UNIVERSIDADE NOVA DE LISBOA

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

An NMR Insight into the Liquid-Liquid Phase Separation of FUS

Philip Guilherme Capelo O'Toole

Dissertação para obtenção do Grau de Mestre em Bioquímica

Orientador: Professor Doutor Eurico José da Silva Cabrita, Professor Associado com Agregação, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Júri:
Presidente: Professor Doctor Ricardo Franco
Arguente: Doctor Douglas Laurents
Vogal: Professor Doutor Eurico Cabrita

Fevereiro 2021

An NMR Insight into the Liquid-Liquid Phase Separation of FUS

Philip Guilherme Capelo O'Toole

2021

An NMR insight into the liquid-liquid phase separation of FUS

“Copyright”

Philip Guilherme Capelo O’Toole

Faculdade de Ciências e Tecnologia

Universidade Nova de Lisboa

Acknowledgements

Firstly, I would like to thank my supervisor, professor Eurico Cabrita, for the opportunity to undergo this work. For the discussions and the assistance regarding the work and dissertation, I deeply thank professor Eurico Cabrita and Sara Félix. An additional special thanks to Sara Felix for all the support regarding the more practical aspect of the work, discussions and for understanding the hardships associated to the project. Additionally, I would like to thank professor Eurico Cabrita for his NMR classes and assistance regarding the setup of NMR experiments, alongside Dr. Aldino Viegas, and Dr. Jorge Dias for answering any doubts I had regarding protein purification.

I would like to thank my friends who had to listen to me discuss the work, even if they did not show their lack of interest so blatantly.

I expressly thank my closest family, for all their love and support, and for trying to push me to do my best throughout my entire life, thus far.

Abstract

The formation of two distinct liquid phases from a homogenous solution of solvent and polymer is termed liquid-liquid phase separation (LLPS). The LLPS phenomenon has been described for polymers and proteins. The LLPS of proteins is crucial for the formation of membraneless organelles, which play an important role in intracellular compartmentalization, and is inherently linked to protein fibrillation found in neurodegenerative diseases.

The protein fused in sarcoma (FUS) is present in protein fibrils found in amyotrophic lateral sclerosis patients and has been extensively studied as a model protein for LLPS. The protein contains a prion-like domain, two folded RNA binding domains, the RNA recognition motif (RRM) and a zinc finger domain (ZnF) and three arginine-glycine-glycine rich regions (RGG). The main molecular drivers and interactions underlying the LLPS phenomenon of FUS have been previously described, particularly for the LC domain. However, the role of the different regions in this structurally complex protein remains to be fully understood.

In this work, the full-length protein was divided into two distinct previously unstudied constructs, namely RGG_1-RRM-RGG_2 and RGG_2-ZnF-RGG_3. The LLPS propensity was examined, together with the molecular and structural drivers underlying the LLPS phenomenon of each construct. These objectives were addressed by performing turbidity assays and nuclear magnetic resonance spectroscopy (NMR).

Based on the turbidity assays, the RGG_1-RRM-RGG_2 construct appears to undergo LLPS when isolated. The folded RRM may play a direct role under certain conditions and RNA binding significantly affects the observed LLPS propensity. Moreover, at least ionic, and π-cation interactions act as drivers for the LLPS phenomenon of this domain.

Through NMR spectroscopy, several residues of the folded RRM domain were significantly affected by temperature, sample cooling and reheating procedures and protein concentration. The specifically affected residues are likely to participate in protein-protein interactions, therefore establishing a hypothetical protein-protein interaction in the folded RRM domain. Due to the structural localization of the interface, the protein-protein interactions established by the folded RRM domain are not directly affected by RNA-binding.

The RGG_2-ZnF-RGG_3 construct, as observed through NMR spectroscopy, presented an apparent cluster of H-bond performing residues within the folded ZnF domain, namely in the vicinity of the residues coordinating the Zn^{2+} ion. Furthermore, the construct did not appear to undergo LLPS under the tested conditions. To fully certify the absence of LLPS in this construct, further assays must be performed under distinct conditions.

A previously unstudied FUS based construct revealed the capacity to undergo LLPS, driven at least by ionic, π-cation interactions and with direct participation of the folded RRM domain. The contribution of the RGG regions on the observed LLPS phenomenon remains to be elucidated. The tested RGG_2-ZnF-RGG_3 did not undergo LLPS under the tested conditions. Further assays are necessary to evaluate the LLPS capacity of the construct.

Keywords: liquid-liquid phase separation; FUS; NMR; protein-protein interactions.

Resumo

A formação de duas fases líquidas distintas partindo de uma solução homogénea de solvente e polímero é designado de separação de fase líquido-líquido (LLPS). O fenómeno LLPS foi previamente descrito para polímeros e para proteínas. O LLPS de proteínas é crucial para a formação de organelos sem membrana, que participam na compartimentação intracelular, e está ligada à fibrilação proteica em doenças neurodegenerativas.

A proteína fused in sarcoma (FUS) está presente em fibrilas proteicas encontradas em doentes de esclerose lateral amiotrófica e foi estudado extensivamente como modelo de LLPS de proteínas. Os determinantes moleculares e interações subjacente ao fenómeno de LLPS foram descritos no passado. Porém, a relevância das diferentes regiões da proteína estruturalmente complexa permanece a ser descrito.

Para este objetivo, a proteína inteira foi dividida em duas regiões distintas, nomeadamente RGG_1-RRM-RGG_2 e RGG_2-ZnF-RGG_3. A propensão de LLPS dos constructos descritos foi estudada, assim como as condicionantes moleculares e interações subjacentes aos fenómenos de LLPS dos respetivos construtos. Estes objetivos foram estudados através de ensaios de turbidez e espetroscopia de ressonância magnética nuclear (RMN).

Com base em ensaios de turbidez, o construto isolado RGG_1-RRM-RGG_2 aparenta realizar LLPS. O domínio estruturado RRM aparenta participar diretamente no fenómeno de LLPS em certas condições e a ligação a RNA afeta o fenómeno. Foi possível identificar a contribuição de interações iónicas e π-catião no fenómeno observado.

Através de NMR, variados resíduos do domínio estruturado RRM foram significativamente afetados por temperatura, o processo de arrefecimento e aquecimento da amostra e pela variação da concentração de proteína. A especificidade dos efeitos observados é indicativa da participação dos resíduos em interações proteína-proteína, tendo sido possível estabelecer uma interface de interações proteína-proteína hipotética. Devido à localização estrutural da interface, foi possível elucidar que a ligação a RNA não deverá interferir diretamente com as interações estabelecidas por esta interface.

O construto RGG_2-ZnF-RGG_3, observado através de NMR, apresenta um agrupamento de resíduos a realizar pontes de hidrogénio no interior do domínio estruturado ZnF, nomeadamente em torno de resíduos que coordenam o iao de Zn^{2+}. Nas condições testadas, o construto não aparenta realizar LLPS isolado. De modo a certificar a hipótese, ensaios futuros em condições distintas serão necessários.

Um construto nunca previamente estudada, baseado na proteína FUS, revelou a capacidade de realizar LLPS, sendo o fenómeno dependente de interações iónicas, π-catião e participação direta do domínio estruturado RRM. A contribuição direta das regiões RGG permanece a ser elucidado. O construto RGG_2-ZnF-RGG_3 não aparenta ser capaz de realizar LLPS nas condições testadas. Futuros ensaios são necessários de modo a confirmar a capacidade de LLPS deste construto.

Palavras-chave: separação de fase líquido-líquido; FUS; RMN; interações proteína-proteína.

TABLE OF CONTENTS

TABLE OF FIGURES

List of abbreviations

6xHis	Polyhistidine tail
OD_{600}	Optical density at 600 nm
ALS	Amyotrophic lateral sclerosis
BMRB	Biological Magnetic Resonance Bank
BSA	Bovine serum albumin
CSP	Chemical Shift perturbation
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded Deoxyribonucleic acid
FET	FUS, EWS, TAF15
FRAP	Fluorescence recovery after photobleaching
FTLD	Frontotemporal lobar degeneration
FUS	Fused in sarcoma
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani medium
LC	Low-complexity domain
LCST	Lower critical separation temperature
LLPS	Liquid-Liquid phase separation
MBP	Maltose binding protein
MWCO	Molecular Weight Cut-off
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NTA	Nitrilotriacetic acid
OD	Optical Density
PAR	Poly adenosine diphosphate ribose
PDB	Protein databank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulphonyl fluoride
PrLD	Prion-like domain
RGG	Arginine-Glycine-Glycine rich region
RNA	Ribonucleic acid
RRM	RNA Recognition Motif
SAXS	Small Angle X-ray scattering
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SGs	Stress Granules
SLiM	Short linear motif
ssDNA	Single-stranded Deoxyribonucleic acid
TEV	Tobacco etch virus
Tris	tris(hydroxymethyl)aminomethane
UCST	Upper critical separation temperature
ZnF	Zinc-finger

Amino Acids One-letter abbreviations

Alanine	A
Arginine	R
Asparagine	N
Cysteine	C
Glutamate	E
Glutamine	Q
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V

1 Introduction

Living organisms have the remarkable task of confining specific reactions in a precise spatial and temporal manner. In eukaryotes, as commonly known, such a feat is achieved through membrane-bound organelles. Recently it has been observed the presence of membraneless organelles. These organelles are macromolecule-rich dynamic structures, generated through liquid-liquid phase separation (LLPS) [1].

1.1 Liquid-Liquid Phase Separation

The demixing phenomenon of a homogeneous solution into separate phases, is termed as liquid-liquid phase separation. Considering a binary mixture, phase separation occurs due to the tendency for the system to lower its free energy by dividing into separate phases with distinct compositions, as depicted in **Figure 1-1** [1].

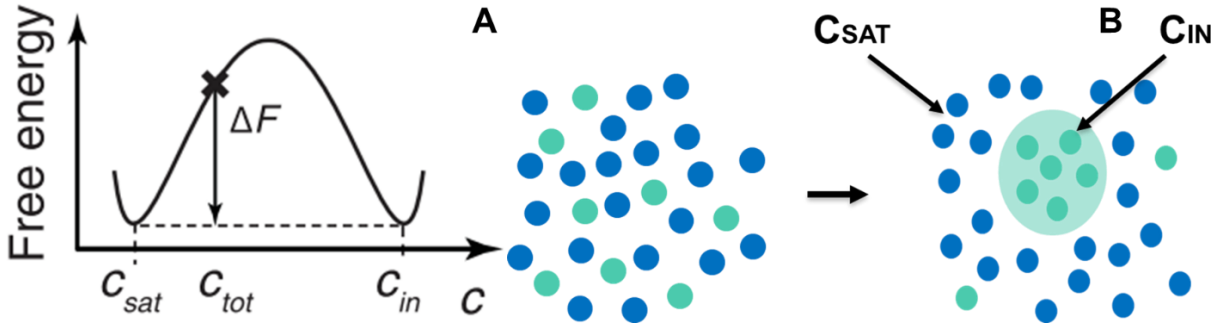


Figure 1-1 Schematic representation of the thermodynamic profile of phase separation (A) and the resulting system (B). The black line in **A** represents the free energy of a system at each total solute concentration C_{tot}. C_{sat} indicates the saturated concentration of the component in bulk solution, whereas C_{in} corresponds to the concentration of the component within the enriched phase, after demixing. The parameter ΔF represents the difference in free energy at a specific C_{tot} upon demixing. Visually represented in **B**, the blue circles represent solvent molecules whereas cyan circles represent solute molecules. Adapted from [1].

The LLPS phenomenon occurs when the concentration of a compound or mixture of compounds trespass above a certain critical concentration in bulk solution (C_{sat} in **Figure 1-1**), becoming supersaturated. The consequence of reaching this concentration is the promotion of the LLPS, where the supersaturated component is enriched in a separate phase (C_{in}). Due to this, the LLPS process is highly tunable by concentration and several other parameters which may affect the free energy profile [2].

The core theoretical approach to LLPS of polymers is the Flory-Huggins theory [3]. The theory accounts for the entropic mixing energy as well as with the enthalpy associated to solvent-solvent, solvent-polymer and polymer-polymer interactions. A crucial parameter of the theory is the energy cost for replacing a polymer interaction site with a solvent molecule, the parameter X [4]. Effectively, the value of X indicates which are the dominant interactions, which translates to the solvent being a poor solvent (X>0) or a good solvent (X<0) [5]. In poor solvent conditions, the polymer will tend to collapse and adopt a compact conformation, such as the case of proteins. In extreme cases however, where X greatly exceeds 0, the polymer will tend to phase separate into a polymer-rich and a distinct polymer-depleted phase [6],[7].

In a thermodynamic context, entropy contributions will generally favor a well-mixed system, and consequentially, the LLPS process is mainly enthalpically driven through intra and

inter-molecular polymer-polymer interactions. A variety of polymer-polymer interactions are possible, the types of interactions are however, inherently linked to the polymer. Different types of interactions driving the LLPS phenomenon will respond differently to certain conditions, e.g. H-bonds tend to decrease in strength with the increase in temperature and the opposite is observed for hydrophobic interactions [8]–[10]. This distinct responses to temperature are responsible for the distinct LLPS behaviour with temperature, as some polymers display an upper critical solution temperature while others display a lower critical solution temperature (UCST and LCST, respectively) **Figure 1-2** [11].

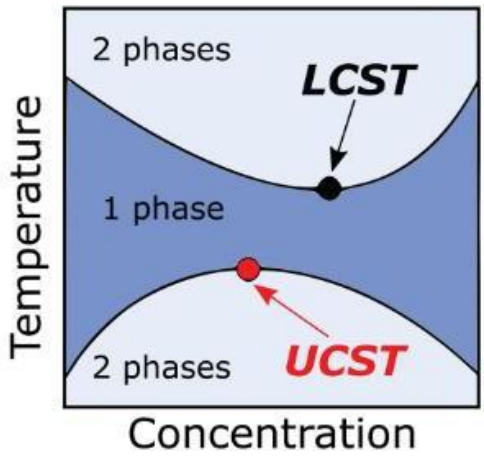


Figure 1-2 Schematic representation of a phase diagram depicting upper critical solution temperature UCST (red) and lower critical solution temperature LCST (black), that represent the maximum and minimum points of the solubility curves, respectively. Adapted from [2]**.**

Accordingly, in the phase diagram presented in **Figure 1-2**, the LCST and UCST represent the minimum and maximum points of the solubility curves, respectively. In respect to phase separation, a solute presenting LCST behaviour will undergo LLPS with the increase in temperature, and the opposite arises with those presenting a UCST profile [2].

The temperature dependence behaviour may be tuned by varying the pH of the solution, the polymer composition or ionic strength [12]. This effect translates directly to the perturbation of the existing solute-solute interactions, e.g. ionic strength largely affects solute-solute ionic interactions [13]. Therefore, the tandem analysis of the temperature dependence and the influence of certain conditions will provide insight on the LLPS driving forces.

The study of synthetic polymer has created the framework for the studies regarding biomolecules, namely proteins and nucleic acids [14]. However, given the complex nature of these molecules, the possible interactions these may perform are intrinsically complex, rendering the complete understanding of the LLPS process of these molecules an arduous and yet unfulfilled task.

1.1.1 Biomolecular LLPS

The elucidation of phase separation using synthetic polymers has paved the path to understanding LLPS of biopolymers, namely protein and RNA. Both types of polymers

have shown the ability to undergo LLPS, either in a mixture of compounds or when isolated [15]–[17].

The key feature to undergo LLPS is the presence of multivalent transient interactions [1],[18]. These are present in multiple folded binding domains and in short linear motifs (SLiMS), being the latter closely related to the presence of intrinsically disordered proteins or regions (IDPs and IDRs, respectively) [19]. Furthermore, the increase in affinity is correlated to the decrease in multivalency and concentration required to undergo LLPS [20]–[22].

A hierarchical organization of these types of interactions performed by proteins has been postulated. Ionic interactions allow long-range interactions, which subsequently promote directional short-range interactions, namely H-bonding, π-π and π-cation interactions [4]. Hence, charge clustering promotes LLPS in comparison to its random distribution [23],[24].

As previously discussed, entropy tends to hinder demixing. However, the enthalpic contribution originated from solute-solute interactions and the release of solvent molecules from the hydration shells of these proteins largely compensate for the entropic contribution. Furthermore, some proteins may adopt a less compact state upon phase separation, providing an additional entropic contribution [25].

IDPs lack a defined structure, presenting a heterogeneous ensemble of conformations, kinetically accessible at a moderate temperature [26]–[28]. However, IDRs may gain a transient folded state, namely kinked β-sheets [29]. These transiently ordered regions may provide a further enthalpic contribution to the phase separation phenomenon.

The LLPS propensity of proteins is sensitive to a variety of conditions. Depending on the driving forces of the process, the conditions will affect differently the LLPS [11],[30]. For example, ionic interactions are inherently sensitive to pH and ionic strength. Furthermore, as discussed in section **1.1**, the effect of temperature on the LLPS process is dependent on the driving force, i.e. H-bonds and hydrophobic interactions are highly temperature sensitive, whereas ionic, π-cation and π-π interactions are relatively insensitive [10],[31].

The LLPS of RNA molecules is also sensitive to several conditions. However, in contrast to what was previously described with proteins, when RNA molecules undergo LLPS in the absence of partner proteins, the process is insensitive to ionic strength, since this factor does not disrupt RNA-RNA interactions [15].

The phase separation process of isolated proteins, RNA and simple mixtures has been extensively studied and characterized. Nevertheless, the complexity increases immensely when considering intracellular conditions, and thus increases the difficulty of completely understanding the membraneless organelles found *in vivo*.

1.1.2 Membraneless organelles

The existence of membraneless organelles with liquid-like properties was first observed in 2009 by Brangwynne *et al.* pertaining to P granules, which are bodies composed of RNA and proteins found in embryos of *C. elegans* [32]. Thereafter, the existence of membraneless organelles has been proven to be ubiquitous [33].

These membraneless organelles have been found to form transiently in response to certain stimuli (e.g. stress granules, SGs) or in an essential house-keeping manner (e.g. nucleoli) [34]. These organelles are mainly composed of proteins and nucleic acids, whereas the variety of different proteins is dependent on the organelle, although it may reach thousands of distinct proteins in certain cases, namely IDPs and nucleic acid-binding proteins [35],[36].

The dynamics of these structures are inherently connected to their material properties/states. These may vary from droplets resembling liquids, hydrogels and solid aggregates. The liquid structures allow the continuous component exchange between the droplet and the surrounding solution, whereas solid aggregates encapsulate the components within the structure, preventing their release [37]. Hydrogel structures, however, allow the selective permeation of components, namely based on size and affinity towards the matrix of the organelles, e.g. nuclear pores [38]. Furthermore, certain organelles possess substructures with different material states, e.g. SGs, which possess an outer liquid layer and a stable gel-like core [39].

The spatial-temporal control of the formation of these structures is regulated through the presence and local concentration of biomolecules which drive the LLPS process. Specifically, biomolecules which present multivalent interaction sites, termed *scaffolds*, the drivers of the LLPS, such as fused in sarcoma (FUS), found in SGs and nuclear paraspeckles [35],[40], and those recruited into these organelles, the *clients* [41],[42]. The components and "lifespan" of these organelles control their material properties, since several entities have been observed to form initially liquid-structures and evolve into essential or pathological solid-structures [43],[44].

The physicochemical properties within membraneless organelles differ significantly from the bulk solution, and therefore, these may directly modulate the kinetic properties of enzymes and chemical reactions, or through the selective partition or exclusion of components [45]. However, as the material state of the organelles transition into solid state, chemical reactions performed by components within the structure are severely limited, resulting therefore in a selective "loss-of-function" of these components [37]. This process is essential in certain solid membraneless organelles, e.g. amyloid bodies, which are believed to be connected to cell dormancy [37],[46]. Nevertheless, it is also a mechanism for pathological solid aggregates, i.e. SGs after uncontrolled maturation which, would prevent their reversibility and sequestration of toxic components [47].

1.2 Fused in Sarcoma

Fused in sarcoma, a protein belonging to the FET family of proteins (FUS, EWS, TAF15), was first identified as a chimeric protein in myxoid liposarcomas, where the N-terminal of FUS was fused to the DNA-binding domain of transcription factors, such as CHOP [48].

FUS is capable of undergoing LLPS in various conditions, namely presenting a UCST, either isolated *in vitro* or *in vivo*. The droplets formed in this process exhibit liquid behavior, namely the ability to fuse and deform upon shearing [49].

1.2.1 Structural Complexity

Fused in sarcoma is composed of two folded domains and several IDRs. The protein contains an N- terminal low-complexity domain (LC); three arginine-glycine-glycine rich regions (RGG), spread throughout the structure; an RNA-recognition motif (RRM) and a zinc-finger domain (ZnF). Finally, it contains a C-terminal nuclear localization signal (NLS). A schematic

representation of the protein, identifying the previously mentioned regions, is presented in **Figure 1-3**.

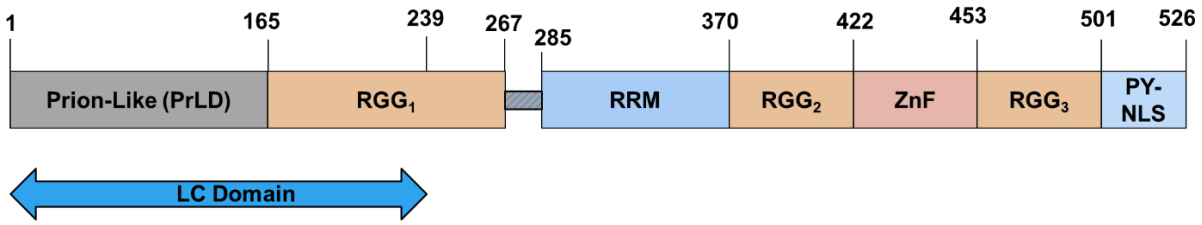


Figure 1-3 Schematic representation of FUS structural organization. The different domains/regions are indicated within the respective region and the last residue of each domain/region is indicated above. Residues 1-165 represent the prion-like domain (PrLD), 166-267 the first arginine-glycine-glycine domain (RGG_1), 286-370 an RNA-recognition motif (RRM), 371-422 the second arginine-glycine-glycine domain (RGG_2), 423-453 a zinc-finger domain (ZnF), 454-501 the third arginine-glycine-glycine domain (RGG_3), 502-526 a proline-tyrosine nuclear localization signal (PY-NLS). The low-complexity domain (LC domain) encompasses residues 1 through 239, i.e. the entire PrLD and a portion of RGG_1.

Despite of the various structural elements identified in FUS, each element plays a crucial role in the biochemical characteristics of the protein and in the performance of LLPS, directly or indirectly.

1.2.1.1 Low-Complexity Domain

The LC^{1-239} domain, is composed of a prion-like domain (PrLD) and a portion of the first RGG region (RGG_1). The LC domain is majorly composed of polar residues, i.e. serine, glutamine, glycine and tyrosine residues (observable in **Figure 1-4A**) [50]. The PrLD presents compositional similarities to known yeast prion proteins, presented in **Figure 1-4C**, observed using the PLAAC software [51].

The attention of various studies was drawn to this domain, having been shown the ability to undergo LLPS, even when isolated from the rest of the protein, displaying, however, a higher saturation concentration and an unchanged UCST profile. The LLPS process of this domain appears to be driven by π-π interaction between tyrosine residues and H-bonds through glutamine residues [20],[52]. The addition of charge to this domain in the scenario of phosphorylation impairs LLPS and aggregation [20],[50],[53].

The LLPS process has been described by means of stickers-and-spacers models. The stickers are residues which drive the phase separation process through multivalent interactions, and, in the case of the LC domain, the stickers are primarily tyrosine residues. The spacers, however, are residues which do not contribute with multivalent interactions and instead provide fluidity to the dense phase, hence mitigating liquid-solid phase transitions. In the present case of the LC domain, the glycine residues are the spacers [54].

The LC domain is able to undergo a liquid-to-solid transition into amyloid fibrils. The structure of these fibrils has been determined (presented in **Figure 1-4B**), with a clear, albeit relatively short, β-strand core while the rest of the domain remains dynamically disordered [55].

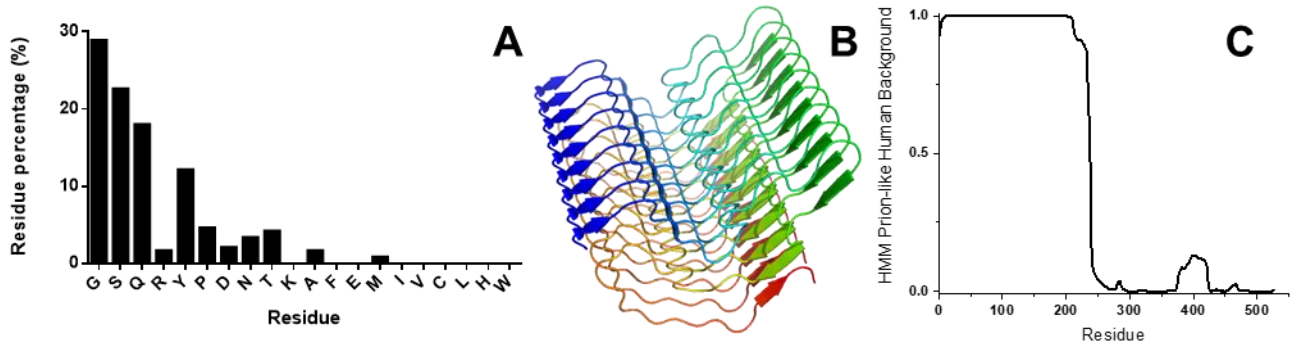


Figure 1-4 Amino acid distribution of LC domain (A), structural representation of FUS PrLD domain fibrils (B) and tendency throughout FUS sequence (C). The model is colored from blue to red, indicating the N and C terminal, respectively. Model obtained from PDB (PDB code: 5W3N [55]). The structure was elucidated through solid state NMR spectroscopy. Tendency to form prion-like structures elucidated through PLAAC program, using the Human background [51].

Typical amyloid fibrils form cross-β-sheets, that are energetically favorable and nearly irreversible. In the case of FUS however, and as was observed alongside other phase-separating proteins, the fibrils formed present kinked cross-β-sheets. These structures are less stable, and the energy for separation of these structures is in the order of thermal energy. Thus, the weak transient kinked cross-β-sheets might contribute to the LLPS process [29],[56].

1.2.1.2 Arginine-Glycine-Glycine Domains

Arginine-Glycine-Glycine (RGG) domains are arginine and glycine rich regions found in various proteins involved in the binding/processing of RNA [57]. Due to its lack of diversity in amino acids, these regions have been identified as generally disordered, although some cases have been reported of the existence of β-turns [57]. A total of three RGG regions are spread throughout FUS, respectively numbered based on their position in the primary structure.

Studies have indicated that these regions are crucial for the LLPS phenomenon, namely through the arginine residues. Indeed, studies have shown that mutating these arginine residues to lysine leads to a decrease in FUS self-assembly [58]. Such observations indicate that other factors, namely the favored participation of the delocalized π bonds in the arginine guanidinium group and additional possible H-bonds, dictate the effect of these regions in the capability of FUS to undergo LLPS [54],[59]. The arginine residues present in these regions contribute directly to the LLPS process through π-cation interactions, namely with the tyrosine residues present in the LC domain [60].

The RGG motifs have been reported to bind to RNA, with affinities depending on primary and/or secondary RNA structure [57],[60]. Furthermore, these motifs have been shown to mediate protein-protein interactions [57], an important factor to account for in the LLPS process *in vivo*, where a plethora of different proteins may be involved in the formation of droplets.

The hypermethylation of arginine residues may induce cytoplasmic FUS accumulation. This effect is due to the direct contacts performed between RGG_3 (the C-terminal RGG) and the associated importins [59],[61].

These domains are observable in many RNA-binding proteins, namely those who present the ability to undergo LLPS. Particularly, RNA helicase LAF-1, who is found in P granules, LAF-1 presents a single RGG domain. Nevertheless, it is essential for the LLPS process, since the elimination of this domain causes the inability of undergoing LLPS [22].

1.2.1.3 RNA-Recognition Motif

The RRM domain is commonly found in RNA-binding proteins of eukaryotes. The RRM domain of FUS contains significant sequence variations in respect to other studied RRM, it retains however, the characteristic structural similarities to other RRM domains [62]. As presented in **Figure 1-5**, the fold of the RRM consists of four β-strands and two perpendicular α-helices [62].

Figure 1-5 Structural representation of FUS RRM domain. The model is colored from blue to red, indicating the N and C terminal, respectively. Model obtained from PDB (PDB code: 2LCW [62]). The structure was elucidated through solution NMR spectroscopy. The characteristic “KK-loop” are represented in grey.

Being the main portion of FUS with a defined structure, the RRM domain has been extensively studied using circular dichroism (CD) and nuclear magnetic resonance (NMR) [63]. Results have indicated that the RRM domain may interact with other regions of the protein and presents the capability to spontaneously form amyloid fibrils [63].

FUS RRM domain presents a characteristic feature: it contains an extended “KK-loop” connecting α1 and β2, observable in **Figure 1-5**, which is a crucial component for the binding between FUS to both RNA and DNA [62].

A variety of RNA and DNA may bind to FUS RRM domain. The nucleic acid recognition is sequence and secondary structure dependent, resulting in a large binding affinity range towards these different molecules [64], although a higher prevalence has been observed for stem-loop RNAs [65]. Nevertheless, the binding affinity for RRM domain does not account for the observed binding affinity of the full-length protein, indicating the mediation by other regions. These regions are account for the previously mentioned RGG motifs and the subsequently addressed ZnF domain [48].

The RRM domain does not appear to directly promote FUS LLPS, however the replacement of this domain with a flexible linker significantly promotes FUS toxicity, and the replacement with a rigid linker appears to inhibit LLPS. Therefore, the flexibility of this domain, related to whether it is folded, could further control the LLPS tendency of FUS [66].

1.2.1.4 Zinc-Finger

Zinc-finger domains are short nucleic acid binding domains, whose structure and function are determined through the coordination of a Zn^{2+} ion. The classical Zn^{2+} ion coordination is achieved through two cysteine and two histidine residues [67]. The ZnF domains are capable of binding to RNA and DNA and are consequentially crucial components in transcriptional regulation [67].

Fused in sarcoma presents a small ZnF domain, which coordinates Zn^{2+} ions through a non-classical four cysteine binding motif. This domain adopts a disordered conformation in the absence of zinc, since the Zn^{2+} ion coordination drives the folding process [68].

Structural studies of this domain presented discrepancies regarding the presence of secondary structure. Based on NMR spectroscopy, these studies either identified solely the presence of β-sheet [65], α-helix [68], or the combination of the two in distinct regions of the domain [69].

Similarly to the previously addressed RRM domain, the ZnF domain is capable of binding to a variety of nucleic acids [64]. The ZnF domain presents however, a higher degree of selectivity, namely regarding the recognition of specific RNA sequences with increased affinity [65],[69]. The binding of the isolated domain has been previously studied, whereas the structure of the domain bound to a target RNA has been elucidated (with the following sequence: UGGUG) **Figure 1-6** [65].

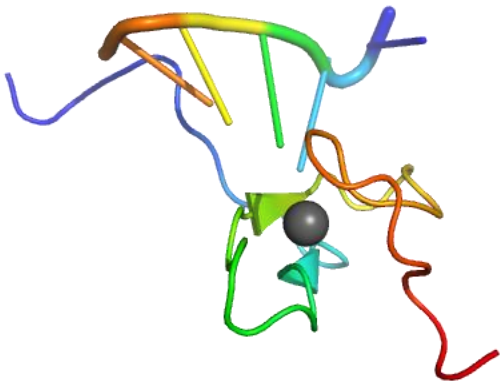


Figure 1-6 Structural representation of FUS ZnF domain bound to UGGUG RNA. The model is colored from blue to red, indicating the N and C terminal, respectively. Model obtained from PDB (PDB code: 6G99 [65]). The structure was elucidated through solution NMR spectroscopy. The Zn^{2+} ion is represented as a sphere. The UGGUG RNA is represented in cartoon.

1.2.2 LLPS of FUS

Fused in sarcoma has been immensely studied has a model for LLPS. Consequentially, the LLPS process has been thoroughly characterized, however, the technical difficulties regarding the *in vitro* study of FUS prevents the complete understanding of the LLPS phenomenon, several advances have, however, been accomplished.

FUS presents a UCST LLPS profile, therefore it undergoes LLPS at lower temperatures. Ionic interactions have been shown to play a crucial role in the LLPS process of FUS, as evidenced by the dependence on ionic strength and the mutation of glutamate residues of the

protein [70]–[72]. Moreover, previous studies have indicated that the main contributors for the LLPS process of FUS lies on the patterning and interaction of aromatic and arginine residues, where π-π and π-cation interactions are dominant driving forces [58],[73],[74]. Tyrosine residues are mainly present in the LC domain and arginine residues in the RGG regions. When these regions are within the same structure or mixed in the sample, a decrease of the saturation concentration required to undergo LLPS is observed [58].

Subsequently to LLPS, FUS droplets may undergo liquid-solid phase transition, where the dynamic nature of the droplets, a crucial *in vivo* feature, is lost. The abundance of glycine residues, as opposed to polar residues namely glutamine and serine residues, contributes to the increased dynamics of the droplets and prevents liquid to solid transition [58],[75].

Many *in vivo* membraneless organelles present RNA in their composition. Therefore, the presence of two nucleic acid binding domains in FUS, implicates the crucial role of RNA in the regulation of the LLPS phenomenon of the protein. Indeed, it has been observed the ability of RNA to induce LLPS at lower concentrations and prevent at higher concentrations [76],[77]. This tendency directly correlates with the higher LLPS propensity of FUS in the cytoplasm and diminished in the nucleus [77].

The *in vitro* simulation of crowding conditions through the addition of crowders such as polyethylene glycol (PEG) increases the LLPS propensity. However, the addition of protein crowders, such as bovine serum albumin (BSA), decreased the LLPS propensity through nonspecific interactions and competition with FUS [78],[79].

1.2.3 Biological and Pathological Pertinence

FUS is primarily present in the nuclei, although small quantities may be found in the cytosol. Its transport is mediated through the binding of its C-terminal NLS to transportin protein. The localization of FUS is dependent on arginine methylation (namely of RGG_3, see **Figure 1-3**) and point disease related mutations [80]. These mutations and arginine hypermethylation may result in the accumulation of FUS in the cytoplasm.

Through the ability to bind to RNA and DNA, FUS exhibits the capability to impact gene expression in several different steps. Starting with the transcription, reports have shown that FUS is able to bind to promoters of various genes and further interact with RNA polymerase II [48],[81]. The transport of mRNA from the nucleus to the cytoplasm is mediated by the binding and subsequent nucleus-cytoplasm shuttle of FUS. Furthermore, FUS affects mRNA stability by binding to 3'UTR, thus preventing the deadenylating of these transcripts, resulting in the increase of the stability of the target transcripts [80].

FUS plays an important role in mRNA processing, namely in the splicing step. The protein may either associate to components of the spliceosome or directly bind to the transcript, such is the case of mRNA encoding for FUS, thus exhibiting an autoregulatory mechanism. FUS contributes to the regulation of the translation of transcripts, namely through the interaction with ribosomal subunits and translation factors [80],[82].

In the event of DNA damage, FUS is readily localized in the vicinity of the damaged site. This is achieved through intermediate action of poly adenosine diphosphate ribose polymerase, which binds to sites of double-strand breaks. The newly polymerized poly adenosine diphosphate

ribose (PAR) chain is recognized by FUS, which in turn recruits several other proteins responsible for the DNA repair [48].

Upon the induction of a certain stress, such as osmotic or heat shock, cells may form stress granules. However, the effect of FUS on the biophysical characteristics of SGs is dependent on the abundance in the cytoplasm. As a result, FUS mutants have shown to cause drastic effects on the dynamics of these stress granules, namely the destabilization and progressive hardening of these structures [83].

The hardening of SGs is believed to cause the aggregation of proteins present in these organelles [34]. This aggregation is further accelerated by mutations related to neurodegenerative diseases, namely amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) [75]. Mutations in the FUS gene account for 5 % of inherited ALS [80]. Therefore, the clear correlation between FUS mutations and the irregular accumulation in the cytoplasm could be a cause of certain pathological conditions.

1.3 Techniques used in the study of LLPS

Membraneless organelles are dynamic bodies whose apparent characteristics derive from the molecular organization from a larger scale of several μm, to the atomic meticulousness at the Å level. Therefore, a multitude of techniques have allowed the biophysical study of the entities, with varying limitations and complementary advantages [84].

Routinely, one of the preliminary methods used in the *in vitro* study of LLPS is the performance of turbidity assays [85]. The technique relies on the measurement of turbidity, arising from the LLPS phenomenon, using wavelengths which are not absorbed by any component of the studied solution. The simplicity of the method provides an opportunity for the screening of different conditions and the ability to inquire on the concentration threshold of the studied molecules [81],[86]. Nevertheless, the perceived turbidity may arise from different phenomena, i.e. protein aggregation. Thus, turbidity assays are commonly complemented by imaging techniques.

Light and fluorescence microscopy are imaging techniques commonly used for the *in vitro* and *in vivo* observation of membraneless organelles [87]. The techniques allow for the deduction of the droplet properties, namely through the direct observation of fission and fusion events. Furthermore, the techniques allow the distinction between sub compartments of droplets and estimation of the concentration of specific components within droplets [42]. These feats are achieved through the selective fluorescent labeling of components [88]. Nevertheless, the requirement for the addition of fluorophores to the sample may directly influence the LLPS phenomenon [84].

Low-resolution techniques have been used to describe the overall size distribution of the molecules in relevant conditions, namely small-angle x-ray scattering (SAXS). SAXS is an increasingly used technique that allows the study of conformational changes within molecules, namely the compaction state, binding and agglomeration events [89]. However, to fully understand the observed LLPS phenomenon, an atomic resolution of the system may be required, and due to the intrinsic mobility of these components, NMR has risen as the leading biophysical technique to study LLPS pertaining components [90].

NMR spectroscopy arises from the interaction between nuclei and an external magnetic field, the nuclear Zeeman splitting. As a result, in NMR, the magnetically active nuclei (nuclei with

a nuclear magnetic moment different than 0) act as probes in this spectroscopy. As the nuclei interact with the external magnetic field, their magnetic moments will resonate at defined frequencies. However, the frequency at which these resonate is dependent on the chemical environment of the corresponding nucleus, providing direct information regarding its surroundings. Consequently, NMR may be used to provide insight regarding secondary and tertiary structure of the components and directly pinpoint residues participating in crucial interactions [81],[91]. The latter may be uncovered through the observation of a change in resonating frequency (the chemical shift), which may uncover the affected residues, or through the identification of the residues which are in close proximity(<5 Å), by means of the nuclear Overhauser effect (NOE) [92] among other NMR parameters.

1.3.1 NMR studies of FUS

The employment of NMR spectroscopy of protein LLPS provides insight at a residue or atomic level, having been prior used in studies regarding the LLPS of FUS. However, due to its structural complexity and size, namely the presence of various IDRs and over 500 residues, hinders the use of standard protein NMR spectroscopy techniques in the study of FUS. This results in an incomplete spectral assignment and the inability to observe the expected number of resonances [60],[63]. Due to this, the use of NMR spectroscopy has been mainly focused on using sections of the full-length protein. Nevertheless, the effect of ATP and RNA on the full-length FUS, using NMR spectroscopy, have been reported [68].

The study of the PrLD using NMR spectroscopy has provided insight regarding the molecular interactions underlying the LLPS of the domain [52], the potential phosphorylation sites and direct structural observation within the macromolecule-dense phase [70]. Through solid state NMR spectroscopy, the structure of the fibril core formed by the LC domain was determined [93].

The residues participating in the RNA binding of the RRM and ZnF domains have been identified through NMR spectroscopy, where the solution structures of these complexes were described [65],[69]. Furthermore, NMR spectroscopy provided direct structural insight regarding the reversibility of the thermal denaturation of the isolated RRM domain [63].

2 Objectives

FUS has significant implications in biological and pathological situations, and the propensity to phase separate and its LLPS process have been immensely studied. However, the structural complexity of FUS prevents a detailed description of the LLPS phenomenon of the protein, i.e. the degree of participation of the different regions of the molecule.

Therefore, smaller portions of the molecule are usually studied in order to truly understand the role of specific regions on the LLPS process. In this context, extensive studies have been reported regarding the LC domain of FUS. However, little information exists regarding the role of the remainder of the molecule.

To study the role of different regions on the LLPS process, the identification of regions self-sufficient to undergo LLPS is crucial. Furthermore, the effect of different conditions and the underlying molecular drivers must be identified.

The objectives of this study were to scrutinize the role of different regions of FUS on the inherent LLPS phenomenon. The goal will be achieved by studying distinct regions of the molecule and the LLPS propensity associated to these. The previously unstudied regions focused herein are the folded RRM and ZnF domains each attached to the adjacent RGG regions, namely RGG_1-RRM-RGG_2 and RGG_2-ZnF-RGG_3. The LLPS phenomenon of the distinct regions will be further studied, namely by verifying the effects of different conditions on the phenomenon and the molecular drivers underlying it, through the performance of turbidity assays and NMR spectroscopy.

3 Methods

3.1 Constructs Used in This Study

The full-length FUS protein was divided in two distinct constructs, the RRM domain attached to the adjacent RGG regions (RGG_1-RRM-RGG_2) and lastly, the ZnF domain attached to the adjacent RGG regions (RGG_2-ZnF-RGG_3). The two protein constructs were fused to a solubility tag (maltose-binding protein, MBP), an N-terminal 6xHis tail and a *Tobacco Etch Virus* protease (TEV) cleavage site. The constructs were inserted into pHTP1 vectors (NZYTech, **Appendix 1**). The constructs are schematically presented in **Figure 3-1**.

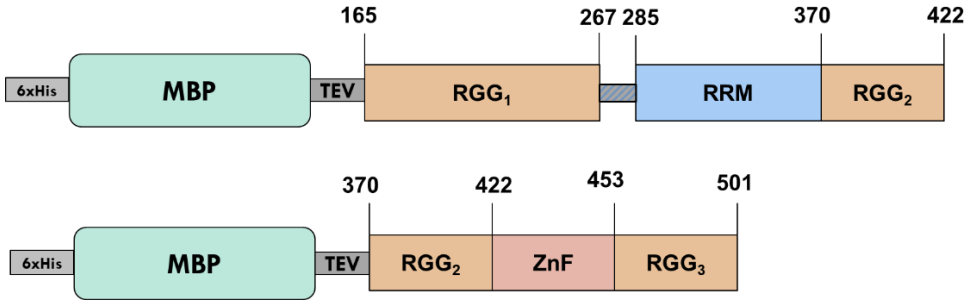


Figure 3-1 Schematic representation of the constructs used in the study. A 6xHistidine tag, MBP and TEV cleavage site were fused to the N-terminal of the protein.

3.2 Protein Expression Optimization

The two constructs herein were expressed in competent *E.coli* BL21(DE3) cells. The competent cells were treated with $CaCl_2$ and transformation was performed through thermal shock.

The temperature, induction time and inductor (Isopropyl β-d-1-thiogalactopyranoside, IPTG, NZYTech) concentration were optimized for the 2 constructs of this study. Three conditions of each variable were tested, temperature: 25 °C, 30 °C and 37 °C; time after induction: 4 hours, 8 hours, 16 hours; IPTG concentration: 0.1 mM, 0.5 mM and 1 mM. The OD_{600} value at which induction occurred remained constant between 0.6 – 0.8. The cell cultures were grown in Luria-Bertani (LB) medium.

Protein expression in each condition was verified through sodium dodecyl sulphate polyacrylamide gel (SDS-PAGE), using Coomassie blue staining. The analysis of the gels allowed the relative quantification of protein expression. Gel analysis was conducted using GelAnalyzer 19.1 [94]. NZYColour protein marker II or low molecular weight protein marker II (NZY Tech) was used in all the SDS-PAGE in this study.

3.3 Expression and Purification of RGG_1-RRM-RGG_2

For the expression of unlabeled protein, the transformed cells were grown in LB medium, supplemented with 2 g/L glucose to inhibit endogenous MBP expression [95], at the optimal temperature of 37 ºC and continuous agitation at 180 rpm, in the presence of 50 µg/mL of kanamycin (NZY Tech). For the expression of isotopically enriched protein, the cells were grown

in minimal 2xM9 medium [96], full composition enlisted in **Appendix 2,** supplemented with 2 g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) for ^{15}N enrichment and 3 g/L ^{13}C glucose (Cambridge Isotope Laboratories) for ^{13}C enrichment.

Protein expression was induced upon the addition of IPTG to a final concentration of 1 mM, at an OD_{600}=0.6-0.8. Upon induction, the cells grew for 4 h at 37 ºC, with continuous agitation of 180 rpm.

The purification procedure of the construct was based on the previously reported purification of FUS RRM [65].

The cells were subsequently harvested through centrifugation at 7,477 x*g* for 15 min (Beckman Coulter centrifuge). The pellet was resuspended in lysis buffer, which consisted of: 50 mM Tris-HCl buffer, 1 M NaCl (Fisher Chemical), 50 mM glycine (Fisher Bioreagents), 10 mM imidazole (PanReac AppliChem), 2 mM β-mercaptoethanol (Sigma Aldrich), 2 mM benzamidine (Fisher Bioreagents) and 0.1 mM phenylmethylsulphonyl fluoride (PMSF, PanReac AppliChem), pH 8.0. Cell lysis was performed through sonication (80 % amplitude, 0.5 cycle 1 s on/1 s off for 10 min, Hielscher). Lysate was cleared by centrifugation at 48,384 x*g* for 30 min.

The MBP-RGG_1-RRM-RGG_2 fusion protein was purified through immobilized metal affinity chromatography (IMAC), using Ni-NTA HiTrap FF Crude (GE Healthcare). The binding buffer consisted of the lysis buffer, whilst the elution buffer had the addition of 500 mM of imidazole (PanReac AppliChem). After sample injection, the column was washed with binding buffer and with 8 % of elution buffer (total imidazole concentration of 40 mM). The elution of the protein was performed through a linear gradient of elution buffer (0-100 %).

Sample buffer was exchanged using HiTrap desalting columns (GE Healthcare) or dialyzed using SnakeSkin dialysis tubing (Thermo Scientific), against a buffer consisting of 50 mM Tris-HCl, 200 mM NaCl, 50 mM glycine, 2 mM β-mercaptoethanol and 0.05 % NaN_3 (PanReac AppliChem) pH 8.0. Cleavage of the fusion protein was performed overnight at room temperature, with the addition of in-house produced TEV protease in a proportion of 1 mg of enzyme/ 2 L of initial culture (TEV production was based on [97]).

A second IMAC step was performed to isolate the cleaved product from the intact fusion protein, TEV protease (containing 6xHis tag) and the cleaved MBP. The elution buffer used was identical to that of the first IMAC step, without protease inhibitors, i.e. PMSF and benzamidine. The protein was collected in the flow-through.

The sample was dialyzed against 20 mM Tris-HCl, 100 mM NaCl, 2 mM β-mercaptoethanol and 0.05 % NaN_3 at pH 7.0. Samples were concentrated using Vivaspin 10 KDa MWCO concentrators (Sartorius), using a temperature-controlled centrifuge (Eppendorf). The temperature was kept at 30 °C during the concentration.

Sample purity was evaluated through SDS-PAGE. The presence of nucleic acids was evaluated through the ratio of absorbance at 260 and 280 nm, with a ratio between 0.6-0.8 indicating the absence of nuclei acids in the sample [98]. Protein was quantified through the measurement of absorbance at 280 nm and using the theoretical extinction coefficient predicted using Protparam Expasy tool [99]. The yield was 6.5 mg and 4 mg per liter of culture for unlabeled and $^{15}N/^{13}C$ RGG_1-RRM-RGG_2, respectively

3.4 Expression and Purification of RGG_2-ZnF-RGG_3

The cell growth of unlabeled and $^{15}N/^{13}C$ RGG_2-ZnF-RGG_3 followed similar protocol of the RGG1-RRM-RGG2 domain. Protein expression was induced upon the addition of IPTG to a final concentration of 0.5 mM, at an OD_{600}=0.6-0.8. Upon induction, the cell culture incubated for 8 h at 30 ºC, with continuous agitation of 180 rpm.

The cells were subsequently harvested through centrifugation at 7,477 x*g* for 15 min. The pellet was resuspended in lysis buffer, which consisted of 50 mM Tris-HCl buffer, 500 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol, 0.1 mM $ZnCl_2$, 2 mM benzamidine and 0.1 mM PMSF, pH 8.0. Cell lysis was performed through sonication (80 % amplitude, 0.5 cycle 1 s on/1 s off for 10 min). Lysate was cleared by centrifugation at 48,384 x*g* for 30 min.

The MBP-RGG_2-ZnF-RGG_3 fusion protein was purified through IMAC, using Ni-NTA HiTrap FF Crude. The binding buffer consisted of the lysis buffer, whilst the elution buffer had the addition of 500 mM of imidazole. The elution of the protein was performed through a linear gradient of elution buffer (0-100 %).

The pooled samples were dialyzed using 3.5 KDa MWCO SnakeSkin dialysis tubing (thermo scientific), against a buffer consisting of 50 mM Tris-HCl, 1 M NaCl and 0.05 % NaN_3 pH 7.4.

Cleavage of the fusion protein was performed overnight at room temperature, with the addition of in-house produced TEV (tobacco etch virus) protease in a proportion of 1 mg of enzyme/ 2 L of initial culture.

A second IMAC step was performed to isolate the cleaved product from the intact fusion protein, TEV protease (containing 6xHis tag) and the cleaved MBP. The elution buffer used was identical to that of the first IMAC step, apart from protease inhibitors, i.e. benzamidine and PMSF. The protein was collected in the flow-through.

The sample was dialyzed against 50 mM Tris-HCl buffer, 100 mM NaCl, 5 µM $ZnCl_2$, 0.05% NaN_3 pH 7.1. Samples were concentrated using Vivaspin 3.5 KDa MWCO concentrators (Sartorius), using a temperature-controlled centrifuge.

Sample purity was evaluated through SDS-PAGE. The presence of nucleic acids was evaluated through the ratio of absorbance at 260 and 280 nm, a ratio within the interval of 0.6-0.8 indicated the absence of nuclei acids in the sample [98]. Protein was quantified through the measurement of absorbance at 280 nm and using the theoretical extinction coefficient predicted using Protparam Expasy tool [99]. Total $^{15}N/^{13}C$ enriched RGG_2-ZnF-RGG_3 yield of 0.5 mg was obtained per liter of culture.

3.5 Turbidity Assays

For each tested condition, three independent samples were prepared and the OD_{600} was measured at room temperature and after defined periods of time incubating on ice. The pH of each sample was measured and titrated to the desired value prior to measurements. OD_{600} measurements were conducted in a Shimadzu UV-1280 spectrophotometer, using a Hellma Analytics 10 mm path length quartz cell.

Unless stated otherwise, the turbidity measurements were performed in samples containing 100 μM RGG_1-RRM-RGG_2, 20 mM Tris-HCl, 100 mM NaCl, 2 mM β-mercaptoethanol and 0.05 % NaN_3, pH 6.4. The tested concentrations of RGG_1-RRM-RGG_2 range from 0 to 400 μM. A range between 5.3 – 9.8 pH values were tested, and the concentration of NaCl ranged from 50 to 500 mM.

Protein:RNA mass ratios up to 1:2 of type VI RNA from torula yeast (Sigma-Aldrich) were tested.

Turbidity measurements in the presence of the following amino acids was performed: sodium glutamate (Riedel de Haen), lysine hydrochloride (Alfa Aeser), arginine hydrochloride (Alfa Aeser). The tested concentration of each amino ranged from 1 mM to 5 mM, which is equivalent to 1:10 and 1:50 in protein:amino acid molar ratios.

All measurements were performed in triplicate and the errors presented represent the standard deviation from the measurements.

3.6 NMR Experiments

3.6.1 RGG_1-RRM-RGG_2 NMR experiments

Unless stated otherwise, all ^1H, ^{15}N HSQC spectra were acquired in a Bruker Avance Ultrashield Plus 600 MHz spectrometer equipped with 5 mm inverse detection triple-resonance z-gradient cryogenic probe head (CP TCI) and a gradient unit capable of producing magnetic field pulsed gradients in the z direction of 56.0 $G.cm^{-1}$ operating at 600.13 MHz for ^1H, 150.94 MHz for ^{13}C and 60.82 MHz for ^{15}N.

^1H, ^{15}N HSQC spectra were acquired with a matrix of 2048 x 128 points with 8, 32 or 64 scans in a spectral window of 8417.509 Hz (centered at 2806.93 Hz) x 2311.070 Hz (centered at 7175.66 Hz), representing the sweep width of ^1H and ^{15}N, respectively. Standard Bruker hsqcetf3gpsi2 experiment was used to acquire the presented ^1H, ^{15}N HSQC spectra.

3.6.1.1 ^1H, ^{15}N HSQC Temperature Dependence

The experiments were performed in either 800 μM samples of $^{15}N/^{13}C$ enriched RGG_1-RRM-RGG_2 containing 20 mM Tris-HCl buffer, 100 mM NaCl, 2 mM β-mercaptoethanol, 0.05 % NaN_3, 50 μM sodium trimethylsilylpropanesulfonate (DSS) and 10 % D_2O pH 7.0. The spectra were acquired at 5 °C, 15 °C, 25 °C, 30 °C and 37 °C.

3.6.1.2 ^1H, ^{15}N HSQC RGG_1-RRM-RGG_2 Concentration Dependence

^1H, ^{15}N HSQC spectra were acquired using 400 μM samples of $^{15}N/^{13}C$ enriched RGG_1-RRM-RGG_2 containing 20 mM Tris-HCl buffer, 100 mM NaCl, 2 mM β-mercaptoethanol, 0.05 % NaN_3, 50 μM DSS and 10 % D_2O, pH 7.0. The spectra were acquired at 30 °C.

3.6.2 Preliminary RGG_2-ZnF-RGG_3 Studies

3.6.2.1 ^1H, ^{15}N-HSQC Temperature Dependence

All ^1H, ^{15}N HSQC spectra were acquired in a Bruker Avance Ascend 500 MHz Spectrometer equipped with 5 mm inverse detection triple-resonance z-gradient cryogenic probe, head (CP TCI) and a gradient unit capable of producing magnetic field pulsed

gradients in the z direction of 56.0 G.cm-1 operating at 500.34 MHz for ^{1}H, 125.83 MHz for ^{13}C and 50.70 MHz for ^{15}N.

^{1}H, ^{15}N HSQC spectra were acquired with a matrix of 2048 x 128 points with 32 scans in a spectral window of 6578.947 Hz (centered at 2351.60 Hz) x 1876.085 (centered at 5982.48 Hz), representing the sweep width of ^{1}H and ^{15}N, respectively. Standard Bruker hsqcetfpf3gpsi2 experiment was used to acquire the presented ^{1}H, ^{15}N HSQC spectra.

The experiments were performed in a 100 µM ^{15}N and ^{13}C enriched RGG_2-ZnF-RGG_3 sample containing 50 mM Tris-HCl buffer, 100 mM NaCl, 5 µM $ZnCl_2$, 0.05% NaN_3, 50 µM DSS and 10% D_2O, pH 7.1. The spectra were acquired at 5 °C, 15 °C, 25 °C and 37 °C.

3.6.3 Data Analysis

All spectra were processed in Topspin 4.0.7. Unless stated otherwise, chemical shifts were tracked using CCPN software [100]. Resonance peak integration, in order to extract peak intensity and line width, was performed using PINT software [101],[102]. Standard errors of the individual parameters were estimated using jackknife resampling, available in PINT. The ^{1}H linewidth of H_2O and DSS resonances and respective standard error were extracted by performing gaussian resonance fitting using Topspin 4.0.7.

The RGG_1-RRM-RGG_2 ^{1}H-^{15}N HSQC relative resonance intensities were obtained through the ratio between a specified condition and a reference condition. Unless specified otherwise, the following condition was used as the reference: 800 µM ^{15}N/^{13}C RGG_1-RRM-RGG_2, 20 mM Tris-HCl, 100 mM NaCl, 2 mM β-mercaptoethanol, 0.05 % NaN_3, 50 µM DSS, 10% D_2O, pH 7.0, 30 °C prior to cooling. The standard errors for the respective ratios were estimated based on error propagation.

Amide $^{1}H_N$ temperature coefficients were estimated by performing linear regression of amide protons chemical shifts as a function of temperature and extracting the resulting slope [103]. Slope and standard error of the linear regression were extracted using the LINEST function of Microsoft Excel. Amide $^{1}H_N$ temperature coefficients estimated for RGG_2-ZnF-RGG_3 were based on spectra acquired at 37, 25, 15 and 5 °C.

Chemical shift perturbation (CSP) was calculated from the following equation [104]:

$$CSP\ (ppm) = \sqrt{\frac{(\Delta\delta\ ^{1}H)^2 + (0.14 \times \Delta\delta^{15}N)^2}{2}}$$

Whereas Δδ represents the difference in chemical shift between two conditions for the specific nuclei. Due to the difference in gyromagnetic ratios between ^{1}H and ^{15}N, the chemical shift difference associated to ^{15}N was multiplied by a weighing factor of 0.14.

4 Results and Discussion

4.1 LLPS Study of the RGG_1-RRM-RGG_2 Construct

4.1.1 Concentration and Temperature Dependence

To investigate the propensity for this construct to undergo LLPS, turbidity measurements were performed at varying protein concentrations **Figure 4-1**. The buffer conditions for the first assays remained unaltered. Furthermore, the temperature effect on LLPS was addressed through the turbidity measurement at room temperature (25 °C) and after incubation in ice for specific periods.

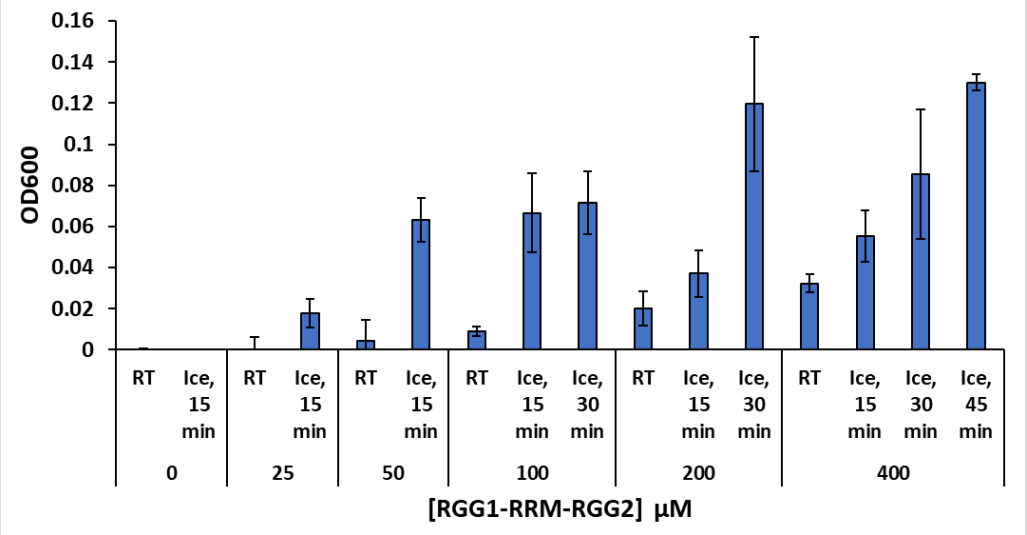


Figure 4-1 Turbidity measurements (OD 600 nm) of RGG_1-RRM-RRG_2 samples at varying concentration. Turbidity measurements were performed at room temperature (RT) and after incubation on ice for specified periods of time. Each condition was measured in triplicate and error bars represent the standard deviation.

The formation of protein-dense droplets causes an increase in sample turbidity, readily measured through OD. The unequivocal confirmation of LLPS is through the direct observation through optical microscopy, which was not performed in this study, and therefore the unequivocal confirmation was not performed [84]. Nevertheless, previous studies using the full-length FUS, provided a direct microscopic observation of protein-dense droplets, whereas the OD_{600} measurements did not reach values of 0.1 [70].

At room temperature, an increase in turbidity was observed only at protein concentrations above 100 µM, which is likely attributable to LLPS, which is expected to present a severe concentration dependence.

Furthermore, a significant increase in turbidity was observed upon sample incubation on ice. This increase in turbidity is likely linked to a temperature dependent LLPS process. Due to the prominent increase in turbidity at lower temperature, the LLPS appears to present an UCST profile.

Appreciable temperature induced LLPS was observed starting at RGG_1-RRM-RGG_2 concentrations of 50-100 µM. These values are comparable to those obtained with an isolated PrLD [58], believed to be a major component for the LLPS process of FUS.

The kinetics of the LLPS process appears to be RGG_1-RRM-RGG_2 concentration dependent. An apparent decrease in LLPS rate was observed at higher protein concentrations.

To examine the influence of different sample conditions on the LLPS process, the concentration of RGG_1-RRM-RGG_2 was set at 100 μM. At this concentration, significant turbidity was observable upon incubation on ice and the turbidity did not increase upon incubation for longer periods of time, suggesting that an equilibrium was achieved.

4.1.2 pH and Ionic Strength Dependence

The influence of sample pH on the LLPS was investigated. A total of 4 pH values were tested **Figure 4-2**. These were selected based on the theoretical isoelectric point (pI) of the isolated RRM domain of 5.3, and of the studied construct of 9.8. Furthermore, the physiological pH value of 7.4 was tested as well as the intermediate value of 6.4.

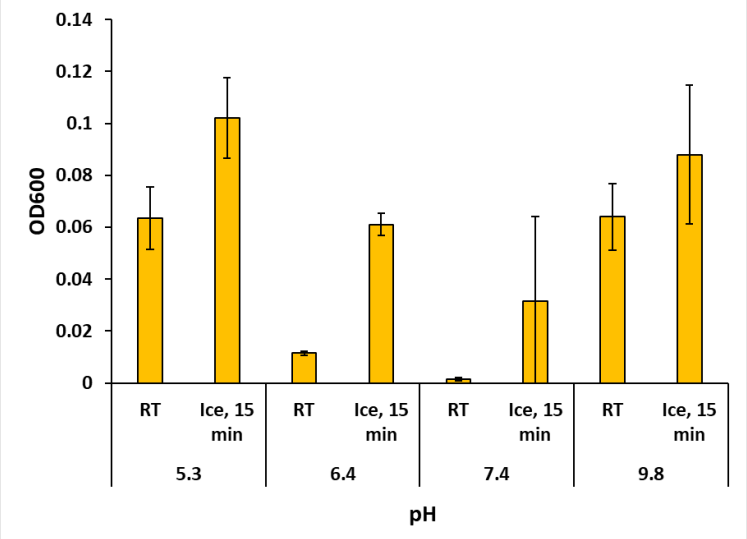


Figure 4-2 Turbidity measurements (OD 600 nm) of RGG_1-RRM-RRG_2 samples at 100 μM, at varying pH values. Turbidity measurements were performed at room temperature (RT) and after incubation on ice for 15 minutes. Each condition was measured in triplicate and error bars represent the standard deviation.

A clear tendency is observable, at both temperatures. The LLPS is favored in the most acidic and basic tested pH. These pH values coincide with the pI of the isolated RRM domain (5.3) and of the tested construct (9.8), suggesting that both might be implicated in the LLPS process.

The LLPS process of IDPs and IDRs is inherently correlated to the spatial overall conformation the molecules tend to adopt in each condition. Generally, these unstructured proteins tend to adopt extended and flexible conformations or collapsed conformations, driven by intramolecular interactions. Under the conditions where these adopt collapsed conformations, the local protein density may increase, due to the decrease in hydrodynamic ratio of the molecules, thus promoting intermolecular protein-protein interactions and consequently promoting LLPS [11]. The theory may be easily extended to polyampholyte proteins, namely proteins which present a significant number of anionic and cationic residues, and the apparent influence of pH on the LLPS process. The studied construct RGG_1-RRM-RGG_2, may be treated as polyampholyte, due to prevalence of positive charges in the RGG regions and negative charges in the RRM domain.

At a pH of 9.8, the overall construct presents a neutral charge. Under these conditions the protein chains tend to present a collapsed state, namely due to the decrease in intramolecular electrostatic repulsions from the RGG regions. The collapsed would inherently promote protein-protein interactions and cause the observed promotion of LLPS [11].

The effect of acidic pH on the LLPS process entails a direct contribution of the RRM domain on the phenomenon. At a pH of 5.3, the RRM domain presents an overall neutral charge. Under these conditions, the RGG regions present an overall positive charge. Therefore, the direct contribution of electrostatic repulsions regarding the RGG regions should remain unaltered. Furthermore, the overall attractive ionic interactions between the RRM domain and the RGG regions should be diminished under these conditions. Therefore, the observed increase in LLPS propensity at pH 5.3 indicates direct contributions of electrostatic interactions from the RRM domain. The RRM domain should play a crucial role with repulsive electrostatic interactions at a higher pH, as observed with the decrease in LLPS at pH 6.4 and 7.4.

The influence of ionic strength on the LLPS propensity allows to directly assess the contribution of electrostatic interactions on the LLPS process. Therefore, turbidity measurements were performed using varying concentrations of NaCl at a constant pH of 6.4 **Figure 4-3**. The use of NaCl to study the effect of ionic strength allows the assurance that the observed effect lies solely from the effect on electrostatic interactions, and not from any salting-in or salting-out effects, since according to the Hofmeister series both Na^+ and Cl^- are considered to have a negligible effect in this regard (at the concentrations used in this study) [105].

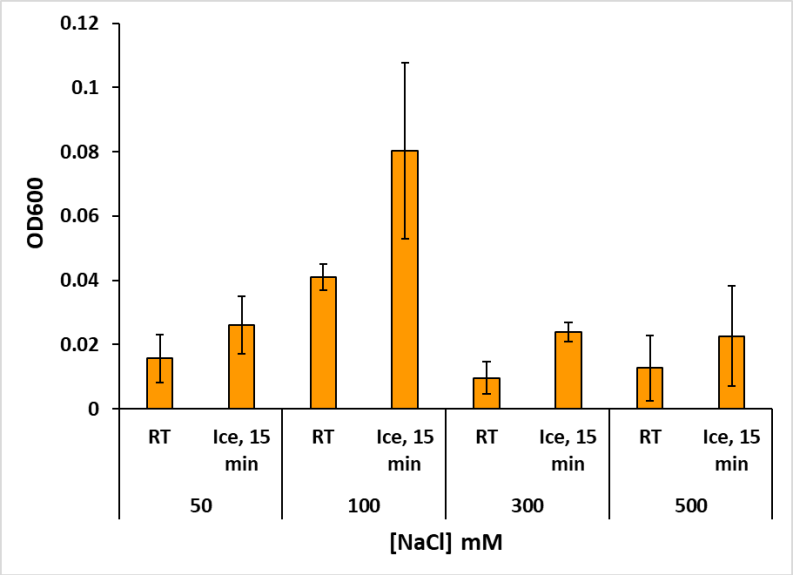


Figure 4-3 Turbidity measurements (OD 600 nm) of RGG_1-RRM-RRG_2 samples at 100 µM, at varying concentrations of NaCl. Turbidity measurements were performed at room temperature (RT) and after incubation on ice for 15 minutes. Each condition was measured in triplicate and error bars represent the standard deviation.

The increase in the concentration of NaCl from 50 mM to 100 mM caused a significant increase in turbidity, more than a 2-fold increase. The turbidity however showed a significant decrease when the concentration of NaCl was increased to 300 mM and remained relatively unaltered upon further increasing to 500 mM.

The increase in turbidity and consequently, LLPS propensity, at 100 mM NaCl entails a possible disruption of electrostatic repulsions, which would be inhibiting LLPS. By nullifying these electrostatic repulsions, further interactions are likely to occur, thus promoting LLPS. Furthermore, the decrease in LLPS at higher NaCl concentrations indicate the disruption of electrostatic attractions, due to the replacement of the ionic pair present in these interactions with either Na^+ or Cl^- abundantly present in solution. The results indicate the easier disruption of electrostatic repulsions, which would then promote stronger electrostatic attractions (lower NaCl sensitivity), which appear to be a major driving force for the observed LLPS.

The lack of any observable increases in turbidity, at the maximum tested range of ionic strength, may indicate the relative insignificance of hydrophobic interactions in the LLPS process of the construct [30],[106].

A compositional analysis of the RGG_1-RRM-RGG_2 sequence reveals a significant number of basic and acid amino acids (a total of 30 and 22, respectively) **Appendix 7A**. Therefore, a significant number of possible ionic interaction sites are present in the construct. Half of the acid residues, however, are located in the RRM domain, where one third of the total basic residues are present in this domain, i.e. lysine residues. This feature entails that if ionic interactions are indeed a primary driving force for LLPS, as the previously presented results indicate, the RRM domain may contribute directly for the LLPS phenomenon of FUS.

Possible driving forces remain to be addressed. Namely H-bonds through polar residues and π-cation interactions between arginine and tyrosine residues. A total of 56 polar residues are present in RGG_1-RRM-RGG_2. Accordingly, LLPS processes driven by H-bond present an UCST profile, therefore no evidence is presented as to disregard the possible driving force, as previously postulated for the PrLD [52]. Lastly, a total of 20 arginine and 8 tyrosine residues are present in RGG_1-RRM-RGG_2, which likely further contribute for the LLPS process. Although, since the possible interaction sites are less abundant in this construct, it may only act synergistically with the previously discussed driving forces.

Previous reports indicated that the RRM domain does not affect the LLPS process when bound to PrLD-RGG_1 [68]. Furthermore, it did not promote LLPS when bound to RGG_2-ZnF-RGG_3-NLS, which did not undergo LLPS when isolated in the study [68]. These observations may indicate that the RRM domain must be bound to both RGG_1 and RGG_2 in order to influence the LLPS.

No effect on the *in vivo* toxicity of FUS was observed upon deletion of the RRM [66]. However, if the RRM domain directly influences LLPS, then it is possible it may influence the toxicity of FUS.

4.1.3 RNA Dependence

The RRM domain is known to be able to bind to RNA, assisted by the adjacent RGG regions. Regarding full-length FUS, the RNA binding causes the protein to undergo significant conformational changes, namely the adoption of an extended conformation [107]. Furthermore, the RNA binding and concentration causes a pronounced effect on the LLPS process. At lower concentrations of RNA, the LLPS of FUS is promoted and inhibited at higher concentrations of

RNA [77]. Therefore, it is expected that the LLPS of RGG_1-RRM-RGG_2 may also be affected by the presence and binding of RNA.

To determine the dependence of RNA on the LLPS phenomenon of RGG_1-RRM-RGG_2, turbidity measurements were performed at various protein:RNA mass ratios **Figure 4-4**.

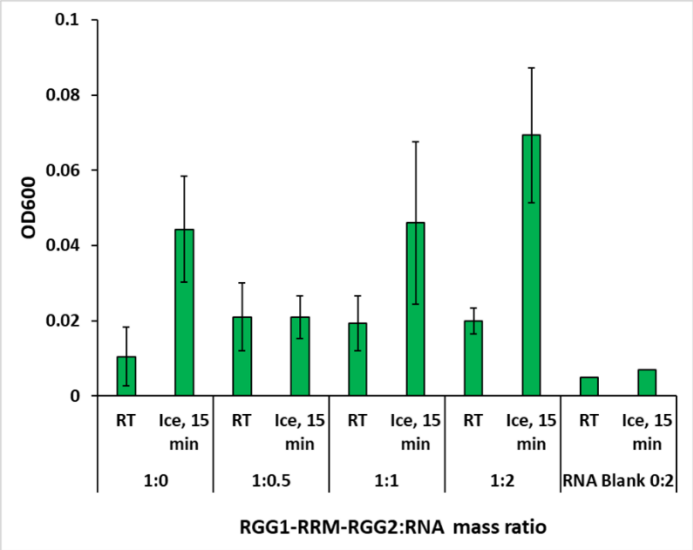


Figure 4-4 Turbidity measurements (OD 600 nm) of RGG_1-RRM-RRG_2 samples at 100 µM, at varying RNA concentration values. Turbidity measurements were performed at room temperature (RT) and after incubation on ice for 15 minutes. Each condition was measured in triplicate and error bars represent the standard deviation. The RNA blank measurement was performed in the absence of protein and RNA at the maximum tested concentration.

At room-temperature, a small increase in turbidity was observed upon addition of RNA. However, the turbidity remained unchanged with increasing RNA concentration. For the series of experiments with incubation in ice, a considerable decrease in turbidity was observed with the addition of a small amount of RNA. However, the addition of larger amounts of RNA (proportion higher than 1:1) led to an increase in turbidity.

These results are in contrast to those reported for the full-length FUS, i.e. an increased LLPS propensity in lower RNA concentrations and a decrease LLPS propensity higher RNA concentrations [68],[77]. Furthermore, since the commercial RNA used was the same as the one used in a previous study [70], the differences on the effect of RNA on LLPS cannot be attributed to differences in RNA samples.

Similarly to previous studies regarding the effect of RNA on LLPS of full-length FUS, the turbidity assays were performed at varying protein:RNA mass ratios. Due to the considerable mass differences between RGG_1-RRM-RGG_2 and full-length FUS (25.2 and 53.4 kDa, respectively), the corresponding RNA mass used in the present study corresponds to about half of the FUS:RNA ratio. If RGG_1-RRM-RGG_2 were to display identical LLPS tendency as the full-length FUS, the experiment with 1:1 RGG_1-RRM-RGG_2:RNA ratio should have provided the highest turbidity measurements. The fact that the turbidity increased even further with a higher amount of RNA indicates that there must be fundamental differences in the interaction of RNA with RGG_1-RRM-RGG_2 when compared to RNA and full-length FUS, to account for the different effect on the LLPS.

Previous studies have proven that full-length FUS protein binds to RNA cooperatively [64]. The cooperative nature of the binding event is due to the contribution of the RRM, ZnF domains and the RGG regions, as the isolated PrLD-RGG_1 construct did not present the ability to bind to RNA [108]. However, the PrLD-RGG_1-RRM construct presented the ability to bind to RNA in a non-cooperative event. The effect of RGG_2 on the cooperativity of the binding event remains to be solved. Hypothetically, if the RGG_1-RRM-RGG_2 in the present study presents non-cooperative binding to RNA, then this difference in the binding reaction could cause variations on the effect of RNA on the LLPS in respect to the full-length FUS, since different fractions of protein may be bound to RNA at identical protein:RNA ratios.

4.1.4 Effect of Amino Acids

The presence of metabolites may affect the LLPS process of proteins. A subset of metabolites rationally capable of affecting this process are amino acids. These metabolites may directly affect intermolecular protein-protein interactions. Furthermore, certain amino acids are present in relatively high concentrations in cells, namely in muscle and brain tissue, and are therefore, more prone to affect this phenomenon [109],[110].

Glutamate is present in high concentration in certain cells, where its concentration in the cytoplasm of neurons is within 5 to 10 mM [111]. More specifically, the concentration of glutamate within axon terminals and synaptic vesicles is several times higher than in the cytoplasm [111]. However, the study of biologically less prevalent amino acids may provide additional insight into the molecular drivers underlying the LLPS process. Therefore, the effect of the amino acids lysine and arginine were likewise assessed **Figure 4-5**.

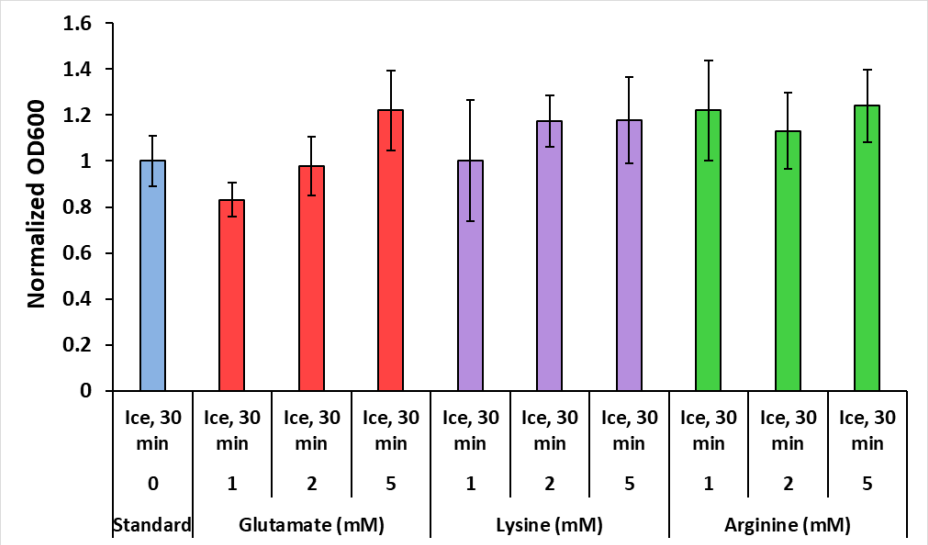


Figure 4-5 Normalized turbidity measurements (OD 600 nm) at varying concentrations of amino acids, namely glutamate (red), lysine (purple) and arginine (green). A concentration range of 1-5 mM was tested. The turbidity values were normalized in respect to the standard value. All measurements were performed in triplicate and error bars represent error propagation based on the standard deviation of each data point.

Concentrations of amino acids were kept at a maximum of 5 mM in order to avoid any considerable contribution from the increase in ionic strength. Having performed the turbidity measurements at a constant RGG_1-RRM-RGG_2 concentration of 100 μM, a maximum protein:amino acid proportion of 1:50 was tested.

The tested concentration range of each amino acid is similar to the intracellular concentration of the free amino acid, i.e. up to 0.8 mM for arginine [112] and 1.15 mM for lysine [109]. The studied concentrations were up to 5 mM for easier comparison to prior glutamate results and likewise, to avoid considerable ionic strength effect on the LLPS propensity.

A slight decrease in turbidity was observed in lower concentrations. However, upon further increasing the concentration of glutamate, the turbidity appeared to increase. Overall, at higher concentrations, the three tested amino acids tend to slightly promote the LLPS process.

At the tested pH (6.4) the three amino acids present global charge. Glutamate is negative, lysine and arginine present a positive charge. The distinction between these amino acids is the respective side chain, and therefore the differentiated effect on protein LLPS is associated to the side chain of the amino acid.

Therefore, these amino acids could be able to mediate intermolecular protein-protein interactions, thus promoting LLPS at higher concentrations. A similar hypothesis was established for the effect of ATP on the LLPS propensity of FUS, due to its anionic and aromatic extremities, both capable of interacting with basic residues, that act as a mediator for intermolecular interactions [60],[113]. Likewise, a similar event could occur with the three tested amino acids. However, more thorough studies at higher amino acid concentrations would be required, in parallel with controls observing the contribution of the ionic strength at each condition.

The turbidity increase appears to reach a plateau at higher concentrations of each amino acid. However, the apparent concentration of arginine required to reach the plateau is lower than that of lysine. This observation entails a differential contribution of arginine on the LLPS of RGG_1-RRM-RGG_2. Both arginine and lysine present an overall positive charge at the tested pH, due to the guanidinium and amine groups in their side chains, respectively. Owing to their cationic side chain functional groups, these amino acids are able to form ionic interactions with acidic amino acids present in the protein. The ionic interactions and/or simple salt bridges formed by these amino acids have similar strengths, and therefore no biased ionic pair formation for lysine or arginine is expected [114]. However, these cationic amino acids are also able to form π-cation interactions with aromatic residues, with a larger preference towards tryptophan residues [115]. The arginine guanidinium group, in contrast to the lysine amine group, is able to form stronger π-cation interactions. The nature of the interactions strength disparity lies on the likelihood of the guanidinium group providing additional possible interactions, namely the additional H-bonds and participation of delocalized π bonds in π- π stacking with aromatic residues [74],[116]. Therefore, by comparing the effect of arginine and lysine on the LLPS propensity of RGG_1-RRM-RGG_2 we may confirm the possible contributions of π-cation interactions on the phenomenon.

While the apparent additional contribution of arginine on the LLPS phenomenon may arise from more significant π-cation interactions with the protein, previous structural data on multimeric proteins have indicated the dominant presence of arginine as a linker residue on complex salt bridges, i.e. 3 or more participating residues. These complex salt bridges contribute for intermolecular interactions and multimerization [114]. Hypothetically, arginine could likewise contribute with transient complex salt bridges on the LLPS of proteins.

Further experiments with larger ranges of concentrations of these amino acids are required to understand in a semi-quantitative manner the contribution of each amino acid on the LLPS process. Moreover, high-resolution techniques would be required to assure the types of

residues affected by each amino acid and thus, validating the hypothesis that the free amino acids act as “bridges” between protein monomers. Nevertheless, the results may entail the intervention of π-cation interactions on the LLPS process of RGG_1-RRM-RGG_2.

4.1.5 NMR Study of RGG_1-RRM-RGG_2 LLPS

The resonances present in a protein ^{1}H, ^{15}N HSQC spectrum arise from amide protons in a slow-exchange regime with the solvent. Therefore, the resonances will arise from backbone amide (1 for each residue, except for the N-terminal residue which presents an amine group and prolines which lack amide protons), and sidechain amides, i.e. 2 for glutamine and asparagine residues, and 1 from tryptophan side chains. The ^{1}H,^{15}N HSQC spectrum presents a structural protein fingerprint, whereas the acquired spectrum is specific to the protein sample and the acquisition conditions [117]. Therefore, the acquisition of an ^{1}H,^{15}N HSQC spectrum allows the qualitative confirmation of proper protein folding.

As seen in **Figure 4-6** a plethora of clustered signals are observable in the ^{1}H,^{15}N HSQC spectrum of RGG_1-RRM-RGG_2, these arise from the RGG regions, whereas the glycine residues present in these regions are clustered at ≈ 8.4 ppm in ^{1}H and 110 ppm in ^{15}N, while the majority of the remaining resonances from these regions are clustered at ≈ 8.25 ppm in ^{1}H and 120 ppm in ^{15}N. A variety of resolved resonances are observable, these mainly arise from the RRM domain. Furthermore, the appropriate resonance dispersion further indicates that the RRM domain is folded.

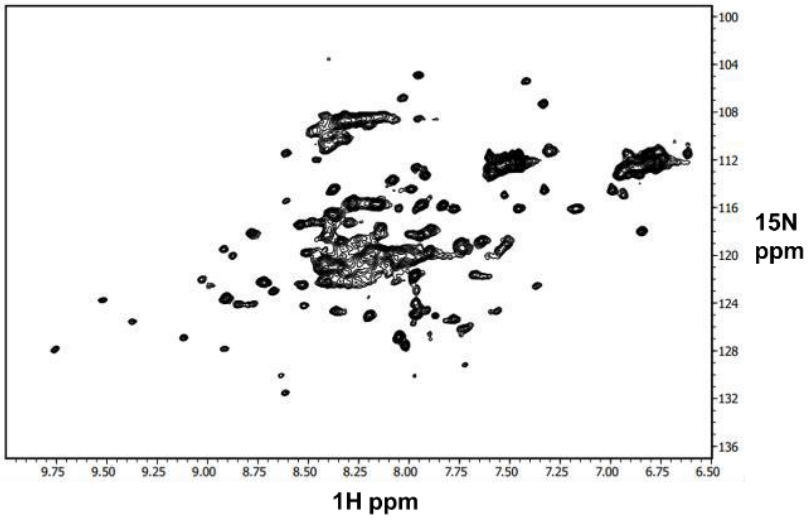


Figure 4-6 ^{1}H,^{15}N HSQC spectrum of 800 µM RGG_1-RRM-RGG_2 at 30 °C in 20 mM Tris-HCl, 100 mM NaCl, 2 mM β-mercaptoethanol and 0.05 % NaN_3, pH 7.0.

The ^{1}H and ^{15}N amide backbone resonance assignment of the RRM domain has been deposited in the biomagnetic resonance bank (BMRB, entry 17508), but the resonance-overlap due to the presence of resonances arising from the RGG regions prevents the complete assignment of the folded RRM domain. However, it is still possible to assign 60 of 85 residues of the folded RRM domain **Appendix 9**.

It would be expected that a decrease in temperature would induce LLPS as previously confirmed. Therefore, a series of ^{1}H,^{15}N HSQC spectra were acquired lowering the temperature from 37 °C to 5 °C, and an additional spectrum at 30 °C upon reheating of the sample, to check for the reversibility of any temperature induced spectral modifications. To facilitate comparison of intensities, the projections of the ^{1}H, ^{15}N HSQC spectra obtained at different temperatures are presents in **Figure 4-7** along the ^{1}H dimension.

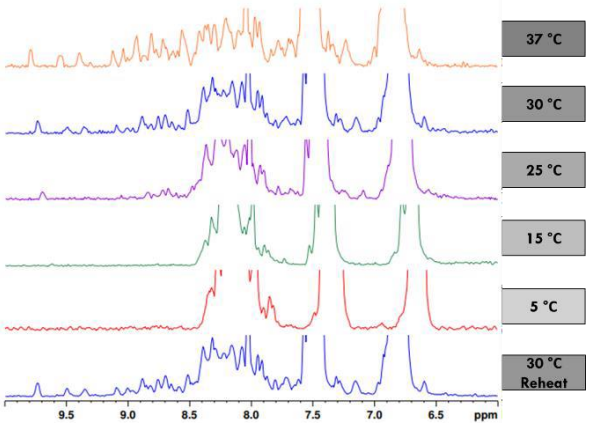


Figure 4-7 1D ^{1}H dimension ^{1}H,^{15}N HSQC projections at varying temperatures. The tested temperatures were: 37 °C (orange, top), 30 °C (blue), 25 °C (purple), 15 °C (green), 5 °C (red) and 30 °C upon reheating (blue, bottom). The spectrum at 30 °C upon reheating was acquired with 4x the number of scans as the remaining spectra, therefore the presented spectrum was scaled 0.25x.

A gradual decrease in intensity of resonances corresponding to the folded region, i.e. downfield resonances (higher ppm), and increase in those corresponding to the IDRs (around 8.2 ppm) was observed. The decrease in intensity of resonances from the folded region and increase of the disordered region could be explained either through the denaturation caused by the decrease in temperature (cold denaturation) [118] or due to LLPS. The resonance arising from tryptophan side-chains do not tend to decrease in intensity upon protein denaturation, solely a variation in chemical shift is commonly observed, which the degree of variation is dependent on solvent exposure of the residue [119]. Therefore, the decrease in intensity of the resonance arising from W353 side-chain (most downfield signal presented in the spectrum), contradicts the hypothesis of cold denaturation. Therefore, the decrease in resonance intensity must arise from LLPS, which would cause a decrease in resonance intensity due to higher viscosity within the dense phase. The increase in IDR resonance intensity should be due to the decrease in exchange rate of amide protons with bulk solvent due to the decrease in temperature [120].

As seen in **Figure 4-7**, upon reheating the sample to 30 °C, the 1D ^{1}H,^{15}N HSQC projection is qualitatively comparable to the spectrum at the same temperature prior to cooling. This fact indicates the reversibility of the temperature effect on the observed spectra. This feature is expected from LLPS of the protein, since the increase in temperature would disrupt the LLPS process and reverse the spectral effects of the phenomenon.

A detailed analysis regarding the resonances most promptly affected by the decrease in temperature could provide crucial information regarding temperature-dependent interactions and the LLPS process. However, due to the disappearance of many resonances at 15 and 5 °C, cross-peak intensities at 30 °C were compared with the respective at 25 °C **Figure 4-8**.

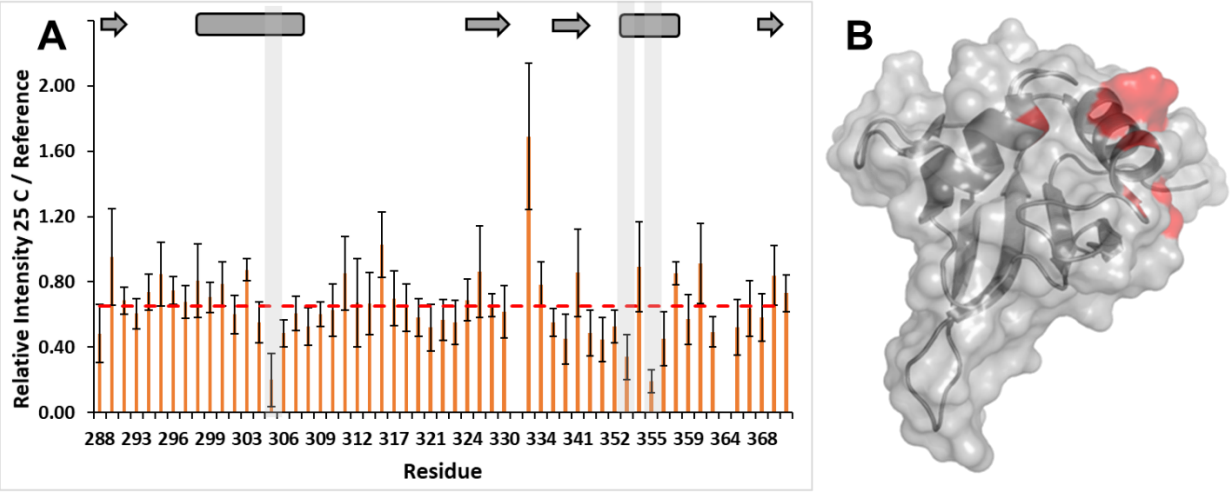


Figure 4-8 $^{1}H,^{15}N$ HSQC cross-peak intensity of RGG_1-RRM-RGG_2 at 25 °C (**A**), relatively to intensities at 30 °C, both measured using 800 µM of RGG_1-RRM-RGG_2 pH 7.0. Error bars represent the standard error associated to the intensity ratio. Residue G331 presented severe overlap with unassigned resonances at lower temperatures and therefore, was excluded from the analysis. Red dashed line represents the average intensity ratio for each plot. Cartoon representation of secondary structure on the top portion of **A**. Residues presenting more pronounced effects are highlighted using grey bars in **A** and highlighted in red on the known structure of the isolated RRM domain (**B**), PDB entry 2LA6.

As presented in **Figure 4-8**, a lower intensity ratio represents a more significant decrease in intensity due to the decrease in temperature. An overall decrease in intensity was observed for most residues, i.e. intensity ratio > 1. Certain residues presented a positive intensity ratio, likely due to the contribution of chemical exchange with solvent molecules at higher temperatures, indicating the inherent mobility of these residues. The decrease in intensity caused by the decrease in temperature is expected, namely due to the increase in solvent viscosity which causes an increase in correlation time [121]. This effect is expected to be uniform for all studied peaks. Therefore, the analysis was performed by identifying resonances more prominently affected in comparison to the average effect, which would represent the direct contribution from the increase in solvent viscosity. The existence of resonances more prominently affected would indicate the contribution from other effects on these resonances, namely the presence of local dynamic or structural alterations, or the prevalence of interactions.

Certain outliers however, present a more significant decrease in intensity, namely residues F305, W353 and D355. The significant decrease in intensity due to the decrease in temperature indicates the participation of these residues in temperature-dependent interactions, which would result in a decrease in resonance intensity upon formation or promotion of these interactions at lower temperatures [122].

The aromatic and anionic nature of these residues indicate the possibility of performing π-cation and ionic interactions with the adjacent RGG regions, specifically with the arginine residues present in these regions.

The identification of RRM residues significantly affected by the variation in temperature may also indicate the direct participation of the RRM domain in protein-protein interactions, hypothetically pertinent for the LLPS phenomenon, which would be promoted at lower temperatures, i.e. upon establishing these interactions.

Interestingly, and contrary to the backbone amide cross-peak the $^{1}H,^{15}N$ HSQC cross-peak arising from W353 side-chain does not present a significant intensity decrease (data point in the far most right in **Figure 4-8A**), in respect to the average value. This may indicate a more direct contribution from the amide group from W353 on the protein-protein interaction, or from adjacent residues whose amide groups are not significantly affected. Furthermore, the exchange regime regarding the resonance arising from W353 side chain may not cause a significant observable variation in intensity.

As previously observed in **Figure 4-7**, the effects of temperature on the spectra appeared to be reversible. However, a more detailed analysis was performed by comparing the intensity of the $^{1}H,^{15}N$ HSQC spectra at 30 °C before cooling and after reheating **Figure 4-9**.

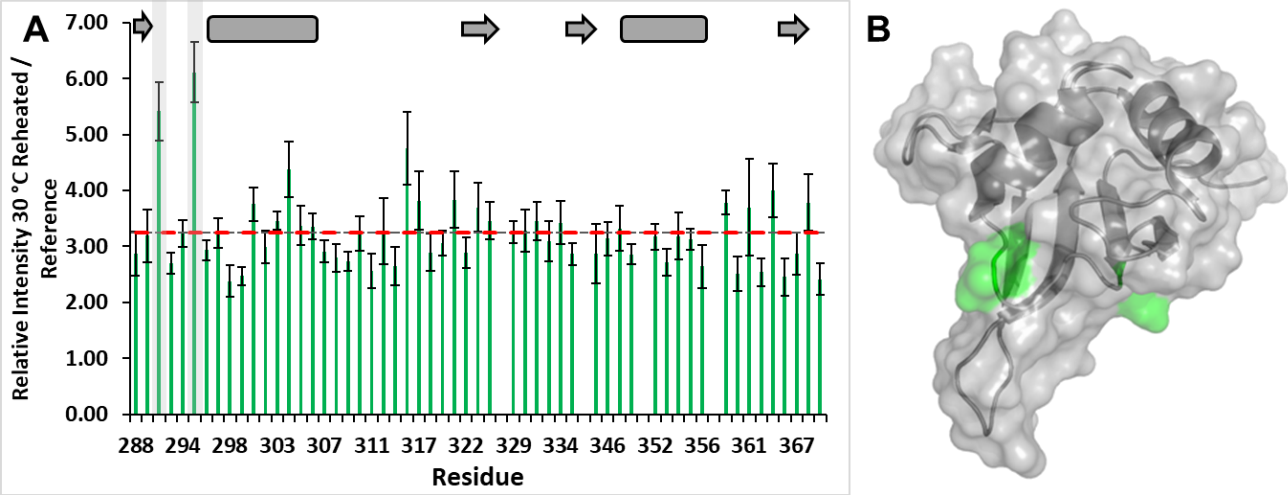


Figure 4-9 $^{1}H,^{15}N$ HSQC cross-peak intensity of RGG_1-RRM-RGG_2 at 30 °C after sample reheating (**A**), relatively to intensities at 30 °C prior cooling, measured using 800 µM of RGG_1-RRM-RGG_2 pH 7.0. Error bars represent the standard error associated to the intensity ratio. Red dashed line represents the average intensity ratio for each plot. Cartoon representation of secondary structure on the top portion of **A**. Residues presenting more pronounced effects are highlighted using grey bars in **A** and highlighted in green on the known structure of the isolated RRM domain (**B**), PDB entry 2LA6.

Both spectra were acquired with 4x the number of scans as the spectrum prior to cooling. It is expected therefore that the intensity ratios to be > 1. However, the intensity variation caused by the difference in number of scans is expectedly uniform for all resonances, therefore the analysis was again performed through comparison with the average intensity ratio value.

Most studied residues do not significantly deviate from the average, indicating that the cooling and reheating process did not specifically affect the spectral properties of these residues. This confirms the overall sample integrity throughout the process.

Two outliers are clearly visible, namely residues Q290 and N295. The spectral properties of these residues appear to have been specifically affected by the cooling and reheating process. Furthermore, to truly understand the extent of the reversal, an additional $^{1}H,^{15}N$ HSQC spectrum was acquired upon incubating the sample for 9 hours at 30 °C, whereas the effect does not appear to have been reversed after incubating for 9 hours at 30 °C **Appendix 10**.

These residues were not specifically affected by the cooling process down to 25 °C **Figure 4-8**. Therefore, the participation of these residues in intra or intermolecular interaction must have

been affected upon the decrease in temperature and was not recovered to the starting state upon reheating, resulting in the observed increase in intensity.

Nevertheless, the nature of the observed effect remains to be elucidated, namely by performing additional cooling and reheating cycles.

However, due to the direct and inherently uneven contribution of temperature on the chemical shift of ^{1}H-^{15}N resonances [123], the study of the effect of these potential interactions on the chemical shift deviation of resonances could not be performed.

4.1.5.1 The effect of RGG_1-RRM-RGG_2 concentration on NMR observables

As previously established, the LLPS process of RGG_1-RRM-RGG_2, similarly to various other proteins, is highly dependent on the concentration of the phase separating protein **Figure 4-1**. Therefore, one could expect the influence of protein concentration on the extent of protein-protein interactions and, consequentially, on the spectral properties of the participating residues.

For this purpose, a separate 1:2 dilute sample was prepared, presenting negligible turbidity at room temperature (OD_{600} = 0.023). To avoid any direct contribution from LLPS, ^{1}H,^{15}N HSQC spectrum was acquired at 30 °C and the resonance chemical shift were compared to the spectrum acquired at the identical temperature, prior to cooling, for the concentrated sample **Figure 4-10**.

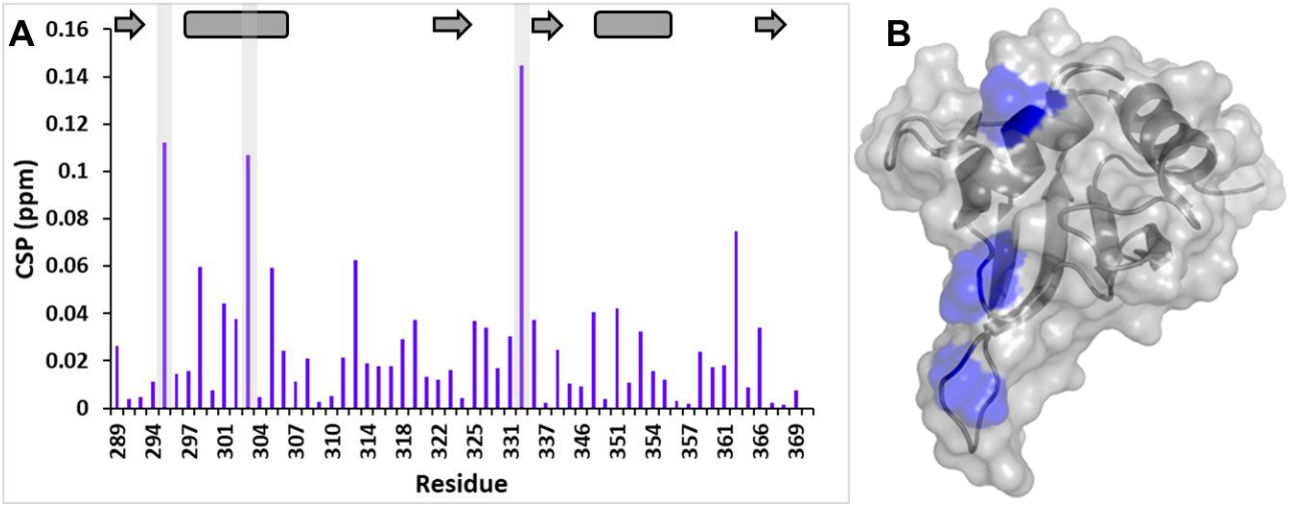


Figure 4-10 ^{1}H,^{15}N HSQC chemical shift perturbation upon increasing concentration, from 400 to 800 µM at pH 7.0 (**A**, CSP). Cartoon representation of secondary structure on the top portion of **A**. Residues presenting more pronounced effects are highlighted using grey bars in **A** and highlighted in blue on the known structure of the isolated RRM domain (**B**), PDB entry 2LA6.

A further analysis focusing on the chemical shift perturbation (CSP) upon increasing concentration revealed the existence of a subset of residues affected by the variation in concentration, namely N295, D303 and K332. Having observed a maximum CSP of ≈ 0.14 ppm, the observed chemical shift deviation is relatively low for protein-protein interactions, indicating the weak nature of the present interactions [104]. The weak and transient nature of protein-protein interactions correlated to LLPS is essential to assure the dynamics of the macromolecule-dense phase [124].

The affected residues possess polar, anionic and cationic side chains, providing therefore a scaffold for several polar, ionic and π-cation interactions.

The asparagine residue was previously observed to be significantly affected by the cooling and reheating process in **Figure 4-9**. The significant observed CSP for N295 highlights an apparent relevance of N295 in protein-protein interactions, within the tested concentration range.

Asparagine contains an amide side-chain and therefore, the primary contribution for protein-protein interactions would arise from the participation in H-bonding and, as previously postulated, π-π interactions [125]. Furthermore, the participation of glutamines on the LLPS process of FUS PrLD has been already established [52],[58].

The estimated protein concentration used in this analysis ranged between 400 and 800 µM, it is therefore impossible to exclude the possibility of residues specifically affected at different ranges of protein concentration.

The residues previously observed to be severely affected at decreasing temperature **Figure 4-8**, were not observed to be affected by the variation in protein concentration. These residues may be specifically affected at protein concentrations other than the ones used in the previous analysis. Alternatively, distinct residues may be predominantly affected by other variables, e.g. temperature and protein concentration.

4.1.5.2 Structural insight concerning affected residues

Throughout the various analysis already performed, a total of 7 residues have been identified to be specifically affected by various conditions. The identified residues may be classified as aromatic (F305 and W353), anionic (D303 and D355), a single cationic residue (K332) and polar (Q290 and N295). The present diversity within these residues promotes a variety of possible interactions, namely previously established ionic and π-cation, as well as π-π and H-bonding interactions.

Hypothetically, this wide range of possible interactions could directly affect the LLPS process of RGG_1-RRM-RGG_2. However, the structural positioning the various affected residues is crucial to verify the significance of the hypothesis that they are relevant for the LLPS process. Therefore, the analysis of the relative positioning of these residues on the folded RRM structure could provide insight regarding the relevance of the affected residues. For this analysis the RNA-RRM complex structure was used **Figure 4-11**, because it allows to locate these affected residues relative to residues involved in protein RNA-interactions.

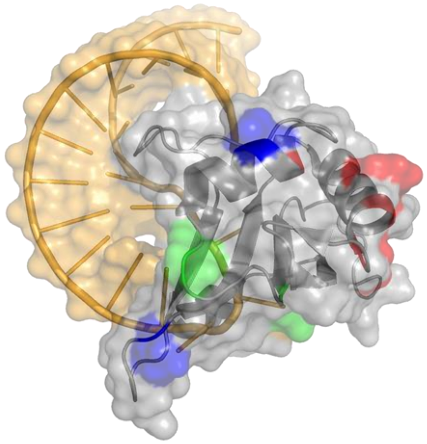


Figure 4-11 RNA-RRM complex structure, PDB entry 6GBM [65]. The identified affected residues from the various analysis are highlighted in the following color scheme: red F305, W353, D355; green Q290, N295; blue D303 and K332. The RNA unit is presented in light orange.

The affected residues appear to be positioned opposed in respect to the RNA-binding region. Furthermore, none of the presented residues coincide with those previously confirmed as participating in the RNA binding [65].

Therefore, a hypothetical protein-protein interaction interface for LLPS may be established, where the affected residues lie in a defined region of the folded RRM domain different from that participating in RNA binding. This suggests that RNA-binding should not directly interfere with the protein-protein interaction interface.

Overall, the data indicates the possibility of the folded RRM domain directly participating in the LLPS process of RGG_1-RRM-RGG_2 and, hypothetically by extrapolation, of full-length FUS. Indeed, turbidity assays at varying pH suggested the direct influence of RRM on LLPS under certain conditions. In contrast, prior studies have revealed the insignificance of the RRM domain on the toxicity associated to full-length FUS [66].

4.1.5.3 Generating a “stable” LLPS sample

As previously observed, the decrease in pH from 7.0 to 6.4 should promote LLPS **Figure 4-2**. The pH of the sample was adjusted to 6.4 and corroborating the previous results regarding the pH dependence of LLPS, the turbidity of the sample visibly increased at room temperature, presenting an OD_{600} = 0.225 (measured at room temperature, **Figure 4-12**).

An analysis of the ^{1}H, ^{15}N HSQC resonance intensities of this turbid sample before and after the 9 h incubation at 30 °C was performed, to analyze the reversibility of the process **Figure 4-12**. The reference spectrum from previous analyses (pH 7.0) was not used, in order to avoid the direct influence of pH on the comparison.

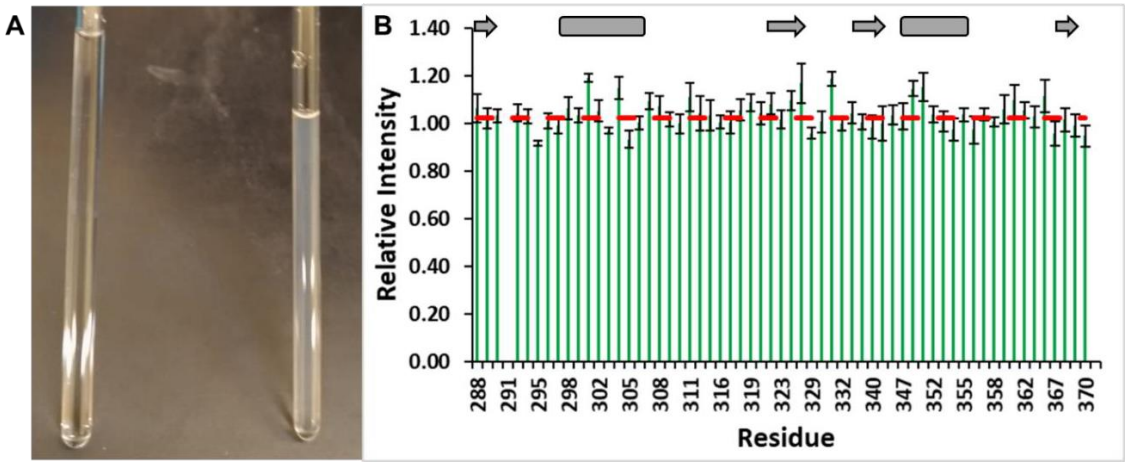


Figure 4-12 Visual comparison of a disperse sample (left, **A**) and a phase separated RGG_1-RRM-RGG_2 (right, **A**) sample. ^{1}H,^{15}N HSQC cross-peak intensity ratios (**B**), spectra acquired at 800 µM pH 6.4 after/ before incubating for 9 hours at 30 °C. Average intensity ratios are presented in red dashed lines. Error bars represent standard errors for the ratios. Cartoon representation of secondary structure on the top portion.

As previously observed in **Figure 4-1**, in the tested conditions, the turbidity increase rate appears to be inherently linked to the LLPS propensity. As a result, the stable phase-separated sample could indeed present a lesser forward reaction rate in comparison to the transiently phase-separated sample.

No notable regions presented significant variations in intensity after incubating the phase separated sample at 30 °C for 9 hours. Furthermore, the average intensity ratio was 1.02, indicating that the intensity did not significantly vary after incubating for 9 hours at 30 °C. In contrast to the reversibility pattern of temperature induced LLPS observed at pH 7.0 **Figure 4-7**, upon decreasing the pH to 6.4, the reversibility appears to have significantly decreased. A more "stable" phase appears to be formed upon lowering pH and therefore, the dissolution rate of the dense phase appears to be inherently connected to the LLPS propensity, where a decrease in propensity causes an increase in dissolution rate.

Furthermore, as observed in the lack of reversibility in **Figure 4-12**, the LLPS propensity could likewise cause a variation in the dissolution rate. These properties provide a future opportunity to modulate the LLPS propensity and attempt to observe the variations in the underlying chemical exchange of the phenomenon [126]. Such modulation could provide a prospect to indirectly study the NMR-invisible states using non-condensed samples [127].

4.2 RGG_2-ZnF-RGG_3 construct

4.2.1 Preliminary NMR Studies

As was already mentioned before, when considering the NMR study of proteins, the standard experiment firstly performed is a ^{1}H, ^{15}N HSQC, since this experiment provides the spectral equivalent of the protein “fingerprint”, with each protein presenting a characteristic ^{1}H, ^{15}N HSQC spectrum. Therefore, a ^{1}H, ^{15}N HSQC of the RGG_2-ZnF-RGG_3 construct was acquired at pH 7.1 and 25 °C **Figure 4-13**.

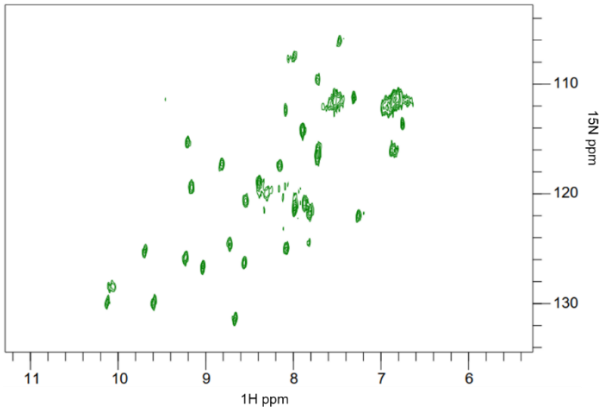


Figure 4-13 ^{1}H,^{15}N HSQC spectrum of 100 µM RGG_2-ZnF-RGG_3 at 30 °C in 50 mM Tris-HCl pH 7.1, 100 mM NaCl, 5 µM $ZnCl_2$ and 0.05 % NaN_3.

As previously described in **4.1.5**, the resonances present in a ^{1}H, ^{15}N HSQC spectrum arise from backbone amide and sidechain amides. Considering the size of the protein in question, approximately 130 residues, the lack of resonances is noticeable.

The high dispersion of resonances in a ^{1}H, ^{15}N HSQC spectrum is characteristic of folded proteins, indicative that the ZnF adopts a folded conformation. Furthermore, the lack of ^{1}H resonances from the protein upfield of 0.8 ppm and the lack of aliphatic residues in the protein **Appendix 11**, indicates that the folding is likely due to Zn^{2+} and not due to the existence of an hydrophobic core [128].

To verify the relative stability of the sample at varying temperatures, a series of ^{1}H, ^{15}N HSQC spectra were acquired at different temperatures **Figure 4-14**.

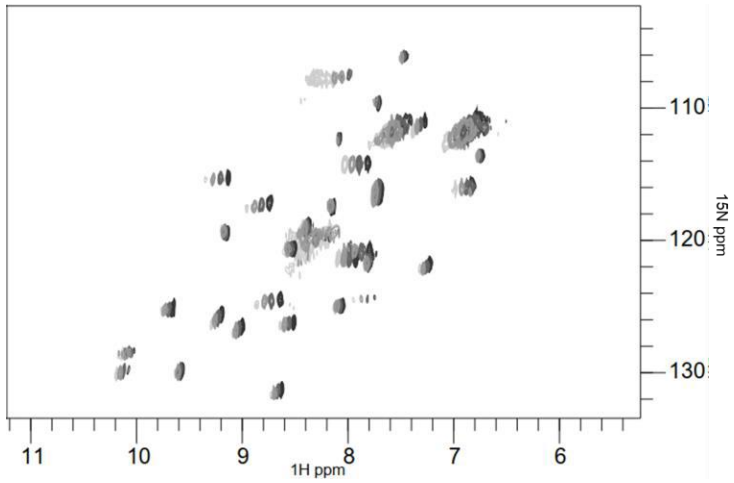


Figure 4-14 ^{1}H, ^{15}N HSQC spectra acquired at varying temperature. The spectra were acquired at 5, 15, 25 and 37 °C, brightest to darkest, respectively.

After lowering the temperature, the appearance of resonances at chemical shift characteristic of IDRs was observable, i.e. approximately 8.5 ppm in ^{1}H. These resonances arise from the RGG regions, which are indistinguishable using this experiment. The appearance of these resonances at lower temperature indicates that these were originally undergoing chemical exchange with the solvent in a fast-exchange regime, and therefore their resonances are masked by that of the solvent. The exchange rate was lowered due to the decrease in temperature, and consequentially, a portion of these amide protons entered the slow-exchange regime.

There is an observable upfield shift (lower ppm) of the resonances with the increase in temperature. This tendency is uniform for all HN resonances, however the degree of chemical shift variation with the increase in temperature (the amide temperature coefficient) may be used to extract structural information of the protein **Figure 4-15**.

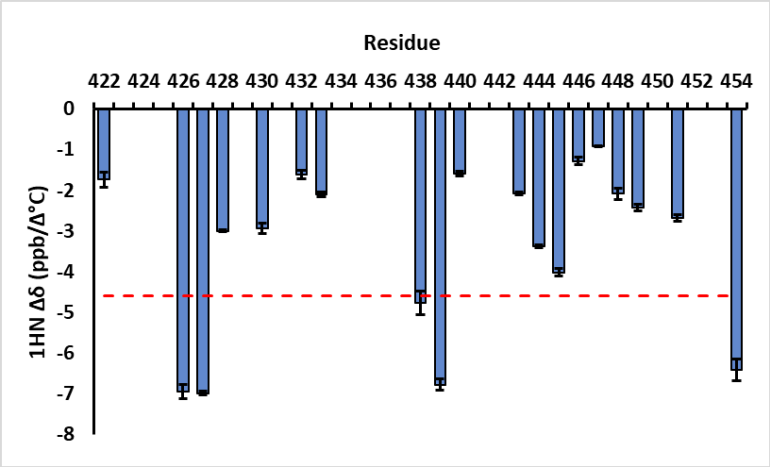


Figure 4-15 $^{1}H_N$ chemical shift variation per residue caused by the variation of temperature. Variation of chemical shift in ppb (parts per billion) per variation of degree Celsius. The threshold of -4.6 ppb/°C is presented in dashed line. Error bars represent the standard error associated with the linear regression performed.

The variation in $^{1}H_N$ chemical shift caused by the variation in temperature is extremely small, presented in **Figure 4-15** in ppb. Nevertheless, the analysis of this variation may pinpoint the participation of the specific HN group in intramolecular H-bonds [129]. To achieve this, an

established threshold of -4.6 ppb/°C is used, and HN groups presenting Δδ/Δ°C values less negative than the threshold are likely to be participating in H-bonds [129].

This analysis allows the observation of a relatively large proportion of residues participating in intramolecular H-bonds. Namely, two sequential groups of residues are identifiable, stretches 428-433 and 440-451. Interestingly, these stretches encompass all 4 cysteine residues: C428, C433, C443 and C447. These four cysteines are presumably collectively coordinating a Zn^{2+} ion, this may provide a driving force for the subsequent clustering of these stretches and result in the intramolecular H-bonding between the HN groups of these residues.

If certain residues were participating in the formation of secondary structure, these will necessarily identify as participating in H-bond. These two previously identified stretches of residues partially encompass previously proposed β-sheet structure for this region, i.e. 441-443 and 448-449 [65]. However, certain regions previously proposed to form β-sheets or α-helices present chemical shift deviations above the threshold [68],[69]. Furthermore, the presence of ^{1}H resonances between 5-6 ppm **Appendix 11**, corresponding to H^{α}, could be indicative of the presence of β-sheet in the structure [128],[130].

Residues known to form secondary structure who presented chemical shift deviations more negative than the threshold were identified, it is however improbable and primarily associated to α-helix structures [129].

A variation in ^{1}H,^{15}N HSQC cross-peak intensity was observed as the temperature decreased **Figure 4-14**. The observed decrease in ^{1}H,^{15}N HSQC cross-peak intensity at lower temperatures, could be due to the expected variation in solvent viscosity and consequential variation in correlation time [121]. However, as previously observed regarding RGG_1-RRM-RGG_2 **4.1.5**, the LLPS phenomenon causes a decrease in intensity, which could be more prominent at lower temperatures.

As previously observed in RGG_1-RRM-RGG_2, the LLPS phenomenon caused a clear overall decrease in intensity. The phenomenon was confirmed by the observation of a pronounced decrease in intensity of the resonance arising from the tryptophan side-chain (W353), which would be nevertheless affected by the variation in correlation time. Therefore, if RGG_2-ZnF-RGG_3 undergoing LLPS would cause the observed variation in intensity, the resonances arising from the 2 tryptophan residues of the ZnF domain (W426 and W440) should be significantly affected at lower temperature, similarly to the observed with W353. No ^{1}H, ^{15}N HSQC spectrum of RGG_2-ZnF-RGG_3 was acquired at 30 °C, and therefore the peak intensities from the 3 tryptophan side-chains were normalized accordingly to their respective intensity at 25 °C, the closest common tested temperature **Figure 4-16**.

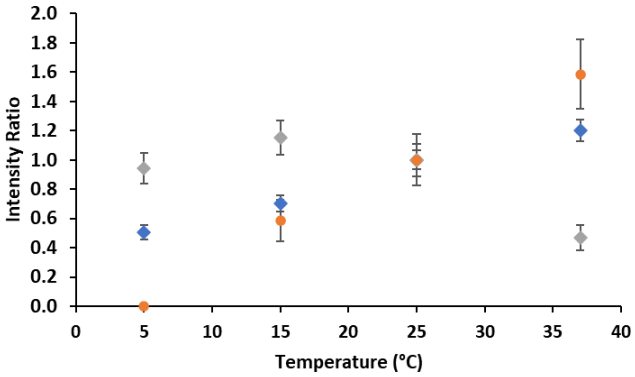


Figure 4-16 Tryptophan Side-chain NH resonance intensity at different temperatures. Intensities normalized in respect to the intensities at 25 °C. From ZnF domain: W426 in blue, W440 in gray and W353 from RRM domain in orange. Error bars represent standard error from the fitting directly extracted from PINT. At 5 °C the resonance from W353 could no longer be observed, being attributed 0 intensity at this temperature.

A clear decrease in intensity was observed at lower temperatures. A significant decrease was observed regarding W426, in contrast however, W440 did not present a large decrease at lower temperatures. However, the resonance arising from W353 was unobservable at 5 °C, indicating a more significant temperature induced intensity decrease, directly caused by the LLPS phenomenon of RGG_1-RRM-RGG_2.

Previous reports have indicated the ability of RGG_2-ZnF-RGG_3 to undergo LLPS under certain conditions [131]. However, the LLPS phenomenon should cause a significant decrease in intensity, as observed in the W353 resonance. Due to the relatively less significant decrease in intensity regarding the side-chain resonances arising from the tryptophan residues in the RGG_2-ZnF-RGG_3 protein, the protein does not appear to undergo significant LLPS under the studied conditions. However, an alternate phenomenon appears to specifically affect W440, resulting in the relatively minor decrease at lower temperatures.

Alternatively, the decrease in cross-peak intensity may translate to a variation in resonance line width. Furthermore, the correlation time decreases at higher temperatures, either through a direct influence of the temperature or through the decrease in H_2O viscosity, correlated to the observed decrease in line width of the solvent resonance **Figure 4-17** [121].

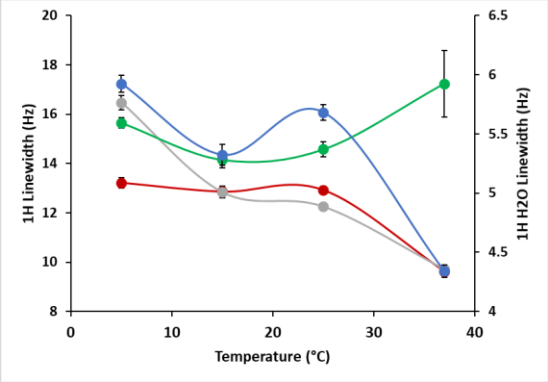


Figure 4-17 Normalized ^{1}H Line Width of H_2O (blue, right-hand axis) and 3 RGG_2-ZnF-RGG_3 residues: C433 (red), K448 (grey), G454 (green). Linewidth values were normalized in respect to 1H linewidth of the reference (DSS). Error bars represent the standard error.

The ^{1}H linewidth of a selected group of residues was measured for each temperature. Based on the previous amide temperature coefficient results **Figure 4-15,** the selected residues were: C433 and K448, with amide groups participating in intramolecular H-bonding, located in the folded ZnF domain, and G454, with an amide group that is not participating in intramolecular H-bonding and located at the start of the disordered RGG_3 region.

A decrease in line width was observed for C433 and K448, most likely due to a decrease in local correlation time. However, G454 deviates from this trend at higher temperatures, most likely due to the contribution of chemical exchange with the solvent, further indicating that this HN group is indeed more sensitive to solvent chemical exchange and therefore not participating in intramolecular H-bonding [103].

The decrease in linewidth with the increase in temperature entails the inexistence of chemical exchange in the slow exchange regime, which would cause line broadening by approaching intermediate exchange regime at higher temperature. The observed line narrowing however, does not confirm the existence of chemical exchange in the fast exchange regime, since it could be likely due to the decrease in correlation time [126].

5 Conclusion and future perspectives

The primary objective of the study was to elucidate the influence of different regions of FUS on the inherent LLPS phenomenon. Thus, the full-length protein was divided in two distinct previously unstudied constructs, which were addressed in the current work.

The direct contribution of these sections on the LLPS phenomenon of FUS could be elucidated by evaluating the LLPS capacity of these sections and the effect of various conditions on the phenomenon.

Using turbidity assays, the effect of various conditions on the LLPS propensity could be evaluated, namely protein concentration, pH, ionic strength, RNA and several amino acids. Also, the use of NMR spectroscopy provided valuable insight regarding the participation of certain residues on protein-protein interactions, which could be crucial for the observed LLPS phenomenon.

The capacity of the isolated RGG_1-RRM-RGG_2 section to undergo LLPS was established. Furthermore, through the evaluation of the effects of ionic strength, pH and addition of distinct amino acids, at least ionic and π-cation interactions appear to drive the observed LLPS phenomenon. The folded RRM domain may play a direct role on the phenomenon under certain conditions, namely in acidic pH. The addition of small amounts of RNA caused a decrease in turbidity, while larger concentrations of RNA caused an increase in turbidity, which is distinct from the observed in assays using full-length FUS. However, the exact reason for this differentiated effect when compared to full-length FUS remains to be fully understood. Through NMR spectroscopy, a subset of aromatic, anionic, polar and cationic residues within the RRM domain, clustered opposed to the RNA binding region, appeared to be significantly affected by variations in temperature, protein concentration and a cooling and reheating cycle. The specific effect on these residues could be indicative of the direct participation on protein-protein interactions, therefore establishing a hypothetical protein-protein interaction interface. Furthermore, due to the structural positioning of the interface, RNA-binding should not directly affect the postulated protein-protein interactions. However, the hypothetical interface remains to be unequivocally confirmed.

Primarily evaluated through NMR spectroscopy, the RGG_2-ZnF-RGG_3 section did not appear to under LLPS under the tested conditions. Further assays, namely in varying conditions, must be performed to truly dismiss the existence of self-sufficiency to undergo LLPS of the section.

The study relied on constructs previously unstudied, namely the structured domains of FUS bound to the flanking RGG regions. The work herein presented allowed the confirmation that the folded RRM domain may directly contribute for the LLPS phenomenon of FUS, while providing a further construct based on FUS capable of undergoing LLPS. The direct contribution of each RGG region for the LLPS phenomenon of the construct remains to be uncovered, along with the observed effect of RNA, amino acids and ionic strength at a residue level. Under the studied conditions, the RGG_2-ZnF-RGG_3 did not undergo LLPS. The LLPS capacity of this construct remains to be confirmed or unequivocally disproved, which could provide further information regarding the minimum requirements for LLPS of the studied system.

6 References

[1] Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* (**2017**) 357**,** eaaf4382.

[2] Niskanen, J. & Tenhu, H. How to manipulate the upper critical solution temperature (UCST) *Polym. Chem.* (**2017**) 8**,** 220–232.

[3] Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* (**1942**) 10**,** 51–61.

[4] Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* (**2015**) 11**,** 899–904.

[5] Berry, J., Brangwynne, C. P. & Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. *Reports Prog. Phys.* (**2018**) 81**,** 046601.

[6] Ribeiro, S. S., Samanta, N., Ebbinghaus, S. & Marcos, J. C. The synergic effect of water and biomolecules in intracellular phase separation. *Nat. Rev. Chem.* (**2019**) doi:10.1038/s41570-019-0120-4

[7] Overbeek, J. T. G. & Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.* (**1957**) 49**,** 7–26.

[8] Martin, E. W. & Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* (**2018**) 57**,** 2478–2487.

[9] Dill, K. A., Alonso, D. O. V. & Hutchinson, K. Thermal Stabilities of Globular Proteins. *Biochemistry* (**1989**) 28**,** 5439–5449.

[10] Lin, Y. H., Forman-Kay, J. D. & Chan, H. S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* (**2018**) 57**,** 2499–2508.

[11] Ruff, K. M., Roberts, S., Chilkoti, A. & Pappu, R. V. Advances in Understanding Stimulus-Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. *J. Mol. Biol.* (**2018**) 430**,** 4619–4635.

[12] Higashi, N., Sonoda, R. & Koga, T. Thermo-responsive amino acid-based vinyl polymers showing widely tunable LCST/UCST behavior in water. *RSC Adv.* (**2015**) 5**,** 67652–67657.

[13] Ukmar-Godec, T., Wegmann, S. & Zweckstetter, M. Biomolecular condensation of the microtubule-associated protein tau. *Semin. Cell Dev. Biol.* (**2019**) 1–13. doi: 10.1016/j.semcdb.2019.06.007

[14] Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U. S. A.* (**2019**) 116**,** 7889–7898.

[15] Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* (**2018**) 115**,** 2734–2739.

[16] Wang, Y., Lomakin, A., Kanai, S., Alex, R. & Benedek, G. B. Liquid-Liquid Phase Separation in Oligomeric Peptide Solutions. *Langmuir* (**2017**) 33**,** 7715–7721.

[17] Wegmann, S. *et al.* Tau protein liquid–liquid phase separation can initiate tau aggregation. *EMBO J.* (**2018**) 37**,** 1–21.

[18] Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* (**2018**) 28**,** 420–435.

[19] Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* (**2016**) 26**,** 668–679.

[20] Lin, Y., Currie, S. L. & Rosen, M. K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* (**2017**) 292**,** 19110–19120.

[21] Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* (**2012**) 483**,** 336–340.

[22] Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* (**2018**) 9**,** 1–12.

[23] Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* (**2015**) 57**,** 936–947.

[24] Somjee, R., Mitrea, D. M. & Kriwacki, R. W. Exploring Relationships between the Density of Charged Tracts within Disordered Regions and Phase Separation. in *Biocomputing 2020* (WORLD SCIENTIFIC, 2019). 218**,** 207–218.

[25] Majumdar, A., Dogra, P., Maity, S. & Mukhopadhyay, S. Liquid–Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* (**2019**) 10**,** 3929–3936.

[26] Howton, T. C., Zhan, Y. A., Sun, Y. & Shahid Mukhtar, M. Intrinsically disordered proteins: Controlled chaos or random walk. *Int. J. Plant Biol.* (**2016**) 6**,** 52–57.

[27] Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta - Proteins Proteomics* (**2013**) 1834**,** 932–951.

[28] Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* (**2017**) 44**,** 18–30.

[29] Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* (**2018**) 359**,** 698–701.

[30] Raut, A. S. & Kalonia, D. S. Pharmaceutical Perspective on Opalescence and Liquid-Liquid Phase Separation in Protein Solutions. *Mol. Pharm.* (**2016**) 13**,** 1431–1444.

[31] Folch, B., Dehouck, Y. & Rooman, M. Thermo- and mesostabilizing protein interactions identified by temperature-dependent statistical potentials. *Biophys. J.* (**2010**) 98**,** 667–677.

[32] Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* (**2009**) 324**,** 1729–1732.

[33] Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* (**2017**) 357**,**

[34] Spannl, S., Tereshchenko, M., Mastromarco, G. J., Ihn, S. J. & Lee, H. O. Biomolecular condensates in neurodegeneration and cancer. *Traffic* (**2019**) 20**,** 890–911.

[35] Sawyer, I. A., Sturgill, D. & Dundr, M. Membraneless nuclear organelles and the search for phases within phases. *Wiley Interdiscip. Rev. RNA* (**2019**) 10**,** 1–20.

References

[36] Aumiller, W. M. & Keating, C. D. Experimental models for dynamic compartmentalization of biomolecules in liquid organelles: Reversible formation and partitioning in aqueous biphasic systems. *Adv. Colloid Interface Sci.* (**2017**) 239**,** 75–87.

[37] Woodruff, J. B., Hyman, A. A. & Boke, E. Organization and Function of Non-dynamic Biomolecular Condensates. *Trends Biochem. Sci.* (**2018**) 43**,** 81–94.

[38] Schmidt, H. B. & Görlich, D. Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends Biochem. Sci.* (**2016**) 41**,** 46–61.

[39] Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* (**2016**) 164**,** 487–498.

[40] Mahboubi, H. & Stochaj, U. Cytoplasmic stress granules: Dynamic modulators of cell signaling and disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* (**2017**) 1863**,** 884–895.

[41] Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* (**2019**) 176**,** 419–434.

[42] Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* (**2016**) 166**,** 651–663.

[43] Peran, I. & Mittag, T. Molecular structure in biomolecular condensates. *Curr. Opin. Struct. Biol.* (**2019**) 60**,** 17–26.

[44] Woodruff, J. B. *et al.* The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell* (**2017**) 169**,** 1066-1077.e10.

[45] Nakashima, K. K., Vibhute, M. A. & Spruijt, E. Biomolecular chemistry in liquid phase separated compartments. *Front. Mol. Biosci.* (**2019**) 6**,**

[46] Audas, T. E. *et al.* Adaptation to Stressors by Systemic Protein Amyloidogenesis. *Dev. Cell* (**2016**) 39**,** 155–168.

[47] Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *Elife* (**2016**) 5**,** 1–25.

[48] Sama, R. R. K., Ward, C. L. & Bosco, D. A. Functions of FUS/TLS From DNA Repair to Stress Response: Implications for ALS. *ASN Neuro* (**2014**) 6**,** 175909141454447.

[49] Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* (**2015**) 162**,** 1066–1077.

[50] Rhoads, S., Monahan, Z., Yee, D. & Shewmaker, F. The Role of Post-Translational Modifications on Prion-Like Aggregation and Liquid-Phase Separation of FUS. *Int. J. Mol. Sci.* (**2018**) 19**,** 886.

[51] Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* (**2014**) 30**,** 2501–2502.

[52] Murthy, A. C. *et al.* Molecular interactions underlying liquid−liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* (**2019**) 26**,** 637–648.

[53] Monahan, Z. *et al.* Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* (**2017**) 36**,** 2951–2967.

[54] Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* (**2018**) 174**,** 688-699.e16.

[55] Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* (**2017**) 171**,** 615-627.e16.

[56] Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* (**2019**) 294**,** 7115–7127.

[57] Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG Motif. *Mol. Cell* (**2013**) 50**,** 613–623.

[58] Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* (**2018**) 174**,** 688-699.e16.

[59] Yoshizawa, T. *et al.* Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell* (**2018**) 173**,** 693-705.e22.

[60] Kang, J., Lim, L., Lu, Y. & Song, J. A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids. *PLOS Biol.* (**2019**) 17**,** e3000327.

[61] Kaneb, H. M., Dion, P. A. & Rouleau, G. A. The FUS about arginine methylation in ALS and FTLD. *EMBO J.* (**2012**) 31**,** 4249–4251.

[62] Liu, X. *et al.* The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site. *Biochim. Biophys. Acta - Mol. Basis Dis.* (**2013**) 1832**,** 375–385.

[63] Lu, Y., Lim, L. & Song, J. RRM domain of ALS/FTD-causing FUS characteristic of irreversible unfolding spontaneously self-assembles into amyloid fibrils. *Sci. Rep.* (**2017**) 7**,** 1043.

[64] Wang, X., Schwartz, J. C. & Cech, T. R. Nucleic acid-binding specificity of human FUS protein. *Nucleic Acids Res.* (**2015**) 43**,** 7535–7543.

[65] Loughlin, F. E. *et al.* The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Mol. Cell* (**2019**) 73**,** 490-504.e6.

[66] Bogaert, E. *et al.* Molecular Dissection of FUS Points at Synergistic Effect of Low-Complexity Domains in Toxicity. *Cell Rep.* (**2018**) 24**,** 529-537.e4.

[67] Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell Death Discov.* (**2017**) 3**,**

[68] Kang, J., Lim, L., Lu, Y. & Song, J. *A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acid*s. *PLoS biology* (2019). 17**,**

[69] Iko, Y. *et al.* Domain architectures and characterization of an RNA-binding protein, TLS. *J. Biol. Chem.* (**2004**) 279**,** 44834–44840.

[70] Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* (**2015**) 60**,** 231–241.

[71] Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* (**2018**) 175**,** 1467-1480.e13.

References

[72] Ribeiro, S. S., Samanta, N., Ebbinghaus, S. & Marcos, J. C. The synergic effect of water and biomolecules in intracellular phase separation. *Nat. Rev. Chem.* (**2019**) 3**,** 552–561.

[73] Holehouse, A., Martin, E., Peran, I., Pappu, R. & Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Disordered Prion-Like Domains. *Bull. Am. Phys. Soc.* (**2020**) 699**,** 694–699.

[74] Chong, P. A., Vernon, R. M. & Forman-Kay, J. D. RGG/RG Motif Regions in RNA Binding and Phase Separation. *J. Mol. Biol.* (**2018**) 430**,** 4650–4665.

[75] Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* (**2015**) 162**,** 1066–1077.

[76] Rhine, K., Vidaurre, V. & Myong, S. RNA Droplets. *Annu. Rev. Biophys.* (**2020**) 49**,** annurev-biophys-052118-115508.

[77] Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* (**2018**) 360**,** 918–921.

[78] Bratek-Skicki, A., Pancsa, R., Meszaros, B., Van Lindt, J. & Tompa, P. A Guide to Regulation of the Formation of Biomolecular Condensates. *FEBS J.* (**2020**) febs.15254. doi:10.1111/febs.15254

[79] Protter, D. S. W. *et al.* Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* (**2018**) 22**,** 1401–1412.

[80] Birsa, N., Bentham, M. P. & Fratta, P. Cytoplasmic functions of TDP-43 and FUS and their role in ALS. *Semin. Cell Dev. Biol.* (**2019**) 1–9. doi: 10.1016/j.semcdb.2019.05.023

[81] Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* (**2015**) 60**,** 231–241.

[82] Ederle, H. & Dormann, D. TDP-43 and FUS en route from the nucleus to the cytoplasm. *FEBS Lett.* (**2017**) 591**,** 1489–1507.

[83] Baron, D. M. *et al.* Amyotrophic lateral sclerosis-linked FUS/TLS alters stress granule assembly and dynamics. *Mol. Neurodegener.* (**2013**) 8**,** 1–18.

[84] Mitrea, D. M. *et al.* Methods for Physical Characterization of Phase-Separated Bodies and Membrane-less Organelles. *J. Mol. Biol.* (**2018**) 430**,** 4773–4805.

[85] Wang, Z., Zhang, G. & Zhang, H. Protocol for analyzing protein liquid–liquid phase separation. *Biophys. Reports* (**2019**) 5**,** 1–9.

[86] Cummings, C. S. & Obermeyer, A. C. Phase Separation Behavior of Supercharged Proteins and Polyelectrolytes. *Biochemistry* (**2018**) 57**,** 314–323.

[87] Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* (**2018**) 361**,**

[88] Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* (**2016**) 165**,** 1686–1697.

[89] Riback, J. A. *et al.* Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* (**2017**) 168**,** 1028-1040.e19.

[90] Murthy, A. C. & Fawzi, N. L. The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *J. Biol. Chem.* (**2020**) 1–19. doi:10.1074/jbc.REV119.009847

[91] Wheeler, R. J. *et al.* Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv* (**2019**) 721001. doi:10.1101/721001

[92] Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U. S. A.* (**2017**) 114**,** E8194–E8203.

[93] Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* (**2017**) 171**,** 615-627.e16.

[94] Jr., Istvan Lazar, I. L. S. GelAnalyzer 19.1 (www.gelanalyzer.com).

[95] Walls, D. & Loughran, S. T. Purification of proteins fused to glutathione S-tranferase. *Methods Mol. Biol.* (**2011**) 681**,** 151–175.

[96] Azatian, S. B., Kaur, N. & Latham, M. P. Increasing the buffering capacity of minimal media leads to higher protein yield. *J. Biomol. NMR* (**2019**) 73**,** 11–17.

[97] Blommel, P. G. & Fox, B. G. A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr. Purif.* (**2007**) 55**,** 53–68.

[98] Liu, J., Sneeden, J. & Heyer, W. In Vitro Assays for DNA Pairing and Recombination-Associated DNA Synthesis Jie Liu, Jessica Sneeden, and Wolf-Dietrich Heyer Abstract. *Methods Mol. Biol.* (**2011**) 745**,** 363–383.

[99] Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* (**2005**) 571–607. doi:10.1385/1592598900

[100] Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* (**2005**) 59**,** 687–696.

[101] Ahlner, A., Carlsson, M., Jonsson, B. H. & Lundström, P. PINT: A software for integration of peak volumes and extraction of relaxation rates. *J. Biomol. NMR* (**2013**) 56**,** 191–202.

[102] Niklasson, M. *et al.* Comprehensive analysis of NMR data using advanced line shape fitting. *J. Biomol. NMR* (**2017**) 69**,** 93–99.

[103] Veltri, T. *et al.* Amide hydrogens reveal a temperature-dependent structural transition that enhances site-II Ca2+-binding affinity in a C-domain mutant of cardiac troponin C. *Sci. Rep.* (**2017**) 7**,** 1–14.

[104] Williamson, M. P. Using chemical shift perturbation to characterize ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* (**2013**) 73**,** 1–16.

[105] Hyde, A. M. *et al.* General Principles and Strategies for Salting-Out Informed by the Hofmeister Series. *Org. Process Res. Dev.* (**2017**) 21**,** 1355–1370.

[106] Wong, L. E., Kim, T. H., Muhandiram, D. R., Forman-Kay, J. D. & Kay, L. E. NMR experiments for studies of dilute and condensed protein phases: Application to the phase-separating protein CAPRIN1. *J. Am. Chem. Soc.* (**2020**) jacs.9b12208. doi:10.1021/jacs.9b12208

References

[107] Hamad, N. *et al.* Direct visualization of the conformational change of FUS/TLS upon binding to promoter-associated non-coding RNA. *Chem. Commun.* (**2020**) 56**,** 9134–9137.

[108] Schwartz, J. C., Wang, X., Podell, E. R. & Cech, T. R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Rep.* (**2013**) 5**,** 918–925.

[109] Bergstrom, J., Furst, P., Noree, L. O. & Vinnars, E. Intracellular free amino acid concentration in human muscle tissue. *J. Appl. Physiol.* (**1974**) 36**,** 693–697.

[110] Danbolt, N. C. Glutamate uptake. *Prog. Neurobiol.* (**2001**) 65**,** 1–105.

[111] Featherstone, D. E. Intercellular glutamate signaling in the nervous system and beyond. *ACS Chem. Neurosci.* (**2010**) 1**,** 4–12.

[112] Shin, S., Mohan, S. & Fung, H. L. Intracellular l-arginine concentration does not determine NO production in endothelial cells: Implications on the 'l-arginine paradox'. *Biochem. Biophys. Res. Commun.* (**2011**) 414**,** 660–663.

[113] Kang, J., Lim, L. & Song, J. ATP enhances at low concentrations but dissolves at high concentrations liquid-liquid phase separation (LLPS) of ALS/FTD-causing FUS. *Biochem. Biophys. Res. Commun.* (**2018**) 504**,** 545–551.

[114] Musafia, B., Buchner, V. & Arad, D. Complex salt bridges in proteins: Statistical analysis of structure and function. *J. Mol. Biol.* (**1995**) 254**,** 761–770.

[115] Acids, A. & Dougherty, D. a. Cation-π Interactions Involving Aromatic. *J. Nutr.* (**2007**) 137**,** 1504–1508.

[116] Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A.* (**1999**) 96**,** 9459–9464.

[117] Teilum, K., Kunze, M. B. A., Erlendsson, S. & Kragelund, B. B. (S)Pinning down protein interactions by NMR. *Protein Sci.* (**2017**) 26**,** 436–451.

[118] Wong, K. B., Freund, S. M. V. & Fersht, A. R. Cold denaturation of barstar: 1H,15N and 13C NMR assignment and characterisation of residual structure. *J. Mol. Biol.* (**1996**) 259**,** 805–818.

[119] Kamatari, Y. O. *et al.* Response of native and denatured hen lysozyme to high pressure studied by 15N1H NMR spectroscopy. *Eur. J. Biochem.* (**2001**) 268**,** 1782–1793.

[120] Korsak, M. & Kozyreva, T. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*. *Advances in Experimental Medicine and Biology* (Springer International Publishing, 2015). 870**,**

[121] Varnay, I. *et al.* Optimized measurement temperature gives access to the solution structure of a 49 kDa homohexameric β-propeller. *J. Am. Chem. Soc.* (**2010**) 132**,** 15692–15698.

[122] Gell, D. A., Kwan, A. H. & Mackay, J. P. *NMR spectroscopy in the analysis of protein-protein interactions*. *Modern Magnetic Resonance* (2018). doi:10.1007/978-3-319-28388-3_121

[123] Baxter, N. J. & Williamson, M. P. Temperature dependence of 1H chemical shifts in proteins. *J. Biomol. NMR* (**1997**) 9**,** 359–369.

[124] Li, H. R., Chiang, W. C., Chou, P. C., Wang, W. J. & Huang, J. rong. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic

residues. *J. Biol. Chem.* (**2018**) 293**,** 6090–6098.

[125] Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* (**2018**) 7**,** 1–48.

[126] Hoklotubbe, S. S. Exchange Processes. in *Fundamentals of Protein NMR Spectroscopy* (Springer-Verlag, 2018). 403–430. doi:10.1007/1-4020-3500-4_18

[127] Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* (**2009**) 5**,** 808–814.

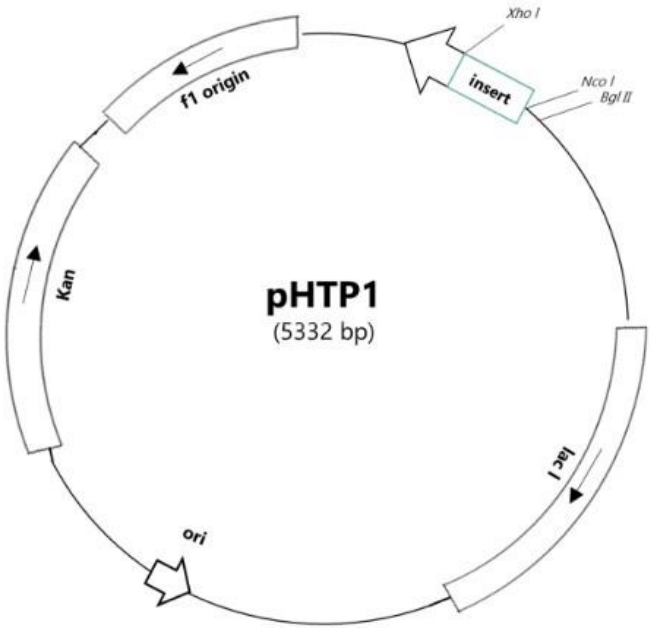
[128] Rehm, T., Huber, R. & Holak, T. A. Application of NMR in structural proteomics: Screening for proteins amenable to structural analysis. *Structure* (**2002**) 10**,** 1613–1618.

[129] Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* (**2001**) 21**,** 249–261.

[130] Wishart, D. S., Sykes, B. D. & Richards, F. M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* (**1991**) 222**,** 311–333.

[131] Gui, X. *et al.* Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* (**2019**) 10**,** 2006

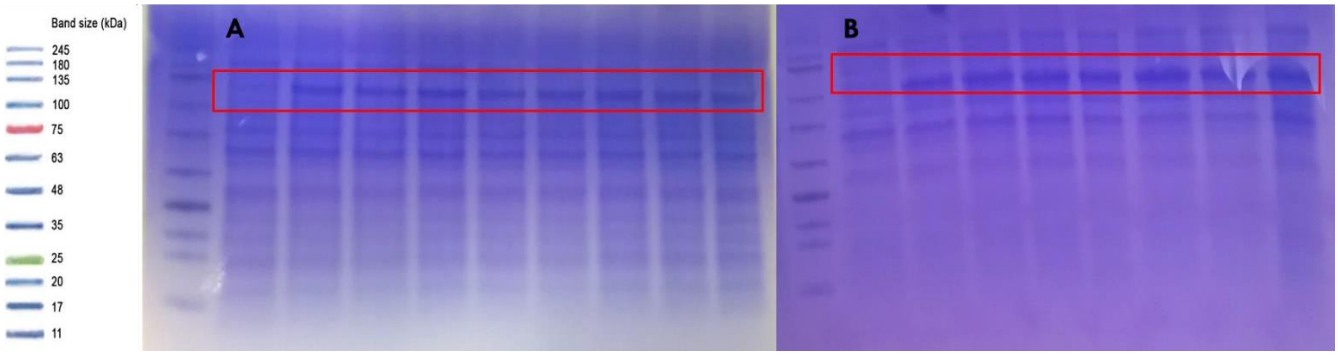
7 Appendix



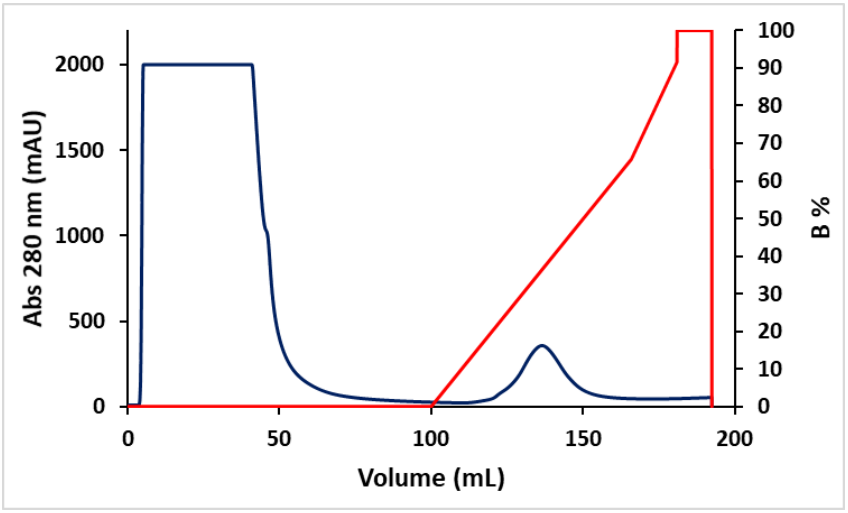
Appendix 1 Vector map of pHTP1 vector (NZYTech).

Appendix 2 Composition of 2x M9 medium for isotopically enriched protein expression.

Supplement	Concentration	Source
Na_2HPO_4	14 g/L	Fisher Chemical
KH_2PO_4	6 g/L	Riedel de Haen
NaCl	1 g/L	Fisher Chemical
$FeSO_4$	100 µM	PanReac Applichem
$CaCl_2$	2 mM	PanReac Applichem
MEM Vitamins	1 %	Sigma Aldrich (Merck)
$MgSO_4$	100 µM	PanReac Applichem
$^{15}NH_4Cl$	2 g/L	Cambridge Isotope Laboratories
^{13}C glucose	3 g/L	Cambridge Isotope Laboratories
^{12}C glucose (^{15}N enrichment only)	4 g/L	Scharlau

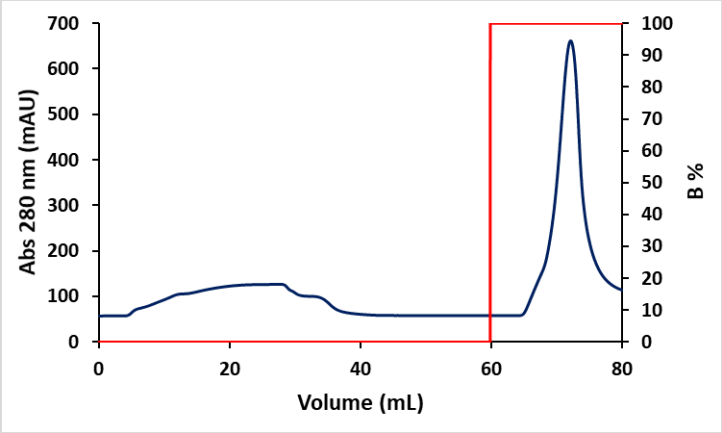


Appendix 3 Protein expression optimization gels (examples). An example of the expression of MBP-RGG$_1$-RRM-RRG$_2$ is presented in **A** and MBP-RGG$_2$-ZnF-RGG$_3$ in **B**. The protein marker used in the SDS-PAGE is presented on the left.

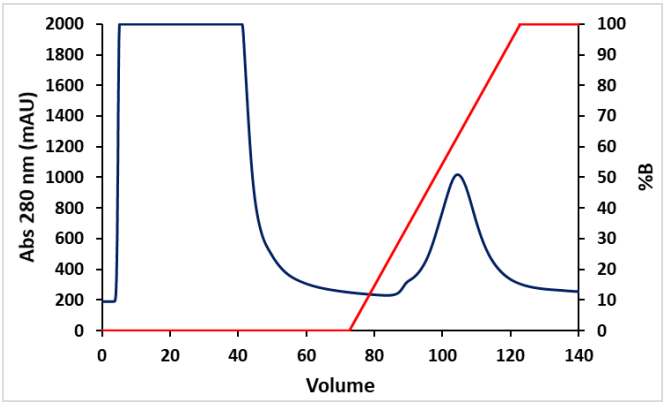


Appendix 4 IMAC Chromatogram of MBP-RGG$_1$-RRM-RGG$_2$. Two 5 mL FF Crude Ni-NTA columns were used in this step. A total of 30 mL of sample was injected.

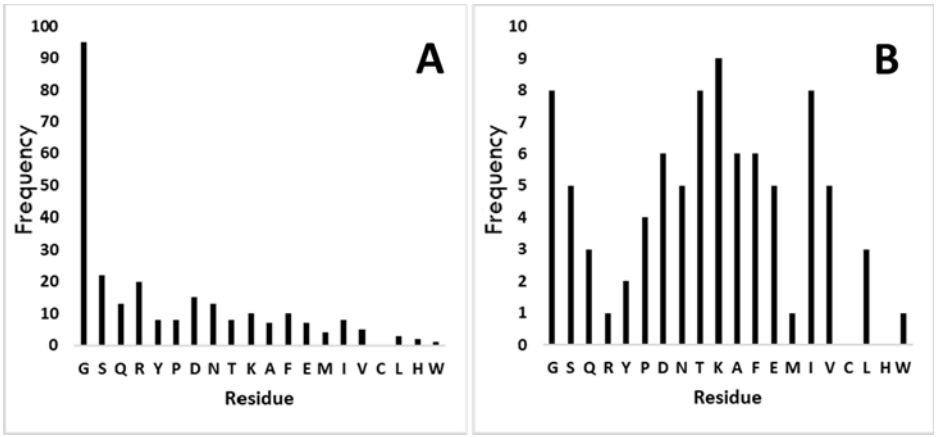
Appendix



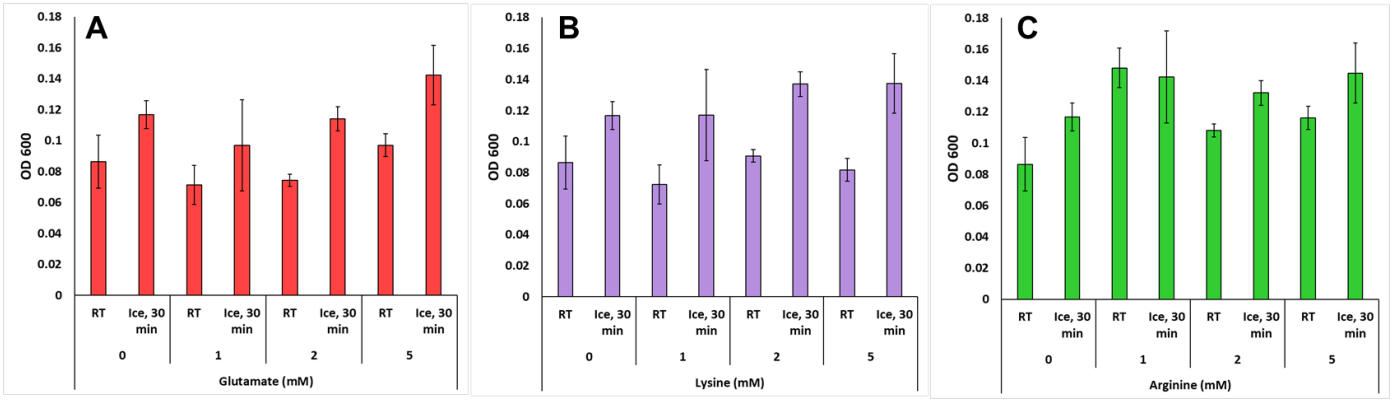
Appendix 5 IMAC chromatogram RGG_1-RRM-RGG_2. Purification step performed after TEV digestion of MBP-RGG_1-RRM-RGG_2. A total of 16 mL was injected. A single 5 mL FF Crude Ni-NTA column was used in this procedure.



Appendix 6 IMAC Chromatogram MBP-RGG_2-ZnF-RGG_3. Two 5 mL FF Crude Ni-NTA columns were used in this step. A total of 30 mL of sample was injected.

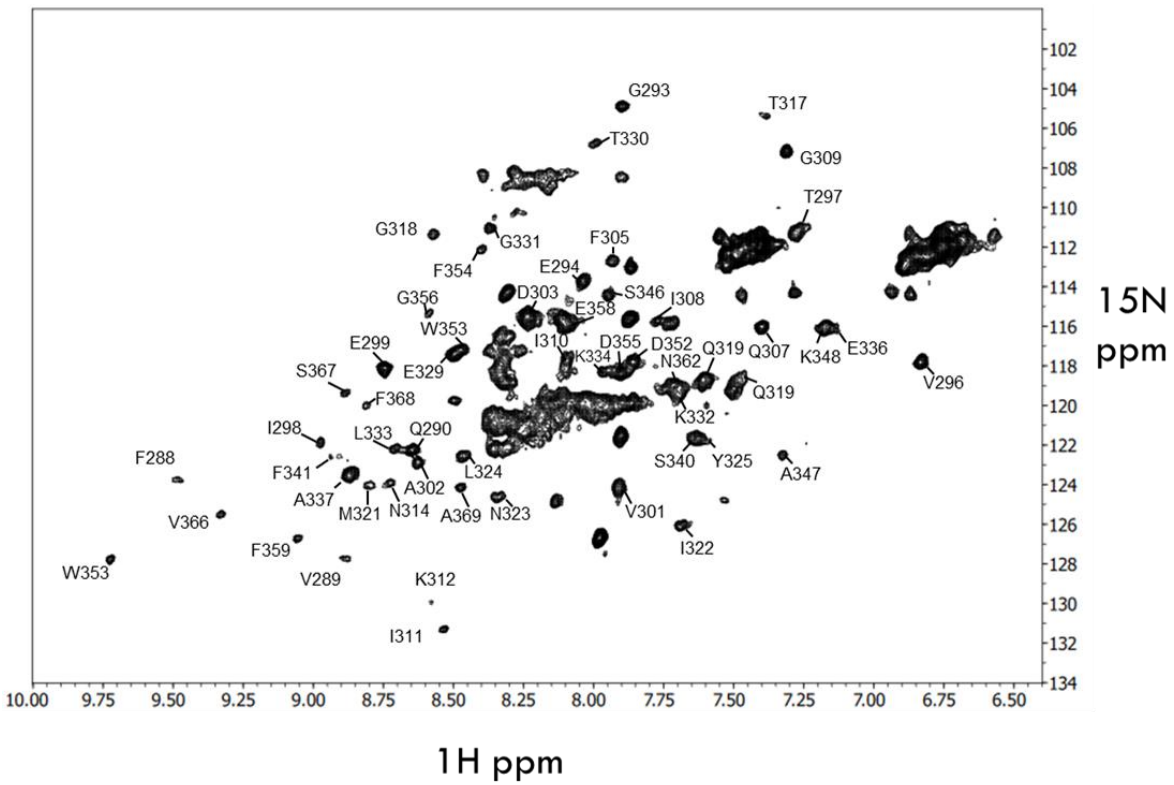


Appendix 7 Frequency of amino acid type in the RGG_1-RRM-RGG_2 construct (**A**) and in the RRM domain (**B**).

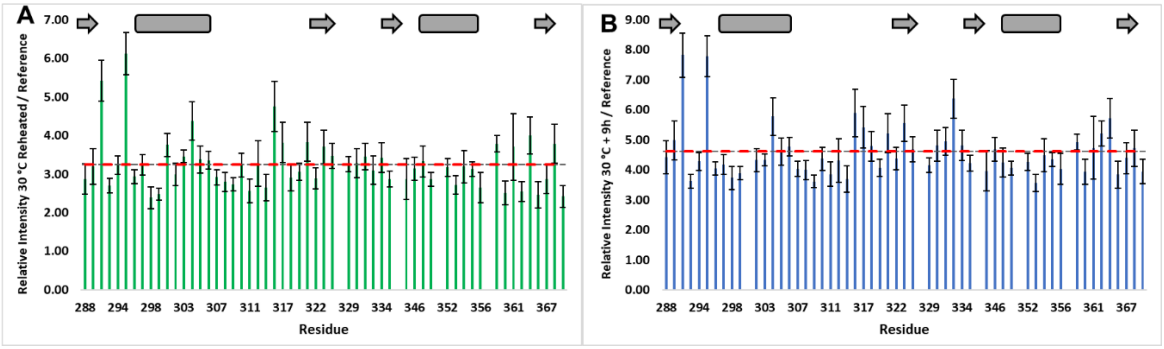


Appendix 8 Turbidity measurements (OD 600 nm) at varying concentrations of glutamate (**A**), lysine (**B**) and arginine (**C**), 0-5 mM. All measurements were performed in triplicate and error bars represent the standard deviation.

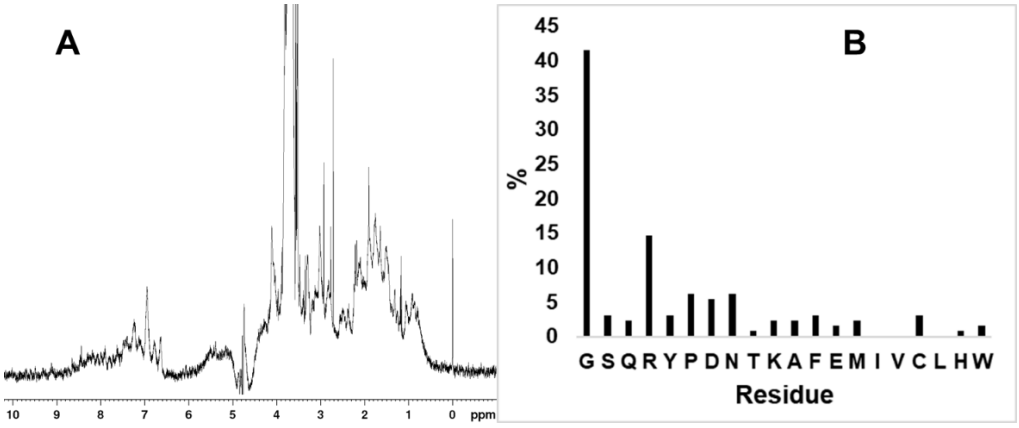
Appendix



Appendix 9 Partially assigned $^{1}H,^{15}N$ HSQC spectrum of 800 µM RGG_1-RRM-RGG_2 at 30 °C in 20 mM Tris-HCl pH 7.0, 100 mM NaCl, 2 mM β-mercaptoethanol and 0.05 % NaN_3.



Appendix 10 $^{1}H,^{15}N$ HSQC cross-peak intensity of RGG_1-RRM-RGG_2 at 30 °C after sample reheating (**A**) and after 9 h incubation at 30 °C (**B**), relatively to intensities at 30 °C prior cooling. Error bars represent the standard error associated to the intensity ratio. Red dashed line represents the average intensity ratio for each plot.



Appendix 11 ^{1}H spectrum of RGG_2-ZnF-RGG_3, was acquired at 25 °C and 50 mM Tris-HCl pH 7.1 (**A**). The percentual abundance of each amino acid in RGG_2-ZnF-RGG_3 (**B**).