,D., Cajigas,I.J., González,C.I., Seedorf,M., Bastians,H. and Krebber,H. (2004) Yeast shuttling SR proteins Npl3p, Gbp2p, and Hrb1p are part of the translating mRNPs, and Npl3p can function as a translational repressor. *Mol. Cell. Biol.*, **24**, 10479–10491.
14. Baierlein,C., Hackmann,A., Gross,T., Henker,L., Hinz,F. and Krebber,H. (2013) Monosome formation during translation initiation requires the serine/arginine-rich protein Npl3. *Mol. Cell. Biol.*, **33**, 4811–4823.
15. Cléry,A., Sinha,R., Anczuków,O., Corrionero,A., Moursy,A., Daubner,G.M., Valcárcel,J., Krainer,A.R. and Allain,F.H. (2013) Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E2802–E2811.
16. Neidhardt,F.C., Bloch,P.L. and Smith,D.F. (1974) Culture medium for enterobacteria. *J. Bacteriol.*, **119**, 736–747.
17. 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.
18. 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.
19. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.

20. 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.
21. 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.
22. 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.
23. Güntert,P., Mumenthaler,C. and Wüthrich,K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
24. Case,D.A., Darden,T.A., Cheatham,T.E.I., Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Merz,K.M., Pearlman,D.A., Crowley,M. *et al.* (2006). University of California, San Francisco.
25. Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.
26. Royer,C.A., Smith,W.R. and Beechem,J.M. (1990) Analysis of binding in macromolecular complexes: a generalized numerical approach. *Anal. Biochem.*, **191**, 287–294.
27. Baudin-Baillieu,A., Guillemet,E., Cullin,C. and Lacroute,F. (1997) Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast*, **13**, 353–356.
28. Puig,O., Rutz,B., Luukkonen,B.G., Kandels-Lewis,S., Bragado-Nilsson,E. and Séraphin,B. (1998) New constructs and strategies for efficient PCR-based gene manipulations in yeast. *Yeast*, **14**, 1139–1146.
29. Rigaut,G., Shevchenko,A., Rutz,B., Wilm,M., Mann,M. and Séraphin,B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, **17**, 1030–1032.
30. Guarente,L., Yocum,R.R. and Gifford,P. (1982) A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 7410–7414.
31. Ito,H., Fukuda,Y., Murata,K. and Kimura,A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.*, **153**, 163–168.
32. Dreumont,N. and Séraphin,B. (2013) Rapid screening of yeast mutants with reporters identifies new splicing phenotypes. *FEBS J.*, **280**, 2712–2726.
33. Riordan,D.P., Herschlag,D. and Brown,P.O. (2011) Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. *Nucleic Acids Res.*, **39**, 1501–1509.
34. Baejen,C., Torkler,P., Gressel,S., Essig,K., Söding,J. and Cramer,P. (2014) Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Mol. Cell*, **55**, 745–757.
35. Lin,J.J. and Zakian,V.A. (1994) Isolation and characterization of two Saccharomyces cerevisiae genes that encode proteins that bind to (TG1–3)n single strand telomeric DNA in vitro. *Nucleic Acids Res.*, **22**, 4906–4913.
36. Kypr,J., Kejnovská,I., Renciuk,D. and Vorlícková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
37. Tripsianes,K., Friberg,A., Barrandon,C., Brooks,M., van Tilbeurgh,H., Séraphin,B. and Sattler,M. (2014) A novel protein-protein interaction in the RES (REtention and Splicing) complex. *J. Biol. Chem.*, **289**, 28640–28650.
38. Oeffinger,M., Wei,K.E., Rogers,R., DeGrasse,J.A., Chait,B.T., Aitchison,J.D. and Rout,M.P. (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods*, **4**, 951–956.
39. James,S.W., Banta,T., Barra,J., Ciraku,L., Coile,C., Cuda,Z., Day,R., Dixit,C., Eastlack,S., Giang,A. *et al.* (2014) Restraint of the G2/M transition by the SR/RRM family mRNA shuttling binding protein SNXAHRB1 in Aspergillus nidulans. *Genetics*, **198**, 617–633.
40. Peña,A., Gewartowski,K., Mroczek,S., Cuéllar,J., Szykowska,A., Prokop,A., Czarnocki-Cieciura,M., Piwowarski,J., Tous,C., Aguilera,A. *et al.* (2012) Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J.*, **31**, 1605–1616.
41. Richardson,D.N., Rogers,M.F., Labadorf,A., Ben-Hur,A., Guo,H., Paterson,A.H. and Reddy,A.S. (2011) Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS One*, **6**, e24542.
42. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Véronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., André,B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387–391.
43. Tian,B. and Graber,J.H. (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscipl. Rev. RNA*, **3**, 385–396.
44. Tuck,A.C. and Tollervey,D. (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell*, **154**, 996–1009.
45. Gewartowski,K., Cuéllar,J., Dziembowski,A. and Valpuesta,J.M. (2012) The yeast THO complex forms a 5-subunit assembly that directly interacts with active chromatin. *Bioarchitecture*, **2**, 134–137.