



OPEN

A combined NMR and EPR investigation on the effect of the disordered RGG regions in the structure and the activity of the RRM domain of FUS

A. Bonucci^{1,2}, M. G. Murrall¹, L. Banci^{1,2}✉ & R. Pierattelli^{1,2}✉

Structural disorder represents a key feature in the mechanism of action of RNA-binding proteins (RBPs). Recent insights revealed that intrinsically disordered regions (IDRs) linking globular domains modulate their capability to interact with various sequences of RNA, but also regulate aggregation processes, stress-granules formation, and binding to other proteins. The FET protein family, which includes FUS (Fused in Sarcoma), EWG (Ewing Sarcoma) and TAF15 (TATA binding association factor 15) proteins, is a group of RBPs containing three different long IDRs characterized by the presence of RGG motifs. In this study, we present the characterization of a fragment of FUS comprising two RGG regions flanking the RNA Recognition Motif (RRM) alone and in the presence of a stem-loop RNA. From a combination of EPR and NMR spectroscopies, we established that the two RGG regions transiently interact with the RRM itself. These interactions may play a role in the recognition of stem-loop RNA, without a disorder-to-order transition but retaining high dynamics.

The “classic” RNA-binding proteins (RBPs) are a group of proteins playing central roles in the formation of ribonucleoprotein complexes as they participate in various gene expression processes. The capability of RBPs to bind RNA has been generally ascribed to the globular domains that possess a defined secondary structure, such as the RNA recognition motif (RRM), the K homology domain (KH), the zinc fingers (ZnF) and the DEAD box helicase motifs^{1–4}.

Some studies revealed that 3–11% of the human proteome is involved in RNA binding, but only a small fraction of proteins could be classified as typical globular RBPs⁵. Indeed, recent reports revealed that intrinsically disordered regions (IDRs) are largely distributed in various RBPs sequences and contribute to RNA binding, cooperate with globular domains, and should play relevant roles in RNA interaction^{6–9}.

The second most common class of intrinsically disordered regions in RBPs is represented by RGG-boxes (or GAR) that contain repeats of arginine and glycine residues. The absence of a folded conformation of the RGG motifs is due to the abundance of hydrophilic and charged amino acids and to the lack of bulky hydrophobic groups that promote the formation of stable three-dimensional structures. RGG repeats are usually associated with other repetitive sequences, including polyalanine, polyaspartate, polyglutamate, polyglutamine, and polyglycine repeats, as well as other diglycine-containing motifs such as FGG, PGG, SGG and YGG. The positively charged guanidinium group of arginine can form different types of interactions with different elements of nucleic acids, including cation- π interactions, π - π stacking, hydrogen bonding, charge-charge/dipole and van der Waals interactions. On the other hand, glycine is the smallest amino acid with a single hydrogen atom as side chain and therefore it represents the key-factor for the conformational flexibility of RGG domains. Moreover, post-translational modifications (PTM), such as arginine methylation, may represent another important factor for modulating the functions of the RGG boxes. PTMs could alter protein-protein and protein-RNA binding capability and could regulate the interaction of the protein with a group of RNA sequences and structures^{6,10–12}.

Fused in Sarcoma (FUS) protein is a nuclear protein implicated in the transcription, splicing and transport of mRNA. The scientific interest for this protein is due to its direct implication in neurodegenerative disorders,

¹CERM - Magnetic Resonance Center, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. ²Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy. ✉email: banci@cerm.unifi.it; roberta.pierattelli@unifi.it

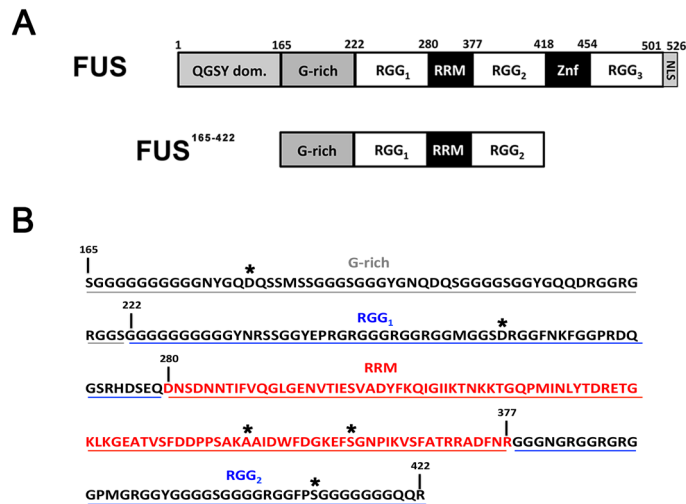


Figure 1. (A) Organization of full-length FUS and FUS^{165–422}; (B) primary sequence of FUS^{165–422}. The residues selected for replacement with cysteine and MTSL-labelling are marked with asterisk.

specifically in Amyotrophic Lateral Sclerosis (ALS)^{13,14}. FUS is a multi-domain protein formed by a QGSY-rich region at the N-terminus, a globular RNA-recognition motif (RRM) with the classical $\beta\alpha\beta\beta\alpha\beta$ fold, a zinc finger domain (ZnF), three different RGG boxes (RGG₁, RGG₂ and RGG₃), a G-rich region and a C-terminal nuclear localization signal (NLS) (Fig. 1). RRM and ZnF represent the principal sites for RNA interaction, even if recent studies suggested a relevant role exerted by the RGG motifs in the RNA interaction and other cellular processes, such as phase-separation and stress granules formation^{15–18}. In addition, the RGG boxes promote the affinity of folded domains for RNA possibly without taking a defined conformation during nucleotides binding. These disordered regions confer also the “degenerate-specificity” character in RNA interaction that is the capability of FUS to bind different nucleotides sequences and structures of RNA¹⁸. Moreover, the RGG₃ and NLS regions are directly involved in the interaction with the nuclear import receptor Transportin/Karyopherin- β 2, which translocates FUS into the nucleoplasm^{16–21}.

In the present study we focused on the core RRM structured domain of FUS and on its two flanking IDRs both containing RGG motifs, to further characterize these parts of the protein and assess whether the interplay between the RGG_{1,2} and the RRM could be of importance in determining the function of the protein, particularly in the contest of RNA binding. Through a combined approach based on NMR and EPR spectroscopies, we provided insights concerning the conformation and dynamics of these domains, but also some hints on the possible roles exerted by RGGs regions in the RRM-RNA binding.

Results

Characterization of FUS^{165–422}. Various studies highlighted the importance of the RGG domains in the physiological RNA-binding mechanism of the RRM of FUS^{17,18}; however, biophysical investigations on a FUS fragment comprising the RRM and the two extended flanking RGG disordered regions (FUS^{165–422}, Fig. 1) are very scarce. For this reason, we designed a plasmid encoding this fragment into pDEST-HisMBP vector and we expressed and purified the system using recombinant methodologies (see “Methods”).

The resulting 258-residues long construct, with a molecular weight of 25 kDa, is composed of about 20% charged residues, 10% polar residues, and 37% glycine residues, with many repeats beside the canonical GG and RGG motifs (i.e. SS, QQ, etc.). This unique content of glycine in FUS^{165–422} primary structure represents the key reason of the intrinsically disordered behavior of the RGG domains. Moreover, the high pI values of both unfolded domains (10.2 and 12.4 for RGG₁ and RGG₂, respectively) underline that these regions of FUS are positively charged under physiological conditions (pH 7.0–7.4), thus being capable of interacting with negative charges of nucleotides or others proteins. The main type of interactions between RBPs and RNA are π - π contacts, exerted by nucleobases and aromatic amino acids. Trp, Tyr and Phe residues in FUS^{165–422} are present in the RRM domain only, consistent with this region being the principal site for RNA interaction²².

CD spectroscopy was used to confirm the secondary structure content present in FUS^{165–422}. The far-UV CD spectrum of the protein in MES buffer solution at pH 6.5 (Fig. 2A) presents two peaks at 207 nm and 225 nm, a hallmark of the presence of helical conformation in the polypeptide, but the overall spectrum indicated the net presence of random coil structures as well. The percentage of secondary structural elements, estimated with the program Dichroweb²³, accounts for 25% \pm 3.1% of folded conformations (16% \pm 3.1% β -sheet and 9% \pm 3.1% α -helix respectively) and 75% \pm 3.1% of disordered conformations. These data are in agreement with the previous work by Song and co-workers²⁴ and also with the estimated content of secondary structural elements based on the disorder predictor GlobPlot2.3²⁵ and the available structural models (see Ref.²⁶ and PDB 2LA6), which account for 11% β -sheet and 7% α -helix and 72% disordered polypeptide.

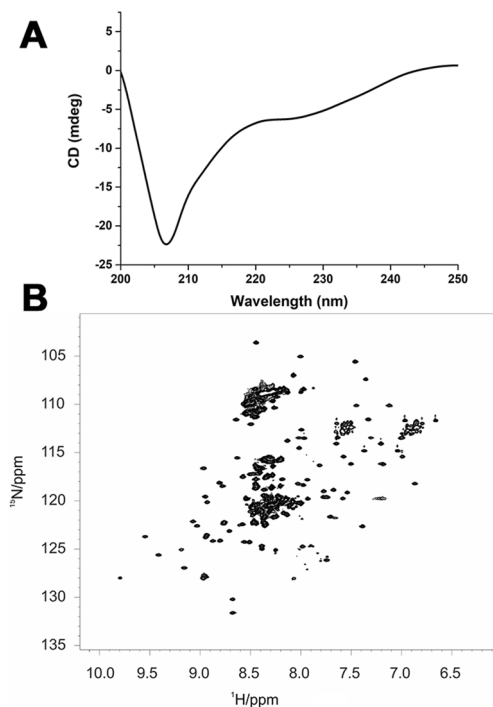


Figure 2. (A) Far-UV CD spectrum (200–250 nm) of FUS^{165–422}; (B) 2D ¹H-¹⁵N HSQC spectra of FUS^{165–422}. The NMR spectrum was acquired at 298 K with a 22.3 T Bruker Avance III spectrometer equipped with a TCI probehead.

The full-length FUS protein has never been characterized in detail by NMR spectroscopy in solution since it is prone to aggregation, but the structure in solution of the RRM domain (residues 278–385, PDB 2LCW, and residues 282–370, PDB 2LA6) was determined²⁶. We thus took advantage of the available sequence specific assignment (BMRB ID 17508) to assign the RRM signals in our experimental conditions by using a sub-set of NMR experiments (see “Methods”). The ¹H-¹⁵N correlation map (Fig. 2B) is consistent with the spectra available in the literature for the well-dispersed signals of the RRM domain^{24,26,27}. The small differences in chemical shift of the signals of our construct with respect to those of the isolated domain are due to differences in buffer composition, as the deposited chemical shift values can be matched by changing the sample conditions. Thus, interactions of the RRM domain with its flanking regions, if present, should be transient and not affecting the NMR signals. The signals of the disordered regions are clustered at the centre of the ¹H-chemical shift region, with particularly high overlap in the so-called glycine region, as expected.

The dynamics of the isolated RRM domain has been already investigated by NMR spectroscopy²⁴. We acquired and analyzed ¹⁵N relaxation properties of the RRM domain to investigate the conformational dynamics of the folded domain in our longer construct. The ¹⁵N longitudinal relaxation rate (¹⁵N R₁) and the ¹⁵N transverse relaxation rate (¹⁵N R₂) confirm the general features of the domain, without major variation of the ns-ps dynamics of the folded domain and with little changes consistent with an overall increase in correlation time induced by the presence of the additional residues of the two dynamical flanking regions (see Supplementary Fig. S1 online). The ¹H-¹⁵N NOE experiments confirm the structured nature of the folded RRM domain, which maintains its overall compactness with a slight increase in dynamics in the region 305–320, in the β-hairpin region between α1 and β2, and in the region 325–335, which comprises the terminal part of β2 and extends to the β3 strand.

The two RGG boxes are completely disordered; EPR spectroscopy is therefore a suitable complement to NMR spectroscopy for obtaining information on their dynamics.

Insights on FUS^{165–422} dynamics features through paramagnetic probes. EPR spectroscopy combined with site-directed spin labeling (SDSL) is a powerful method to investigate protein dynamics. From the analysis of the lineshape of the EPR spectrum of a spin label (i.e. nitroxide radicals) selectively grafted on an appropriate residue (i.e. cysteine, tyrosine, etc.) and the determination of its mobility, protein conformation can be characterized²⁸. Sharp EPR signals are due to a fast tumbling rate of the spin label indicating a high degree of protein flexibility (small τ_C value), while a decreased mobility determines broader lines and higher τ_C values. This methodology has been extensively applied to characterize intrinsically disordered proteins and to follow possible disorder-to-order transitions^{29,30}.

To monitor the structural behaviour of FUS, including the G-rich, the RGG₁ and the RGG₂ regions, as well as the RNA-binding site, five different mutants of FUS^{165–422} to insert a cysteine residue in the protein sequence were designed (see “Methods”). The choice of the positions was driven by the occurrence in the primary sequence of many glycine residues as well as by the presence of amino acids repeats (GRS) which limited the accessible

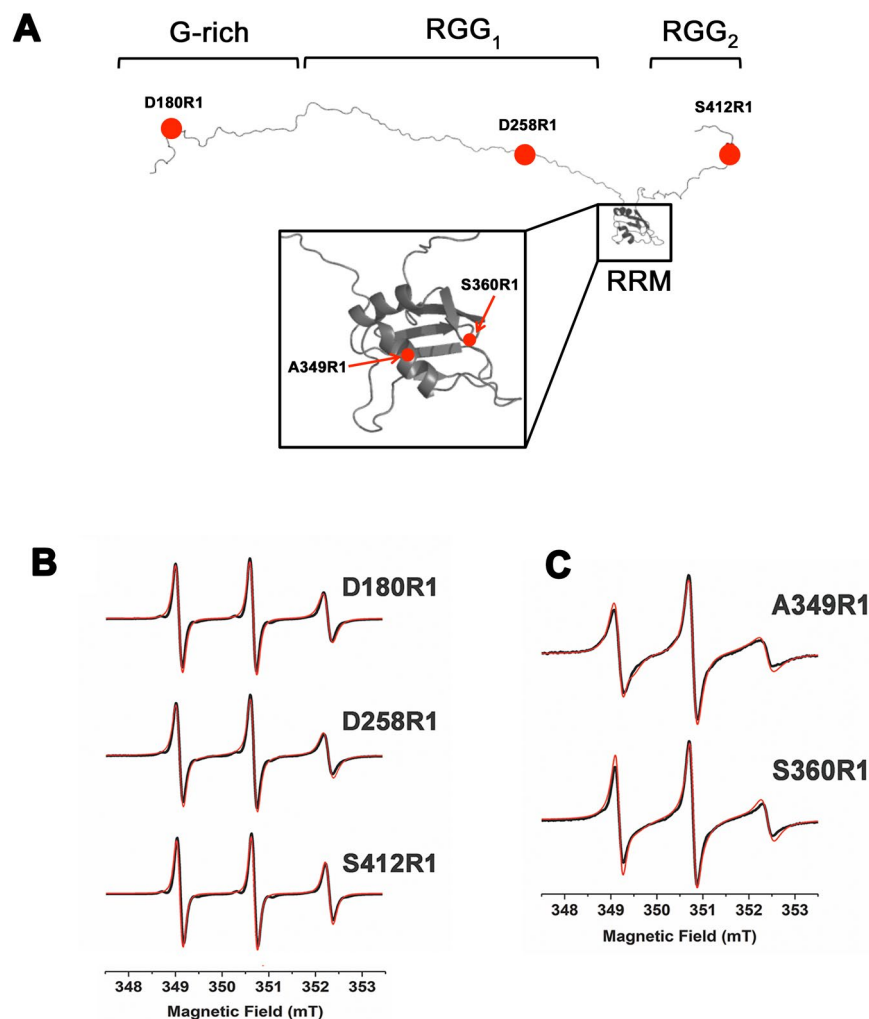


Figure 3. (A) Labelling positions selected for G-rich, RGG_{1,2} and RRM domains of FUS^{165–422}. (B,C) Experimental (black line) and simulated (red line) X-band (9.8 GHz) CW-EPR spectra of D180R1, D258R1, S412R1, A349R1 and S360R1 spin labelled mutants of FUS^{165–422}.

mutation positions. In the end, we were able to mutate to cysteine the amino acids D180, D258, and S412. The two positions in the RRM domain, A349 and S360, were selected in a region opposite to the site for nucleotides interaction to avoid interference with the RNA-binding activity of FUS^{165–422} (Figs. 1B and 3A). Each of these mutants was labelled with MTSL (see “Methods”, the resulting mutants are marked R1) and X-band (9.8 GHz) CW-EPR spectra were recorded at room temperature (Fig. 3B,C).

The D180R1 and S412R1 variants, bearing the spin label in the G-rich and in the RGG₂ region respectively, show a narrow EPR signal typical of a mobile stretch (Fig. 3B). The high mobility of these regions is confirmed by simulations of the spectra, which reveal the presence of a single spectral component with short τ_C value (0.22 ± 0.08 ns and 0.16 ± 0.08 ns respectively, Table 1). These results indicate that these two regions of FUS (G-rich and RGG₂) have a large degree of fluctuations and lack a defined conformation. The CW-EPR spectrum of D258R1 (Fig. 3B) is instead composed by two different components: a sharp signal associated to a fast tumbling rate of the grafted spin probe (component 1 = $74 \pm 3\%$, $\tau_C = 0.22 \pm 0.08$ ns) and a minor component with a higher τ_C value (component 2 = $26 \pm 3\%$, $\tau_C = 1.60 \pm 0.08$ ns) (Table 1), revealing the occurrence of different dynamic states in solution. The broad component can be ascribed to restrictions induced by the proximity of the RRM domain to the tag position and to possible inter-domains interactions between these regions, considering the absence of significant signals’ linewidth broadening in the NMR spectrum of the mutant that can be associated to a conformation change due to the Asp to Cys substitution or to the presence of the MTSL tag (see PRE experiments below).

CW-EPR spectra were also recorded for the A349R1 and S360R1 mutants bearing the spin label on the $\alpha 2$ helix and on the flexible loop connecting $\alpha 2$ and $\beta 4$ of the RRM domain, respectively (Fig. 3A). The EPR spectrum of A349R1 (Fig. 3C) is composed by three different components, indicating the coexistence of various dynamic states for the system (Table 2). In particular, two components indicate the presence of conformations with a high degree of flexibility (component 1 = $15\% \pm 4\%$, with $\tau_C = 0.40 \pm 0.09$ ns and component 2 = $50 \pm 4\%$, with $\tau_C = 0.72 \pm 0.09$ ns) together with a conformation with reduced dynamics (component 3 = $35 \pm 4\%$ with

MTSL-labelled mutants	Component 1		Component 2	
	%	τ_c (ns)	%	τ_c (ns)
D180R1	100	0.22		
D180R1 + RNA	100	0.16		
D258R1	74	0.22	26	1.60
D258R1 + RNA	58	0.27	42	1.92
S412R1	100	0.16		
S412R1 + RNA	100	0.16		

Table 1. Component percentages (%) and correlational times (τ_c) calculated from simulations (SimLabel software, ref.⁴⁴) for each spin labeled FUS^{165–422} mutants in the disordered domains, alone and in presence of hnRNPA2/B1 stem-loop RNA. Standard deviations on components percentages and τ_c were estimated equal to $\pm 3\%$ and ± 0.08 ns respectively.

MTSL-labelled mutants	Component 1		Component 2		Component 3	
	%	τ_c (ns)	%	τ_c (ns)	%	τ_c (ns)
A349R1	15	0.40	50	0.72	35	1.40
A349R1 + RNA	65	0.30			35	1.40
S360R1	72	0.48			28	2.17
S360R1 + RNA	27	0.57	35	0.78	38	2.13

Table 2. Component percentages (%) and correlational times (τ_c) calculated from simulations (SimLabel software, ref.⁴⁴) for each spin labeled FUS^{165–422} mutants in the RRM domain, alone and in presence of hnRNPA2/B1 stem-loop RNA. Standard deviations on components percentages and τ_c were estimated equal to $\pm 4\%$ and ± 0.09 ns respectively.

$\tau_c = 1.40 \pm 0.09$ ns). Similarly, the EPR analysis of the S360R1 mutant (Fig. 3C) indicate that also this part of the RRM motif experiences multi-dynamic states. Interestingly, the broad EPR signal component (component 3 of S360R1) shows a τ_c value (2.17 ± 0.09 ns) higher than that determined for the same component in A349R1 ($\tau_c = 1.40 \pm 0.09$ ns). These differences in correlational time values between the two labelling positions could be caused by the spin label position. In fact, the less restricted environment for the label grafted on the α_2 helix should confer a larger tumbling rate (smaller τ_c) to MTSL compared to the S360R1 mutant.

In order to obtain further information on the influence of the RGG boxes on the RRM domain, we exploited Paramagnetic Resonance Enhancement (PRE) NMR experiments to determine possible long-range contacts between different protein regions^{31,32}. Using the same five labelled mutants of FUS^{165–422} (D180R1, D258R1, A349R1, S360R1 and S412R1, Fig. 3A) previously characterized by EPR, we acquired pairs of ¹H-¹⁵N-HSQC spectra of the MTSL-labeled and of the unlabeled mutants monitoring the effect of the unpaired electron of the spin label on signals broadening (Fig. 4).

PREs data obtained on D180R1 (Fig. 4A,F) confirm that this region of the G-rich domain is not in direct contact with RRM. The analysis of the PRE observed on the D258R1 (Fig. 4B,G) mutant shows some minor effects on the signals of residues in the 306–311, 342–346 and 357–360 regions, suggesting the occurrence of some transient interactions between RGG₁ and the RRM domain. On the RGG₂ side, the paramagnetic effects produced by the label in position S412R1 (Fig. 4E,J) are comparable to the ones of D258R1 mutant and affect several positions on the protein surface confirming the high flexibility of this harm.

The same PRE NMR experiments were acquired for the A349R1 and S360R1 mutants. The presence of a spin label on the α_2 helix bearing A349R1 cause severe broadening of the signals of a stretch encompassing residues 308–312 (Fig. 4C,H), which is a loop juxtaposed to the α_2 helix, and a slight effect on some signals of residues located in the strand immediately preceding the helix and towards the end of the RRM domain. The effect of the spin label on S360R1 (Fig. 4D,I) results in the selective broadening of the signals of two regions: the region around the paramagnetic tag (residues 357–364) and a region inside the RRM domain encompassing residues 291–308. The PRE pattern reflects the spatial proximity of the label to the hairpin between β_1 and α_2 and the helix itself; the extension of this segment underlines the high dynamics of the region.

Interaction of FUS^{165–422} with stem-loop RNA. In order to test the capability of FUS^{165–422} to interact with RNA, we performed experiments in the presence of hnRNPA2/B1 stem-loop. Since the RRM of FUS binds a NYNY single strand (N = C or U, Y = A, C, G and U)²⁷, we performed all the spectroscopic measurements with a RNA sequence containing these specific nucleotide quartets. Previous studies measured the affinity of the full length FUS and various protein segments for different nucleotide sequences, revealing that the RGGs motifs strongly influence the capability of the protein to bind RNAs¹⁸.

Upon addition of hnRNPA2/B1 stem-loop RNA to FUS^{165–422}, at 1:1 molar ratio, no major changes in the structure of the protein occur. The far-UV CD spectrum showed a minor shift of the maximum absorbance band (from 207 to 210 nm) and a weak variation in the spectral lineshape (Fig. 5A), suggesting the absence of

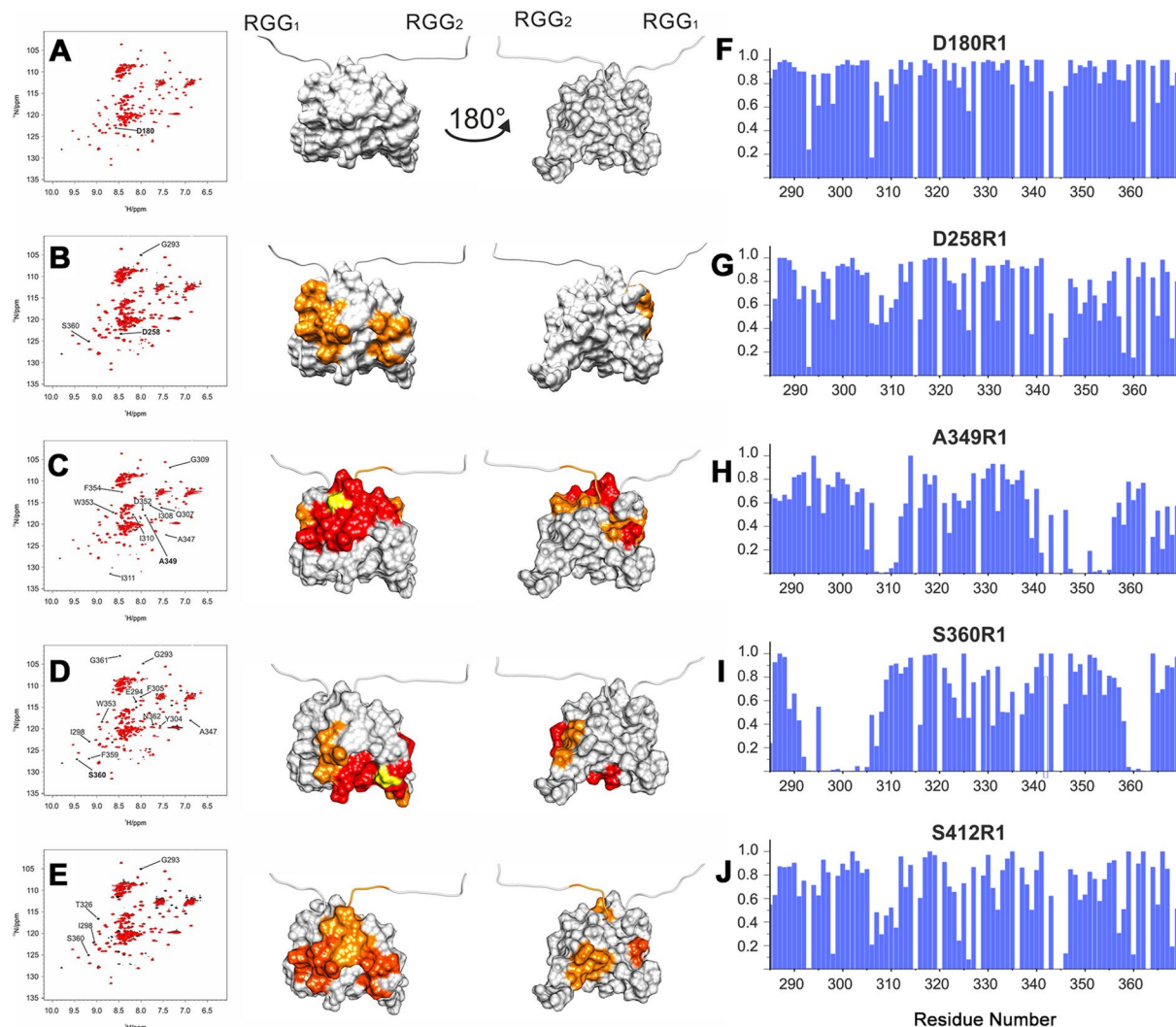


Figure 4. Left panels— ^1H - ^{15}N HSQC NMR spectra of MTSL-labelled (red) and unlabelled (black) FUS^{165–422} mutants (A) D180R1, (B) D258R1, (C) A349R1, (D) S360R1 and (E) S412R1 with the assignment of the relevant cross peaks; Central panels—mapping on the protein surface of the RRM domain (PDB ID 2LA6) of the relevant changes in signal intensity. Residues are coloured from red to orange according to PRE magnitude (F–J); unaffected residues are coloured in grey. The labelled position in the RRM is coloured in yellow. Right panels: Paramagnetic relaxation enhancement values reported as normalized intensity ratio (I/I^0) for FUS^{165–422} mutants (F) D180R1, (G) D258R1, (H) A349R1, (I) S360R1 and (J) S412R1.

a major structural reorganization upon RNA-binding. The far-UV CD spectrum of stem-loop RNA alone is shown in Fig. 5B.

Consistently with the CD spectrum, ^1H - ^{15}N correlation NMR spectrum does not show major variations due to conformational changes occurring upon RNA interaction (Fig. 5C). The comparison with the free-form spectrum shows several chemical shift changes due to RNA binding for signals of residues in the RRM domain. The largest changes occur in the regions around residues 286, 326 and 369, confirming the location of the binding site in the proximity of these amino acids²⁷. The spectral features of the disordered regions of the two RGG boxes were instead unchanged. We cannot rule out a fuzzy interaction that cannot be detected by NMR experiments, but RNA binding does not produce any stable conformational change in the disordered parts of FUS^{165–422}.

The CW-EPR spectra of D180R1 and S412R1 do not show any relevant variations in the spectral lineshape (see Supplementary Fig. S2 online) and the τ_c value (Table 1) in the presence of RNA, indicating the absence of a direct interaction of the G-rich and the RGG₂ regions with nucleotides.

On the other hand, the broadening of the EPR spectrum of the D258R1 labelled mutant (Fig. 6A) and the variation in the ratio of the components constituting the signal suggest an interaction of the RGG₁ region with the nucleotides. However, the increase in τ_c for the broader component is small ($\Delta\tau_c = 0.32$ ns Table 1) suggesting that the variation in RGG₁ dynamics is also small and does not induce a relevant structural effect.

More dramatic effects occurred in the RRM region. For A349R1, the narrowing of the corresponding EPR lines (Fig. 6B) indicates fast motion of the α_2 helix, which surprisingly increases in the presence of stem-loop RNA. The effect on the dynamic of this FUS region induced by nucleotides interaction appears also observing

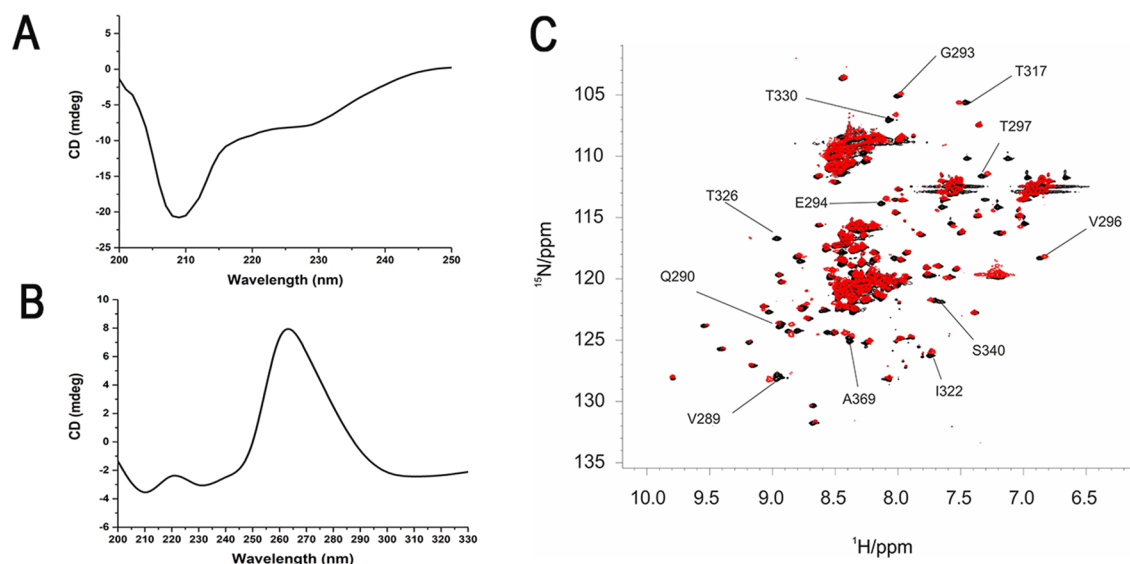


Figure 5. (A) Far-UV CD spectra (200–250 nm) of FUS^{165–422} in the presence of hnRNPA2/B1 stem-loop RNA; (B) Far-UV CD spectra (200–250 nm) of hnRNPA2/B1 stem-loop RNA; (C) 2D ¹H–¹⁵N HSQC spectrum of FUS^{165–422} in the presence of hnRNPA2/B1 stem-loop RNA (red) superimposed with the spectrum of the free protein in solution (black) with the assignment of selected cross peaks.

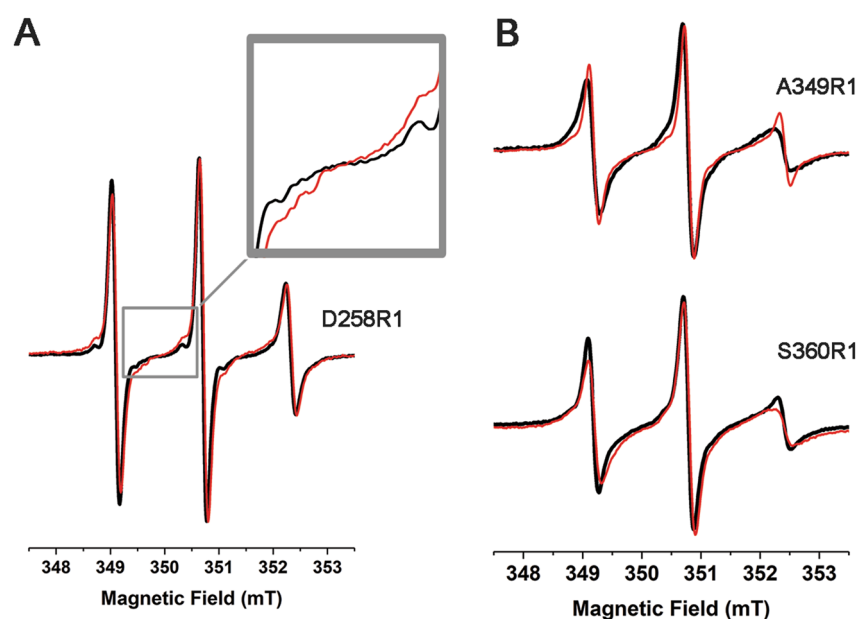


Figure 6. X-band (9.8 GHz) CW-EPR spectra without (black line) and in the presence of 1:1 hnRNPA2/B1 stem-loop RNA (red line) recorded spin labelled FUS^{165–422} mutants (A) D258R1 and (B) A349R1 and S360R1.

the increase of the sharpest component percentage (from $15 \pm 4\%$ in solution to $65 \pm 4\%$, with stem-loop RNA, Table 2). Since the tumbling of the RRM domain is influenced by its transient interactions with the RGG boxes, as revealed by CW-EPR and NMR experiments, the enhancement in dynamics suggests a reduction in these interactions upon RNA binding.

In the presence of RNA the EPR spectrum of S360R1 (Fig. 6B) results broader than the one recorded in its absence. The simulation suggests significant differences in the spectral parameters mainly for component 1, but also the appearance of a new sharp component (component 2, see Table 2).

To get insights on the structural variations of FUS^{165–422} upon RNA interaction, Q-band DEER experiments were performed. This pulse-EPR technique permits to measure the distance between multiple spin probes attached to a biomolecule in various experimental conditions^{28,30,33}. Two different FUS^{165–422} mutants bearing two cysteine residues were produced and labeled with MTSL (D258R1/A349R1 and A349R1/S412R1). On each of these mutants the 4-pulse Q-band DEER traces were recorded in the absence and in the presence of stem-loop

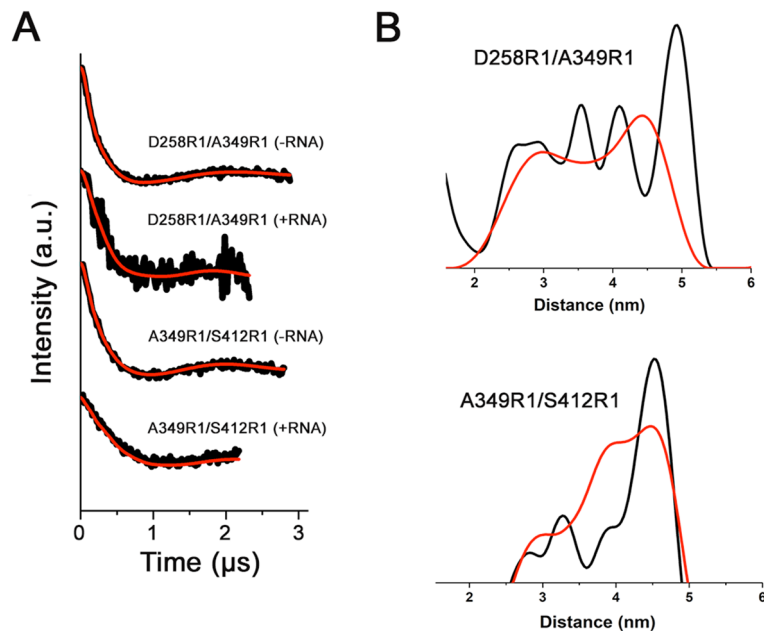


Figure 7. (A) Background corrected Q-band (9.8 GHz) DEER traces (black line) and corresponding fitted curve (red line) for D258R1/A349R1 and A349R1/S412R1 mutants of FUS^{165–422} in without and in the presence of hnRNP A2/B1 stem-loop RNA; (B) corresponding distance distribution without (black line) and in the presence of stem-loop RNA (red line) recorded for D258R1/A349R1 (top panel) and A349R1/S412R1 (bottom panel) mutants of FUS^{165–422}.

RNA to understand the spatial arrangement of the RGG boxes compared to the RRM region and to identify their role in the RNA interaction process.

Background corrected DEER time traces, recorded for both doubly labelled mutants with and without RNA, are reported in Fig. 7A. The inter-spin distances for D258R1/A349R1 and A349R1/S412R1 mutants in the absence of RNA (Fig. 7B black lines, top and bottom panel respectively) showed a similar profile for both samples with peak maxima at 5.0 nm and 4.7 nm Å (± 0.4 nm), respectively. The broad profiles of distributions is related to the large mobility of the RGG regions in solution due to their intrinsically disordered behaviour.

Upon RNA interaction, changes on spin–spin distances distributions are mainly observed for the D258R1/A349R1 sample (Fig. 7B red line, top panel), even if the DEER traces present a lower signal-to-noise ratio than that recorded in absence of nucleotides. In this case, a shift to lower distances (from 5.0 to 4.3 nm) of the maximum peak occurred, highlighting a weak rearrangement of RGG₁ and RRM induced by RNA binding. Analogous Q-band DEER experiments performed on the A349R1/S412R1 system with stem-loop RNA (Fig. 7B red line, bottom panel) reveal a very broad inter-spin distance distribution. A less evident main peak compared to that obtained in the absence of nucleotides is present in the range ~4–5 nm. These data suggest that both RGG regions do not undergo significant conformational changes in the presence of stem-loop RNA, keeping their native disordered behaviour in the presence of nucleobases.

Discussion

RGG motifs are continuous Arg-Gly-Gly repeats in the primary structure of different types of RBPs and this unique amino acid composition seems to be a key factor in the mechanisms of protein–RNA and/or protein–protein interaction.

The FUS protein possesses three different RGG domains; the C-terminal RGG₃ domain has been well characterized and demonstrated to be important for the interaction with Transportin, in phase-separation and stress granule formation of FUS^{19–21}. RGG₁ and RGG₂ regions were independently characterized in constructs including the RRM but they were never characterized together with the intervening RRM domain. We thus decided to focus on the FUS^{165–422} construct comprising both the RGG₁ and RGG₂ regions, and on their interaction with RNA.

The RGG regions have been predicted to be completely disordered thus with a high degree of flexibility^{15,17,18,27}. Far-UV CD and NMR experiments performed on FUS^{165–422} confirmed that only part of the polypeptide is structured. In addition, X-band CW-EPR, DEER and NMR experiments on different spin labeled FUS variants revealed the great mobility of the G-rich, RGG₁, and RGG₂ regions typical of IDPs, as previously suggested^{26,34}. IDPs can be characterized by ensembles of rapid interconverting extended and compact conformations of protein stretches³⁵. This definition perfectly fits with the properties of the RGG₁ and the RGG₂ regions. The DEER inter-spin distances obtained for doubly MTSL-labelled FUS^{165–422} mutants have a broad distribution, suggesting that both motifs simultaneously possess both elongated and compact forms in solution. These concomitant “open” and “close” states have been also found in TDP-43, which is a protein related to FUS. NMR spectroscopy showed

that TDP-43 has dynamic inter-domains interactions mediated by IDRs that could control the functions and the proteinopathy of the protein³⁶.

We can speculate that RGG₁ and RGG₂ interact with RRM, due to the multimodal and broad shape of the DEER inter-spin distance profile. However, these contacts between domains should be transient interactions in the ps-ns range since NMR experiments recorded for FUS^{165–422} mutants bearing MTSL in RGG boxes did not reveal strong PRE effects on the RRM domain. This is also in agreement with CW-EPR data reflecting that only a small fraction of the RGG₁ conformation ensembles possessed partial restrict motions due to steric hindrance exerted by the globular domain.

The RGG regions result particularly important for the physiological RNA-binding mechanism of FUS. Even if the main sites for RNA interaction of FUS are the RRM and the ZnF domains, the RGG repeating motifs coupled to adjacent folded domains showed affinities approaching those of full-length FUS, suggesting that the RGG boxes might act as primary mediators in the nucleotide binding process¹⁸. In fact, Allain and co-workers recently assessed that the RGG₂ domain of FUS is implicated in stem-loop RNA structural destabilization favoring the interaction of the RRM. In terms of dynamics, no evident change upon RNA interaction was detected by CW-EPR experiments for the RGG₂ domain suggesting that this region of FUS does not undergo any conformational variation in the presence of RNA. This behaviour was also supported by Q-band DEER measurements since a broad shape of inter-spin distance distributions was observed for the A349R1/S412R1 system with and without RNA, evidencing that the RGG₂ motif could retain its flexibility with the hnRNPA2/B1 stem-loop in solution. Considering that various studies observed a synergistic RNA-binding mechanism of protein domains, this discrepancy of results could be related to the absence of the ZnF domain in the FUS^{165–422} construct used here^{17,27}.

Although the RGG₁ motif resulted to be very flexible, it displays some small variations upon RNA-binding compared to the other disordered regions of FUS^{165–422}. In particular, partial decrease in mobility was detected for stem-loop RNA binding by CW-EPR while DEER experiments indicated a minimal reduction in distances between RGG₁ and the RRM domain. These findings could be related to the formation of a less-extended conformation (i.e. β -turns) during RNA-coordination, as previously determined for others RRG motifs of FUS and similar proteins^{34,37,38}. However, no drastic variations have been observed for RGG₁ in complex with RNA in this study thus we can infer that also this IDR of FUS does not undergo a defined disorder-to-order transition in presence of nucleotides. Summarizing, the experiments performed in this work were not able to detect any direct interaction of the RGG regions with stem-loop RNA.

Recently, several studies evidenced that a large class of proteins that contain IDR connecting globular domains form fuzzy complexes with partners (i.e. protein or nucleic acids) with different topologies³⁹. Flanking models are a category of complexes that lack a defined conformation where IDRs increase the propensity of the folded regions toward binding of competent form and regulate the entropy of the process⁴⁰. We speculate that this type of conformational model could be adopted by RGG₁ upon coordination of stem-loop RNA, as also previously demonstrated for another splicing factor complex (U2AF⁶⁵/SF1) sharing various similarities with the FUS protein⁴¹.

Various RBPs have evolved a great variety of pathways to remodel RNA. These proteins, which generally contains a large number of cationic residues, transiently interact with nucleotides enabling the RNA backbone to assume wide conformations⁴². Moreover, Allain and co-workers speculated about the role exerted by the RGG regions of FUS in RNA destabilization and remodelling, without assuming a folded conformation²⁷. Thus, the possibility that RGG₁₋₂ of FUS protein could bind transiently the hnRNPA2/B1 stem-loop RNA cannot be ruled out.

The presence of the adjacent RGG motifs significantly increases the affinity of the RRM domain for different nucleotide sequences, and structured RNA in particular^{18,27}. We thus investigated the structural effects of stem-loop RNA coordination on the RRM region of FUS. In solution, this domain has a more compact conformation compared to RGG₁ and RGG₂, but it experiences a more pronounced mobility than generally observed for a normal globular protein⁴³. Based on CW-EPR spectra recorded in the presence of RNA, the RRM domain revealed concomitant increase and a decrease of local flexibility, underlining that remodelling of the conformation takes place upon nucleotide binding. This capability to modulate its structure, in combination with the RNA destabilization exerted by RGG motifs, supports the recent speculation that RRM nucleotide recognition is more based on RNA shape than on nucleobase sequence. However, further studies are needed to confirm this finding.

Summarizing, our results provide additional structural information on the RGG domains of FUS and their implication in RNA binding activity. Through an integrated approach of EPR and NMR spectroscopies, the degree of disordered of the RGG boxes of FUS have been characterized even if future experiments aimed to individuate their role in RNA-binding are necessary. The experimental data revealed also a dynamic effect of these flanking domains exerted on the folded region and a variation in terms of tumbling rates for the α_2 helix and the flexible loop connecting α_2 and β_4 of the RRM domain during stem-loop RNA coordination. RGG motif should represent a key-factor for determining the physiological and also pathological pathways of different human proteins related to fatal neurodegenerative diseases, such as ALS and FTL. Further biochemical and biophysical efforts on the characterization of the role exerted by these domains in protein activities should be performed in order to contribute to the development of possible therapies for the treatment of the related pathologies.

Methods

Protein expression and purification. The gene encoding for FUS^{165–422} fragment (GENEWIZ) was cloned in the pENTR vector for the Gateway cloning technology and subsequently in the pDEST-HisMBP vector (which adds a His tag fused with the Maltose Binding Protein at the N-terminus of the protein), utilizing the pENTR/TEV/D-TOPO cloning kit (Invitrogen). *E. coli* BL21 (DE3) cells were transformed with the plasmid pDEST-HisMBP-FUS^{165–422}, grown in LB medium (or in ¹⁵N M9 medium for NMR experiments) to mid-log phase at 37 °C, induced with 0.50 mM IPTG, and then they were kept at 20 °C for at least 16 h. For protein purifi-

cation we adapted the protocol established by Ozdilek et al.¹⁸. In brief, the cells were harvested and re-suspended in 50 mL of binding buffer (20 mM sodium phosphate buffer at pH 6.5, 5 mM of DTT, 5 mM of Imidazole, 10% v/v of glycerol and 1.5 M of NaCl) supplemented with protease inhibitor tablets (Bayer) and successively lysed by sonication. The lysate was passed through a 5 mL HisTrap FF affinity column (GE Healthcare Life Sciences) and washed with the elution buffer (20 mM sodium phosphate buffer at pH 6.5, 5 mM DTT, 1.5 M of NaCl, 500 mM imidazole and 10% v/v of glycerol). The His-MBP tag was cleaved incubating the protein solution overnight with TEV protease and successively separate from the fusion protein with a second step of purification with 5 mL HisTrap FF affinity column using the same elution buffer reported above.

As final purification step, a gel filtration using a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences) was performed to separate FUS^{165–422} from possible contaminants and to transfer the protein in the final buffer (20 mM MES at pH 6.5 with 150 mM NaCl). The corresponding MALDI-TOF spectrum (see Supplementary Fig. S3A online) was acquired as control, reporting the peak at about 25,000 m/z, corresponding to the correct molecular weight of FUS^{165–422}. In addition a SDS-page gel was performed to check the purity of the protein at the end of the purification protocol (see Supplementary Fig. S3B online).

Site directed mutagenesis for cysteine mutants. The five single cysteine mutations (D180C, D258C, A349C, S360C and S412C) were inserted into FUS^{165–422} sequence using the kit Quick Change II Site-Directed Mutagenesis Kit (Agilent). The primers of each variant (see Supplementary Table S4 online) were placed in solution with pENTR-FUS^{165–422} vector and successively subjected to 18 steps of PCR reaction. The final products, after DNA sequencing controls, were cloned into pDEST-HisMBP and the relative protein variants were produced and purified as described previously.

The plasmids encoding the protein bearing the double cysteine variations (D258C/A349C and A349C/S412C) were obtained as described previously using the pENTR-FUS^{165–422} vector carrying already the A349C mutations and then introducing D258C and S412C mutations respectively.

Site-directed spin labeling reaction. Each protein solution, containing ~ 100 nmol of FUS^{165–422} single and double cysteine variants, was incubated with a tenfold molar excess of DTT at room temperature in order to reduce the cysteine thiol group. After 30 min, the reaction mixture was eluted into a PD-10 desalting column to remove the reducing agent using MES buffer (20 mM MES at pH 6.5 with 150 mM NaCl) as mobile phase. Successively S-(1-oxyl-2,2,5,5,-tetramethyl-2,5,-dihydro-1H-pyrrol-3-yl) methylmethane-sulfonothiolate (MTSL) was added to the solution with 1:10 protein:label molar ratio; the reaction was kept at room temperature under gentle stirring and in absence of light for at least 10 h. The unreacted MTSL was removed using a PD-10 desalting column using MES buffer. The fractions containing single (D180R1, D258R1, A349R1, S360R1 and S412R1) and double (D248R1/A349R1 and A349R1/S412R1) MTSL-labelled positions of FUS^{165–422} were individually checked recording the relative X-band CW-EPR spectrum, to assure the complete removal of the free tag. Successively all the fractions containing labelled samples were pooled together and the solution was concentrated using AMICON ultra centrifugal filters (MWCO=10 kDa). The final protein concentration was calculated observing the absorbance at 280 nm ($\epsilon=17,420 \text{ M}^{-1} \text{ cm}^{-1}$) with a UV/Vis Varian Cary 50 spectrophotometer. The labelling yield of each sample (see Supplementary Table S5 online) was calculated from the double integral of CW-EPR spectrum and it was compared with a calibration curve obtained for free MTSL in water solution at different molar concentrations.

For PRE experiments, the FUS^{165–422} mutants bearing cysteine substitution without MTSL grafted on the thiol group were used as diamagnetic samples.

hnRNPA2/B1 stem-loop RNA preparation. hnRNPA2/B1 stem-loop RNA (GGCAGAUUACAAUUC UAUUUGCC) was synthesized by Eurofins Genomics. The lyophilized product has been dissolved in RNAase free buffer (20 mM MES at pH 6.5 and 150 mM NaCl) and the concentration of stock solution was checked measuring the absorbance at 260 nm with UV/vis spectroscopy. For CD, EPR and NMR analysis of FUS^{165–422} in presence of hnRNPA2/B1 stem-loop RNA, the buffer solution containing the protein has been prepared in RNAase free solution in order to avoid possible RNA degradation.

Circular Dichroism spectroscopy. Far-UV Circular Dichroism (CD) experiments were performed at room temperature with a Jasco CD-J-815 spectropolarimeter using a quartz cuvette with a path length of 1 mm length. CD spectrum of FUS^{165–422} in solution was recorded with a final protein concentration either of 15 μM or 50 μM in 20 mM of MES buffer (pH 6.5) and 150 mM of NaCl from 200 to 250 nm. At wavelength below 200 nm the absorbance due to the presence of NaCl, necessary for protein stability, saturate the photomultiplier and the data were discarded. Experiments in the presence of RNA were recorded placing in solution the protein with 1:1 FUS: hnRNPA2/B1 RNA molar ratio. Each spectrum was accumulated at least 10 times in order to improve signal-to-noise ratio. Secondary structure elements of the protein in different conditions were derived from CD spectra using Dichroweb analysis tool²³. The prediction of the percentage of folded and disordered region based on the primary sequence of FUS^{165–422} were realized through the PDB structural model of the RRM (PDB ID 2LA6) and GlobPlot 2.3 program²⁵.

X-band CW-EPR spectroscopy. Each CW-EPR spectrum of MTSL-labeled FUS^{165–422} variants (D180R1, D258R1, A349R1, S360R1 and S412R1) in solution and with RNA was recorded with an ELEXYS E580 spectrometer equipped with a Super High Q resonator (SHQE) operating at X-band (9.8 GHz) at room temperature. Each CW-spectrum for spin labeled FUS^{165–422} variants were obtained with a final protein concentration

of approximately 60 μM in 20 mM of MES buffer (pH 6.5) and 150 mM of NaCl. For the experiments in the presence of hnRNPA2/B1 RNA, a 1:1 protein:RNA ratio was used.

The following spectroscopic parameters were applied to record each spectrum: microwave power = 10 mW; magnetic field amplitude = 1 G; field sweep = 150 G; receiver gain = 60 dB; modulation frequency = 100 kHz.

All EPR spectra were successively simulated with the program SimLabel (a GUI of Easyspin software)⁴⁴ in order to obtain spectral component percentages and the corresponding correlational times (τ_c). The error on the components' percentages and the correlational times values for each spectrum were obtained performing three independent simulations of the same spectra. The correlation times and the component percentages reported in the text for each labelling position represent the average values arising from the spectral simulations.

Q-band DEER spectroscopy. DEER experiments on FUS^{165–422} doubly spin labeled mutants (D258R1/A349R1 and A349R1/S412R1) in buffer solution (20 mM MES at pH 6.5 with 150 mM of NaCl) or in the presence of 1:1 protein:RNA were performed on Elexys E580 spectrometer coupled with the standard resonator EN5107D2 operating at Q-band (34 GHz) and a 3 W microwaves amplifier (see Supplementary Fig. S6 online). The spectrometer was equipped with an Oxford helium temperature regulation unit and the relative experiments were carried out at 50 K. Each sample was prepared with a final protein concentration of ~80–90 μM in the presence of glycerol (final concentration equal to 10% v/v).

DEER experiments were performed using the four-pulse dead-time free DEER sequence ($\pi/2 - \pi - \pi_{\text{PUMP}} - \pi$) setting $\pi/2 = 20$ ns and $\pi = 40$ ns. The inter-pulse t_1 was set to 200 ns and the t_2 was adjusted based on the phase memory constant time recorded for each labelled mutant. The pump ELDOR pulse ($\pi_{\text{PUMP}} = 36$ ns) was positioned at the center of the MTSL absorption spectrum and the frequency difference between the pump and the observer pulses was 65 MHz. The DEER traces were accumulated at least 14 h in order to minimize the signal to noise ratio.

DeerAnalysis 2019 software was used for the correction of background echo decay and consequently to obtain the inter-spin distances involving the Tikhonov regularization. The regularization factor (α) was selected into the corner, according to L-curve criteria.

NMR spectroscopy. The 2D ¹H-¹⁵N HSQC NMR spectra of a 100 μM ¹⁵N-FUS alone or in complex with hnRNPA2/B1 stem-loop RNA (1:1 FUS:RNA) were acquired at 298 K with a 22.3 T Bruker Avance III NMR spectrometer operating at 950.06 MHz for ¹H and equipped with a cryogenically cooled triple resonance probehead (TCI).

Each 2D ¹H-¹⁵N HSQC NMR spectrum of unlabelled and MTSL-labelled ¹⁵N-FUS^{165–422} variants (D180R1, D258R1, A349R1, S360R1 and S412R1, 200 μM) in solution were acquired at 298 K with a 22.3 T Bruker Avance III NMR spectrometer operating at 950.06 MHz for ¹H and equipped with a cryogenically cooled triple resonance probehead (TCI).

All the heteronuclear relaxation experiments (R_1 , R_2 and ¹H-¹⁵N NOEs) were acquired using ¹⁵N labeled FUS at about 200 μM . The spectra were recorded at 298 K on a 16.4 T Bruker Avance NEO spectrometer operating at 700.06 MHz for ¹H and equipped with a cryogenically cooled triple resonance probehead (TCI). The ¹⁵N R_1 and R_2 experiments were acquired with 8 scans per increment (2048 \times 320 points). To determine the ¹⁵N R_1 the following delays were used: 20 ms, 60 ms, 120 ms, 180 ms, 240 ms, 400 ms, 600 ms, 800 ms, 1 s, 1.2 s. To determine the ¹⁵N R_2 the following delays were used: 32 ms, 64 ms, 128 ms, 160 ms, 190 ms, 220 ms, 260 ms, 320 ms, 360 ms, 440 ms, 500 ms. The relaxation delay for ¹⁵N R_1 and R_2 experiments was 3.0 s. The ¹H-¹⁵N NOEs experiments were acquired with 64 scans (2048 \times 320 points) and a relaxation delay of 6.0 s, used for ¹H saturation when necessary.

NMR data sets were processed using either Bruker TopSpin 3.5pl7 or TopSpin 4.0.6 software. CARA⁴⁵ was used to analyze and annotate the spectra.

The ¹⁵N relaxation rates (R_1 and R_2) were determined by fitting the cross-peak intensity measured as a function of variable delay, to single-exponential decay using the Bruker Dynamics Center 2.4, available as stand-alone ancillary software of TopSpin by Bruker. The error bars are generated by the software using the peak intensities under the assumption that their uncertainties are unknown but equal for all Y values and the non-linear fit determines errors for the fitted parameters. ¹H-¹⁵N NOEs values were obtained as a ratio between peak intensity in spectra recorded with and without ¹H saturation. The error can be estimated about 10% measuring the noise level of the spectrum without saturation.

The secondary structure propensity (SSP, see Supplementary Fig. S1 online) from heteronuclear chemical shifts was calculated by using the neighbour corrected structural propensity calculator (ncSPC) tool⁴⁶ available online at <http://nmr.chem.rug.nl> and the BMRB 17508 data.

Chemical shift mapping was depicted on the RRM structure (PDB ID 2LA6) by using UCSF Chimera⁴⁷, a molecular graphics program developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

Received: 22 July 2020; Accepted: 18 November 2020

Published online: 01 December 2020

References

1. Lunde, B. M., Moore, C. & Varani, G. RNA-binding protein: modular design for efficient function. *Nat. Rev. Mol. Cell. Biol.* **8**, 479–490 (2007).
2. Valverde, R., Edwards, L. & Regan, L. Structure and function of KH domains. *FEBS J.* **275**, 2712–2726 (2008).
3. Linder, P. & Jankowsky, E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell. Biol.* **12**, 505–516 (2011).

4. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
5. Anyanaharan, V., Koonin, E. V. & Aravind, L. Comparative genomics and evolution of protein involved in RNA metabolism. *Nucl. Acid Res.* **30**, 1427–1464 (2002).
6. Järvelin, A., Noeremberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. *Cell Commun. Signal.* **14**, 9 (2016).
7. Srivastava, A., Ahmad, S. & Gromiha, M. M. Deciphering RNA-recognition patterns of intrinsically disordered proteins. *Int J. Mol. Sci.* **19**, 1595 (2018).
8. Basu, A. & Bahadur, R. P. A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell. Mol. Life Sci.* **73**, 4075–4084 (2004).
9. Calabretta, S. & Richard, S. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).
10. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG motif. *Mol. Cell.* **50**, 613–623 (2016).
11. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell.* **33**, 1–13 (2009).
12. Qamar, S. *et al.* Fus phase separation is modulated by molecular chaperone and methylation of arginine cation- π . *Cell* **173**, 720–734 (2018).
13. Da Cruz, S. & Cleveland, D. W. Understanding the role of tdp-43 and FUS/TLS in ALS and beyond. *Curr Opin. Neurobiol.* **21**, 904–919 (2011).
14. Chen, C., Ding, X., Akram, N., Xue, S. & Luo, S. Z. Fused in sarcoma: properties, self-assembly and correlation with neurodegenerative diseases. *Molecules* **24**, 1622 (2019).
15. Efimova, A. D. *et al.* The FUS protein: physiological functions and a role in amyotrophic lateral sclerosis. *Mol. Biol.* **51**, 341–351 (2017).
16. Chong, P. A., Vernon, M. R. & Forman-Kay, J. D. RGG/RG motif regions in RNA binding and phase separation. *J. Mol. Biol.* **430**, 4650–4665 (2018).
17. Wang, X., Schwartz, J. C. & Cech, T. R. Nucleic acid binding specificity of human FUS protein. *Nucl. Acc. Res.* **43**, 7535–7543 (2015).
18. Ozdilek, B. A. *et al.* Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucl. Acc. Res.* **45**, 7984–7996 (2017).
19. Dormann, D. *et al.* Arginine methylation next to the PY-NLS modulates transportin binding and nuclear import of FUS. *EMBO J.* **31**, 4258–4275 (2012).
20. Hofweber, M. *et al.* Phase separation is suppressed by its nuclear import receptor and arginine methylation. *Cell* **173**, 406–419 (2018).
21. Darovic, S. *et al.* Phosphorylation of the C-terminal tyrosine 526 of FUS impairs its nuclear import. *J. Cell Sci.* **128**, 4151–4159 (2015).
22. Wilson, K. A., Holland, D. J. & Wetmore, S. D. Topology of RNA-protein nucleobase-amino acid π - π interactions and comparison to analogous DNA-protein π - π contacts. *RNA* **22**, 696–707 (2016).
23. Withmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and references databases. *Biopolymers* **89**, 392–401 (2007).
24. Lu, Y., Liangzhong, L. & Song, J. RRM domains of ALS/FTD-causing FUS characteristic of irreversible unfolding spontaneously assemble to amyloid fibril. *Sci. Rep.* **7**, 1043 (2017).
25. GlobPlot2.3. <http://globplot.embl.de>
26. Liu, X. *et al.* The RRM domain of human fused in sarcoma reveals a non canonical nucleic acids binding site. *Biochim. Biophys. Acta* **1832**, 375–385 (2013).
27. Loughlin, F. E. *et al.* The solution structure of FUS binds to RNA reveal a bipartite mode of RNA recognition with both shape and sequence specificity. *Mol. Cell* **73**, 490–504 (2018).
28. Klare, J. P. Site-directed spin labeling EPR spectroscopy in protein research. *Biol. Chem.* **394**, 1281–1300 (2013).
29. Le Breton, N. *et al.* Exploring intrinsically disordered protein using site-directed spin labeling electron paramagnetic spectroscopy. *Front. Mol. Biosci.* **2**, 21 (2015).
30. Weickert, S., Cattani, J. & Drescher, M. Intrinsically disordered protein studied by EPR and in-cell. *Electron Paramagn. Reson.* **26**, 1 (2018).
31. Kosol, S., Contreras-Martos, S., Cadeno, C. & Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802–10828 (2013).
32. Konart, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Res.* **241**, 74–85 (2014).
33. Jeschke, G. DEER distance measurements on proteins. *Annu Rev. Phys. Chem.* **63**, 419–446 (2013).
34. Yagi, R., Miyazaki, T. & Oyoshi, T. G-quadruplex binding ability of TLS/FUS depends on the β -spiral structure of the RGG domain. *Nucl. Acc. Res.* **46**, 5894–5901 (2018).
35. Kurzbach, D. *et al.* Cooperative unfolding of compact conformations of the intrinsically disordered protein Osteopontin. *Biochem.* **52**, 5167–5175 (2013).
36. Wei, Y., Lim, L., Wang, L. & Song, J. Inter-domain interactions of TDP-43 as decoded by NMR. *Biochem Biophys. Res. Commun.* **473**, 614–619 (2016).
37. Ramos, A., Hollingworth, D. & Pastore, A. G-quadruplet dependent recognition between the FMRP RGG box and RNA. *RNA* **9**, 1198–1207 (2003).
38. Takahama, K., Sugimoto, C., Arai, S., Kurokawa, R. & Oyoshi, T. Loop length of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in EWS. *Biochem.* **50**, 5369–5378 (2011).
39. Sharma, R., Raduly, Z., Miskei, M. & Fuxreiter, M. Fuzzy complexes: specific binding without complete folding. *FEBS Lett.* **589**, 2533–2542 (2015).
40. Tompa, P. & Fuxreiter, M. Fuzzy complex: polymorphs and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).
41. Selenko, P. *et al.* Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* **11**, 965–976 (2003).
42. Doetsch, M., Schroeder, R. & Fürtig, B. Transient protein-RNA interactions in RNA folding. *FEBS J.* **278**, 1634–1642 (2011).
43. Longhi, S., Belle, V., Fournel, A., Guigliarelli, B. & Carriere, F. Probing structural transitions in both structured and disordered proteins using site-directed spin labelling EPR spectroscopy. *J. Pept. Sci.* **17**, 315–328 (2011).
44. Etienne, E., Le Breton, N., Martinho, M., Mileo, E. & Belle, V. Simlabel : a graphical user interface to simulate continuous wave EPR spectra from site-directed spin labelling experiments. *Magn. Reson. Chem.* **55**, 714–719 (2017).
45. Keller, R., The computer aided resonance assignment tutorial. *Cantina Verlag* (2004)
46. Tamiola, K. & Mulder, F. A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40**, 1014–1020 (2012).
47. Pettersen, E. F. *et al.* UCSF Chimera, a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

The support and the use of resources of the CERM/CIRMMP center of Instruct-ERIC is gratefully acknowledged. This work was supported by Instruct-ULTRA (Grant 731005), an EU H2020 project to further develop the services of Instruct-ERIC, and by the Italian Ministry for University and Research (FOE funding). AB is grateful to Regione Toscana for a post-doctoral grant (POR-FSE 2014-2020). Bruno Rizzuti and Leonardo Gonnelli are acknowledged for the stimulating discussions.

Author contributions

A.B., L.B. and R.P. conceived the project. A.B. did the molecular biology work. M.G.M. recorded and analyzed the NMR experiments. A.B. recorded and analyzed the EPR experiments. All the authors reviewed and discussed the results and contribute to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-77899-x>.

Correspondence and requests for materials should be addressed to L.B. or R.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020