Published in final edited form as:

Methods. 2017 August 15; 126: 3–11. doi:10.1016/j.ymeth.2017.06.011.

Cross- β polymerization and hydrogel formation by low-complexity sequence proteins

Masato Kato*, Yi Lin, and Steven L. McKnight*

Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas TX 75390-9152

Abstract

Low-complexity (LC) sequences, typically believed to be incapable of assuming structural order, are abundant constituents of the proteomes of all eukaryotic organisms. These sequences have emerged as critical components for formation of meso-scaled, sub-cellular organelles not invested by surrounding membranes, exemplified by RNA granules. We have observed that LC domains of many RNA binding proteins known to be constituents of RNA granules readily form labile cross- β polymers under physiological conditions. Several lines of experimentation have shown that formation of labile, cross- β polymers assembled from LC domain monomers is important for formation of RNA granules. Among the various experiments we have carried out, hydrogel binding assays have evolved as a versatile technique allowing a reliable means of assessing polymer formation and the binding of heterotypic cellular components integral to the formation of RNA granules. This article presents methods allowing for the production of hydrogel droplets composed of LC domain polymers. We further describe methods allowing straightforward assessment for binding of test LC domains to hydrogel droplets by fluorescence microscopy.

1. Introduction

Among the human genome, more than 10% of our encoded proteins contain regions/ domains characterized by a low-complexity (LC) of amino acid composition [1, 2]. Instead of using a normal distribution of the 20 amino acids required to specify structurally ordered proteins, LC domains are typically composed of only 3–5 types of amino acids. One of the major groups of proteins carrying LC domains are the RNA and DNA binding proteins required for proper information flow from gene to message to protein [3]. Heretofore, LC domains have been thought to be intrinsically disordered and devoid of three dimensional structure. Despite being assumed to function *sans* molecular structure, extensive evidence has confirmed the biologic utility of LC domains [4–6]. It has long been known, for example, that the activation domains of transcription factors are of low complexity with respect to amino acid sequence [7].

*Corresponding authors. masato.kato@utsouthwestern.edu and steven.mcknight@utsouthwestern.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Over the past 5–10 years, information coming primarily from human genetics studies of patients suffering from neurodegenerative disease have identified numerous disease-causing mutations that often map within in LC domains [8, 9]. Given obvious biological and medical significance, we offer that mechanistic studies of LC domains should be of value as a prelude to understanding their function and involvement in both normal biology and neurological disease. Historically, biochemical studies of LC domains have been hampered because of their susceptibility to proteolysis and predilection to aggregation. Here we describe methods allowing production and purification of large amounts of numerous LC domains. When purified, concentrated and incubated under physiological conditions, LC domains readily form labile cross- β polymers (Figure 1A) [10–12]. Polymerization triggers a phase transition of otherwise soluble protein to a hydrogel state composed of a uniform network of cross- β polymers. Although morphologically indistinguishable from irreversible, pathogenic amyloid fibers, LC domain polymers are readily labile to disassembly [10–12]. Four independent lines of evidence favor the biologic utility of labile LC domain polymers. First, mutagenesis experiments have shown that mutations affecting formation of LC domain polymers correlatively affect their incorporation into RNA granules in living cells [10, 13]. Second, the molecular architecture of LC domain polymers produced in test tubes using recombinant protein closely resembles that found in the nuclei of mammalian cells [14]. Third, the hierarchical melting of LC domain polymers by aliphatic alcohols, as deduced in test tube reactions composed of polymers formed from purified, recombinant protein, matches the ability of the same alcohols to melt RNA granules in living cells [11]. Fourth, a substantive proportion of mutations causative of neurodegenerative disease have been found to enhance polymer stability and map to the core-forming regions of LC domain polymers [9, 11, 14].

A number of research groups published reports starting in 2015 giving evidence that LC domains can become phase separated into liquid-like droplets upon incubation in vitro [15–21]. We had long observed this “oiling out” or “clouding” of LC domain protein samples prior to hydrogel formation. By use of an N-acetylimidazole (NAI) footprinting method, it was possible to show that LC domains begin to form cross- β polymers even before liquid-like droplet formation [14]. NAI footprinting provided direct evidence of enhanced polymerization of LC domains as a function of the transition from soluble to liquid-like droplet state to hydrogel state. We have parsimoniously concluded that the same structural forces, centered upon the formation of highly specific cross- β polymers, are responsible for the ability of LC domains to undergo phase transitions into either the liquid-like or hydrogel state [14]. Importantly, this conclusion is in stark contrast to that of almost all other authors of papers describing LC-driven formation of liquid-like droplets [15–21]. The latter papers uniformly conclude that LC domains do not adopt molecular structure as a function of transition from solubility to liquid-like droplets. These scientists instead conclude that LC domains interact via “fuzzy”, “molten” or “slithering” interactions devoid of protein structure, and that the labile, cross- β polymers we have characterized are of nothing more than pathologic significance. This controversy could hardly be more stark or contentious. It is hoped that a combination of time and additional experimentation will lead to its resolution.

As a tool to study LC domains, hydrogel droplets offer several distinct advantages. First, hydrogels are stable to solution. It is thus easy to develop and perform binding assays. Hydrogels can be used for binding assessment of purified, recombinant test proteins – such as the CTD of RNA polymerase II, the SR domains associated with pre-mRNA splicing factors [22, 23], or the LC domain that was itself used to form the hydrogel. Likewise, hydrogels can be used to selectively retrieve macromolecules from complex cellular lysates. As an example, mCherry:FUS hydrogel droplets exposed to mouse brain lysates were, by use of RNAseq, observed to selectively bind the mRNAs known to be incorporated into neuronal granules [13]. The fact that hydrogel droplets are capable, in an unbiased manner, of binding both the protein and RNA constituents of RNA granules gives some level of credence to both their biologic validity and experimental utility.

2. Preparation of purified LC domains

The major challenge associated with expression and purification of LC domain proteins is that they are labile to proteolysis and susceptible to amorphous aggregation. To overcome these technical challenges, we have accumulated know-how summarized below that is helpful in obtaining substantial quantities of purified LC domain fusion proteins using bacterial expression systems. Experimental variables including length of test LC sequence, solubilizing fusion tag, expression vector, overexpression methods and purification conditions are discussed hereafter point-by-point.

2.1 Construction of bacterial overexpression plasmids for LC domains

To both solubilize LC domains and facilitate visualization by fluorescence microscopy, we routinely fuse each test LC domain to mCherry or GFP. Prototypically, mCherry fusions are used to make hydrogel droplets, and test proteins for binding experiments are fusions between GFP and the LC domain of interest (see below). Based on our experience, these fluorescent proteins are nearly as good a protein tag to increase solubility as maltose binding protein (MBP) or glutathione-S-transferase (GST). Another factor that influences the solubility of an expressed LC domain is its length. If a full-length LC domain fused to mCherry or GFP can be purified at high yield without significant proteolysis, and also can be concentrated to a range of 50–100 mg/mL, it can be routinely used to make hydrogel droplets.

LC domains longer than 200 amino acids tend to be prone to aggregation and proteolysis. Thus, if unusually long LC domains (>300 aa) fail to be expressed and purified in an intact and soluble state, we often subdivide it into shorter fragments of 100–200 amino acids. To our knowledge, no reliable algorithm exists to accurately predict the polymer-forming core sequence within a test LC domain. Therefore, we routinely prepare constructs covering the entire LC domain segmented into stretches of 100–200 amino acids.

Structural studies of both pathogenic and labile LC domain polymers have given evidence that relatively short segments of 50–60 amino acids are sufficient for polymer formation. By definition, the A β _{1–40} and A β _{1–42} prion-like amyloids require less than 50 amino acids to polymerize [24, 25]. Likewise, pathogenically stable amyloids formed from the α -synuclein protein require only 40–50 residues to polymerize [26]. Finally, labile polymers formed

from the LC domain of FUS involve a segment of the LC domain roughly 60 residues in length (Murray, D., Kato, M., Lin, Y., McKnight, S.L. and Tycko, R., manuscript submitted). These various data give evidence that subdivision of unusually long LC domains may represent an acceptable experimental approach.

To obtain homogeneous hydrogel droplets, we concentrate the purified mCherry:LC domain fusion protein to 50–100 mg/mL. Given typical loss of the purified fusion proteins due to either proteolysis or aggregation, we begin with a large amount protein (at least 10 mg). We preferentially use pHis-parallel1 vector, which was developed by Sheffield et al. [27]. We insert the coding sequence of either mCherry or GFP at the NcoI site of this vector to construct pHis-parallel1-mCherry or pHis-parallel1-GFP vectors. The NcoI site is the first restriction site in the multi-cloning site of this vector, leaving plenty of restriction sites to insert LC domain coding DNA for the protein of interest. We typically employ the BL21(DE3) strain of *E. coli*. Overexpression of proteins using this vector in BL21(DE3) cells regularly yields 50–100 mg of soluble mCherry:FUS LC domain from 1 L of LB culture.

2.2 Overexpression of LC domains

If the expression and purification properties of an LC domain are unknown, it is best to start experiments at a small scale using the following protocol. All culture media described hereafter should be supplemented with appropriate antibiotics.

1. Start culture of transformed BL21(DE3) cells from a single colony in 1 mL LB media.
2. Shake the culture at 37°C for 3–4 hours.
3. Induce protein expression by adding IPTG at a final concentration of 0.5–1 mM (no IPTG for a control culture).
4. Continue to shake at 37°C for 2–3 hours.
5. Centrifuge 0.5 mL culture in a microcentrifuge tube.
6. Resuspend cell pellet in 100–200 μ L of B-PER reagent (Thermo Scientific).
7. Sonicate the cell suspension with a micro probe (1% power, 1 sec $\times$ 3 times with Fisher Scientific Model FB705).
8. Centrifuge at maximum speed on a bench-top microcentrifuge for 10 min.
9. Separate the supernatant and pellet, and prepare samples of each for SDS-PAGE.
10. Run a SDS-PAGE gel to see which fraction contains the LC domain fusion protein.

As described above, LC domains are susceptible to proteolytic degradation. The routine use of a protease inhibitor cocktail significantly limits degradation of LC domains during subsequent purification steps. If we recognize degradation of the test LC domain fusion protein after purification, it is likely that some or most of the degradation took place during expression in bacterial cells. One way to mitigate in-cell degradation is to perform

overexpression of the LC domain fusion protein at lower temperature, such as 16°C or 20°C. However, at lower temperature, speeds of both cell growth and protein expression are decreased. To compensate and enhance the final yield of LC domains, protein expression is carried out for overnight. The standard protocol is outlined below:

1. Transform BL21(DE3) cells with the plasmid of interest.
2. Prepare 1 L LB media. Depending on expression level, scale up the medium volume. We recommend a 4–6 L culture scale for mCherry:LC domain fusions to make hydrogels, and a 1 L scale for test GFP:LC domain fusions.
3. Inoculate 10 mL of LB media with a single colony. Incubate at 37 °C overnight without shaking.
4. Following morning - inoculate 1 L LB media with 10 mL of the pre-culture. Shake at 37 °C (180–220 rpm).
5. After 3–4 hours, check OD at 600 nm. If $OD_{600} = 0.6–0.8$ (no more than 1), decrease temperature to 20°C or 16°C. Wait 20–30 min, then add 0.5 mL of 1 M IPTG (final 0.5 mM). Shake at 20°C or 16°C for overnight.
6. Next morning - harvest cells at 4000 rpm for 20 min. Resuspend cells in 1xPBS, harvest by centrifugation and store cell pellet at –80°C.

After purification (described below), the purity of the LC domain fusion protein should be checked by SDS-PAGE (Figure 2). If severe degradation is observed, overexpression conditions need to be modified. An alternative modification is to shorten the overexpression time to 2–5 hours. In this case, an overexpression temperature of 37°C is preferred to compensate in enhancing the final yield of the expressed protein.

2.3 Purification of mCherry:LC domains

Purification protocols are slightly different between mCherry:LC domain fusions and GFP:LC domain fusions. We first describe purification of mCherry:LC domain fusions. If the mCherry:LC domain fusion protein is soluble, we purify it with Ni-NTA affinity resin under native conditions. Our standard purification protocol is described below.

1. Make a lysis buffer 50–200 mL depending on the culture volume.
 - 50 mM Tris-HCl pH 7.5
 - 500 mM NaCl
 - 1% Triton X-100
 - 20 mM β -mercaptoethanol (BME)
 - 1 tablet of protease inhibitors (Sigma S8830, 1 tablet is made for 100 mL)
 - 2 M urea or guanidine hydrochloride (GdnHCl), in the case of less soluble LC domain fusion proteins (see below)

2. Suspend cell pellet in lysis buffer. Use 30–40 mL of lysis buffer per 1–2 L culture. Add lysozyme (100 mg/mL, final 0.2 mg/mL) and incubate on ice for 20–30 min.
3. Sonicate the cell suspension on ice for 2–3 min (~65% power, 10 sec on + 30 sec off). Centrifuge at 35,000 rpm for 1 hour at 4°C.
4. Mix the supernatant with 20–40 mL of Ni-NTA resin in a bottle. Gently shake the bottle in the cold room for 10–20 min.
5. Pour the resin in an empty glass column.
6. Wash the resin in the column with 250–300 mL of wash buffer supplemented with 20 mM imidazole.

Wash Buffer (filtered with a 0.2 µm bottle-top filter)

20 mM Tris-HCl (pH 7.5)

500 mM NaCl

20 mM BME

0.1 mM phenylmethanesulfonyl fluoride (PMSF)

2 M urea or GdnHCl, in the case of less soluble LC domains (see below)

7. Elute the bound protein with wash buffer supplemented with 250 mM imidazole (~50 mL).
8. Add 0.5 M EDTA to the elution sample at a final concentration of 0.5 mM.
9. Add urea powder to the eluted protein to a final concentration of 2 M in case no urea is added in lysis or wash buffer.
10. Concentrate the protein with centrifugal ultrafiltration, such as Amicon Ultra, at 4°C. While concentrating, check whether the protein precipitates or not (by visual inspection for turbidity). When precipitation is first observed, discontinue concentration. Otherwise, concentrate to 30–40 mg/ml of mCherry:LC domain fusion protein.
11. Spin down particles/precipitate at the maximum speed of a bench-top microcentrifuge.
12. Aliquot in microcentrifuge tubes with 300–400 µL each and store at –80°C.
13. Check for fusion protein purity and integrity by SDS-PAGE (Figure 2).

At step 9, we usually add 2 M urea to the eluted protein. Urea helps keep LC domain fusion proteins soluble during concentration and storage. The mCherry tag is materially stable in 2 M urea. We have never detected loss of mCherry color of the protein solution during prolonged storage at –80°C.

We have also successfully prepared hydrogels from insoluble LC domain fusion proteins. To contend with precipitation-prone fusion proteins, we routinely perform a small-scale

solubility test, wherein less-soluble LC domain fusion proteins typically partition to the pellet fraction. In this case, we attempt purification under semi-denaturing conditions. To do so, we add 2 M urea in the lysis and wash buffers of the protocol described above. Urea helps solubilize LC domain fusion proteins in the lysis step, and keeps them soluble throughout purification. Again, the mCherry tag is materially stable in 2 M urea and not detectably denatured during purification.

2.4 Preparation of GFP:LC domains to be tested in hydrogel binding assays

LC domains to be tested for hydrogel binding should be constructed with a fluorescent protein tag other than mCherry, and we usually use GFP. We regularly use the pHis-parallel1-GFP vector described above. Once expression plasmids are constructed, a small-scale solubility test is performed. Overexpression and purification protocols are basically the same as those described for mCherry:LC domain fusion proteins described above. A slight modification to the purification protocol is as follows. For soluble LC domains, follow the protocol described in section 2.3 up to step 8, and then substitute the following steps:

- 9 Concentrate the protein by centrifugal ultrafiltration at 4°C. Certain GFP:LC domain fusion proteins form soluble aggregates that clog the ultrafiltration membrane at lower temperatures. If this is recognized, perform concentration step at room temperature. While concentrating, check whether protein solution shows evidence of precipitation.
- 10 When precipitation is observed, cease concentration and remove precipitated protein by microfuge centrifugation.
- 11 Transfer the supernatant into a microcentrifuge tube or a conical tube (15 mL).
- 12 Add an equal volume of 100% glycerol to the supernatant and mix well.
- 13 Store at −20°C.
- 14 Check fusion protein purity by SDS-PAGE (Figure 2).

For less soluble LC domains, we use 2 M urea throughout the purification protocol as described above. Follow the above protocol up to the step 12 (adding 100% glycerol). This step leads to a decrease in urea concentration from 2 M to 1 M. If any precipitate or cloudiness is observed, add urea powder to the solution to bring the urea concentration back to 2 M. Precipitates or cloudiness often disappear. If not, spin down the particles, and then store the supernatant at −20°C.

Particularly insoluble LC domains may not be solubilized by 2 M urea in the lysis step. In this case, we add 2 M guanidine hydrochloride (GdnHCl), which is a stronger denaturing reagent than urea. GFP and mCherry can survive in 2 M GdnHCl for a certain period of time. Add 2 M GdnHCl, instead of urea, to the lysis and wash buffers of the purification protocol (see above). Separately prepare additional wash buffer containing 2 M urea. Using the buffer supplemented with 2 M GdnHCl, follow the protocol in section 2.3 until step 6 (resin wash). After step 6, follow the modified protocol as described below.

- 7 Wash the resin further with 30–50 mL of 2 M urea wash buffer supplemented with 20 mM imidazole to replace GdnHCl with urea.

- 8 Elute the bound protein with the 2 M urea wash buffer supplemented with 250 mM imidazole (~50 mL).
- 9 Add 0.5 M EDTA to the elution at a final concentration of 0.5 mM.
- 10 Concentrate the protein by centrifugal ultrafiltration at 4°C. When signs of precipitation are observed, stop concentration and remove insoluble debris by centrifuge.
- 11 Transfer the supernatant into a microcentrifuge tube or a conical tube (15 mL).
- 12 Add an equal volume of 100% glycerol to the supernatant and mix well. If necessary, add urea powder to bring the urea concentration back to 2 M.
- 13 Store at -20°C.
- 14 Check protein purity by SDS-PAGE.

3. Hydrogel binding assay

As described in the introductory section, hydrogels are composed of labile cross- β polymers formed from LC domains (Figure 1A). The cross- β polymers in the hydrogel droplets effectively trap test LC domain applied to hydrogels as soluble monomers. Homotypic trapping of the same LC domain that was employed to produce the hydrogel is attributable to co-polymerization of the test protein to polymer ends within the hydrogel (Figure 1B) [10]. Therefore, the hydrogel binding assay is a facile system to study the polymerization capacity of test LC domains. Summarized below are protocols to make homogeneous hydrogel droplets and to assess binding of test LC domains to hydrogel droplets by fluorescent microscopy.

3.1 Preparation of hydrogel droplets

Cross- β polymerization depends on the identity of the test LC domain, protein concentration and time. Therefore, we usually start this process with overnight dialysis of the purified mCherry:LC domain fusion protein against gelation buffer (see below). The reason for concentrating the eluted, purified protein to 30–40 mg/mL at the step 10 of section 2.3 is to facilitate polymer nucleation during dialysis. We prefer overnight dialysis because it gives time for the mCherry:LC domain fusion protein to experience as many nucleation events as possible. During dialysis, some of the mCherry tag is cleaved within its fluorophore center (perhaps a side reaction of fluorescence excitement) (Figure 2), and the cleaved mCherry:LC domain fusion protein precipitates as white particles. Considerable protein can be lost at this step, and we often re-concentrate the sample after removal of the precipitate. Preparation of useful hydrogel droplets requires concentration of the fusion protein to 50 mg/mL or more. A second important step is sonication of the dialyzed protein solution. Sonication breaks existing polymers into many short fragments (seeds). These seeds grow into long polymers upon incubation with high concentrations of the soluble fusion protein. Thus, sonication significantly boosts polymer number and it is a key step required to obtain homogeneous hydrogel droplets. Over the years, we have realized that sonication can lead to amorphous aggregates if over-applied. Therefore, we minimize both the time and power of sonication (1 sec with power level 1% on our sonication machine, Fisher Scientific Model FB705) for one

sonication cycle. The number of sonication cycles requires optimization depending on the speed of polymer formation of test LC domains. Formation of FUS LC domain polymers is relatively slow. Thus, we apply three cycles within a 2–3 hour interval so that the number of polymers are amplified each cycle. Certain other LC domains, such as hnRNPA2 and TAF15, require only one cycle of sonication because polymerization is more rapid than that observed for the FUS LC domain. We summarize the protocol for formation of mCherry:FUS LC hydrogel droplets below:

1. Dialyze overnight a solution of mCherry: FUS LC domain fusion protein (30–40 mg/mL, 300–400 μ L) against a gelation buffer (1 L) to remove salt and imidazole.

Gelation buffer:

20 mM Tris-HCl pH7.5

200 mM NaCl

20 mM BME

0.1 mM PMSF

0.5 mM EDTA

2. After overnight dialysis, the protein solution may be partially cloudy.
3. Transfer the solution to a microcentrifuge tube and sonicate with a micro sonication probe (1% power, 1 sec $\times$ 3 times with Fisher Scientific Model FB705).
4. Spin down the solution at maximum speed of a table-top microcentrifuge for 1 min to remove white precipitate.
5. Concentrate the supernatant to 50–60 mg/mL by Amicon Ultra (0.5-mL device). The volume of the protein solution is usually 100–150 μ L.
6. Incubate for 2–3 hours at room temperature.
7. While waiting, prepare glass bottom dishes. Place a paper strip around the dish perimeter (10 cm $\times$ 0.5 cm) as shown in Figure 3.
8. Apply sonication with a micro probe (1% power, 1 sec $\times$ 2 times).
9. Incubate for 30 min – 1 hour at room temperature or until the solution becomes slightly viscous.
10. While waiting, wet the paper strip in the dishes with 35 μ L of gelation buffer.
11. Apply sonication one last time with micro probe (1% power, 1 sec $\times$ 1 times).
12. Spin down the solution at the maximum speed for 1 min.
13. Immediately, make 0.5- μ L drops on the glass bottom of the prepared dishes. Multiple droplets can be made in a single dish.

14. Seal the dish with parafilm and avoid a sudden movement that can cause deformation of droplets.
15. Leave the dishes at room temperature for two to four days. Keep leftovers of the protein solution in the original microcentrifuge tube. When the leftover protein gels, the droplets in the dishes will also have gelled.

It is straightforward to test whether useful hydrogels have been formed by adding gelation buffer (1 mL) to the hydrogel-containing dish. If droplets are of poor quality, we visually observe release of mCherry color from the droplets. If droplets are stable, the hydrogel binding assay can then be performed (see below). One problem with the current protocol is that we cannot store hydrogel dishes for prolonged periods of time because sealing by parafilm is imperfect. As a result, hydrogels droplets will eventually dry out after roughly 1 week. Once dried out, hydrogel droplets suffer surface cracks, and pieces of the cracked portion dissociate from the droplet.

We have prepared hydrogels from many different LC domains using the protocols described herein (Table 1). If an LC domain can polymerize into labile, cross- β polymers deduced by either electron microscopy or fluorescent light microscopy, that preparation is nearly certain to form experimentally useful hydrogel droplets.

3.2 Preparation of test GFP:LC domain solution for hydrogel binding

Once hydrogel droplets and GFP:LC test proteins are prepared, proceed to performing hydrogel binding assay. The protocol to prepare binding mixture is as follows.

1. Measure the protein concentrations of the GFP:LC domain to be tested. Dilute the GFP:LC domain in 1 mL of gelation buffer to a final concentration of 1 μ M.
2. Open the sealed hydrogel dish, remove wetted paper strip.
3. Slowly apply the GFP:LC domain solution into the dish so that hydrogel droplet(s) are covered by the solution.
4. Seal the dish with parafilm.
5. Place the dish at 4°C overnight.

3.3 Visualization of hydrogel binding by fluorescence microscopy

After incubation, the hydrogel droplets are examined by confocal fluorescence microscopy. We regularly use Leica SP5 or Zeiss LSM510 (now updated to LSM880) confocal microscopes available within the live cell imaging core facility at UT southwestern. We scan a horizontal section of a hydrogel droplet (Figure 4) to visualize the distinct signals of mCherry (for the hydrogel droplet) and GFP (for the bound test LC domain). Important microscope parameters we use are summarized below.

- Objective lens: use a 10 $\times$ lens to capture an entire droplet.
- Pin hole size: adjust to obtain a 15- μ m thickness of a confocal section.
- 2 channel mode for mCherry and GFP. Use standard settings (excitation, emission, filter, etc).

- Laser power: for mCherry 5–10%, for GFP ~20%.

To scan the hydrogel droplets, follow the protocol below.

1. Place hydrogel dish on a microscope stage. We use a standard stage for a culture dish.
2. Check the droplets through eye pieces with filter tuned to mCherry color. If the dish contains multiple droplets, choose a homogeneous, round-shaped droplet.
3. Visualize the droplet through a camera tuned to the mCherry channel in real-time mode.
4. Center the droplet in a visible frame (X and Y directions), and then adjust a focal point (Z direction) to maximum intensity of the hydrogel droplet. This point is usually slightly above the glass surface.
5. Adjust the gain parameter so that no saturated pixel is present.
6. Change the channel to GFP in the real-time mode.
7. Adjust the gain parameter so that no saturated pixel is present. However, if comparison between different samples is required, the GFP gain must be the same between all samples.
8. Capture high-resolution images from both channels.

3.3 Evaluation of hydrogel binding

After securing digitized images, they can be quantitatively evaluated. As described earlier, the binding mechanism of this assay may represent extension of existing polymers in the hydrogel by the test GFP:LC domain molecules (Figure 1B). Therefore, the intensity of the GFP signal correlates with polymerization capacity of the test LC domain into existing polymers of which the hydrogel droplet is composed. Typically, we observe three different binding patterns (Figure 5).

- Pattern 1: homotypic binding

When we test the same LC domain that is used to make the hydrogel droplet, GFP signals are typically localized to the rim of the droplet (Figure 5A). This peripheral binding signal is due to rapid polymerization of the test protein into polymers located at the outer rim of the hydrogel droplet. Rapid capture of the test protein at the hydrogel perimeter prevents the GFP fusion protein from penetrating into the central region of the droplet.

- Pattern 2: heterotypic binding

Co-polymerization of a different LC domain with an existing hydrogel polymer tends to be slower than that of the homotypic case. Therefore, the test LC domain can diffuse into the center of the hydrogel droplet and the GFP signal tends to be evenly detectable across the droplet (Figure 5B). Depending on the efficiency of co-polymerization, the GFP signal intensity is varied.

- Pattern 3: no binding

If a test LC domain does not have the capability to form cross- β polymers, we do not see any accumulation of GFP signals on the hydrogel droplet (Figure 5C). At high gain, GFP signals inside and outside of the hydrogel droplet are virtually the same, indicating that the test GFP:LC domain freely diffuses in the droplet, but is not bound. In some cases, the GFP signal inside of the droplet may be weaker than that observed on the outside. The Stokes radius of such an LC domain may be so large that it may be excluded from the hydrogel droplet.

We have confirmed that intensities of GFP signal increase linearly as a function of time up to 24 hours. The intensities may still rise up until 48–72 hours, but we are uncertain that the increase is linear after 24 hours. Therefore, when we compare different samples, we measure intensities after no more than 24 hour incubation. To quantify observed intensities, we use the graphics program ImageJ [28]. Methods to quantify the intensities are different between the patterns described above. When we observe homogeneous GFP signals across the hydrogel droplet (Pattern 2, Figure 5B), we quantify the intensity of the GFP signal by using the area measurement method in ImageJ as follows (Figure 6A).

1. Choose an oval selection mode.
2. Embrace the droplet with an oval selection. Try to fit the oval selection to the droplet as closely as possible.
3. Measure the mean intensity (Analyze -> Measure).

For pattern 1 (Figure 5A), however, we cannot use the area measurement method. GFP signals are localized at the hydrogel rim, and the majority of the central area of the droplet is dark. In this case, we use the profile plot mode as follows (Figure 6B).

1. Choose a straight line mode.
2. Draw a straight line across the hydrogel droplet passing through the droplet center.
3. Draw a profile (Analyze -> Plot Profile).
4. Measure the peak values at either rim of the droplet.

We have used both of the above measurement methods to study effects of tyrosine-to-serine mutations of LC domains from FUS and hnRNP A2 for hydrogel binding [14, 23]. The results were compared with other functional measurements of those mutants, and we have obtained strong correlation between two results, indicating that both measurement methods are reliable.

4. Discussion

LC domains have been difficult to study biochemically because of their tendency to form amorphous aggregates and their susceptibility to proteolytic degradation. Here we describe methods for the study of LC domains capable of forming labile, cross- β polymers. Having optimized a set of readily defined conditions (length of amino acid sequence, solubilizing tag, expression vector, overexpression and purification conditions, etc), we can express and

purify large amounts of LC domains as fusion proteins attached to mCherry, GFP or other solubilizing tags.

We often use 2 M urea in preparation of LC domains linked to mCherry or GFP. The supplemented urea might cause partial denaturing of these fluorescent mCherry or GFP tags, with unfolded regions causing the formation of amyloid-like polymers. Indeed, some globular proteins such as myoglobin, lysozyme or histones have been reported to form amyloid fibers after exposure to moderate concentration of denaturants, low pH, high pH, high temperature, or a combination of these conditions [29–31]. These conditions partially unfold the structure of these globular proteins, and unfolded regions can polymerize into amyloid fibers. We have never observed that the fluorescent mCherry or GFP protein tags alone form polymers/hydrogels after exposure to 2 M urea. However, the LC domains fused to mCherry or GFP form polymers/hydrogels after the same treatment. We are, therefore, confident that polymer/hydrogel formation is specifically mediated by LC domains.

Unlike the globular proteins mentioned above, we have never observed that LC domains polymerize in 2 M urea. LC domains remain soluble in 2 M urea even though they are largely unfolded. When LC domains are dialyzed against physiological buffers at neutral pH in the absence of denaturant, they begin a transition from a disordered state to a polymeric state. Therefore, LC domains and globular proteins are similar in terms of proceeding towards polymer formation from an unfolded state. One distinct difference in the polymers formed from denatured globular proteins and LC domains is polymer stability. Polymers formed from denatured globular proteins are prion-like in their extreme stability to chaotropic reagents and extreme denaturants including SDS. Polymers formed from LC domains are instead fully labile to disassembly, even upon simple dilution [10–12].

The transition of LC domains from a disordered state to a polymeric state upon dialysis is reminiscent of the refolding step of fully-denatured globular proteins. Denatured globular proteins start refolding upon dialysis in buffers lacking denaturants. In this case, each single polypeptide intramolecularly folds back to its own structure because the amino acid sequence has enough information to direct proper folding [32]. Polymerization of LC domains upon dialysis in physiological buffer is analogous to this refolding step. However, LC domains do not have enough sequence information to fold within a single molecule. Thus, LC domains require intermolecular contact to fold into polymers. Therefore, we predict that polymerization represents the obligate and natural folding step for LC domains. We emphasize, however, that LC domain polymers are likely to be short and dynamic when employed *in vivo*. Polymeric cross- β interactions of as little as two or three associated polypeptides may be fully adequate to serve the biologic function of LC domains in living cells.

The methods we have developed here have assisted many groups now actively involved in the study of LC domain-containing proteins. Indeed, subsequent to our description of LC polymerization and hydrogel formation in 2012 [10, 13], many research groups have begun and extended work on LC domains, including the highly active field of liquid-like droplets and phase separation [15–19, 21, 33]. Our data, disputed by others, have given evidence that both liquid-like droplets and hydrogels are driven by the same molecular forces – the

formation of labile, cross- β polymers. We propose that the simple difference between liquid-like droplets and hydrogels can be accounted for by polymer length. We predict that initial polymer nucleation, followed by the generation of short polymers, leads to the phase separation of LC domains and formation of liquid-like droplets. Prolonged incubation of solutions containing LC domains leads to the formation of much longer polymers that trigger the transition of liquid-like droplets into the hydrogel state [15].

Different cellular demands may prefer the use of shorter or longer cross- β polymers. An RNA granule meant to be moved the length of a motorneuron axon, as much as a meter in length [34], or a granule deposited at the pole of an insect or amphibian embryo for long periods of time [35], might utilize longer cross- β polymers. By contrast, much more dynamic puncta might rely on cross- β polymers composed of only a handful of polymerized subunits. Whether this simplistic thinking holds up to the test of time remains to be determined. Short of this, it is clear that the hydrogels described in this methods paper do offer experimental utility. They allow for clear and easy binary measures of LC binding; they allow complex mixtures of cellular lysate to be probed for the selective binding of either protein or RNA constituents that match the hydrogel/LC domain of interest [13, 14]; they have allowed development of the N-acetylimidazole footprinting technique [14]; they allow analysis of the effects of aliphatic alcohols [11, 12]; and they are enabling the elucidation of the molecular structures of labile, cross- β polymers by solid state NMR spectroscopy (Murray, D., Kato, M., Lin, Y., McKnight, S.L. and Tycko, R., manuscript submitted). We close by offering that the reductionist approaches described herein may help provide a mechanistic understanding of how LC domains function in living cells.

Acknowledgments

We thank the Live Cell Imaging and Electron Microscopy Core Laboratories at UTSMC for technical assistance. This work was supported by grant #5UO1-GM107623-02 awarded SLM from the National Institute of General Medical Sciences and unrestricted funds provided to SLM by an anonymous donor.

References

1. Rado-Trilla N, Alba M. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins. *BMC Evol Biol.* 2012; 12:155. [PubMed: 22920595]
2. Toll-Riera M, Rado-Trilla N, Martys F, Alba MM. Role of low-complexity sequences in the formation of novel protein coding sequences. *Mol Biol Evol.* 2012; 29(3):883–6. [PubMed: 22045997]
3. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* 2012; 1462:61–80. [PubMed: 22445064]
4. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* 2005; 6(3):197–208. [PubMed: 15738986]
5. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* 1999; 293(2):321–31. [PubMed: 10550212]
6. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry.* 2002; 41(21):6573–82. [PubMed: 12022860]
7. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. *Biochemistry.* 2006; 45(22):6873–88. [PubMed: 16734424]
8. Taylor JP, Brown RH Jr, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature.* 2016; 539(7628):197–206. [PubMed: 27830784]

9. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J.* 2017; 474(8):1417–1438. [PubMed: 28389532]
10. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell.* 2012; 149(4):753–67. [PubMed: 22579281]
11. Lin Y, Mori E, Kato M, Xiang S, Wu L, Kwon I, McKnight SL. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell.* 2016; 167(3):789–802 e12. [PubMed: 27768897]
12. Shi KY, Mori E, Nizami ZF, Lin Y, Kato M, Xiang S, Wu LC, Ding M, Yu Y, Gall JG, McKnight SL. Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proc Natl Acad Sci U S A.* 2017
13. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell.* 2012; 149(4):768–79. [PubMed: 22579282]
14. Xiang S, Kato M, Wu LC, Lin Y, Ding M, Zhang Y, Yu Y, McKnight SL. The LC Domain of hnRNP A2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. *Cell.* 2015; 163(4):829–39. [PubMed: 26544936]
15. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, Alberti S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell.* 2015; 162(5):1066–77. [PubMed: 26317470]
16. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowitz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell.* 2015; 57(5):936–47. [PubMed: 25747659]
17. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* 2015; 163(1):123–33. [PubMed: 26406374]
18. Lin Y, Protter DS, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell.* 2015; 60(2):208–19. [PubMed: 26412307]
19. Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC, Eckmann CR, Myong S, Brangwynne CP. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A.* 2015; 112(23):7189–94. [PubMed: 26015579]
20. Altmeyer M, Neelsen KJ, Teloni F, Pozdnyakova I, Pellegrino S, Grofte M, Rask MB, Streicher W, Jungmichel S, Nielsen ML, Lukas J. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun.* 2015; 6:8088. [PubMed: 26286827]
21. Burke KA, Janke AM, Rhine CL, Fawzi NL. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol Cell.* 2015; 60(2):231–41. [PubMed: 26455390]
22. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science.* 2014; 345(6201):1139–45. [PubMed: 25081482]
23. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell.* 2013; 155(5):1049–60. [PubMed: 24267890]
24. Paravastu AK, Leapman RD, Yau WM, Tycko R. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc Natl Acad Sci U S A.* 2008; 105(47):18349–54. [PubMed: 19015532]
25. Xiao Y, Ma B, McElheny D, Parthasarathy S, Long F, Hoshi M, Nussinov R, Ishii Y. Abeta(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol.* 2015; 22(6):499–505. [PubMed: 25938662]
26. Tuttle MD, Comellas G, Nieuwkoop AJ, Covell DJ, Berthold DA, Kloepper KD, Courtney JM, Kim JK, Barclay AM, Kendall A, Wan W, Stubbs G, Schwieters CD, Lee VM, George JM,

- Rienstra CM. Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat Struct Mol Biol.* 2016; 23(5):409–15. [PubMed: 27018801]
27. Sheffield P, Garrard S, Derewenda Z. Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. *Protein Expr Purif.* 1999; 15(1):34–9. [PubMed: 10024467]
28. Rasband, WS. ImageJ. 1997–2011. <http://imagej.nih.gov/ij/>
29. Fandrich M, Fletcher MA, Dobson CM. Amyloid fibrils from muscle myoglobin. *Nature.* 2001; 410(6825):165–6. [PubMed: 11242064]
30. Swaminathan R, Ravi VK, Kumar S, Kumar MV, Chandra N. Lysozyme: a model protein for amyloid research. *Adv Protein Chem Struct Biol.* 2011; 84:63–111. [PubMed: 21846563]
31. Topping TB, Gloss LM. The impact of solubility and electrostatics on fibril formation by the H3 and H4 histones. *Protein Sci.* 2011; 20(12):2060–73. [PubMed: 21953551]
32. Anfinsen CB, Haber E, Sela M, White FH Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A.* 1961; 47:1309–14. [PubMed: 13683522]
33. Murakami T, Qamar S, Lin JQ, Schierle GS, Rees E, Miyashita A, Costa AR, Dodd RB, Chan FT, Michel CH, Kronenberg-Versteeg D, Li Y, Yang SP, Wakutani Y, Meadows W, Ferry RR, Dong L, Tartaglia GG, Favrin G, Lin WL, Dickson DW, Zhen M, Ron D, Schmitt-Ulms G, Fraser PE, Shneider NA, Holt C, Vendruscolo M, Kaminski CF, George-Hyslop P. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron.* 2015; 88(4):678–90. [PubMed: 26526393]
34. Kiebler MA, Bassell GJ. Neuronal RNA granules: movers and makers. *Neuron.* 2006; 51(6):685–90. [PubMed: 16982415]
35. Boke E, Ruer M, Wuhr M, Coughlin M, Lemaitre R, Gygi SP, Alberti S, Drechsel D, Hyman AA, Mitchison TJ. Amyloid-like Self-Assembly of a Cellular Compartment. *Cell.* 2016; 166(3):637–50. [PubMed: 27471966]

Highlights

- Conditions to produce and purify large amounts of recombinant LC domains.
- Methods to make homogeneous LC domain hydrogels.
- Methods of deploying hydrogel binding assays.

Figure 1. Cross- β polymers and hydrogels formed from mCherry:FUS LC domain

(A) Left: Electron micrograph of cross- β polymers of mCherry:FUS LC domain fusion protein. Scale bar: 500 nm. Right: At higher protein concentration, polymers interact each other, build a mesh-network structure, and eventually become hydrogels.

(B) Schematic drawings of a hydrogel binding assay. Left: A pre-formed hydrogel droplet is soaked in solution of test GFP:LC domain fusion protein. Right: Hydrogel droplets accumulate test GFP:LC domain as it extends from existing mCherry:LC domain polymers.

Figure 2. SDS-PAGE gels of purified mCherry: or GFP: FUS LC domain fusions
FUS LC domain fused to mCherry or GFP is purified by Ni-NTA resin and then analyzed by SDS-PAGE. Left: mCherry:FUS LC domain (12% gel). Right: GFP:FUS LC domain (15% gel). Black arrows indicate intact fusion proteins. Grey arrow indicates mCherry:FUS LC domain truncated at the fluorophore center (Y72) of mCherry.

Figure 3. Preparation of a glass-bottom dish for hydrogel-binding assays

(A) A picture of a glass-bottom culture dish and a paper strip. The part number of the culture dish is P35G-1.5-14-C (MatTek Corporation). The paper strip is cut from Whatman chromatography paper.

(B) A picture of the dish with the paper strip.

Figure 4. Confocal microscope setting for hydrogel binding assays

A schematic drawing illustrates that a horizontal section of a hydrogel droplet is scanned by a confocal microscope.

Figure 5. Typical binding patterns observed by hydrogel binding assays

All hydrogel droplets were made from mCherry:FUS LC domain fusion protein (red: right panels). (A) GFP:FUS LC domain binds at the rim of the hydrogel droplet (pattern 1: homotypic binding). (B) GFP:hnRNPA1 LC domain uniformly binds to the hydrogel droplet (pattern2: heterotypic binding). (C) GFP:FUS LC domain with all tyrosine residues mutated to serine does not bind to the hydrogel droplet (Pattern 3: no binding).

Figure 6. Intensity measurements of hydrogel binding assays
(A) When GFP:LC domain uniformly binds to a hydrogel droplet (Fig. 5B), the GFP intensity is measured by an area mode. First, an oval selection (yellow circle) is fit to the hydrogel droplet (left). A mean intensity is measured by ImageJ (right, red rectangle).
(B) For the binding pattern 1 in Fig. 5A, a profile plot method is used. A straight line is drawn across a hydrogel droplet (left). A profile of the section at the straight line is plotted (right). Gray intensity values at the peaks (red arrows) are measured.

Table 1

Test LC domains that form cross- β polymers or hydrogels.

Protein[*]	Residue range of LC domain used	Purification protocol	Citation
FUS	2–214	Native	Kato et al., 2012 [10]
EWS	47–266	Native	Kwon et al., 2013 [23]
TAF15	2–208	Native	Kwon et al., 2013 [23]
hnRNP A1	186–320	Native or 2M Urea	Lin et al., 2016 [11]
hnRNP A2	181–341	Native or 2M Urea	Kato et al., 2012 [10]
hnRNP AB-1	242–332	Native	Unpublished
hnRNP D2	284–355	Native	Unpublished
hnRNP DL	323–420	Native	Lin et al., 2016 [11]
CIRBP	90–172	Native	Unpublished
TDP-43	266–403	2M Urea	Unpublished
Nup54	2–114	Native or 2M Urea	Shi et al., 2017 [12]
Nup98	2–127	Native or 2M Urea	Shi et al., 2017 [12]
Vimentin	2–95	2M Urea	Lin et al., 2016 [11]
Desmin	2–108	2M Urea	Unpublished
Peripherin	2–99	2M Urea	Lin et al., 2016 [11]
Neurofilament-L	2–100	2M Urea	Lin et al., 2016 [11]
Neurofilament-M	2–104	2M Urea	Lin et al., 2016 [11]
Neurofilament-H	3–100	2M Urea	Lin et al., 2016 [11]
Synaptophysin	247–313	Native	Unpublished

* All proteins are from human.