Published online 23 June 2014

Nucleic Acids Research, 2014, Vol. 42, No. 13 ***8705–8718***
doi: 10.1093/nar/gku468

Structural basis of nucleic acid binding by *Nicotiana tabacum* glycine-rich RNA-binding protein: implications for its RNA chaperone function

Fariha Khan[1,2], Mark A. Daniëls[1], Gert E. Folkers[1], Rolf Boelens[1], S. M. Saqlan Naqvi[2,*] and Hugo van Ingen[1,*]

[1]NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands and [2]Department of Biochemistry, PMAS Agriculture University Rawalpindi, 46300 Rawalpindi, Pakistan

Received October 15, 2013; Revised April 30, 2014; Accepted May 12, 2014

ABSTRACT

Glycine-rich RNA-binding proteins (GR-RBPs) are involved in cold shock response of plants as RNA chaperones facilitating mRNA transport, splicing and translation. GR-RBPs are bipartite proteins containing a RNA recognition motif (RRM) followed by a glycine-rich region. Here, we studied the structural basis of nucleic acid binding of full-length *Nicotiana tabacum* GR-RBP1. NMR studies of *Nt*GR-RBP1 show that the glycine-rich domain, while intrinsically disordered, is responsible for mediating self-association by transient interactions with its RRM domain (*Nt*RRM). Both *Nt*GR-RBP1 and *Nt*RRM bind specifically and with low micromolar affinity to RNA and single-stranded DNA. The solution structure of *Nt*RRM shows that it is a canonical RRM domain. A HADDOCK model of the *Nt*RRM–RNA complex, based on NMR chemical shift and NOE data, shows that nucleic acid binding results from a combination of stacking and electrostatic interactions with conserved RRM residues. Finally, DNA melting experiments demonstrate that *Nt*GR-RBP1 is more efficient in melting CTG containing nucleic acids than isolated *Nt*RRM. Together, our study supports the model that self-association of GR-RBPs by the glycine-rich region results in cooperative unfolding of non-native substrate structures, thereby enhancing its chaperone function.

INTRODUCTION

Gene expression is regulated to ultimately direct growth, development and stress responses. RNA-binding proteins (RBPs) are known to perform multifunctional roles in post-transcriptional RNA metabolism, including pre-mRNA splicing, RNA export, polyadenylation, degradation and translation. RBPs contain RNA-binding domains such as RNA recognition motifs (RRMs) commonly referred to as RNA-binding domain or ribonucleoprotein (RNP), Piwi/Argonaute/Zwille, K-homology, double-stranded RNA-binding domain, arginine–glycine–glycine and RNA helicase DEAD box (1). The RRM domain is a very versatile and widespread RNA-binding domain in eukaryotes (2,3). Glycine-rich RNA-binding proteins (GR-RBPs, sometimes referred to as hnRNPs-like proteins or abbreviated as GRPs or RGPs) are RBPs containing an N-terminal RRM domain and a C-terminal glycine-rich domain that also contains arginine and aromatic residues. Members of the GR-RBP family were first reported in maize (4) and subsequently identified in various organisms from yeast and algae to mammals (5). Ubiquitously present in both angiosperm and gymnosperm flowering plants, their expression levels are regulated in response to a variety of biotic and abiotic stimuli such as drought, salinity, flooding, wounding, pathogen attack and cold (6–10). In *Arabidopsis thaliana* eight GR-RBPs are identified that fulfil different roles (11). *At*GR-RBP7 is a circadian regulator (12, 13) that regulates its own expression and that of *At*GR-RBP8 (14) through a negative feedback mechanism, where elevated protein levels promote alternative splicing through pre-mRNA binding. These splicing variants are then degraded in nonsense mediated decay pathway (14), which likely also holds true for some other *At*GR-RBP7 target genes (15). In addition, *At*GR-RBP7 is reported as a flow-

*To whom correspondence should be addressed. Tel: +31 30 2533641; Fax: +31 30 2537623; Email: hugo.van.ingen@gmail.com
Correspondence may also be addressed to S. M. Saqlan Naqvi. Tel: +92 51 9290467; Fax: +92 51 9290160; Email: saqlan@uaar.edu.pk
Present address:
Fariha Khan, Department of Biochemistry, Pir Mehr Ali Shah, Arid Agriculture University Rawalpindi, 46300 Rawalpindi, Pakistan.

ering timing gene (16) and a stress-related gene involved in defence (17) and cold response (18, 19). In particular, response to cold shock is a complex event in higher plants, in which not only canonical cold shock proteins (CSPs) but also GR-RBPs are involved. Canonical CSPs function as RNA chaperones (20) and GR-RBPs are thought to function in a similar manner. This is evidenced by several studies in which deletion of GR-RBP genes causes reduced growth rates under cold stress that can be rescued upon their heterologous expression (19, 21). Furthermore, GR-RBPs can functionally substitute for *Escherichia coli* CSPs (22). Together, these data suggest that GR-RBPs can act as RNA chaperones to facilitate mRNA transport, and ultimately translation under stress conditions.

Here, we present the first high-resolution structural study of a plant GR-RBP. Using a combination of nuclear magnetic resonance (NMR) spectroscopy, computational modelling and DNA melting assays, we studied the structural basis of the nucleic acid interaction of *Nicotiana tabacum* GR-RBP1 (abbreviated as *Nt*GR-RBP1), a homologue of *At*GR-RBP7. We solved the solution structure of the RRM domain of *Nt*GR-RBP1 (designated as *Nt*RRM), characterized in detail its interaction with RNA and single-stranded DNA (ssDNA), and constructed a structural model for *Nt*RRM–RNA complex. Both *Nt*RRM and full-length *Nt*GR-RBP1 bind specifically and with low micromolar affinity to nucleic acids, resulting from a combination of stacking and electrostatic interactions. Studies with full-length *Nt*GR-RBP1 further indicate that the glycine-rich domain is responsible for self-association. Correspondingly, *Nt*GR-RBP1 is more efficient in melting DNA than isolated *Nt*RRM. In conclusion, we find that the glycine-rich region of *Nt*GR-RBP1 binds transiently and weakly to its canonical RRM domain. *Nt*GR-RBP1 self-association provides a mechanism for cooperative unfolding of non-native substrate RNA structures, thereby enhancing its chaperone function.

MATERIALS AND METHODS

Recombinant protein expression

The RRM domain of GR-RBP1 of *Nicotiana tabacum*, herein referred to as *Nt*RRM (aa 1–85, GenBank accession number ADG03637.1), was cloned into pDEST-HisMBP as a dual tagged construct. Full-length *Nt*GR-RBP1 (aa 1–156) was cloned as a glutathione-S-transferase (GST) fusion into pLICHISGST (23). Transformed *E. coli* BL21(DE3) were grown in M9 minimal medium with $^{15}NH_4Cl$ and/or ^{13}C-glucose to OD_{600} of 0.6 at 37°C, induced with 1 mM IPTG, after which proteins were expressed for 5 h at 25°C.

Protein purification

Cells were lysed by repeated freeze thawing and sonication. The HisMBP-*Nt*RRM fusion protein was bound to Ni-NTA-agarose matrix (Qiagen) in 20 mM Tris pH 8, 100 mM NaCl, 5% glycerol and 1 mM β-mercaptoethanol (BME), washed with the same buffer containing 1% glycerol and 10 mM imidazole, and subsequently eluted in the same buffer with 300 mM imidazole. The tobacco etch virus cleavage site was used to free the RRM domain from HisMBP by digestion for overnight along with dialysis to 20 mM Tris pH 7, 100 mM NaCl, 1 mM BME. The *Nt*RRM domain was subsequently purified over a Sephadex-75 (HiLoad 16/60) column, equilibrated in the same buffer. Fractions containing *Nt*RRM were concentrated and dialysed to NMR buffer (20 mM NaPi pH 7, 100 mM NaCl, 1 mM BME). The final product contains one additional residue (G0) at the N-terminus from the cleavage site. Full-length GST-*Nt*GR-RBP1 was purified at room temperature by binding to a Glutathione agarose column (Sigma) in 20 mM Tris pH 8, 200 mM KCl, 1% glycerol, 1 mM BME and eluted in the same buffer with 300 mM KCl and 50 mM reduced glutathione (Sigma). After 3 h of thrombin digestion, *Nt*GR-RBP1 was purified over a Sephadex-75 (HiLoad 16/60) column, equilibrated with 20 mM Tris pH 7, 100 mM KCl, 1 mM BME. Fractions containing *Nt*GR-RBP1 were concentrated and dialysed to NMR buffer (20 mM KPi pH 7, 100 mM KCl, 1 mM BME). The final product contains two additional residues (GS) at the N-terminus from the cleavage site.

NMR spectroscopy and resonance assignment

Samples for assignment and structure determination contained ~400 μM U-^{15}N/^{13}C labelled *Nt*RRM domain in 90/10% H_2O/D_2O in NMR buffer. Backbone assignment was based on HNCACB, CBCACONH, HNCA, HNCOCA and HNCO spectra and side chain resonances were assigned with HBHANH, CCH-TOCSY, HCH-TOCSY, CBHD and NOESY spectra, essentially as described in (24). Overall assignment completeness for the unbound *Nt*RRM domain was 97.6% for all non-labile protons. Assignment of the 6-nt ssDNA was based on homonuclear NOESY and TOCSY spectra in combination with natural abundance ^{1}H-^{13}C-HSQC. Assignment of the ssDNA bound *Nt*RRM domain was based on the titration data, and overlays of the unbound and bound ^{1}H-^{13}C-HSQC spectra in combination with ^{15}N- and ^{13}C-edited 3D-NOESY spectra (24). Intermolecular distance restraints were derived from a 2D NOESY with ^{15}N/^{13}C filter in F2 (25).

NMR sample of full-length *Nt*GR-RBP1 contained 80 μM protein in 90/10% H_2O/D_2O in NMR buffer. Partial assignments were obtained based on overlays with assigned spectra from isolated *Nt*RRM domain. All NMR spectra were recorded at 298 K on either 600 or 750 MHz Bruker Avance II spectrometers, processed using the NMRPipe package (26) and analysed using CcpNmr Analysis 2.2 (27).

Structure calculation and refinement of *Nt*RRM

Backbone dihedral angle restraints were derived using TALOS+ (28). Distance restraints were derived from ^{13}C- and ^{15}N-edited 3D NOESY spectra (mixing time 100 ms). The NOE cross peaks were assigned and converted into distance restraints using CYANA 3.0 (29,30). First, 10 ensembles of 100 structures were calculated by using CYANA using different random number seeds. Out of the 10 resulting distance restraint lists, only the restraints that were reproduced

in all cases were retained to produce a final restraint list. This final list was then used to calculate 100 structures in CNS 1.2 (31), which were subsequently refined in explicit water by using the RECOORD protocol (32). The final ensemble contained the 20 lowest-energy structures, contained neither distance violations > 0.5 Å nor dihedral angle violation > 5°, and was validated by using the iCing validation suite (33). Open-source PyMOL (The PyMOL Molecular Graphics System, Version 1.4 Schrödinger, LLC) was used to generate molecular graphics.

Titration experiments and data analysis

Interaction between *Nt*RRM and ssDNA or RNA was studied using a 6-nt probe (5′-TTCTGG-3′ for DNA and 5′-UUCUGG-3′ for RNA; Eurofins MWG Operon) that was previously identified as a minimum binding sequence for homologue *At*GR-RBP7 (34). The RNA and DNA probes were either ordered as HPLC purified (RNA) or purified over a G10 column (DNA), lyophilized and dissolved in NMR buffer to a stock concentration of 0.18 mM (RNA) or 1 mM (DNA). NMR spectra (1D; natural abundance ^{13}C-HSQC and TOCSY) of the unlabelled oligos confirmed the purity of the material, without detectable amounts of truncated species. Titrations with ssDNA and RNA were carried out at a protein concentration of 150 and 35 μM, respectively. Observed line shapes in the titration were fitted using MatLAB scripts (MATLAB version 7.13.0, The MathWorks Inc., 2011) using explicit evaluation of the exchange matrix for a 1:1 binding model, and subsequent calculation of the free induction decay and processing to a final spectrum (see Supporting Materials in Kato *et al.* (35) for details). The ^{1}H and ^{15}N line shapes of nine residues with the largest changes in chemical shifts between free and bound, and without overlap during the titration were simultaneously fitted to a single dissociation constant K_D and dissociation rate k_{off}.

The interaction between *Nt*RRM and NtGR-RBP1 was studied in 20 mM KPi pH 7, 100 mM KCl, 1 mM BME, 20°C at 128 and 48 μM ^{15}N-labelled *Nt*RRM using a 600 MHz spectrometer. Half of the sample was removed and replaced by equal concentration of unlabelled *Nt*GR-RBP1 in the same buffer. In a control experiment, the extracted volume was replaced by buffer.

Modelling of *Nt*RRM–RNA complex

The experimental chemical shift perturbations (CSPs), intramolecular and intermolecular NOESY data obtained for the *Nt*RRM–ssDNA complex were used to create a structural model for the *Nt*RRM–RNA complex with Haddock version 2.1 (36, 37) and CNS 1.3 (31). In order to sample the conformational space of the RNA reliably, the RNA conformation in six homologous RRM–RNA complexes (PDB-id 2KM8 (38), 2KXN (39), 2RS2 (40), 2LEB, 2LEC (41) and 4F02 (42) was used to build the 5-UUCUGG-3′ fragment. The observed ssDNA NOEs were translated into the corresponding RNA restraints (40 in total) and subsequently used to refine the RNA conformations and as input for docking to the unbound *Nt*RRM domain (see Supplementary Figure S5b), using the ensemble of 20 structures. This procedure ensures sampling of large conformational space for the RNA. The docking was driven using CSP-derived ambiguous interaction restraints and three intermolecular NOEs involving G5, of which two were unambiguously assigned (I38δ1-G5-H1′ and I38δ1-G5-H4′) and one ambiguously assigned (F9/F49/F51 aromatic protons to G5-H1′). RRM or DNA residues with CSPs larger than the 10% trimmed mean + 2σ and a surface accessible area of more than 33% were defined as active residues. Neighbouring solvent exposed residues were defined as passive residues. In the rigid body docking phase, 600 solutions were calculated (5 per combination of *Nt*RRM and RNA input structure), of which the best 200 structures according to their HADDOCK score were refined in the (semi-)flexible phase. Finally, the 100 structures with the lowest HADDOCK score were further refined in explicit solvent. The resulting models were clustered using 7.5 Å ligand RMSD cut-off and analysed for violations against the intermolecular NOEs. In total 85% of the solutions were clustered into four sets of structures, containing respectively 54, 20, 7 and 4 structures. The representative structures of each cluster are compared in Supplementary Figure S5c. Since the overall scores are rather similar for the four clusters, the models in the dominant cluster were selected to represent the *Nt*RRM–RNA structure. These solutions did not contain intermolecular NOE violations > 0.5 Å. The best scoring 20 models were used for calculation of residue-specific contribution to the intermolecular binding energy.

Electrophoretic mobility shift assay

The ssDNA probes 5′-ATTTTGTTCTGGT-3′ (ssP1) and 5′-ATTTTGTTCTGGTATTTTGTTCTGGT-3′ (ss-dP1) were radioactively labelled with ^{32}P-γATP using T4 polynucleotide kinase. Two nanomolar of gel purified DNA probe was incubated with the indicated amount of full-length *Nt*GR-RBP1 or the *Nt*RRM domain in a buffer containing 10 mM Tris 7.0, 100 mM NaCl, 10 % glycerol (w/v) and 1 mM BME for 30′ on ice. Best separation of protein-DNA complexes was obtained on a 0.5×TBE (40 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.3) buffered 16% acrylamide gel running at 4°C at 160 V for 3 h. After drying, the gel was exposed to a phosphor imager screen and analysed using a GE healthcare Typhoon FLA 7000. Quantification was performed using ImageJ (http://imagej.nih.gov/ij/). Irrespective of the gel system or acrylamide concentration, the *Nt*GR-RBP1–ssDNA complex dissociated during electrophoresis. Decreasing the NaCl concentration to 10 mM in the binding buffer, decreasing the TBE in the electrophoresis buffer to 0.25×, or performing the binding and electrophoresis at room temperature still resulted in formation of a complex with significant dissociation during electrophoresis. The apparent binding affinity was estimated by quantification of the remaining amount of free DNA.

DNA melting studies

Annealed fully complementary DNA oligos (probe I—corresponding to the *At*GR-RBP7 binding site in its 3′UTR (34): 5′-ATT TTG TTC TGG T-3′; probe II—a

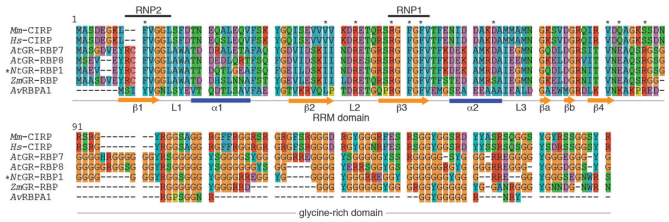

Figure 1. *Nt*GR-RBP1 is highly conserved from bacteria to human. The alignment is generated by CLUSTAL W and displayed by Seaview with colour coding according to amino acid properties. The location of the RNP motifs and the glycine-rich region is indicated. Secondary structure elements as present in the structure of the *Nt*RRM domain of *Nt*GR-RBP1 are indicated below the alignment and labelled as in Figure 2 (orange arrows: β-strand; blue bars: α-helix). Residues that were found to be in the nucleic acid interaction surface are indicated with *. CIRP = cold inducible RNA-binding protein; GR-RBP = glycine-rich RNA-binding protein; RBP = RNA-binding protein.

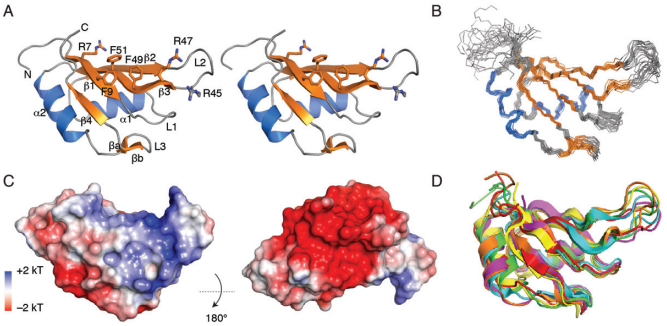

Figure 2. Solution structure of *Nt*RRM domain. (**A**) Cross-eye stereo view in cartoon representation of the lowest-energy structure. Colour coding: orange—β-sheet; blue—α-helix; grey—other. Secondary structure elements were assigned by DSSP. Side chains of exposed residues of the RNP motif are shown as sticks. (**B**) Ensemble of 20 best structures. Colour coding as in (A). (**C**) van der Waals surface of the *Nt*RRM domain colour coded according to electrostatic potential. (**D**) Structural superposition of *Nt*RRM (green) with five most homologous structures in the PDB: 1 × 5S (yellow): human CIRP (cold inducible RNA-binding protein); 2CQD (red): human RBM38 (RNA-binding motif protein 38); 2RS2 (40) (magenta): mouse Msi1 (Musashi RNA-binding protein1); 3S7R (orange): human HNRPAB (Heterogeneous nuclear ribonucleoprotein A/B); 2DH8 (cyan): human DAZAP1 (deleted in azoospermia associated protein1).

homologous sequence in *Nt*GR-RBP1: 5′-ATT TAT GGT TCT AAG-3′; probe III—randomly selected sequence: 5′-AGA CGA GAT ACT A-3′; Eurogentec, Belgium) were lyophilized and dissolved in NMR buffer to a concentration of 2 μM, either with or without 3 M equivalents of *Nt*RRM or full-length *Nt*GR-RBP1 added. As a control bovine serum albumin (BSA) was used. Temperature melting curves were measured in a UV/Vis Spectrophotometer (Perkin-Elmer Lambda 18, equipped with a Peltier heating element) by monitoring A_{260nm} every 2 s from 15°C to 55°C with temperature ramp of 0.5°C/min and 0.5 s integration time in a 1 cm quartz cuvette with stirrer. Experimental curves were smoothed and down-sampled by retaining only the minimum value in a 11 point window (corresponding to 0.18°C increase in temperature) in order to remove the noise from the stir-bar, blanked against the corresponding control curve without DNA, normalized and fitted as a function of temperature (T) using LmFit (http://newville.github.io/lmfit-py/index.html) and SciPy (http://www.scipy.org) packages for Python (http://www.python.org) to a generalized-logistic function including a linear term compensating for drift:

$$A_{260} = a + \frac{b}{\left(1 + e^{(T_m - T)/c}\right)^d} + eT$$

where *a, b, c, d, e* and T_m are fitting parameters.

RESULTS

Tobacco *Nt*GR-RBP1 is a ∼16 kDa protein comprised of an *Nt*RRM domain (85 residues) followed by a glycine-rich region of roughly the same length. Sequence alignment shows that *Nt*GR-RBP1 is highly conserved with orthologous in *Arabidopsis* and *Zea mays* sharing 76% and 73% amino acid identity, respectively, and ∼40% homology to mouse, human and bacterial counterparts (Figure 1). Sequence conservation is highest in the *Nt*RRM domain which features the two canonical RNP motifs that have been shown to be required for RNA binding in other RRM domains (2).

*Nt*GR-RBP1 contains a canonical *Nt*RRM domain

We first solved the solution structure of the *Nt*RRM domain of *Nt*GR-RBP1 (assignments of the ^{1}H-^{15}N-HSQC are given in Supplementary Figure S1). The backbone traces of the lowest-energy structure and the ensemble of structures of the free *Nt*RRM domain are shown in Figure 2A and B. Overall, the structure is well defined and has favourable structural statistics (Table 1 and Supplementary Table S1). It is folded as a canonical RRM domain (2) of which the core is formed of two α-helices (α1 and α2) packed across a four-stranded (β1–β4) anti-parallel β-sheet in a βαββαβ topology. Additionally, the *Nt*RRM contains three loops (L1–L3) and a short two-stranded β-sheet in loop L3. Analysis of the backbone chemical shifts according to the random coil index (43) indicate that the core is rigid with predicted order parameters S^2_{pred} of 0.8–0.9 (Supplementary Figure S2). The terminal residues and residues 45–48 of loop L2 have S^2_{pred} of 0.1–0.7, indicating increased flexibility, which correlates well with the lower structural definition of these regions in the ensemble of structures. The central β-sheet features the conserved RNP1 and RNP2 motifs with its aromatic residues F9, F49 and F51 exposed to the solvent (Figure 2A). This surface is furthermore characterized by positive electrostatic potential, whereas the rest of the protein has mostly negative potential (Figure 2C). Structural superposition of *Nt*RRM with structures of five RRM domains with the highest homology shows that it conforms well to the established RRM fold despite the low overall sequence identity of ∼40% (Figure 2D and Supplementary Figure S3).

*Nt*RRM domain binds with low micromolar affinity to nucleic acids

The *in vivo* target RNA sequence of most GR-RBPs, including *Nt*GR-RBP1, remains to be established. For *At*GR-

Table 1. Structural statistics for the free *Nt*RRM domain

A. Number of restraints	
Total number of NOE-based distance restraints	1386
Intra-residual/sequential/medium/long	351/353/200/464
TALOS derived dihedral angle restraints ϕ/ψ	69/69
B. Average RMS deviation from experimental restraints	
All experimental distance restraints (Å)	0.013 ± 0.002
All dihedral angle restraints (°)	0.36 ± 0.07
C. Coordinate RMS deviation (Å)	
Average RMSD to mean	
Ordered backbone atoms	0.62 ± 0.13
Ordered heavy atoms	1.28 ± 0.16
Global backbone atoms	1.31 ± 0.29
Global all heavy atoms	1.77 ± 0.23
D. iCing ROG score (red/orange/green%)	
Ordered regions	5/31/64
Global	10/30/60
E. Ramachandran quality parameters (%)	
Residues in most favoured regions	86.6
Residues in allowed regions	12.0
Residues in additionally allowed regions	0.9
Residues in disallowed regions	0.4

[a]Statistics are given for residues 1–85 of *Nt*RRM. Ordered regions are residues 6–83 of the *Nt*RRM domain, excluding the disordered termini and loop (res. 45–47).

RBP7, the closest homologue of *Nt*GR-RBP1, the best defined target is its own pre-mRNA as part of its auto-regulation mechanism (13, 44). In this particular case, *At*GR-RBP7 recognizes a 6-nt sequence in its 3′UTR (5′-UUCUGG-3′) and has highly similar affinity for RNA and ssDNA (34). This RNA sequence is not part of the 3′UTR of *Nt*GR-RRBP1, for which it is yet unknown whether a similar auto-regulation mechanism also occurs. Given the high degree of sequence identity between the two proteins (76% overall, 84% RRM domain), we chose to use the 5′-UUCUGG-3′ RNA or the corresponding ssDNA oligonucleotide for binding studies, to derive the interaction surface, and a structural basis for nucleic acid binding by the *Nt*GR-RBP1 RRM domain.

Titration of ssDNA to the *Nt*RRM domain resulted in large changes in chemical shifts for several residues, with most resonances in fast-to-intermediate exchange between free and bound states (Figure 3A). Residues that show the largest chemical shift perturbation (CSP) all locate to the solvent exposed side of the central β-sheet and include aromatic residues F9, F49, F51 from RNP1/2 as well as positively charged residues R41, R47, R85 (Figure 3B). Most of these residues are highly conserved (see also alignment in Supplementary Figure S3a) and have also been shown to be involved in RNA binding for other RRM domains (see alignment and superposition in Supplementary Figure S3b and S3c). In particular, the corresponding residue of R47 in *At*GR-RBP7 has been shown to be critical for interaction with its 3′UTR and its *in vivo* activity (44), whose ADP-ribosylation reduces its RNA ability (45). Fitting the observed line shapes for each titration point to a 1:1 binding model, it was found that the dissociation constant K_D for ssDNA binding is 4 ± 3 μM and the dissociation rate of the complex k_{off} is (8.7 ± 3)·10^2 s^{-1} (Figure 3C and Supplementary Figure S4). The lifetime of the complex is thus ∼1 ms, indicating a rather dynamic binding. To compare these results, spectra of *Nt*RRM domain in the absence or presence of 1 M equivalent of the related 6-nt RNA fragment (5′-UUCUGG-3′) were recorded. Due to the low amount of RNA available, this experiment was carried out at 35 μM protein concentration. Again, the same set of protein residues shows clear changes in peak positions as identified for ssDNA (Figure 3D and E). These observations indicate that the RNA is bound in the same binding region and through similar interactions. Overall, the magnitude of the CSP upon addition of RNA is circa one-third of that observed when adding the corresponding amount of ssDNA, and ∼30% of the CSP for the saturated ssDNA complex. This difference may point to a lower affinity for RNA, such that only 30% of *Nt*RRM would be bound to RNA. Taking the experimental protein and RNA concentrations into account, this implies that the K_D is 50 μM. However, reasoning that such decrease in affinity (∼12-fold) would be due to a corresponding increase in k_{off}, one would expect a pronounced shift to fast exchange and concomitant reduced peak broadening. However, since several resonances are significantly broadened (such as R47 in Figure 3D), we argue that the k_{off} values for ssDNA and RNA cannot be very different. Based on these observations, we estimate that the K_D for RNA binding can be at most two-fold higher (8 μM) than that for ssDNA, in agreement with observations for *At*GR-RBP7 (34). We therefore attribute the major part of the observed difference in CSP to over-estimation of the RNA concentration, possibly due to a UV-absorbing impurity.

*Nt*RRM binds CTG/CUG through base-specific contacts and electrostatics interactions

A reverse titration experiment was performed in which unlabelled *Nt*RRM domain was added to (unlabelled) ssDNA in order to determine the binding interface on the nucleic acid. Both the base and sugar protons of nucleotides C3-T4-G5 showed large chemical shift changes, indicating that this part is the main interaction site of this oligo with the *Nt*RRM domain (Figure 4A). Next, isotope-filtered NOESY spectra were recorded on a 1:1 complex of $^{15}N/^{13}C$ labelled *Nt*RRM and unlabelled ssDNA. Several ssDNA resonances, especially for C3, are broadened in the complex limiting the number of intramolecular distance restraints that could be derived for the ssDNA, and the amount of intermolecular restraints between *Nt*RRM and ssDNA. Two intermolecular NOEs were identified unambiguously between the I38 methyl group and sugar protons of G5 (Figure 4B), which demonstrates the predominance of a highly specific complex. In addition, one intermolecular NOE involving G5 was identified at a chemical shift typical for aromatic protons and was assigned ambiguously to the aro-

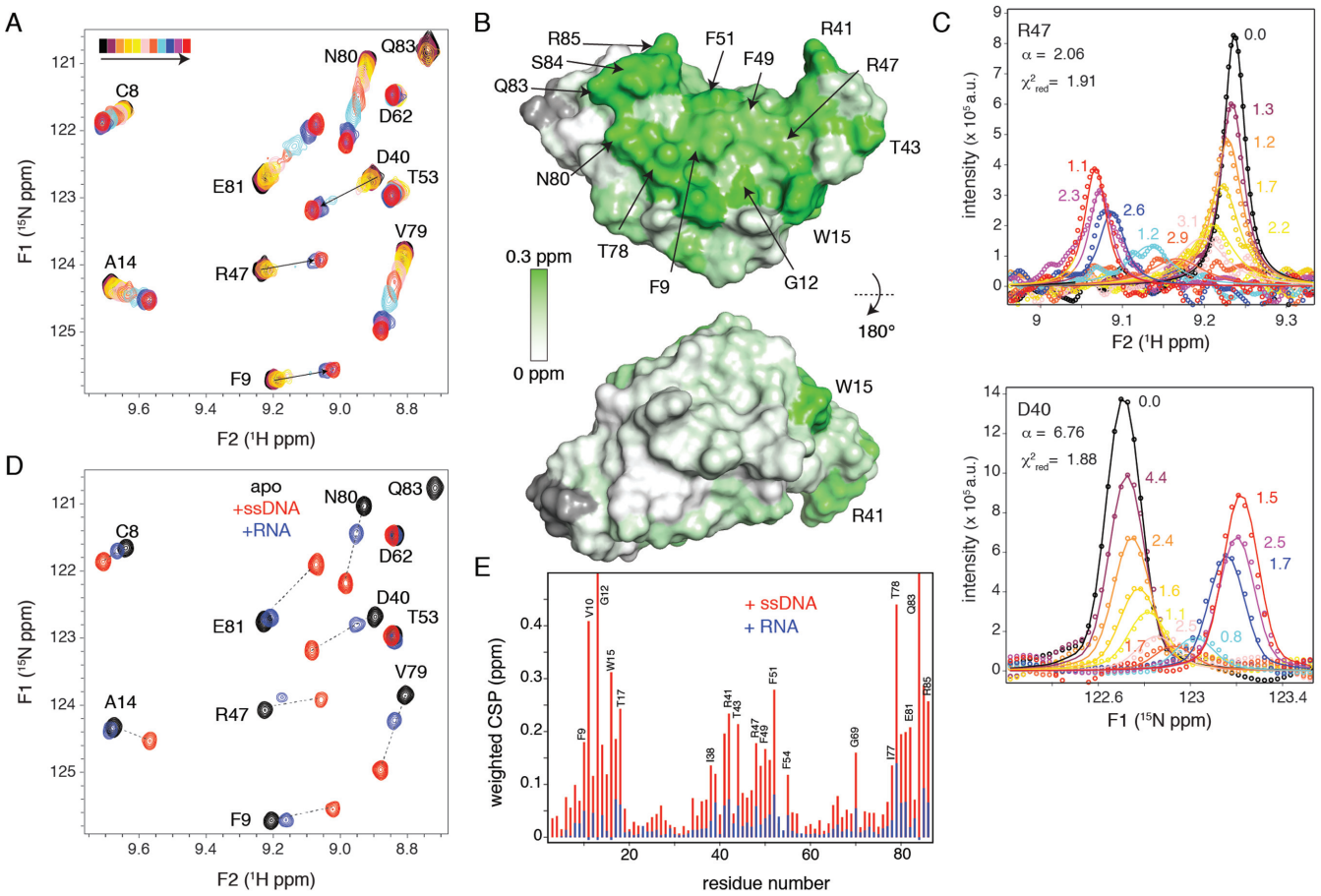

Figure 3. *Nt*RRM interacts strongly and specifically with nucleic acids (**A**) NMR titration results for ssDNA (5′-TTCTGG-3′), showing an overlay of a section of the ^{15}N-HSQC spectra for each titration point. Colour coding of the spectra is indicated at the top, the free *Nt*RRM spectrum is in black, fully bound spectrum is in red. Molar ratio *Nt*RRM:ssDNA at the end of the titration is 1:1.50. Assignments of resonances of interest are indicated. (**B**) Interaction surface for ssDNA binding. CSP colour coded on the van der Waals surface. Grey is used for residues without data; residues with CSP larger than 10% trimmed mean + 2 σ are labelled. (**C**) Experimental (points) and fitted (lines) line shapes during the titration for two selected residues. Fits for all residues including error analysis are shown in Supplementary Figure S4. Best fit was obtained with K_D of 4.2 μM and k_{off} of 860 s^{-1}. Goodness-of-fit in terms χ^2_{red} is indicated for each titration point as well as for all points combined. The exchange regime α ($2{\cdot}k_{off}/\Delta\omega$) is also indicated; fast α > 10; intermediate 10 < α < 0.1; slow α < 0.1. (**D**) Overlay of spectra of *Nt*RRM domain alone (black), with 1 M equivalent of ssDNA (red) or RNA (blue). Same section of the ^{1}H-^{15}N HSQC is shown as in (A). (**E**) Comparison of CSP observed upon addition of 1 M equivalent ssDNA (red) or RNA (blue). Selected residues are indicated. Residues with CSP larger than 10% trimmed mean + 1 σ for RNA have a CSP that is 32% of that for ssDNA on average.

matic protons of either F9, F49 or F51, based on their proximity to I38 (Figure 4B). Since we have too little experimental restraints to define the DNA-RRM interface and DNA conformation completely, high-resolution structure determination of the complex proved intractable. Thus, the available data was used to drive docking of a 6-nt RNA to the *Nt*RRM domain as a model for a *Nt*RRM–RNA chaperone complex using the HADDOCK program. As NOESY data for the complex indicate that the *Nt*RRM domain maintains its structure in the bound state (Supplementary Figure S5A), the NMR ensemble of unbound *Nt*RRM was used as input. As input structure for the RNA, we took six RNA conformations from homologous RRM–RNA complexes and refined them against the identified intramolecular NOEs (Supplementary Figure S5B). Out of the final set of 100 docking solutions, 83% were grouped into four clusters based on the RNA conformation in the complex (Supplementary Figure S5C). More than half of the solutions cluster into a single set of structures (Supplementary Figure S5C and E), showing a high degree of convergence of the docking. The overall orientation of the RNA of the dominant cluster is also found in two of the other clusters of solutions (Supplementary Figure S5C). The most representative model is shown in Figure 4C. The experimental restraints firmly anchor G5 to the *Nt*RRM surface near the edge of the β-sheet. The RNA runs over the β-sheet and is packed close to loop L2, with its 3′-end close to loop L3. Notably, removal of the intermolecular NOE information in the docking results in a complete loss of definition of the RNA orientation in the complex (Supplementary Figure S5D). Analysis of the interface in all models of the dominant cluster shows that mostly bases G5 and C3, but also U4 and U2, form aromatic–aromatic contacts with F9, F49 and F51. In addition, there are favourable electrostatic interactions between basic residues in loops L2 and L3 (R41, R47, R75) and the phosphate backbone. Interestingly, chemical shifts of residues in loop L2 indicate increased backbone dynamics in the bound state.

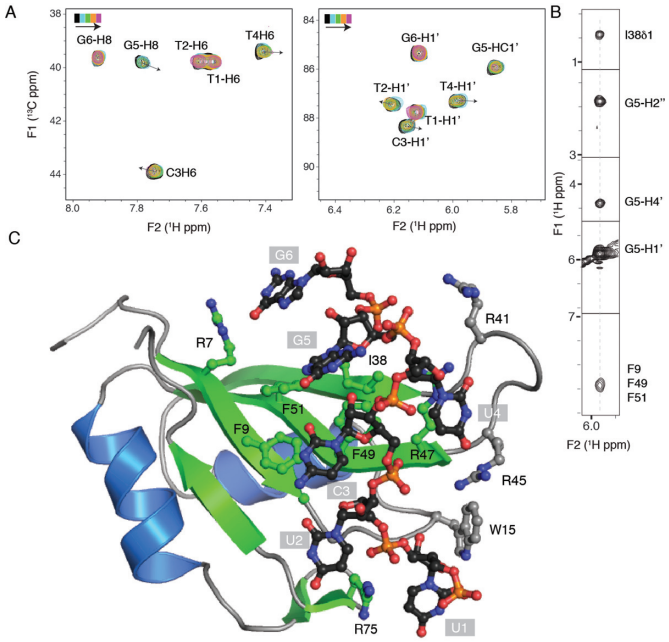

Figure 4. Nucleic acid binding mode. (**A**) Identification of CTG as binding surface for the *Nt*RRM domain. Overlay of natural abundance ^{1}H-^{13}C-HSQC spectra showing C6-H6/C8-H8 and C1′-H1′ regions. Final spectrum corresponds to ssDNA:*Nt*RRM 1:0.3 molar ratio. Significant exchange induced peak broadening and interference from *Nt*RRM resonances towards the end of the titration prohibited tracking the ssDNA resonances to the final bound state. Arrows indicate the direction of the CSP. (**B**) Sections of F2 ^{15}N/^{13}C-filtered 2D NOESY showing intense NOEs observed for G5 H1′. Two intermolecular NOEs involving this resonance could be identified, of which one (to Ile38δ1) could be unambiguously assigned, based on the corresponding peak in the ^{13}C-edited NOESY. The peak at 7.3 ppm is an ambiguously assigned intermolecular NOE involving the aromatic protons of F9, F49 and F51. (**C**) HADDOCK model of RNA–RRM complex. Selected residues from RNP motifs and residues that mediate intermolecular interactions are shown in ball-and-stick representation. Rotated view from Figure 3A, looking down on the β-sheet.

The *Nt*RRM domain and glycine-rich region are not independent modules

We next turned to the full-length *Nt*GR-RBP1 protein to assess the importance of the glycine-rich region (GR) on the structure and function of the RRM domain and in particular for its nucleic acid binding. The full-length protein was expressed as a GST-fusion. The fusion protein showed tendency to aggregate, which could be alleviated by increasing the ionic strength from 100 to 300 mM and keeping the protein at room temperature. After removal of the GST-tag, *Nt*GR-RBP1 was purified under standard ionic strength conditions (20 mM phosphate buffer pH 7, 100 mM KCl) without signs of aggregation, at least at concentrations below 150 μM. Denaturing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis confirmed the purity and integrity of the full-length protein (Figure 5B).

Figure 5A shows a comparison of ^{1}H-^{15}N-HSQC NMR spectra of full-length *Nt*GR-RBP1 and its isolated *Nt*RRM domain. The peak positions and line widths in such spectra provide a 'finger print' of the protein and directly relate to protein structure and dynamics (24). The *Nt*GR-RBP1 spectrum is characterized by the presence of both very broad and very sharp peaks, as well as pronounced clustering of peaks in the central and glycine region of the spectrum. Also visible are a number of spurious peaks from protein degradation products. Such peaks, which in this case most likely originate from minor degradation of the glycine-rich region, have very narrow line width and tend to easily dominate the spectrum, even though most of the protein is still intact. In sharp contrast to the full-length protein, the isolated *Nt*RRM domain shows a very homogenous distribution of peak intensities and excellent dispersion of peak positions.

The quality of the *Nt*GR-RBP1 spectrum precludes determination of the structure or dynamics of the full-length protein. Instead, comparison of domain and full-length spectra can give qualitative information of the properties of *Nt*GR-RBP1. Overlay of both spectra shows that the *Nt*RRM peak pattern is also present in the full-length protein (Figure 5A, right panel), indicating that the *Nt*RRM domain preserves its fold in the full-length protein. The correspondence in peak patterns between the *Nt*RRM and *Nt*GR-RBP1 spectra was used to transfer the RRM assignments of non-overlapping peaks to the full-length protein (see Figure 5C). Also considering the correspondence for the ssDNA bound states of the two proteins (see below), in total 32 backbone resonances (∼38% of the RRM domain) could be tentatively assigned in the apostate of *Nt*GR-RBP1. A detailed analysis of these assigned resonances shows that many have significant chemical shift changes and/or reduced peak intensities in the full-length compared to the *Nt*RRM domain spectra (for instance F9, R47, E81 in Figure 5C and all labelled residues in Figure 5D). In addition, for several non-overlapping peaks of the RRM domain there is no nearby peak in the full-length spectrum (for example A14, D62, Q83 in Figure 5C). When mapped to the solution structure of the *Nt*RRM domain, all residues with either significant changes in peak position, reduction of peak intensity or missing peaks in *Nt*GR-RBP1, predominantly reside on one side of the domain (Figure 5E). Changes near the C-terminus of the domain, including the final β-strand and the α-helix that packs against it, may be due to subtle structural rearrangements as a consequence of the connection to the glycine-rich region. Many residues that are perturbed in the full-length protein compared to the domain are, however, rather distant from the direct point of connection between the GR and the RRM domain. This includes many residues from the ssDNA binding interface (Figure 5E).

Compared to the RRM domain, the *Nt*GR-RBP1 spectrum shows major additional intensity around 8–8.5 ppm in the ^{1}H dimension that most likely originates from resonances in the GR. In the upper part of the spectrum, around 110 ppm in the ^{15}N dimension, there are roughly 10 additional peaks, including a number with very high peak intensity, for the 47 glycines in the GR. This indicates that many glycines experience similar chemical environments and overlap. Excluding glycine and side chain resonances, ∼131 peaks can be counted in the full-length spectrum, whereas 98 are expected (total construct size is 158 amino acids, minus 57 glycines and the 3 N-terminal residues that are also missing in the RRM spectrum). This

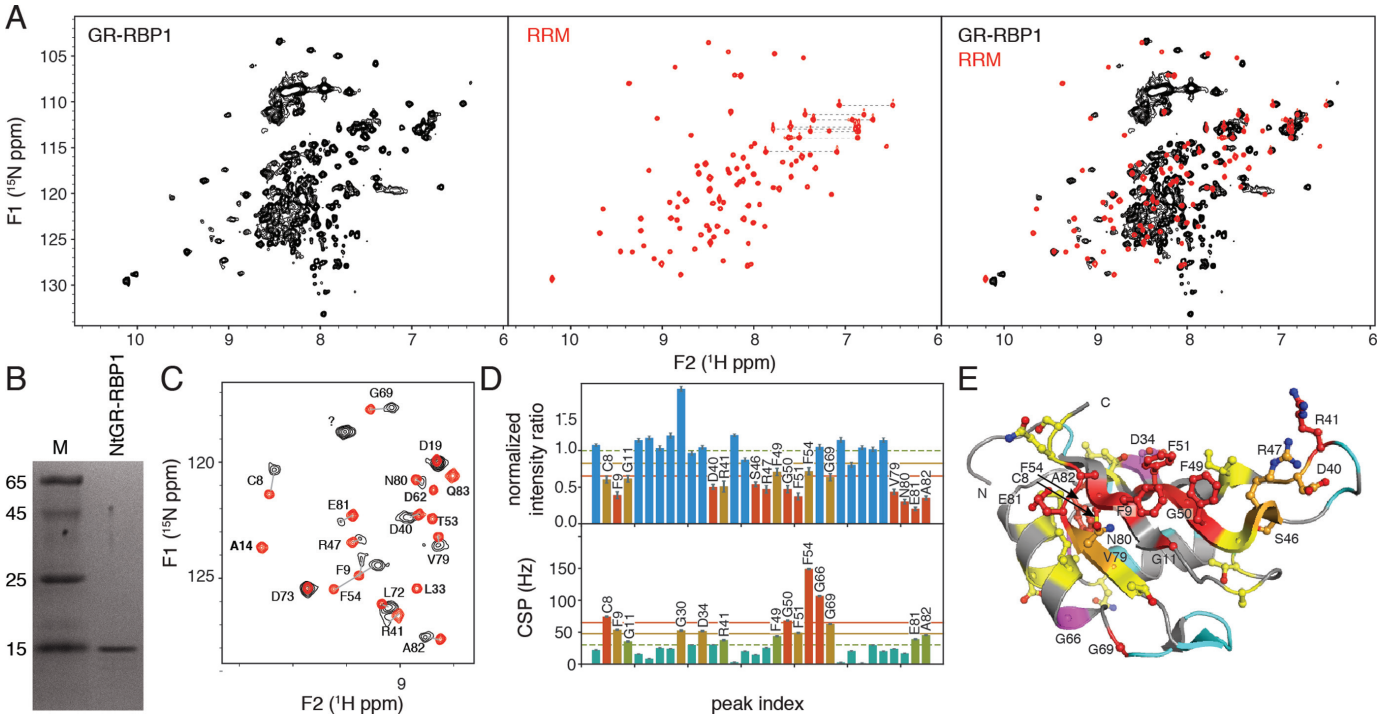

Figure 5. The RRM domain and glycine-rich region are not independent modules. (**A**) Comparison of ^{1}H-^{15}N-HSQC 'fingerprint' spectra of full-length *Nt*GR-RBP1 (left panel) and isolated *Nt*RRM domain (middle panel), both recorded at 293 K. An overlay of the two spectra is shown in the right panel. Spectra were recorded and processed with identical parameters. Peaks around 8/125–130 ppm (^{1}H/^{15}N) in the spectrum of *Nt*GR-RBP1 result from minor partial degradation, presumably from the disordered glycine-rich region. (**B**) SDS-PAGE analysis of the full-length *Nt*GR-RBP1 sample after purification. (**C**) Zoom of *Nt*RRM (red) and *Nt*GR-RBP1 (black) spectrum with peaks connected by their assignment. Labels in bold face indicate that a corresponding nearby peak in *Nt*GR-RBP1 is missing. (**D**) Analysis of normalized peak intensity ratios and CSPs between *Nt*RRM and *Nt*GR-RBP1 for all 32 tentatively assigned backbone resonances. Intensity ratio of two different samples was normalized such that peaks without significant CSP and without broadened line shapes have a ratio of 1, see for instance D19 and D73 in (B). Residues with significant reduced intensity in the full-length spectrum or significant CSP are labelled. (**E**) Results of (D) plotted on the structure of *Nt*RRM. Colour coding: grey—no data due to overlap in full-length; cyan—no significant change in intensity or peak position; orange—significantly reduced peak intensity; magenta—significant CSP; red—significant CSP and reduced intensity; yellow—missing peak in *Nt*GR-RBP1.

significant amount of peak doubling suggests that the non-glycine residues in the GR experience different chemical environments in relatively stable different conformations of the GR. While the lack of ^{1}H chemical shift dispersion for the GR resonances is indicative of a lack of structure, these signals are not sharp as would be expected for a dynamically disordered region.

Together, the peak doubling and broadened appearance of the GR and the perturbations of the RRM resonances suggest that the GR is not simply an independent, disordered tail attached to the RRM. Instead, these observations point to the possibility that the GR may transiently interact with the *Nt*RRM domain, either intra- or intermolecularly.

The glycine-rich region can mediate intermolecular self-association

To test whether GR can associate intermolecularly with the *Nt*RRM domain, we performed an NMR titration experiment in which the effect of addition of unlabelled *Nt*GR-RBP1 on ^{15}N-labelled *Nt*RRM domain was monitored. After recording the spectrum of *Nt*RRM alone, half of the sample was removed and replaced by the same volume of *Nt*GR-RBP1 at identical concentration in identical buffer. Since the total RRM domain concentration is constant, whether part of the full-length protein or not, any perturbation in the spectrum of *Nt*RRM must be due to intermolecular interactions with *Nt*GR-RBP1. Overlay of the *Nt*RRM spectra before and after addition of unlabelled full-length protein shows dramatic reduction in intensity for a distinct set of peaks, as well as small chemical shift changes (Figure 6A and see for example A14, R41, A82 and Q83 in Figure 6B). Systematic analysis of peak intensity ratios and chemical shift changes for all residues shows that residues with strongly reduced intensities mostly also have relatively large chemical shift changes and vice versa (Figure 6C, left panel). In addition, residues without significant chemical shift changes show a fairly homogenous intensity of ~35% of their original value. When mapped onto the structure (Figure 6D), the affected residues (either by intensity, peak position or both) cover one side of the molecule, that includes many of the RRM residues that were also perturbed within the full-length protein (compare Figure 5E), as well as the nucleic acid binding surface (compare Figure 4C). Together, these observations strongly suggest that the GR of the full-length protein can associate specifically to the RRM domain in an intermolecular fashion. To further validate these findings, the experiment was repeated at lower protein concentration (48 μM instead of 128 μM). Again, mostly the same set of residues shows reduced peak inten-

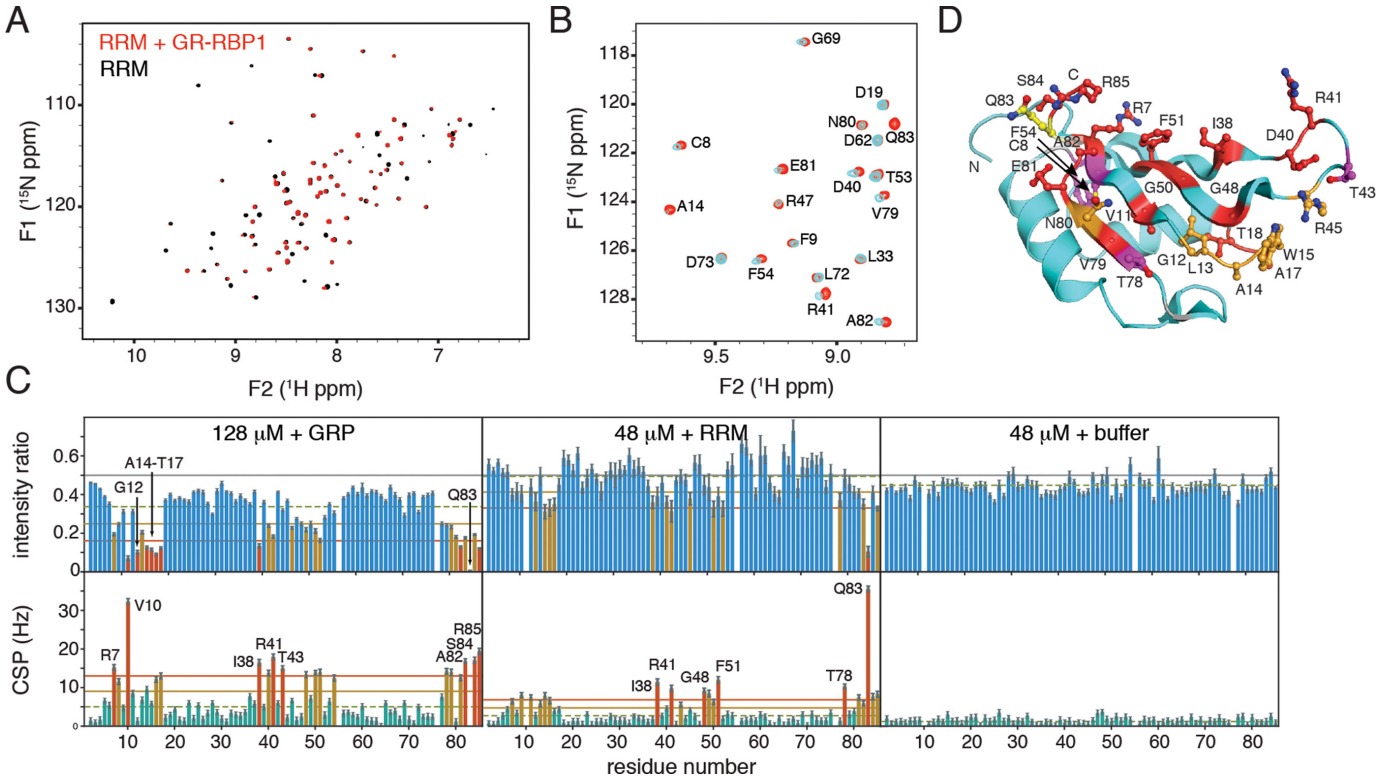

Figure 6. Intermolecular interaction between the glycine-rich region and the RRM domain. (**A**) Comparison of ^{1}H-^{15}N-HSQC spectra of *Nt*RRM domain in absence (black) or presence (red) of 1 M equivalent of unlabelled *Nt*GR-RBP1, at total protein concentration of 128 μM. (**B**) Zoom of overlay in (A) plotted at lower contour threshold to highlight the severe peak broadening and chemical shift changes for selected residues. (**C**) Analysis of peak intensities ratios (top) and CSPs (bottom) upon addition of full-length protein at total protein concentration 128 μM (left) 48 μM (middle) and the addition of buffer only (right). Residues that deviate more than 2 SD (red line) from the 10% trimmed mean (green broken line) are labelled, residues that deviate more than 1 SD (orange line) are highlighted in yellow. Expected peak intensity ratio due to dilution of RRM is 0.5, shown as a grey line. (**D**) Results of (C) plotted on the structure of *Nt*RRM. Colour coding: grey—no data due to overlap; cyan—no significant change in intensity or peak position; orange—significantly reduced peak intensity; magenta—significant CSP; red—significant CSP and reduced intensity; yellow—missing peak.

sities and chemical shift changes (Figure 6C, middle panel). In accordance with an intermolecular interaction, the observed perturbations are smaller. As a control, only buffer was added to *Nt*RRM without significant changes in peak intensities or positions (Figure 6C, right panel).

Full-length *Nt*GR-RBP1 has similar nucleic acid binding mode as *Nt*RRM

We next turned to the nucleic acid binding properties of full-length *Nt*GR-RBP1, using the 6-nt TTCTGG ssDNA fragment. Spectra without and with 1 M equivalent of ssDNA show that significant chemical shift changes occur for resonances attributed to the *Nt*RRM domain (Figure 7A and B, respectively). Strikingly, peaks that show very pronounced broadening in the free state, for instance residues F9, A14, R41 and E81, appear sharper and have higher peak intensities upon ssDNA binding (Figure 7B and C). This up to 5-fold increase in intensity is contrary to the 1.1-fold dilution of the sample due to addition of the ssDNA. Residues with increased peak intensities mostly also experience significant changes in chemical shift (Figure 7C) and map to the ssDNA binding surface of the RRM domain (Figure 7D). Thus we identify the same binding surface as for the isolated *Nt*RRM domain. Moreover, comparison of the CSPs observed for the domain and full-length protein show that most peaks shift in the same direction by a similar amount (Figure 6B). This indicates that *Nt*GR-RBP1 and *Nt*RRM bind the oligonucleotide in the same structural manner and with comparable binding affinities. The peak intensity increase signifies that saturation of the protein with ssDNA locks it in a single, well-defined state and thereby removes the line broadening present in the apostate of this protein. This also rules out partial degradation as a cause of the poor spectral quality of the apostate.

The spectrum of the ssDNA bound *Nt*GR-RBP1 resembles that of the *Nt*RRM domain more than in the free state, resulting in the additional assignment of 12 peaks. Comparison of peak intensities for the 44 backbone resonances that are tentatively assigned in the bound states of *Nt*RRM and *Nt*GR-RBP1 no longer shows strongly broadened peaks as in the free state (Supplementary Figure S6B). Also, there are no significant chemical shift differences for much of the ssDNA interface, including RNP residues F50, F52 and R47 (Supplementary Figure S6C). Still, there are significant chemical shift changes between the two proteins that mostly locate to the direct vicinity of the RRM C-terminus where the GR is linked.

Analysis of the impact of ssDNA binding on GR is hampered by extensive overlap with RRM resonances in the central part of the spectrum. Focusing only on the glycines in

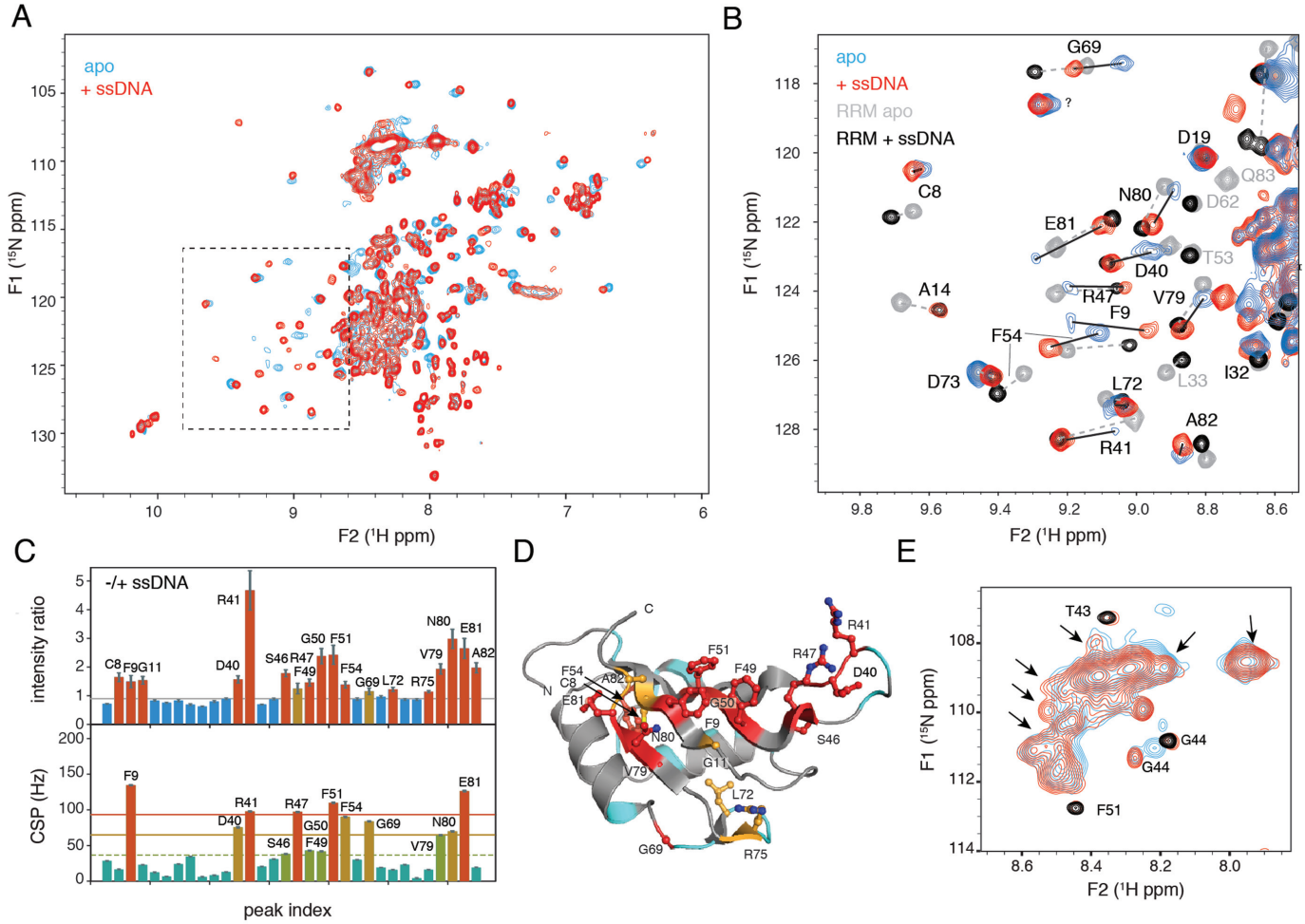

Figure 7. Full-length *Nt*GR-RBP1 and isolated *Nt*RRM share similar nucleic acid binding mode. (**A**) Overlay of *Nt*GRP-RBP1 in absence (blue) and presence (red) of 1 M equivalent of 6-nt ssDNA. Significant chemical shift changes occur for resonances at the edges of the spectrum that are attributed to the *Nt*RRM domain (compare with Figure 2). (**B**) Detail of (A) including an overlay of isolated *Nt*RRM domain in absence (grey) and presence (black) of ssDNA. Selected residues are indicated. (**C**) Analysis of peak intensities ratios (top) and CSPs (bottom) upon addition of ssDNA. Grey solid lines indicate expected intensity ratio due to dilution (0.9). Residues with significant intensity increase upon ssDNA addition and those with CSPs more than 2 SD (red line) from the 10% trimmed mean (green broken line) are labelled, residues that deviate more than 1 SD (orange line) are highlighted in yellow. (**D**) Results of (C) plotted on the structure of *Nt*RRM. Colour coding: grey—no data due to missing assignment; cyan—no significant change in intensity or peak position; orange—significantly reduced peak intensity; magenta—significant CSP; red—significant CSP and reduced intensity. (**E**) Detail of (A) focusing on the glycine region of the spectrum with unassigned peaks from GR. Arrows point to small changes in peak positions. Labelled peaks are from RRM domain in ssDNA bound state (shown in black).

the GR, a number of minor chemical shift changes can be discerned (Figure 7E). These may be due to direct interactions with the ssDNA or due to an indirect effect of ssDNA binding to the RRM domain.

Glycine-rich region stimulates formation of higher-order complexes with nucleic acids

To further evaluate the role of the GR in nucleic acid binding, we performed electrophoretic mobility shift assays (EMSA) for both *Nt*RRM and *Nt*GR-RBP1 using a ssDNA probe (ssP1) having the CTG-containing mRNA-binding site of homologue *At*GR-RBP7 (34) in either a single (13-nt) or double (26-nt) copy (designated as ssP1 or ss-dP1, Figure 8A). Due to the small size of the single probe ssP1, only very limited shift in mobility for both proteins was obtained even at high gel concentrations (see also Figure 8B), precluding estimation of the apparent binding affinity. Nucleic acid binding is nevertheless apparent from the loss of free probe for the RRM domain and from the slight shift and band smearing for the full-length protein. For the longer ss-dP1 probe, a clear mobility shift is obtained for *Nt*RRM at the two highest protein concentrations, together with significant band smearing at the highest RRM concentration. Both observations indicate nucleic acid binding, as expected. Using the loss of free probe to estimate the fraction bound, an apparent dissociation constant $K_{D,app}$ in the low micromolar range was estimated (~10 μM, Figure 8C). Similarly, a $K_{D,app}$ of ~5 μM was estimated for NtGR-RBP1, in correspondence with the NMR-based finding that RRM domain and full-length protein have comparable affinities for ssDNA. Strikingly, for *Nt*GR-RBP1, a very slow migrating band is visible at the

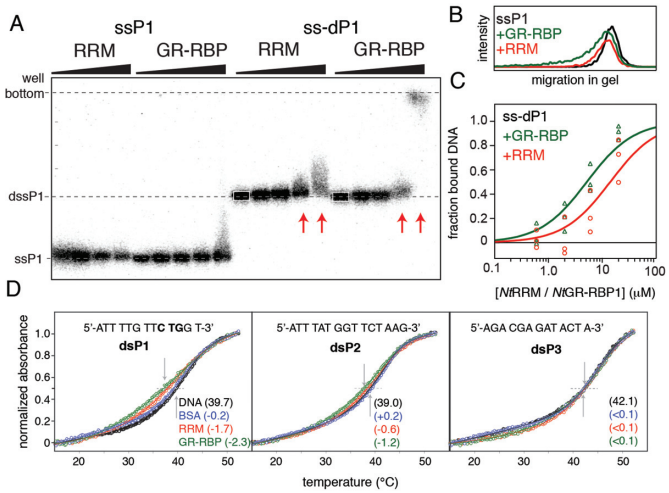

Figure 8. Glycine-rich region stimulates higher-order complex formation and dsDNA unfolding (**A**) Representative DNA binding experiment by EMSA for the binding of 0, 0.7, 2, 7, 20 μM *Nt*RRM or *Nt*GR-RBP1 to a single or double ssDNA binding element (ssP1 or ss-dP1). Dashed lines indicate the positions of the wells and the free mobilities of ssP1 and ss-dP1, and serve to guide the eye. The arrows indicate lanes of interest with clear band shifts. Several lanes show a smeared, asymmetric band appearance caused by significant dissociation during electrophoresis. The selection box for quantification of the free DNA probe is indicated in white on the first lane. (**B**) 1D traces of the lanes with 20 μM *Nt*RRM or *Nt*GR-RBP1 added to ssP1 show a small but reproducible and concentration-dependent shifts in mobility. (**C**) Quantification of the fraction of bound ssDNA for the ss-dP1 probe at the indicated concentration of *Nt*GR-RBP1 and *Nt*RRM, the line represents the calculated binding curve based on three independent experiments. (**D**) UV melting curves of indicated oligonucleotides in presence of 3 M equivalents of *Nt*GR-RBP1, *Nt*RRM or BSA. Absorbance (at 260 nm) versus temperature curves in 20 mM KPi, 100 mM KCl, 1 mM BME at pH 7.0. Temperatures (°C) at the transition midpoint, T_m, are indicated for the free DNA probes, with changes in T_m listed in the presence of the three proteins.

highest protein concentration, after an initial smaller shift. This suggests formation of a higher-order complex. While this complex travelled only slightly through the gel, substantial migration of this band could be established at lower gel concentrations, but at the expense of resolution for the initial smaller shift (data not shown). The diffuse appearance of the band indicates substantial dissociation during electrophoresis. Importantly, no discrete stable higher order complex is visible at the highest concentration of *Nt*RRM. There, band smearing above the shifted probe suggests formation of transient higher-order complexes that dissociate during electrophoresis. The lower apparent binding affinity and lack of stable higher-order complex formation for *Nt*RRM suggest that the GR contributes to DNA binding affinity, possibly by stimulating dimerisation or multimerisation of *Nt*GR-RBP1.

*Nt*GR-RBP1 induces melting of CTG-containing DNA more efficiently than *Nt*RRM

Finally, we studied DNA temperature melting curves in the absence and presence of *Nt*RRM domain and *Nt*GR-RBP1. It is expected that upon binding to nucleic acid, they could stabilize the single-strand conformation and thus promote DNA melting, thereby mimicking their RNA chaperone activity. *Nt*GR-RBP1 facilitated melting of a 13 bp dsDNA probe corresponding to the mRNA-binding site of homologue *At*GR-RBP7 (34) (dsP1) evidenced by the increasing proportion of ssDNA at temperatures below the native melting temperature T_m (Figure 7, left panel). The *Nt*RRM domain alone also showed an increase in melting, although to a lesser extent, whereas negative control BSA did not induce any change in the melting curve. For non-CTG-containing probes dsP2 (a homologous region in 3′UTR of NtGR-RBP1 mRNA), both *Nt*GR-RBP1 and *Nt*RRM exert a marginal effect, whereas for probe dsP3 (random sequence) no change was detected (Figure 7, middle and right panels). These observations further substantiate the importance of CTG sequence for efficient binding and demonstrate that full-length *Nt*GR-RBP1 is more efficient in melting secondary structures in longer nucleic acids than isolated *Nt*RRM domain (Figure 9).

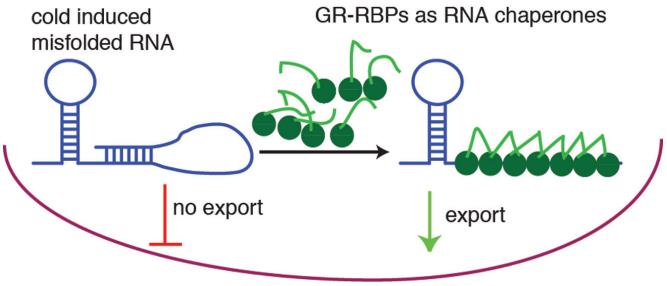

Figure 9. Model of GR-RBP function as RNA chaperones. mRNAs adopt secondary structures upon low temperature. GR-RBPs show RNA chaperone activity by facilitating folding of mRNAs, hence contribute to efficient translation in cold. Glycine-rich tails enhance chaperone function by mediating self-association, resulting in cooperative unfolding of non-native substrate structures.

DISCUSSION

We have shown that *Nt*RRM domain is a canonical RRM domain that binds a CTG/CUG containing nucleic acid with low micromolar affinity and short complex lifetime. Our NMR data allow the guanosine to be unambiguously anchored to the *Nt*RRM domain. An NMR data-driven model of the *Nt*RRM–RNA complex shows that the interaction is mediated by both base/aromatic ring stacking and electrostatic interactions with the phosphate backbone. Observed broadening of crucial base and aromatic protein resonances, as well as increased dynamics in loop L2 of *Nt*RRM, suggests that residual motion of the bases is allowed in the complex. This study shows that the CTG element within a 5′-TTCTGG-3′ ssDNA (derived from the 3′-UTR of *At*GR-RBP7) is the *Nt*RRM binding site. Notably, while the sequence 5′-TTCTGG-3′ is not present in the tobacco mRNA, there are four occurrences of CTG in the 3′ UTR of *Nt*GR-RBP1, suggesting that *Nt*GR-RBP1 could also bind to its own pre-mRNA. Furthermore, in the absence of a thorough evaluation of RNA-binding specificity for *Nt*RRM, we cannot exclude the possibility that the RRM domain also binds other RNA sequences with similar affinities. In fact, studies on homologous GR-RBPs have shown that the RRM domain has a base preference, but no strong sequence specificity: *At*GR-RBP7 prefers U/G

rich RNA (18), and hnRNP A1 binds with narrow range of affinities (30–300 nM) to a wide range of A/G rich RNA (46).

The *Nt*RRM domain gives a beautiful NMR spectrum when isolated, as has also been observed for other RRM domains (38–40,47), including that from a homologous plant GR-RBP (48). The latter study focuses on the resonance assignment of the RRM domain, but also reports that the full-length protein is prone to aggregation at high protein concentrations and that the glycine-rich region is likely unstructured. Here, we obtained NMR data for the full-length *Nt*GR-RBP1 at low protein concentrations. Interestingly, we find that many RRM resonances are severely broadened and displaced in context of full-length *Nt*GR-RBP1. Most of the affected residues are also broadened and shifted upon addition of full-length *Nt*GR-RBP1 to *Nt*RRM, evidencing an intermolecular association between GR tail and RRM domain. This finding also implies the potential for an intramolecular GR–RRM association. The broadened peak shapes in the full-length protein suggest that this intermolecular and intramolecular interaction between GR and RRM is relatively weak such that there is continuous interchange between free and GR-bound states of the RRM domain. Accordingly, the *Nt*GR-RBP1–*Nt*RRM interaction experiment shows smaller changes at lower protein concentrations. The large interaction surface on the RRM domain and the presence of peak doubling for the GR tail suggest that there are multiple bound states, in which the GR tail interacts in different conformations to different parts of RRM domain. This is further supported by the broad appearance of the peaks from the GR, suggesting continuous interconversion between different states. In contrast, very sharp peaks would be expected for an intrinsically and dynamically disordered protein. As such the GR tail is intermediate between a fully unfolded and a structured polypeptide, akin to a molten globule. All these observations, both from the perspective of the RRM domain and the GR, point to the presence of weak, dynamic interactions between the GR and RRM domain in *Nt*GR-RBP1.

While the residues responsible for these transient interactions are yet unidentified, it is notable that the glycine-rich region not only has a high percentage of glycines (65%) but is also enriched in arginines (11%) and tyrosines (10%). Thus, it may be better described as a glycine-tyrosine-arginine rich (GYR) region. Of note, GR and GYR regions have been implicated in protein–protein interactions in many other systems. For instance, a similar GYR region in hnRNP H interacts with the transportin receptor, facilitating shuttling of the mRNA to the cytosol (49). The GYR region of *At*GR-RBP7 also facilitates transportin-mediated nuclear import (50), which could mean that the GYR in *Nt*GR-RBP1 has similar function. Nucleolin has a glycine-arginine rich domain that facilitates interaction with p53 (51). In hnRNP A1 the glycine-rich region also mediates self-association (52). A recent paper showed that it is dispensable for RNA binding and likely involved in functional protein–protein interactions in a ternary splicing complex (47).

In case of *Nt*GR-RBP1 the GR is dispensable for nucleic acid binding as both our NMR and EMSA data show that full-length *Nt*GR-RBP1 and *Nt*RRM have comparable affinities for ssDNA. This implies that the GR tail is unable to compete with ssDNA for the nucleic acid binding surface, in accordance with a weak affinity for the GR–RRM interaction. Nevertheless, the origin of the small chemical shift changes of the GR upon ssDNA remains yet unclear. These could be due to direct interaction with ssDNA, release from the RRM surface and return to an unbound state, or a rearrangement of the GR on the RRM surface. Interestingly, EMSA data indicate that *Nt*GR-RBP1 can form higher-order complexes on longer DNA probes. This suggests that even in ssDNA bound state the GR tail can mediate intermolecular self-association through protein–protein interactions and argues against direct interaction with ssDNA. The binding site involved remains to be established but could include GR tail itself or another part of the RRM surface. Notably, the C-terminal side of the RRM domain shows significant chemical shift changes between the ssDNA bound states of *Nt*RRM and *Nt*GR-RBP1, which could reflect binding of the GR tail. It is also of interest to note that peaks of the GR tail do not become sharper upon nucleic acid binding, further indicating that it remains associated.

Using DNA melting experiments as a proxy for their RNA chaperone activity, we found that both *Nt*RRM and *Nt*GR-RBP1 can induce melting of CTG-containing DNA. While previous studies showed that the RRM domain is sufficient for the RNA chaperone activity in bacteria (18), we found that *Nt*GR-RBP1 is more efficient in DNA melting than *Nt*RRM alone. Reasoning that increased melting is due to stabilisation of the ssDNA by complex formation with *Nt*RRM, the stronger activity of *Nt*GR-RBP1 must be ascribed to enhanced binding affinity to the substrate. Given that 13 bp DNA fragments were used in these melting studies, there is enough space to accommodate two RRM domains. As only one CTG-site is present in the dsP1 probe, the second RRM domain would be bound to another sequence. Considering that homologue *At*GR-RBP7 generally prefers U/G rich RNA, it is of interest to note that dsP1 contains a longer continuous T/G stretch compared to dsP2 and dsP3 (Figure 7B). Intermolecular associations mediated by the GR region provide a molecular mechanism whereby *Nt*GR-RBP1 can effectively increase its local concentration and its affinity for nucleic acids. Thus, our data favour a model where the GR region enhances chaperone function by mediating self-association, resulting in cooperative unfolding of non-native substrate structures (Figure 8).

CONCLUSION

In summary, we described a detailed structural characterization of an intact plant GR-RBP. We showed that the structure of the *Nt*RRM domain of *Nt*GR-RBP1 has a canonical RRM fold that binds both dTTCTGG and rUUCUGG with micromolar affinities. Using NMR-based docking, a 3D model for the RRM–RNA complex could be established. Our study provides strong evidence that the glycine-rich region of *Nt*GR-RBP1 promotes self-association, through a specific interaction with the RRM domain. While the glycine-rich region does not interfere with single-stranded oligonucleotide binding of the RRM domain, it promotes the formation of higher-order com-

plexes on longer nucleic acid substrates and enhances its ds-DNA melting activity. Based on these findings, we propose that the glycine-rich region enhances RNA chaperone activity of GR-RBPs by mediating self-association, resulting in cooperative unfolding of non-native substrate structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACCESSION NUMBERS

The coordinates of the free *Nt*RRM domain have been deposited under accession code 4C7Q in the Protein Data Bank, chemical shift assignments are deposited in the BMRB, accession code 19525.

ACKNOWLEDGEMENT

We thank Dr Hans Wienk, Klemen Dolinar and Johan van der Zwan for their kind assistance.

FUNDING

Higher Education Commission (HEC) of Pakistan through an Indigenous and IRSIP fellowships (to F.K.); The Netherlands Organisation for Scientific Research (NWO) through an NWO-CW Veni fellowship (700.59.401) (to H.v.I.). Funding for open access charge: NWO Stimuleringsfonds Open Access.
Conflict of interest statement. None declared.

REFERENCES

1. Glisovic,T., Bachorik,J.L., Yong,J. and Dreyfuss,G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.
2. Maris,C., Dominguez,C. and Allain,F.H.T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
3. Califice,S., Baurain,D., Hanikenne,M. and Motte,P. (2012) A single ancient origin for prototypical serine/arginine-rich splicing factors. *Plant Physiol.*, **158**, 546–560.
4. Gomez,J., Sanchez-Martinez,D., Stiefel,V., Rigau,J., Puigdomenech,P. and Pages,M. (1988) A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature*, **334**, 262–264.
5. Maruyama,K., Sato,N. and Ohta,N. (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: Probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. *Nucleic Acids Res.*, **27**, 2029–2036.
6. Khan,F., Sultana,T., Deeba,F. and Naqvi,S.M. (2013) Dynamics of mRNA of glycine-rich rna-binding protein during wounding, cold and salt stresses in *Nicotiana Tabacum*. *Pak. J. Bot.*, **45**, 297–300.
7. Nicaise,V., Joe,A., Jeong,B.-R., Korneli,C., Boutrot,F., Westedt,I., Staiger,D., Alfano,J.R. and Zipfel,C. (2013) Pseudomonas HopU1 modulates plant immune receptor levels by blocking the interaction of their mRNAs with GRP7. *EMBO J.*, **32**, 701–712.
8. Mangeon,A., Junqueira,R.M. and Sachetto-Martins,G. (2010) Functional diversity of the plant glycine-rich proteins superfamily. *Plant Signal. Behav.*, **5**, 99–104.
9. Naqvi,S.M., Park,K.S., Yi,S.Y., Lee,H.W., Bok,S.H. and Choi,D. (1998) A glycine-rich RNA-binding protein gene is differentially expressed during acute hypersensitive response following Tobacco Mosaic Virus infection in tobacco. *Plant Mol. Biol.*, **37**, 571–576.
10. Lee,M.-O., Kim,K.P., Kim,B.-G., Hahn,J.-S. and Hong,C.B. (2009) Flooding Stress-Induced Glycine-Rich RNA-Binding Protein from Nicotiana tabacum. *Mol. Cells* **27**, 47–54.
11. Lorković,Z.J. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.*, **14**, 229–236.
12. Heintzen,C., Nater,M., Apel,K. and Staiger,D. (1997) AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in Arabidopsis thaliana. *Proc. Natl Acad. Sci. U.S.A.*, **94**, 8515–8520.
13. Staiger,D., Zecca,L., Wieczorek Kirk,D.A., Apel,K. and Eckstein,L. (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.*, **33**, 361–371.
14. Schöning,J.C., Streitner,C., Meyer,I.M., Gao,Y. and Staiger,D. (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis. *Nucleic Acids Res.*, **36**, 6977–6987.
15. Streitner,C., Köster,T., Simpson,C.G., Shaw,P., Danisman,S., Brown,J.W.S. and Staiger,D. (2012) An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. *Nucleic Acids Res.*, **40**, 11240–11255.
16. Streitner,C., Danisman,S., Wehrle,F., Schöning,J.C., Alfano,J.R. and Staiger,D. (2008) The small glycine-rich RNA binding protein AtGRP7 promotes floral transition in Arabidopsis thaliana. *Plant J.*, **56**, 239–250.
17. Fu,Z.Q., Guo,M., Jeong,B.-R., Tian,F., Elthon,T.E., Cerny,R.L., Staiger,D. and Alfano,J.R. (2007) A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature* **447**, 284–288.
18. Kim,J.S., Park,S.J., Kwak,K.J., Kim,Y.O., Kim,J.Y., Song,J., Jang,B., Jung,C.-H. and Kang,H. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from Arabidopsis thaliana can promote the cold adaptation process in Escherichia coli. *Nucleic Acids Res.*, **35**, 506–516.
19. Kim,J.Y., Kim,W.Y., Kwak,K.J., Oh,S.H., Han,Y.S. and Kang,H. (2010) Glycine-rich RNA-binding proteins are functionally conserved in Arabidopsis thaliana and Oryza sativa during cold adaptation process. *J. Exp. Bot.*, **61**, 2317–2325.
20. Nakaminami,K., Karlson,D.T. and Imai,R. (2006) Functional conservation of cold shock domains in bacteria and higher plants. *Proc. Natl Acad. Sci. U.S.A.*, **103**, 10122–10127.
21. Kim,M.K., Jung,H.J., Kim,D.H. and Kang,H. (2012) Characterization of glycine-rich RNA-binding proteins in Brassica napusunder stress conditions. *Physiol. Plant.*, **146**, 297–307.
22. Kwak,K.J., Park,S.J., Han,J.H., Kim,M.K., Oh,S.H., Han,Y.S. and Kang,H. (2011) Structural determinants crucial to the RNA chaperone activity of glycine-rich RNA-binding proteins 4 and 7 in Arabidopsis thaliana during the cold adaptation process. *J. Exp. Bot.*, **62**, 4003–4011.
23. de Jong,R.N., Daniëls,M.A., Kaptein,R. and Folkers,G.E. (2006) Enzyme free cloning for high throughput gene cloning and expression. *J. Struct. Funct. Genomics*, **7**, 109–118.
24. Cavanagh,J., J Fairbrother,W., Palmer,A.G. III, J Skelton,N. and Rance,M. (2007) .*Protein NMR spectroscopy: principles and practice*. Academic Press, New York.
25. Zwahlen,C., Legault,P., Vincent,S., Greenblatt,J., Konrat,R. and Kay,L. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.*, **119**, 6711–6721.
26. 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.
27. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinás,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.
28. 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.
29. 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.

30. Herrmann,T., Güntert,P. and Wüthrich,K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227.
31. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
32. Nederveen,A.J., Doreleijers,J.F., Vranken,W., Miller,Z., Spronk,C.A.E.M., Nabuurs,S.B., Güntert,P., Livny,M., Markley,J.L., Nilges,M. *et al.* (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
33. Doreleijers,J.F., Sousa da Silva,A.W., Krieger,E., Nabuurs,S.B., Spronk,C.A.E.M., Stevens,T.J., Vranken,W.F., Vriend,G. and Vuister,G.W. (2012) CING: an integrated residue-based structure validation program suite. *J. Biomol. NMR*, **54**, 267–283.
34. Schüttpelz,M., Schöning,J.C., Doose,S., Neuweiler,H., Peters,E., Staiger,D. and Sauer,M. (2008) Changes in conformational dynamics of mRNA upon AtGRP7 binding studied by fluorescence correlation spectroscopy. *J. Am. Chem. Soc.*, **130**, 9507–9513.
35. Kato,H., van Ingen,H., Zhou,B.-R., Feng,H., Bustin,M., Kay,L.E. and Bai,Y. (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc. Natl Acad. Sci. U.S.A.*, **108**, 12283–12288.
36. Dominguez,C., Boelens,R. and Bonvin,A.M.J.J. (2003) HADDOCK: A protein−protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
37. de Vries,S.J., van Dijk,A.D.J., Krzeminski,M., van Dijk,M., Thureau,A., Hsu,V., Wassenaar,T. and Bonvin,A.M.J.J. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
38. Leeper,T.C., Qu,X., Lu,C., Moore,C. and Varani,G. (2010) Novel protein-protein contacts facilitate mRNA 3′-processing signal recognition by Rna15 and Hrp1. *J. Mol. Biol.*, **401**, 334–349.
39. Cléry,A., Jayne,S., Benderska,N., Dominguez,C., Stamm,S. and Allain,F.H.T. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nat. Struct. Mol. Biol.*, **18**, 443–450.
40. Ohyama,T., Nagata,T., Tsuda,K., Kobayashi,N., Imai,T., Okano,H., Yamazaki,T. and Katahira,M. (2012) Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. *Nucleic Acids Res.*, **40**, 3218–3231.
41. Daubner,G.M., Cléry,A., Jayne,S., Stevenin,J. and Allain,F.H.T. (2012) A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.*, **31**, 162–174.
42. Safaee,N., Kozlov,G., Noronha,A.M., Xie,J., Wilds,C.J. and Gehring,K. (2012) Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol. Cell*, **48**, 375–386.
43. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.
44. Schoening,J.C., Streitner,C., Page,D.R., Hennig,S., Uchida,K., Wolf,E., Furuya,M. and Staiger,D. (2007) Auto-regulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *Plant J.*, **52**, 1119–1130.
45. Jeong,B.-R., Lin,Y., Joe,A., Guo,M., Korneli,C., Yang,H., Wang,P., Yu,M., Cerny,R.L., Staiger,D. *et al.* (2011) Structure function analysis of an ADP-ribosyltransferase type III effector and its RNA-binding target in plant immunity. *J. Biol. Chem.*, **286**, 43272–43281.
46. Abdul-Manan,N. and Williams,K.R. (1996) hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res.*, **24**, 4063–4070.
47. Tavanez,J.P., Madl,T., Kooshapur,H., Sattler,M. and Valcárcel,J. (2012) hnRNP A1 proofreads 3′ splice site recognition by U2AF. *Mol. Cell*, **45**, 314–329.
48. Mason,K.E., Tripet,B.P., Parrott,D., Fischer,A.M. and Copié,V. (2014) (1)H, (13)C, (15)N backbone and side chain NMR resonance assignments for the N-terminal RNA recognition motif of the HvGR-RBP1 protein involved in the regulation of barley (Hordeum vulgare L) senescence. *Biomol. NMR Assign.*, **8**, 149–153.
49. Van Dusen,C.M., Yee,L., McNally,L.M. and McNally,M.T. (2010) A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Mol. Cell. Biol.*, **30**, 2552–2562.
50. Ziemienowicz,A., Haasen,D., Staiger,D. and Merkle,T. (2003) Arabidopsis transportin1 is the nuclear import receptor for the circadian clock-regulated RNA-binding protein AtGRP7. *Plant Mol. Biol.*, **53**, 201–212.
51. Bhatt,P., d'Avout,C., Kane,N.S., Borowiec,J.A. and Saxena,A. (2011) Specific domains of nucleolin interact with Hdm2 and antagonize Hdm2-mediated p53 ubiquitination. *FEBS J.*, **279**, 370–383.
52. Zu,K., Sikes,M.L. and Beyer,A.L. (1998) Separable roles in vivo for the two RNA binding domains of Drosophila A1-hnRNP homolog. *RNA*, **4**, 1585–1598.