



RESEARCH PAPER

 OPEN ACCESS 

The RBS1 domain of Gemin5 is intrinsically unstructured and interacts with RNA through conserved Arg and aromatic residues

Azman Embarc-Buh ^a, Rosario Francisco-Velilla ^a, Sergio Camero ^b, José Manuel Pérez-Cañadillas ^b, and Encarnación Martínez-Salas ^a

^aCentro de Biología Molecular Severo Ochoa, CSIC-UAM, Madrid; ^bInstituto de Química Física Rocasolano, CSIC, Madrid

ABSTRACT

Gemin5 is a multifaceted RNA-binding protein that comprises distinct structural domains, including a WD40 and TPR-like for which the X-ray structure is known. In addition, the protein contains a non-canonical RNA-binding domain (RBS1) towards the C-terminus. To understand the RNA binding features of the RBS1 domain, we have characterized its structural characteristics by solution NMR linked to RNA-binding activity. Here we show that a short version of the RBS1 domain that retains the ability to interact with RNA is predominantly unfolded even in the presence of RNA. Furthermore, an exhaustive mutational analysis indicates the presence of an evolutionarily conserved motif enriched in R, S, W, and H residues, necessary to promote RNA-binding via π - π interactions. The combined results of NMR and RNA-binding on wild-type and mutant proteins highlight the importance of aromatic and arginine residues for RNA recognition by RBS1, revealing that the net charge and the π -amino acid density of this region of Gemin5 are key factors for RNA recognition.

ARTICLE HISTORY

Received 2 June 2021
Revised 19 July 2021
Accepted 27 July 2021

KEYWORDS

Gemin5; RNA-protein interaction; non-canonical RNA binding site; intrinsically disordered regions; NMR

Introduction

Gemin5 is a predominantly cytoplasmic protein involved in small nuclear ribonucleoproteins (snRNPs) assembly and translation control [1]. The protein was initially reported as the RNA-binding protein (RBP) of the survival of motor neurons (SMN) complex [2]. More recently, the protein has been implicated in translation control, and gene expression reprogramming [3,4,5].





Separate domains of Gemin5 are responsible for the recognition of distinct targets, either RNAs or proteins. In particular, different regions of the protein recognize the Sm site and stem-loops (SL) of small nuclear RNAs (snRNAs), or the internal ribosome entry site (IRES) element of foot-and-mouth disease virus (FMDV) genomic RNA as well as an internal region of Gemin5 mRNA (reviewed in [6]). Beyond the WD40 repeats domain located at the N-terminal region involved in the recognition of snRNAs [7], the protein harbours a robust dimerization domain (tetratricopeptide (TPR)-like) in the central region [8] and a bipartite non-conventional RNA-binding site (designated RBS1-RBS2) [9] towards the C-terminus. Furthermore, Gemin5 associates through its N-terminal domain to the ribosome down-regulating global protein synthesis [10].

The RBS1 domain is involved in the recognition of viral IRES elements [9] and cellular RNAs [11], including an

internal region of Gemin5 mRNA (designated H12). This mutual recognition results in a positive feedback loop that counteracts the negative effect of Gemin5 on global protein synthesis. Nonetheless, the RNAs recognized by RBS1 domain do not contain a consensus sequence. Instead, RNAs are enriched in secondary structured elements, in agreement with previous studies suggesting that RNA secondary structure affects Gemin5-RNA interaction [7,12,13].


Computational methods developed to predict the coevolution between a protein and its RNA partner [14] allowed the identification of coevolving pairs between the RBS1 residues and the H12 RNA sequence [15]. The coevolving residues, which are centred around the PXSS motif, are evolutionarily conserved, suggesting that the inherent sequence diversity of this region is neutralized by the need for conservation of functional elements. Consistent with this notion, mutant RBS1 proteins carrying deletions or substitutions on the PXSS motif within the predicted coevolving residues drastically reduced RNA-binding capacity, suggesting that selection of variants during RNA-protein coevolution contributes to fine-tune the expression levels of this multitasking factor [16].

The modular architecture of RBPs and the spatial arrangement of the RNA-binding domains (RBDs) are thought to be important for the specificity of target RNA binding. Studies carried out over the years have established a number of

CONTACT Encarnación Martínez-Salas  emartinez@cbm.csic.es  Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Nicolás Cabrera 1, Madrid 28049, Spain; José Manuel Pérez-Cañadillas  jmperez@iqfr.csic.es  Instituto de Química Física Rocasolano, CSIC, Serrano 119, Madrid

This article has been republished with minor changes. These changes do not impact the academic content of the article.

*These authors contributed equally to this work

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

conventional RBDs according to their structural composition and RNA recognition features [17]. However, recent global procedures have discovered numerous RBPs harbouring previously unknown RBDs [18], many of which contain intrinsically disordered regions (IDRs) [19,20]. Non-conventional RBDs generally consist of heterogeneous sequences, hampering the identification of novel RBPs lacking canonical RBDs by conventional methodologies.

Previous attempts to characterize in solution the RBS1 polypeptide by NMR suggested that the three-dimensional structure behaves as an ensemble of flexible conformations rather than having a defined tertiary structure [9]. Remarkably, IDRs lack a defined tertiary structure in the native state, but play important roles in many biological processes involving the assembly of macromolecular complexes [21]. However, RBS1 differs from typical IDRs of other RBPs in the absence of RGG boxes, RS dipeptides, GY motifs and G-rich tracts, as well as lacking high content of aromatic residues (F, W, Y, H) [22]. Noncovalent interactions of aromatic rings play a key role in DNA- and RNA-protein interactions, stabilizing the structure of the macromolecular complex. More specifically, numerous studies have demonstrated the prevalence of π - π stacking interactions in complexes involving proteins and RNAs. These interactions can form between any nitrogenous base ring and a π -containing amino acid, which includes the aromatic residues Trp, His, Phe, and Tyr, as well as the charged residues Arg, Glu, and Asp [23].

A prominent feature of the RBS1 domain is that the coevolving amino acids reside on the most conserved motif of the IDR, suggesting that these residues are important for RNA-binding [15]. Accordingly, deletion of these residues as well as substitution of the PXSS motif to amino acids with different chemical properties resulted in a decrease in Gemin5 binding to H12 mRNA. To better understand the RNA binding features of RBS1 we have attempted the structural characterization by NMR in connection to RNA-binding activity in solution. The results obtained indicate that this protein is largely unfolded even in the presence of RNA, and demonstrate the presence of a R, S, W, H-rich motif, necessary to promote RNA-binding via π - π interactions.

Results

The RBS1 domain of Gemin5 adopts a flexible unfolded structure in solution

We studied the conformational properties of Gemin5 RBS1 by NMR, a technique highly suitable to analyse intrinsically disordered proteins. The original domain definition spans residues 1287–1412 (Fig. 1A). The RBS1 region of Gemin5 is more variable than other domains of the protein. Nonetheless, sequence alignment reveals a short stretch of conserved amino acids towards the N-terminus of the predicted IDR (Fig. 1A). In particular, the most N-terminal region (positions 1294–1307) is conserved among mammalian

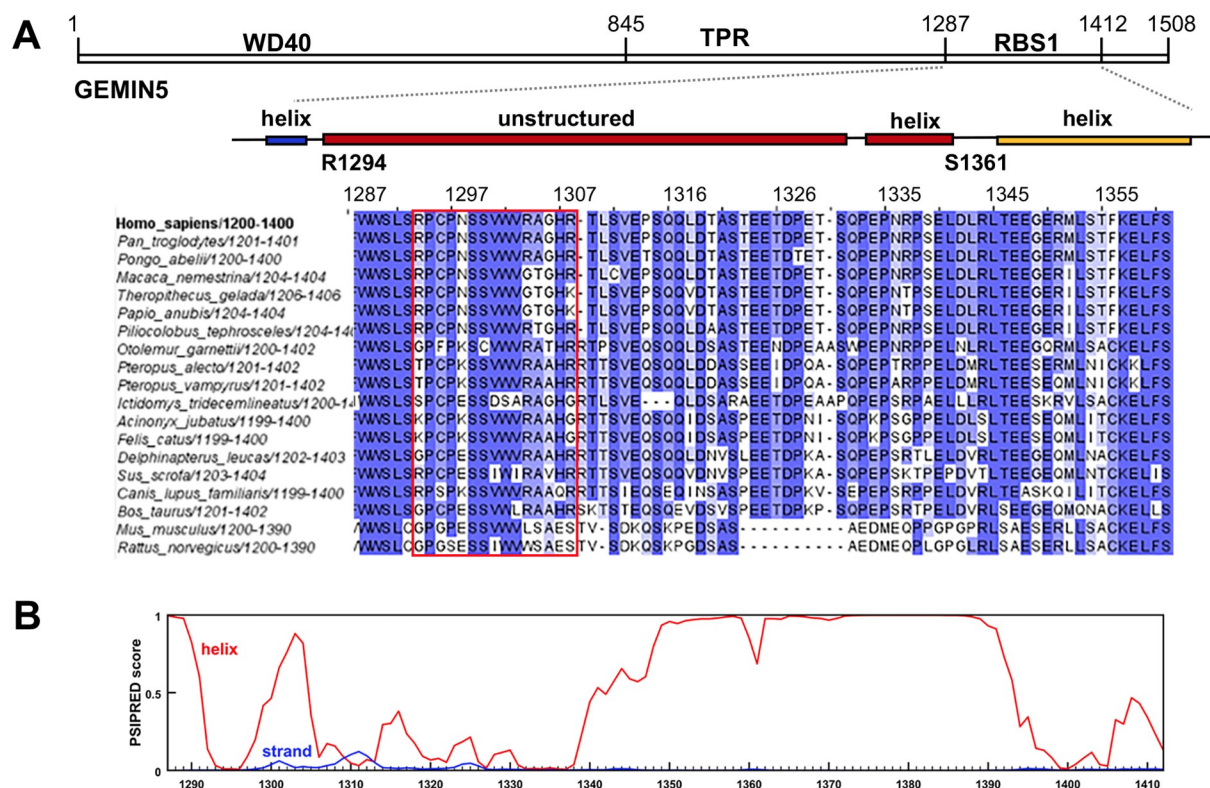


Figure 1. Main features of Gemin5 and conservation of the RBS1 domain. A) Schematic of Gemin5 protein. Numbers indicate the amino acids flanking the WD40 repeats domain, the central region comprising the TPR dimerization module, and the RBS1 domain of the protein (top panel). Alignment of Gemin5 sequences from mammals spanning the RBS1 domain. Residues are coloured according to their identity. Numbers above the sequence denote residue number. The predicted α -helices of RBS1 and the intrinsically disordered region (IDR) are depicted above the sequence. The red rectangle across the amino acid sequences depicts the most conserved zone of the IDR. B) Predicted structure of the RBS1 domain using the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/>).

species, and comprises all the coevolving amino acids, including the PXSS motif, shown to be important for H12 RNA binding [15].

The earlier NMR spectra of RBS1₁₂₉₇₋₁₄₁₂ evidenced some aggregation/oligomerization that prevented its assignment [9]. To overcome this problem, we designed two shorter constructs that remove part of the C-terminal helix of RBS1, as predicted by PSIPRED (Fig. 1B) preserving the sequences that coevolve with Gemin5 RNA in the N-terminus [6]. These new constructs: HIS-RBS1₁₃₆₁ and RBS1₁₃₆₁-HIS show suitable NMR spectra devoid of aggregation problems (see below). Thus, we proceeded to validate their RNA binding activity first.

The RNA encompassing domain 5 (d5) of the FMDV IRES element is a well-known target of Gemin5 [4]. It consists of 46-mer RNA folding into a conserved hairpin followed by a single-strand region (d5ss) [24,25]. RNA binding studies conducted with the HIS-RBS1₁₄₁₂ construct using d5 and

d5ss probes indicated that the protein interacts with both RNAs to a similar extent (Fig. 2A). In both cases two retarded complexes were observed, a high mobility one up to 500 nM of protein, and a slow mobility one above this concentration of protein (Fig. 2B). Previous data have shown that the HIS-RBS1₁₄₁₂ protein can recognize RNAs differing in length and secondary structure [9,11]. However, in support of the RNA-binding specificity of RBS1, similar binding assays using the synthetic oligoribonucleotide U(5) failed to produce a retarded complex (Fig. 2B, bottom panel). Furthermore, previous binding assays conducted with a pyrimidine-rich synthetic RNA, and the long H34 RNA as well, yielded negative results [15].

Interestingly, the protein RBS1₁₃₆₁-HIS encompassing mostly the predicted unstructured region of RBS1 retained RNA-binding activity (Fig. 2C), although with moderate affinity relative to HIS-RBS1₁₄₁₂ (Table 1). Similar results were observed with HIS-RBS1₁₃₆₁ (Fig. 2D). Further reinforcing the

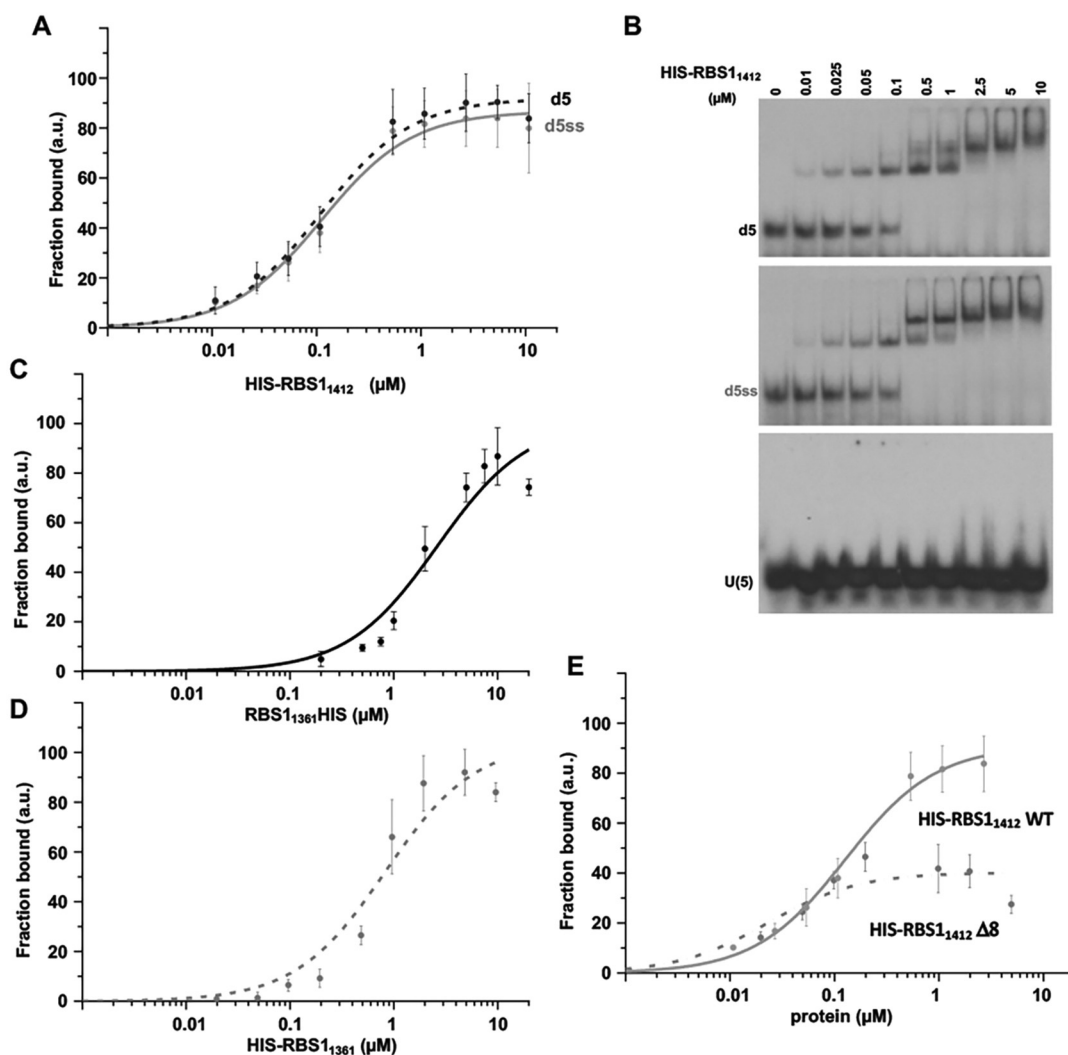


Figure 2. RNA binding studies of RBS1. A) Graph representing the adjusted curves obtained from the quantification (mean \pm SEM) of three independent gel-shift assays using d5 and d5ss probes (broken black line and grey line, respectively) incubated with increasing amounts of HIS-RBS1₁₄₁₂ protein. B) Representative examples of the gel-shift assays conducted with HIS-RBS1₁₄₁₂ protein and labelled d5, d5ss RNA and a synthetic U(5) RNA (top, medium, and bottom panels, respectively). C) Graph representing the adjusted curves obtained from the quantification of three independent gel-shift assays using d5ss labelled RNA incubated with increasing amounts of RBS1₁₃₆₁-HIS (C), or HIS-RBS1₁₃₆₁. (D). E) Graph representing the adjusted curves obtained from the quantification (mean \pm SEM) of three independent assays using d5ss labelled RNA incubated with increasing amounts of HIS-RBS1₁₄₁₂ or HIS-RBS1₁₄₁₂ Δ 8 proteins.

Table 1. RNA-binding affinity of the RBS1 constructs.

Protein	d5ss RNA	d5 RNA
	$K_D \pm SD$ (μ M)	$K_D \pm SD$ (μ M)
RBS1 ₁₃₆₁ -HIS WT	2.63 \pm 0.52	0.98 \pm 0.13
RBS1 ₁₃₆₁ -HIS R1294K	27.55 \pm 12.86	7.72 \pm 2.30
RBS1 ₁₃₆₁ -HIS R11304 K	3.20 \pm 0.57	2.28 \pm 0.30
RBS1 ₁₃₆₁ -HIS R11308 K	2.82 \pm 0.84	4.82 \pm 1.16
RBS1 ₁₃₆₁ -HIS W1302A	5.06 \pm 1.26	6.12 \pm 1.29
RBS1 ₁₃₆₁ -HIS H1307A	4.15 \pm 1.21	2.48 \pm 0.29
RBS1 ₁₃₆₁ -HIS P1296G	1.51 \pm 0.29	0.60 \pm 0.08
RBS1 ₁₃₆₁ -HIS SS-AA	1.74 \pm 0.43	0.77 \pm 0.19
RBS1 ₁₃₆₁ -HIS SS-TT	4.38 \pm 1.14	2.27 \pm 0.41
HIS-RBS1 ₁₄₁₂ WT	0.11 \pm 0.02	0.11 \pm 0.01
HIS-RBS1 ₁₃₆₁ WT	0.83 \pm 0.17	0.73 \pm 0.21
HIS-RBS1 ₁₃₆₁ P/E	2.56 \pm 0.46	1.09 \pm 0.22
HIS-RBS1 ₁₃₆₁ SS/DD	2.11 \pm 0.61	1.28 \pm 0.29

involvement of the N-terminal region of RBS1 in RNA interaction, a deletion construct (HIS-RBS1₁₄₁₂ Δ 8) that lacks residues 1297–1304, revealed a strong RNA binding decrease (Fig. 2E).

Given that attempts to study of RBS1₁₄₁₂ by NMR showed aggregation problems at the concentrations needed for these

experiments, we used RBS1₁₃₆₁ instead. For this, we studied the two versions of the protein with the HIS tag at N- or C-terminus. The ¹H-¹⁵N HSQC of the RBS1₁₃₆₁-HIS protein (Fig. 3A) showed sharp cross peaks, homogeneous in their linewidths and poorly dispersed in the proton dimension, all features characteristic of intrinsically unstructured proteins. We obtained its complete NMR assignment using the standard triple resonance experiments recorded on a ¹³C/¹⁵N labelled sample. The NMR spectra of HIS-RBS1₁₃₆₁ highly overlap with the RBS1₁₃₆₁-HIS in the Gemin5 RBS1 region (Supplementary Fig. 1), indicating that the position of the tag has little effect on its conformation. We conclude that these versions of the RBS1 protein are soluble at high concentration and display RNA-binding capacity to allow structural studies by NMR.

To further explore the conformational propensities of RBS1₁₃₆₁ region we analysed if there are residual secondary structure propensities using ¹³C chemical shift deviations from the random coil values. The percentages of regular secondary structure were calculated with the program δ 2d

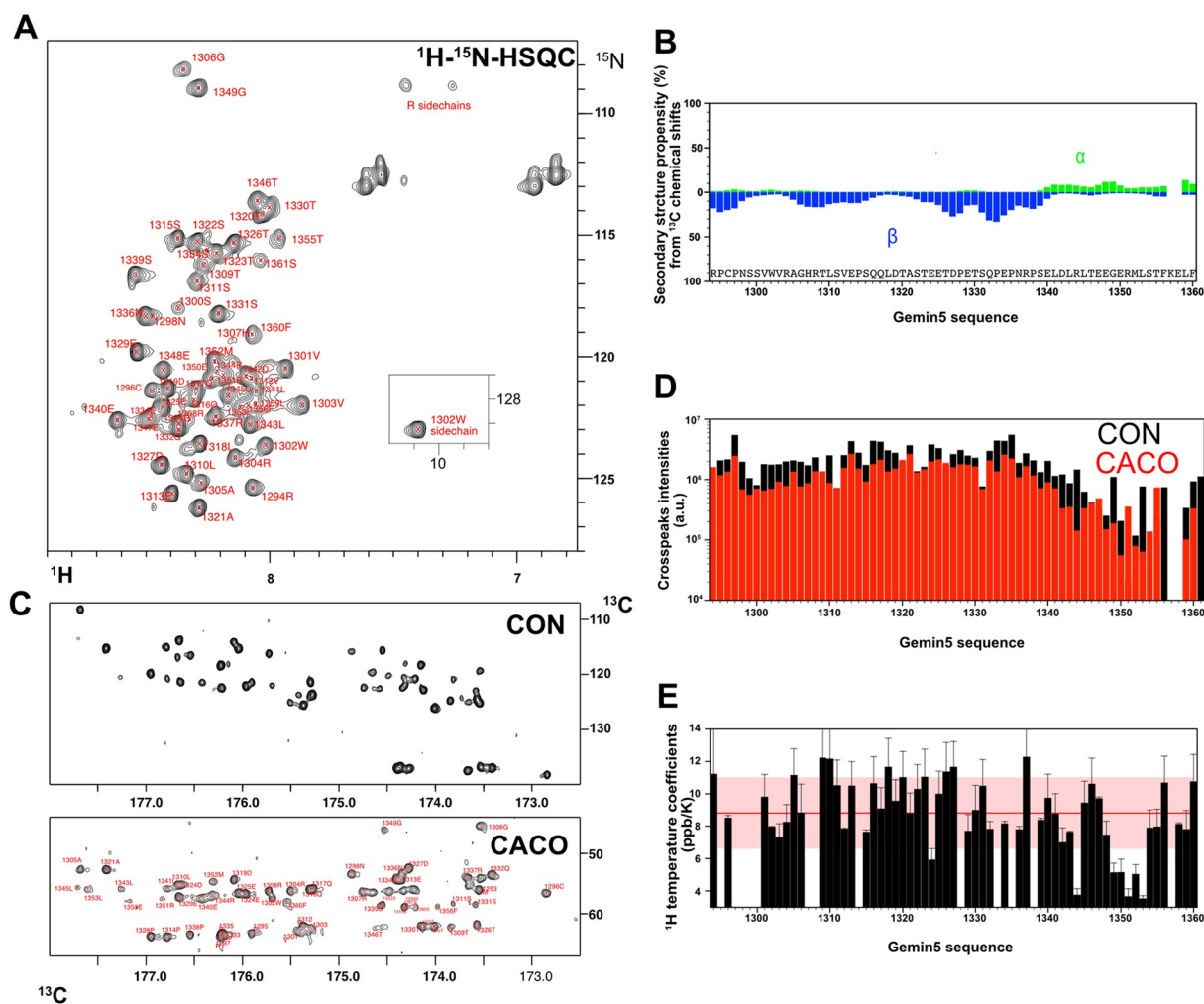


Figure 3. NMR structural characterization of RBS1₁₃₆₁-HIS. A) ¹H-¹⁵N HSQC with assignments. The inset corresponds to the side chain of Trp1302. B) Percentages of secondary structure calculated from experimental ¹³C chemical shifts of backbone atoms. The sequence of the protein constructs is shown below. C) Carbon-detected 2D CON and CACO spectra. As in other proteins containing IDRs these spectra offer a better dispersion than the HSQC. D) Intensity of the peaks in the spectra in C showing a systematic decrease towards the C-terminus of the construct. E) Temperature coefficients of the amide protons, the pink area highlights the values normally expected for unprotected amides (i.e. like those in unfolded regions). Few consecutive residues at the C-terminal region show values below this area. The sequence coordinates of the histograms B, D, and E have been aligned for better comparison of the different NMR-derived data along the sequence.

[26] (Fig. 3B). Residues from 1294 to 1340 showed some residual β -strand tendency, whereas the last part, from 1340 to 1361 indicated a slight propensity for α -helix. However, these residual structural elements displayed percentages below 30%. We also obtained and assigned ^{13}C -detected data: 2D NCO and 2D CACO [27] (Fig. 3C). These experiments are particularly sensitive to the existence of residual structural elements in IDR of proteins. The C-terminal (1340–1361) residues of RBS1₁₃₆₁-HIS showed lower intensity cross peaks in these spectra or even undetectable (Fig. 3D), suggesting some source of chemical exchange affecting this region. In contrast, the rest of the polypeptide exhibited high-intensity signals typical of highly dynamic disordered regions.

Finally, we monitored temperature-related changes in the proton amide signals and obtained the corresponding temperature coefficients. These coefficients take values around 8.8 ppb/K (Fig. 3E), for the N-terminal part of the construct, but are lower for the segment 1349–1353, coinciding with part of the helical prediction (Fig. 1), suggesting some level of protection.

In summary, the gel-shift data show that Gemin5 RBS1 constructs up to residue 1361 retain the RNA-binding ability, although with slightly lower affinity than the complete RBS1₁₂₉₇₋₁₄₁₂. The NMR data show that these constructs are mainly unstructured but with a low percentage of α/β

secondary structure propensities. The C-terminal part of the constructs, coinciding with the low helical population, shows evidences of interaction: decreased signals on CACO and CON spectra and lower temperature coefficients, perhaps due to an incipient self-association process that becomes more important in the complete RBS1 construct (residues 1297–1412).

Conserved RSWH residues within the RBS1 domain confer RNA recognition

As shown in Fig. 2A,B,2C, the protein RBS1₁₃₆₁-HIS interacts with d5ss RNA in a dose-dependent manner. This result prompted us to identify which specific sequence of RBS1₁₃₆₁ is directly involved in this interaction by following the changes in its NMR spectrum upon titration with d5ss. Titrations were performed until 5-fold excess of RNA (Fig. 4A). The changes in the NMR signal occurred in the fast exchange regime, which is typical of weak binding (Fig. 4B). The chemical shift mapping showed that the interaction is mostly located at the N-terminus: the peaks experiencing the larger changes ($\Delta\delta > 0.075$ ppm) were Arg1294, Trp1302 side-chain and His1307 (Fig. 4C). Residues showing moderate changes ($0.075 > \Delta\delta > 0.05$ ppm) are Cys1296, Asn1298, Ser1299, Gly1306, Arg1308, Thr1309, Leu1310

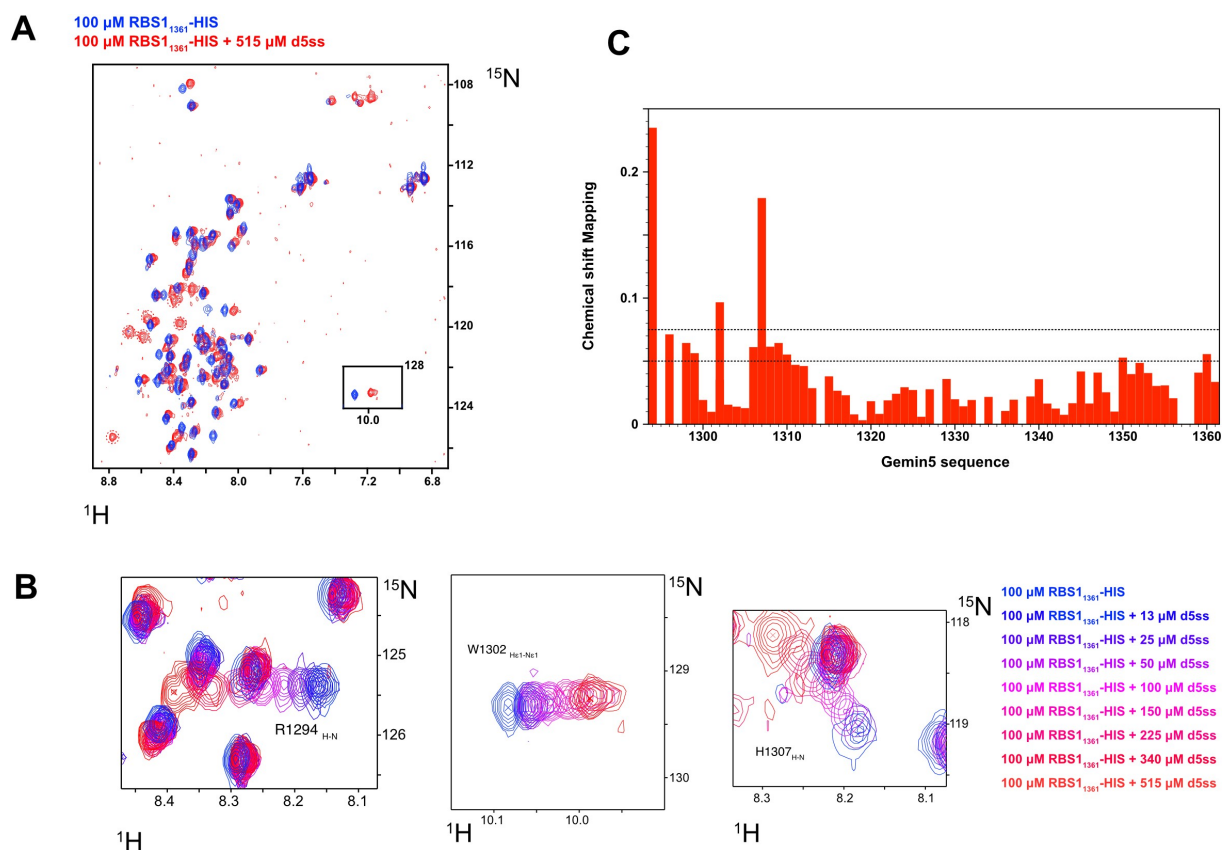


Figure 4. RNA binding studies of RBS1₁₃₆₁-HIS followed by NMR. A) Superposition of the ^1H - ^{15}N HSQC spectra of the free (blue) and RNA-bound (pink) states of the protein upon binding the d5ss RNA. Signals uncycled correspond to the backbone resonances of the C-terminal tag that only appears upon titration with the highest RNA concentration. B) Detailed views of the variation of the signals in the HSQC spectra during the titration. The spectra have been coloured according to the amount of RNA (d5ss) in the titration, following the code on the right. The first and last point corresponds to the spectra shown in full in A). The three signals showing the larger variations are shown. C) Chemical shift perturbation obtained by comparison of the free and bound forms of the spectra in A). Dashed lines mark the limits for residues showing high and medium perturbation (see text for further details).

Glu1350, Met1352, and Phe1360. Of note, most of these residues are located between Arg1294-His1307, coinciding with the region that contains coevolving pairs of RBS1 with Gemin5 mRNA [15].

We also observed a new set of signals appearing at the highest RNA concentration point (Fig. 4A). These correspond to the C-terminal HIS-tag and probably showed up due to small variations in the pH that alter the complex chemical exchange equilibrium in this part of the construct, or to weak RNA interactions. An equivalent NMR study titrating the d5 probe showed similar results (Supplementary Fig. 2) consistent with the hypothesis that Gemin5 RBS1 interacts with the unstructured part of this RNA.

We then focused on Arg1294, Trp1302 and His1307 whose side chains contain chemical groups capable of potential π - π interactions. In the human sequence, there are two other arginine residues Arg1304 and Arg1308 flanking the conserved PXSS motif (Fig. 1A). We made mutations on all of these elements: replacements of arginine by lysine, to analyse the effect of the guanidinium group while keeping the positively charged character, and to alanine in all the other cases. The mutations seek to remove π - π interactions with RNA bases but also to reduce the potential hydrogen bonds. Electrophoretic mobility shift assays done with RBS1₁₃₆₁-HIS wild type and mutant constructs showed binding of all tested proteins to d5ss and d5 RNAs (Table 1). The R1294K mutant showed the biggest decrease in d5ss binding affinities (Fig. 5A). Similar results were observed with d5 (Supplementary Fig. 2B). The other two R to K substitutions also caused affinity drops, but are not as important as in the R1294K case. Considering that charge-charge interactions might be similar in arginine and lysine, these results strongly

suggest that protein-RNA interactions are guided by π - π contacts, presumably with the unpaired bases of the d5 and d5ss RNAs. Alternatively, the higher hydrogen bond potential of Arg *versus* Lys might also contribute to the lower affinity of these mutants. Regarding the aromatic to alanine substitutions, both W1302A and H1307A exhibited lower affinity (Fig. 5B, Table 1), further reinforcing the role of π - π contacts in RNA recognition.

Two residues from the conserved PXSS motif revealed moderate chemical shift perturbations (Fig. 4C). Considering that a previous study replacing this motif with acidic residues resulted in a significant drop of binding to H12 RNA [15] we performed a systematic mutational analysis of the conserved motif PXSS replacing P1297 to G and the Ser-Ser pair to Ala-Ala and Thr-Thr. In the first case, we sought to investigate the possible role of proline *cis* conformation in recognition, whereas in SS mutants we aimed to look at the role of the side-chain hydroxyl. We observed a similar affinity of the P1297G mutant for d5ss than the wild-type protein (Fig. 5C), indicating that the proline *cis* conformation is not determinant in RNA binding. The Ser to Thr double mutants bound slightly less efficient than Ser to Ala according to the calculated apparent K_D values (Table 1), suggesting that the inclusion of the methyl group might cause some steric hindrance. The hydroxyl seems to be dispensable as SS-AA mutant displayed nearly identical binding curve for d5 probe.

Taken together, the mutations in the PXSS to Ala and Thr appear to have less impact than R1294K and W1302A mutations, but their high level of conservation suggests that this motif might play yet unknown additional roles in Gemin5 function. Moreover, considering that the double SS substitution for negatively charged residues (DD) reduced the RNA-

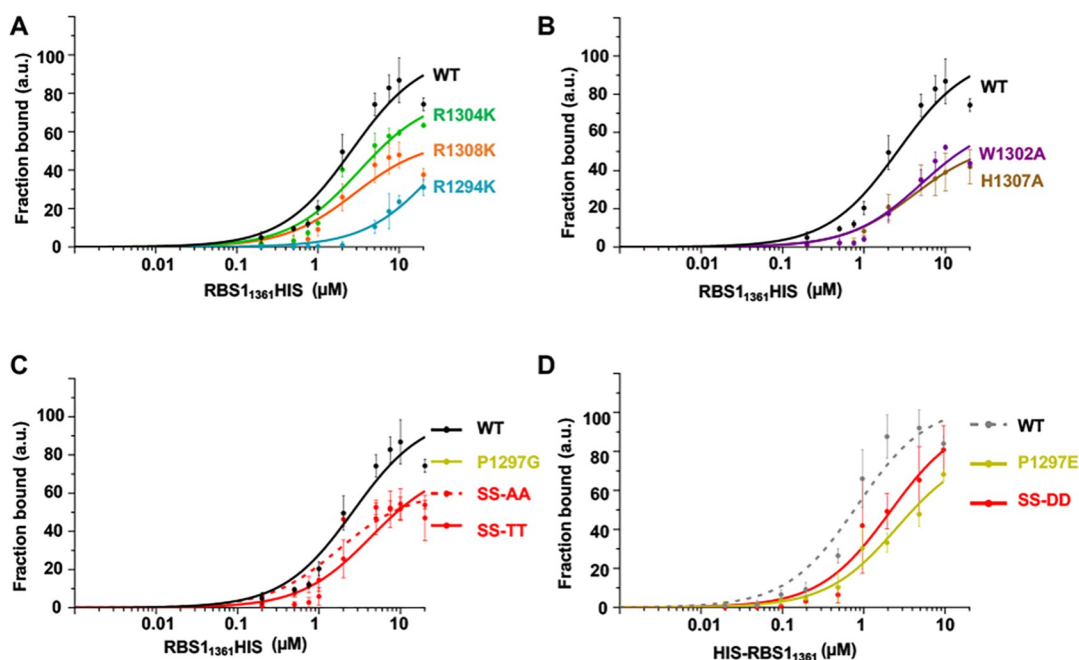


Figure 5. RBS1₁₃₆₁-HIS substitution mutants display lower RNA-binding capacity. A) Graph representing the adjusted curves obtained from the quantification (mean \pm SEM) of triplicate independent assays using d5ss probe incubated with increasing amounts of RBS1₁₃₆₁-HIS WT (black line), and mutants R1304K (green), R1308K (orange), or R1294K (blue). B) Similar set of independent assays performed with W1302A (violet) and H1307A (brown). C) RNA-binding assays conducted with P1297G (light green), SS-AA, and SS-TT (red broken or filled lines, respectively). D) RNA-binding results obtained for P1297E (light green) and SS-DD (red) inserted in HIS-RBS1₁₃₆₁ construct (broken grey line).

binding ability of the protein (Fig. 5D), we hypothesize that the net charge and the π -amino acid density of this region are important for RNA binding. In line with this view, the RBS1₁₄₁₂SS-DD mutant exhibited a significant drop in binding affinity for H12 RNA [15]. This result is consistent with the lower affinity of the SS-TT mutant as both aspartic and threonine are bulkier than Ser or Ala. Therefore, the data suggest that at least one of the two serines of the PXSS motif requires a small side-chain residue for efficient RNA recognition.

We also noticed that the region around 1352, at the C-terminus, displayed subtle changes in the presence of RNA (Fig. 4C). Additionally, several new signals showed up in the up-right corner of the spectra, which can correspond to folded N ϵ -H ϵ cross peaks of arginine (from ~80 ppm in ¹⁵N). RBS1₁₃₆₁-HIS have six arginine residues, but there are only four cross peaks that might correspond to side-chain correlations of residues 1294, 1304, 1308 and 1351, which are located in regions that show chemical shift perturbations. Whichever the case, the appearance of the N ϵ -H ϵ cross peaks suggests that the interaction involves the guanidium moiety of these residues.

Collectively, the reduced RNA binding activity of the individual mutations and the deletion suggest the presence of a novel RNA-binding motif where the R, S, W, and H composition provides a flexible architecture enabling RNA-binding.

Discussion

Gemin5 earlier reported function was as a component of the SMN complex, a macromolecular entity involved in the assembly of snRNPs [28,29]. The recognition of snRNAs resides in the N-terminal half of the protein containing 14 WD40 repeats domain that make base-specific contacts with RNA [30,31,32]. Besides, Gemin5 C-terminal half comprises several domains, including a TPR-like, that form a stable homodimer [8] and two non-canonical RNA binding domains (RBS1 and RBS2) following it [9]. The fact that Gemin5 C-terminal segment is proteolyzed by the Leader protease of FMDV during infection yielding the p85 fragment that enhances viral IRES-dependent translation [12] further supports the notion that the C-terminal domains of the protein have different functions to the N-terminus. Here, we analysed the conformation and RNA binding properties of a protein fragment of the non-canonical RNA binding domain RBS1 at residue level.

According to the NMR data, the Gemin5 RBS1₁₂₉₄₋₁₃₆₁ fragment is preferentially unstructured with some residual secondary structure tendencies that should not be neglected. The C-terminal part that coincides with a surplus α -helix shows some oligomerization/aggregation tendency. It is possible that these remnant structures could be stabilized in the context of RNA or protein binding, but in our interaction studies with d5 and d5ss RNA probes we did not notice such behaviour.

Mapping of the RBS1 residues involved in RNA-binding pointed to a conserved RSWH-rich motif at the N-terminus of the unfolded region. Remarkably, this short stretch is

conserved among mammals, although the degree of conservation decreases in other vertebrata, such as birds, reptiles, amphibians, and fishes. Whether the higher conservation in mammals is connected to evolutionary selection of RNA-binding activity involved in RNA-dependent processes shared by this group of chordata, such as spliceosome assembly and translation regulation, needs to be studied in the future.

The combined results of NMR and RNA gel shift on wild type and several mutants highlight the importance of aromatic and arginine residues on RNA recognition by Gemin5 RBS1. These types of residues are able to interact through π -stackings with RNA bases [23]. In the case of positively charged residues Arg and His (typically a pH < 6.8), the interaction is theoretically stronger because they add the π -cation effect. Besides, the ability of the guanidium group to make multiple hydrogen bonds simultaneously makes Arg a highly versatile residue in RNA and protein recognition (Fig. 6A) [33]. Moreover, several computational studies show that Arg guanidium group interacts preferentially with guanine and cytosine by making two simultaneous hydrogen bonds with Watson-crick (in cytosine) and Hoogsteen (in guanine) faces [34,35,36].

Interestingly, our previous genome-wide meta-analysis [11] identified G/C rich RNA sequences as preferential targets for Gemin5 RBS1, reinforcing the view of Arg as a key residue for RNA recognition. Indeed, the observed K_D for the interaction of RBS1₁₄₁₂ with H12 RNA (0.99 ± 0.01) is within the same range of d5 and d5ss RNAs (Table 1). Therefore, the conservation of Arg, flanking the PXSS motif and in other places of RBS1, further highlights the importance of this type of residue (Fig. 1B). The mutations to lysine remove the possibility to make π -stacking and multiple hydrogen bonds interactions, although they maintain the π -cation interactions with RNA bases, or charge-charge interactions with the phosphate backbone.

The fact that mutant proteins interact with less affinity supports that the π -cation interactions are important in RNA recognition by Gemin5 RBS1. However, it is also clear from our mutagenesis analysis that different Arg residues have dissimilar contribution to binding, showing that the sequence context might enable some sort of diffuse RNA-binding selectivity that is evolutionary selected [15]. It is possible that the tract of RSWH conserved residues interspersed at the N-terminus of RBS1 recognizes RNA sequences using a combinatorial approach of transient interactions similar to those represented in Fig. 6A, favouring some specific ribonucleotide sequences over others. This RNA binding mode is more selective than the simple interactions with phosphodiester backbone, but less than the recognition modes that folded protein domains can achieve, alone or in tandems [17].

Heterotypic π - π interactions and hydrogen bond interactions with protein amino acids require the RNA bases to be accessible. Single-strand segments, bulged-out bases and internal or apical loops in the secondary structure of RNA seem to be targets of Gemin5 RBS1, in agreement with RBS1-H12 RNA footprint data [15]. Other factors like the number of consecutive unpaired bases or their 3D arrangements in the context of the RNA structure would be

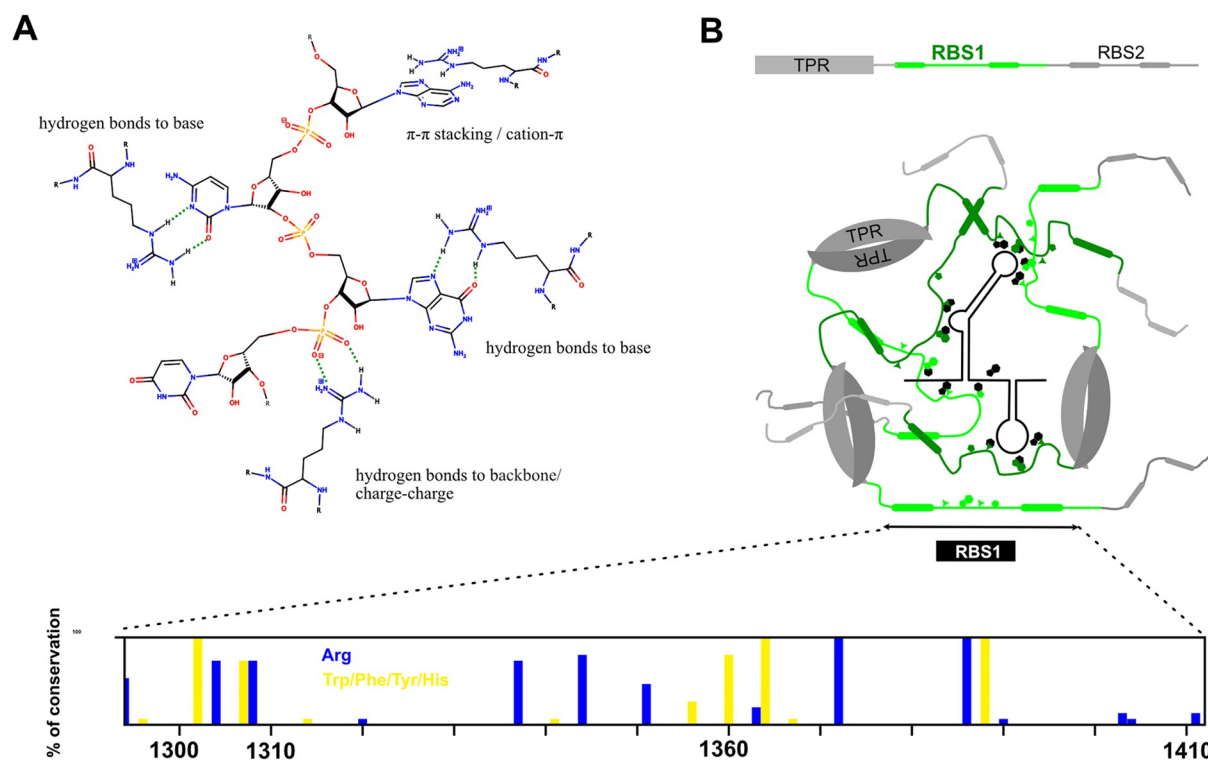


Figure 6. Proposed model of Gemin5 C-terminal domain RNA recognition mode. A) Examples of transient interactions involving Arg side chains that might use Gemin5 RBS1 to recognize RNA. The guanidinium group can make dual hydrogen bond interactions with RNA bases of backbone phosphates and π - π stackings with aromatic rings in the RNA. For more exhaustive analysis of Arg-RNA/DNA interactions refer to [23,34,35,36], and the references therein. B) Schematic proposed model of Gemin5 p85 recognition of IRES RNA. The number of Gemin5 molecules (depicted in greys and greens) and the RNA (depicted in black) molecules represent a simplified model enabling multiple Gemin5 molecules to recognize a single IRES simultaneously. The secondary structure of the RNA denotes a short IRES region containing different structural elements. In the model Gemin5 binds exposed bases in RNAs using aromatic and Arg residues of the RBS1 region, presumably with some sequence selectivity as shown previously [11,15]. π -stacking and specific hydrogen bonds, similar to those in panel (A), could play a leading role in RNA recognition, without excluding other sources of protein-RNA interactions. The intrinsic flexibility of the RBS1 domain would be essential to get access to different RNA elements. These weak and transient interactions would be combined thanks to homodimerization processes mediated by TPR-like domain [8] and/or predicted coiled-coil interactions between RBS1. Combined, the adaptability and multiplicity of interactions provided by Gemin5 RBS1 will make possible the recognition of the IRES, and possibly of other RNAs.

favourable targets of Gemin5 RBS1. It is possible that Gemin5 RBS1 itself could use these weak interactions to guide the RNA folding process itself or by recognizing specific features of the RNA fold. This could be particularly important in the recognition of viral RNAs, including IRES elements like the one present in FMDV genomic RNA [37]. Remarkably, the p85 fragment resulting from Gemin5 cleavage during FMDV infection [12] comprises the TPR-like homodimerization domain [8], followed by the non-canonical RNA-binding domains RBS1 and RBS2 [9] (Fig. 6B). As shown here, the RBS1 moiety recognizes the IRES element through domain 5, capable to interact with its Arg and aromatic residues. We hypothesize that the presence of helical regions within RBS1, capable of forming coiled-coil dimers (or oligomers), would boost this mechanism of recognition (Fig. 6B), at least in part explaining the multiband pattern shown in the RNA gel-shift experiments (Fig. 2B). Hence, the RBS1 domain is possibly assisted by homodimerization domains in p85 (such that canoe shaped TPR-like module) makes possible a sophisticated structure-selective recognition of the IRES element. Along this line, in our recent coevolution study of Gemin5 and Gemin5 mRNA [15], we proposed a mechanism of activation/

repression of Gemin5 translation that is based in the interaction with selective partners.

The purified RBS1₁₃₆₁-HIS protein forms two retarded complexes of different mobility with the RNAs used in this study, d5 and d5ss. Currently, we do not know if these complexes reflect a transient interaction involving 1:1 molecule at low protein concentration, followed by a cooperative effect as the concentration of protein in the reaction increases. However, it is remarkable that the longer form of the protein RBS1₁₂₉₇₋₁₄₁₂ exhibit the same properties, and also a similar observed K_D , further supporting that this is an intrinsic feature of the RBS1 domain.

Understanding protein-RNA recognition and RNA-binding specificity is a prerequisite for obtaining mechanistic insights into how RBPs regulate RNA lifespan. In the work reported here, we discovered that Gemin5 contains an unfolded flexible region within the RBS1 domain, which plays a critical role in RNA binding. The ultimate verification of our hypothesis will require structural characterization of the full-length protein in complex with its target RNA. However, given the challenges of obtaining high amounts of stable full-length Gemin5 samples, investigations of this complex will likely need to be done in the context of a Gemin5-dependent RNP.

The biological implications of IDR proteins in normal physiology are still poorly understood. In combination with the plasticity of RNA molecules, the flexibility of these unfolded regions may increase the possibilities to form macromolecular assemblies, remodelling protein networks impacting on RNA-driven processes. Given the abundance of IDRs in RNA-binding proteins, it is paramount to understand the hidden RNA recognition code involving them and model systems like Gemin5 RBS1 might be useful tools to advance in this way. The unusual composition of the RNA-binding motif identified in the RBS1 domain of Gemin5 would also allow the discovery of similar motifs on poorly characterized IDR proteins, likely expanding the repertoire of non-conventional RBPs.

Methods

DNA cloning, protein expression and purification

The constructs encoding the HIS-RBS1₁₄₁₂ domain of Gemin5 (pETM-11-RBS1), the N-terminal deletion HIS-RBS1₁₄₁₂Δ8 protein, the RNAs corresponding to d5 of FMDV IRES and its single-stranded region were previously described [11,12,15]. Constructs expressing HIS-RBS1₁₃₆₁ and the substitution mutants HIS-RBS1₁₃₆₁P1297E and HIS-RBS1₁₃₆₁SS-DD were generated by Quikchange site directed mutagenesis (Agilent Technologies) on pETM-11-RBS1 according to manufacture instructions using the oligonucleotides described in Table 1. The construct HIS-RBS1₁₃₆₁ contains the segment 1297–1361 preceded by a methionine (ATG starting codon) 6xHIS tag and TEV cleavage site.

The RBS1₁₃₆₁-HIS construct was prepared in various steps. First, the sequence encoding Gemin5₁₂₈₇₋₁₅₀₈ was amplified by PCR from pcDNA3Xpress-G5 [10] using specific oligonucleotides (Table 1) and inserted into the BamHI and XhoI of pET28-txAHTEV, a vector previously prepared in our lab [38] resulting on the pET28-txAHTEV-Gemin5₁₂₈₇₋₁₅₀₈ construct. Next, the segment 1362–1508 was removed by QuikChange (Agilent) site-directed mutagenesis using DNA oligos designed to fuse the C-terminal 6xHis (already present in the vector backbone) in-frame with the RBS1 protein, yielding the construct pET28-txAHTEV-Gemin5₁₂₈₇₋₁₅₆₁-6xHIS. Then, to remove the N-terminal tag, we made use of the single NdeI site at the first codon of the ORF and engineered a second NdeI site at the codon 1293 of the Gemin5 sequence by Quikchange (Agilent). Digestion with NdeI followed by religation with T4 ligase (Takara) rendered the final construct RBS1₁₃₆₁-HIS. The construct contains the segment 1294–1361 preceded by a methionine (ATG starting codon) and followed by a 6xHIS tag. The substitution mutants on the RBS1₁₃₆₁-HIS construct were obtained by Quikchange site directed mutagenesis kit with the corresponding oligonucleotides. Oligonucleotides (Table 1) were purchased from Sigma and Macrogen, and all the plasmids were confirmed by DNA sequencing (Macrogen or Stab-vida).

In all the cases, proteins were expressed in LB media or KMOPS minimal media [39] using ¹⁵N ammonium chloride and/or ¹³C labelled glucose as sole nitrogen and carbon sources. Plasmids were transformed in BL21(DE3) *E. coli*

cells and expressions were induced at 37°C for 2–4 hours upon reaching 0.6 OD with Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested and resuspended in binding buffer (20 mM sodium phosphate (pH 7.4), 500 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol) that contains protease inhibitors (Roche) and processed immediately by sonication and centrifugation at 16,000 g 30 min at 4°C. The supernatant, containing the protein, was loaded on a Ni-NTA column (GE Healthcare) previously equilibrated with binding buffer, washed with 5 column volumes of binding buffer and eluted in a similar buffer but with 500 mM imidazole. Protein was then dialysed with an 8 kDa cut-off membrane against the final buffer depending on its later use. All procedures were performed at 4°. Proteins were quantified by UV using extinction coefficients at 280 nm and/or 205 nm (calculated according to [40]).

NMR

Experiments were acquired at 25°C on Bruker AV800 MHz spectrometer with at TCI cryoprobe. Gemin5 samples (100 μM to 200 μM) were prepared in NMR buffer (25 mM Potassium Phosphate pH 6.5, 100 mM NaCl, 1 mM EDTA, 0.1 mM DTT and 10% D₂O). Fresh soluble proteins HIS-RBS1₁₃₆₁ and RBS1₁₃₆₁-HIS (1.8 mg/ml and 1.5 mg/ml, respectively) were used. Backbone assignments were obtained with triple resonance 3D experiments (HNCA, HNCOC, CBCA (CO) NH, HNCACB) [41]. The analysis of the ¹³C conformational shifts to obtain the residual secondary structure contain was made with the program δ2d [26]. RNA titrations were monitored on the protein ¹H-¹⁵N HSQC and the chemical shift perturbations (CPS) were calculated according to the equation: $\Delta\delta^{av} = (1/2 \cdot ((\Delta\delta_H)^2 + (0.2 \cdot \Delta\delta_N)^2))^{1/2}$. All the NMR spectra were processed with nmrPipe [42] and/or Bruker Topspin 4.1 (Bruker) and analysed with the ccpnmr Analysis software [43].

RNA electrophoretic mobility shift assay

For RNA-binding studies, proteins were dialysed against phosphate buffer pH 6.8, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, and stored at –20°C in 50% glycerol. The purified proteins were analysed by SDS-PAGE and quantified using extinction coefficients at 205 nm rendering the following concentrations: HIS-RBS1₁₄₁₂ 53.8 μM, HIS-RBS1₁₄₁₂Δ8 59 μM, HIS-RBS1₁₃₆₁WT 169.6 μM, HIS-RBS1₁₃₆₁P1297E 149.6 μM, HIS-RBS1₁₃₆₁SS-DD 165.5 μM, RBS1₁₃₆₁-HIS WT 169.5 μM, RBS1₁₃₆₁-HIS R1294K 535.5 μM, RBS1₁₃₆₁-HIS P1297G 105.6 μM, RBS1₁₃₆₁-HIS SS-AA 116.8 μM, RBS1₁₃₆₁-HIS SS-TT 260.7 μM, RBS1₁₃₆₁-HIS W1302A 484.4 μM, RBS1₁₃₆₁-HIS R1304K 170.6 μM, RBS1₁₃₆₁-HIS H1307A 59.7 μM, RBS1₁₃₆₁-HIS R1308K 205.7 μM.

In vitro transcription of FMDV IRES d5 and its single stranded region d5ss were prepared as described [44]. Briefly, RNA probes were uniformly labelled using α³²P-CTP (500 Ci/mmol), T7 RNA polymerase (10 U), and linearized plasmid (1 μg). The newly synthesized RNA was purified

through MicroSpin G-25 columns (GE Healthcare) and ethanol precipitated and resuspended in TE (10 mM Tris-HCl pH 8, 1 mM EDTA) to a final concentration of 0.04 pmol/ μ l. RNA integrity was examined in 6% acrylamide 7 M urea denaturing gel electrophoresis. RNA U(5) (5'-UUUUU-3') was labelled at the 5' using T4-polynucleotide kinase and γ -ATP as described [15].

RNA-binding reactions were carried out as described [15] in 10 μ l of RNA-binding buffer (40 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.1% (w/v) β ME) for 15 min at room temperature using serial increased concentration of protein with a constant concentration of 32 P-labelled RNA (\square 2 nM). Electrophoresis was performed in non-denaturing 6.0% (29:1) polyacrylamide gels at 4°C, run in TBE buffer (90 mM Tris-HCl pH 8.4, 64.6 mM boric acid, 2.5 mM EDTA) at 100 V. The 32 P-labelled RNA and retarded complexes were detected by autoradiography of dried gels. The percentage of the retarded complex was calculated relative to the free probe, run in parallel. GraphPad Prism Software (version 6.01) was used to plot the binding curves and estimate the values for dissociation constants (K_D) by nonlinear regression using the one-site specific binding equation.

Acknowledgments

NMR experiments were performed in the “Manuel Rico” NMR laboratory (LMR) of the Spanish National Research Council (CSIC), a node of the Spanish Large-Scale National Facility (ICTS R-LRB). We thank Jorge Ramajo for technical assistance.

Disclosure statement

No potential conflict of interest were disclosed.

Funding

This work was supported by grants from MINECO (CTQ2018-84371) to JMP-C, BFU2017-84492-R (to EMS), and B2017/BMD-3770 cofinanced by Autonomous Community of Madrid and FEDER funds to EMS and JMP-C.

Author contributions

AE-B, RF-V, SC, and JMP-C made all of the clones, mutants and recombinant proteins. JMP-C obtained and analysed NMR data. AE-B and RF-V carried out all RNA-binding assays, and analysed the data. The project was conceived by JMP-C and EMS with contributions from all authors.

ORCID

Azman Embarc-Buh  <http://orcid.org/0000-0002-1980-942X>
 Rosario Francisco-Velilla  <http://orcid.org/0000-0002-4328-8732>
 Sergio Camero  <http://orcid.org/0000-0001-5283-6510>
 José Manuel Pérez-Cañadillas  <http://orcid.org/0000-0002-8266-5502>
 Encarnación Martínez-Salas  <http://orcid.org/0000-0002-8432-5587>

References

- Francisco-Velilla R, Embarc-Buh A, Martínez-Salas E. Impact of RNA-protein interaction modes on translation control: the versatile multidomain protein Gemin5. *Bioessays*. 2019;41:e1800241.
- Battle DJ, Lau CK, Wan L, et al. The Gemin5 protein of the SMN complex identifies snRNAs. *Mol Cell*. 2006;23:273–279.
- García-Moreno M, Noerenberg M, Ni S, et al. System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol Cell*. 2019;74(1):196–211.e11.
- Pineiro D, Fernández N, Ramajo J, et al. Gemin5 promotes IRES interaction and translation control through its C-terminal region. *Nucleic Acids Res*. 2013;41:1017–1028.
- Pei W, Xu L, Chen Z, et al. A subset of SMN complex members have a specific role in tissue regeneration via ERBB pathway-mediated proliferation. *NPJ Regen Med*. 2020;5:6.
- Martínez-Salas E, Embarc-Buh A, Francisco-Velilla R. Emerging roles of Gemin5: from snRNPs assembly to translation control. *Int J Mol Sci*. 2020;21(11):3868.
- Yong J, Kasim M, Bachorik JL, et al. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell*. 2010;38:551–562.
- Moreno-Morcillo M, Francisco-Velilla R, Embarc-Buh A, et al. Structural basis for the dimerization of Gemin5 and its role in protein recruitment and translation control. *Nucleic Acids Res*. 2020;48:788–801.
- Fernández-Chamorro J, Pineiro D, Gordon JM, et al. Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res*. 2014;42:5742–5754.
- Francisco-Velilla R, Fernández-Chamorro J, Ramajo J, et al. The RNA-binding protein Gemin5 binds directly to the ribosome and regulates global translation. *Nucleic Acids Res*. 2016;44:8335–8351.
- Francisco-Velilla R, Fernández-Chamorro J, Dotu I, et al. The landscape of the non-canonical RNA-binding site of Gemin5 unveils a feedback loop counteracting the negative effect on translation. *Nucleic Acids Res*. 2018;46:7339–7353.
- Pineiro D, Ramajo J, Bradrick SS, et al. Gemin5 proteolysis reveals a novel motif to identify L protease targets. *Nucleic Acids Res*. 2012;40:4942–4953.
- Workman E, Kalda C, Patel A, et al. Gemin5 binds to the survival motor neuron mRNA to regulate SMN expression. *J Biol Chem*. 2015;5:528–544.
- Mallik S, Kundu S. Coevolutionary constraints in the sequence-space of macromolecular complexes reflect their self-assembly pathways. *Proteins*. 2017;85:1183–1189.
- Francisco-Velilla R, Embarc-Buh A, Rangel-Guerrero S, et al. RNA-protein coevolution study of Gemin5 uncovers the role of the PXSS motif of RBS1 domain for RNA binding. *RNA Biol*. 2020;17:1331–1341.
- Pineiro D, Fernández-Chamorro J, Francisco-Velilla R, et al. Gemin5: a multitasking RNA-binding protein involved in translation control. *Biomolecules*. 2015;5:528–544.
- Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007;8:479–490.
- Trendel J, Schwarzl T, Horos R, et al. The human RNA-binding proteome and its dynamics during translational arrest. *Cell*. 2019;176:391–403.e319.
- Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet*. 2014;15:829–845.
- Jarvelin AI, Noerenberg M, Davis I, et al. The new (dis)order in RNA regulation. *Cell Commun Signal*. 2016;14:9.
- Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol*. 2015;16(1):18–29.
- Wang J, Choi JM, Holehouse AS, et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell*. 2018;174:688–699.e616.
- Wilson KA, Holland DJ, Wetmore DS. Topology of RNA-protein nucleobase-amino acid π - π interactions and comparison to analogous DNA-protein π - π contacts. *RNA*. 2016;22:696–708.

- [24] Fernandez-Chamorro J, Lozano G, Garcia-Martin JA, et al. Designing synthetic RNAs to determine the relevance of structural motifs in picornavirus IRES elements. *Sci Rep.* **2016**;6:24243.
- [25] Lozano G, Martinez-Salas E. Structural insights into viral IRES-dependent translation mechanisms. *Curr Opin Virol.* **2015**;12:113–120.
- [26] Camilloni C, De Simone A, Vranken WF, et al. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry.* **2012**;51(11):2224–2231.
- [27] Bermel W, Bertini I, Felli IC, et al. Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *J Magn Reson.* **2006**;178(1):56–64.
- [28] Meister G, Eggert C, Fischer U. SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol.* **2002**;12(10):472–478.
- [29] Matera AG, Raimer AC, Schmidt CA, et al. Composition of the Survival Motor Neuron (SMN) complex in *Drosophila melanogaster*. *G3 (Bethesda).* **2019**;9:491–503.
- [30] Jin W, Wang Y, Liu CP, et al. Structural basis for snRNA recognition by the double-WD40 repeat domain of Gemin5. *Genes Dev.* **2016**;30:2391–2403.
- [31] Tang X, Bharath SR, Piao S, et al. Structural basis for specific recognition of pre-snRNA by Gemin5. *Cell Res.* **2016**;26:1353–1356.
- [32] Xu C, Ishikawa H, Izumikawa K, et al. Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. *Genes Dev.* **2016**;30:2376–2390.
- [33] Chong PA, Vernon RM, Forman-Kay JD. RGG/RG motif regions in RNA binding and phase separation. *J Mol Biol.* **2018**;430(23):4650–4665.
- [34] Rozas I, Alkorta I, Elguero J. Modelling protein-RNA interactions: an electron density study of the guanidinium and formate complexes with RNA bases. *Org Biomol Chem.* **2005**;3(2):366–371.
- [35] Kagra D, Prabhakar PS, Sharma KD, et al. Structural patterns and stabilities of hydrogen-bonded pairs involving ribonucleotide bases and arginine, glutamic acid, or glutamine residues of proteins from quantum mechanical calculations. *ACS Omega.* **2020**;5(7):3612–3623.
- [36] Blanco F, Kelly B, Sánchez-Sanz G, et al. Non-covalent interactions: complexes of guanidinium with DNA and RNA nucleobases. *J Phys Chem B.* **2013**;117(39):11608–11616.
- [37] Pacheco A, Lopez de Quinto S, Ramajo J, et al. A novel role for Gemin5 in mRNA translation. *Nucleic Acids Res.* **2009**;37:582–590.
- [38] Santiveri CM, Mirassou Y, Rico-Lastres P, et al. Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. *PLoS One.* **2011**;6(9):e24481.
- [39] Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *J Bacteriol.* **1974**;119(3):736–747.
- [40] Anthis NJ, Clore GM. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **2013**;22(6):851–858.
- [41] Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc.* **1999**;34:93–158.
- [42] Delaglio F, Grzesiek S, Vuister GW, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* **1995**;6(3):277–293.
- [43] Vranken WF, Boucher W, Stevens TJ, et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* **2005**;59(4):687–696.
- [44] Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, et al. RNA-protein interaction methods to study viral IRES elements. *Methods.* **2015**;91:3–12.