Cross- β Polymerization of Low Complexity Sequence Domains

Masato Kato and Steven L. McKnight

Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390-9152

Correspondence: steven.mcknight@utsouthwestern.edu

Most transcription factors and RNA regulatory proteins encoded by eukaryotic genomes ranging from yeast to humans contain polypeptide domains variously described as intrinsically disordered, prion-like, or of low complexity (LC). These LC domains exist in an unfolded state when DNA and RNA regulatory proteins are studied in biochemical isolation from cells. Upon incubation in the purified state, many of these LC domains polymerize into homogeneous, labile amyloid-like fibers. Here, we consider several lines of evidence that may favor biologic utility for LC domain polymers.

Many neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS), are causally associated with the formation of protein aggregates. Extensive genetic evidence favors the central involvement of A β aggregation in the cause of AD (Hardy and Selkoe 2002; Benilova et al. 2012). Likewise, aggregation of the microtubule binding protein tau has also been strongly implicated in the initiation and spread of AD (Ballatore et al. 2007).

Parallel to these foundational studies on A β and tau aggregates, numerous studies have implicated protein aggregation as a causative event in the etiology of ALS, frontotemporal dementia, Huntington's disease, and other categories of neurodegenerative disease. Human genetic studies have revealed familial inheritance of neurodegenerative disease that trace lesions back to the genes encoding TDP-43, FUS, and several different heterogeneous nuclear ribonu-

cleoproteins (hnRNPs) (Bentmann et al. 2013; Kim et al. 2013). The latter proteins are commonly understood to perform functions associated with RNA biogenesis, and the causative mutations leading to neurodegenerative disease have been shown to trigger the aberrant aggregation of these very proteins.

Disease-causing mutations in TDP-43, FUS, and hnRNPs often map to protein regions of unknown biologic function. All three proteins contain well-folded RNA binding domains, yet aggregation-promoting mutations do not affect this well-understood aspect of how these proteins function in the context of RNA biogenesis. Common to the protein domains mutated as the cause of neurodegenerative disease are regions of molecular disorder, variously termed intrinsically disordered, prion-like, or low complexity (LC) domains.

LC domains are typified by the use of only a subset of the 20 amino acids normally deployed

M. Kato and S.L. McKnight

to facilitate the proper folding of structurally ordered proteins. The term “low complexity” derives from this very feature. LC domains might contain scores of contiguous glutamine residues to the exclusion of all of the other 19 amino acids found in typical proteins. Other LC domains have been described to be rich in proline, glycine, or asparagine residues. Disease-causing mutations often expand the size of LC domains, such as the repeat expansions that cause the polyglutamine domain of the Huntington’s protein to be increased in size (Blum et al. 2013). Alternatively, single amino acid substitution mutations have been found in the LC domains of TDP-43, FUS, and several hnRNPs (Bentmann et al. 2013; Kim et al. 2013).

Strong evidence has accumulated confirming that mutations causing either the size expansion of LC domains or missense mutations within LC domains function by increasing the probability that these domains aggregate within cells (Perutz et al. 1994; Chen et al. 2002; Johnson et al. 2009; Kim et al. 2013; Nomura et al. 2014). This work has evolved effectively in the absence of any concrete understanding of the normal biologic role of LC domains.

In the context of gene expression, the LC domains associated with gene-specific transcription factors have been categorized as “activation domains.” The properly folded DNA binding domains of transcription factors, including homeoboxes, zinc fingers, and leucine zippers, guide the proteins to the proper regulatory sites on DNA. In contrast, despite three decades of work, we have little mechanistic understanding of how LC domains facilitate gene activation—they are indeed vital for gene activation, yet how they work has been a mystery. A significant impediment to this line of research can be attributed to the fact that LC domains exist in an unfolded state upon biochemical isolation. If a protein is unfolded, how can it work? Simply put, understanding the function of a protein without understanding its form is problematic.

Here, we outline an unconventional series of studies leading to a very simple hypothesis. We have found that certain LC domains can, in a purified state, polymerize into cross- β fibers.

Unlike prototypical amyloid fibers, cross- β fibers formed from the LC domains of FUS, EWS, TAF15, and many other RNA regulatory proteins are labile to dilution. Our unconventional thinking is that properly controlled polymerization allows regulatory proteins to organize intracellular puncta, including nuclear speckles, “transcription factories,” and other dynamic bodies within the nucleus, as well as RNA granules, P-granules, and neuronal granules within the cytoplasm. We hypothesize that these puncta help organize, specialize, and optimize aspects of transcription and translation in living, eukaryotic cells, and that polymerization of LC domains is integral to how these puncta are formed.

STARTING WITH AN EXPERIMENTAL ARTIFACT

All of the studies reviewed here are built upon an experimental artifact. A high throughput drug screen performed at the University of Texas Southwestern Medical Center identified an isoxazole chemical that was observed to prompt mouse embryonic stem cells to differentiate into progenitors of cardiomyocytes (Sadek et al. 2008). This chemical was modified to contain a biotin moiety such that its molecular target might be identified by affinity chromatography.

When incubated with cellular lysates, the biotinylated isoxazole (b-isox) chemical led to the precipitation of hundreds of proteins. This precipitation reaction had nothing to do with the proper binding of the isoxazole moiety to its intracellular target. Indeed, precipitation did not require addition of strep-avidin coated beads. A flocculent precipitate could be observed in cell lysates immediately on addition of 10–30 μ M levels of the b-isox chemical.

Whereas it was recognized that this line of investigation held little promise for deciphering the molecular target of the parent chemical, we noticed that the proteins precipitated by the b-isox probe were unusual. The pattern of Coomassie staining of mammalian cell lysates before and after the b-isox precipitation reaction was indistinguishable and clearly showed that the b-isox chemical did not precipitate any of the

major, abundant proteins of the cell lysate. For this reason, we used shotgun mass spectrometry to identify the rarer cellular proteins precipitated by the chemical, leading to the discovery that the majority of b-isox precipitated proteins correspond to RNA binding proteins (Kato et al. 2012). The list of b-isox precipitated proteins corresponded closely to the list of proteins known to associate with various forms of cytoplasmic RNA granules, including polar granules from fly and worm embryos, and stress granules from yeast and mammalian cells. For this reason, we considered that the perplexing properties of the b-isox chemical might reveal information about the organization of RNA granules.

Using methods of molecular biology, we dissected several of the b-isox precipitated proteins in search of determinants required for precipitation. These efforts led us to recognize that the determinants commonly required for b-isox precipitation were unfolded LC domains. When the LC domain was removed from a protein, such as TIA-1, it would no longer be precipitated by the b-isox chemical. Moreover, when an LC domain was appended onto a folded domain that was itself not precipitated—such as green fluorescent protein (GFP)—the chimeric protein was fully susceptible to precipitation.

Our understanding of these perplexing observations was clarified, in part, upon recognizing that the b-isox chemical rapidly forms crystals when introduced into cold aqueous buffer. We found that the flocculent precipitate initially observed upon addition of the b-isox chemical to cell lysates was seeded by microcrystals. By solving the X-ray structure of these crystals at high resolution, it was possible to offer a conceptual description of this unusual form of precipitation. The surface of the b-isox crystals contains waves of peaks and troughs separated by 4.7 Å. Given a trough width of this dimension, and knowing that the width of β -strands is the same size, we hypothesized that unfolded LC domains might transition from a random coil state in cellular lysates to an extended β -strand conformation lying along the 4.7 Å troughs displayed on the crystal surface (Kato et al. 2012).

Cross- β Polymerization of LC Sequence Domains

LC DOMAINS POLYMERIZE INTO LABILE FIBERS

Knowing that LC domains of RNA binding proteins were the determinant of b-isox precipitation, we began to consider the very crude idea that these domains might in some way be important to the organization of intracellular puncta, including RNA granules. We unexpectedly observed that certain of these LC domains, when expressed in bacterial cells, purified, and incubated at a high concentration, transitioned into a gel-like state. A combination of transmission electron microscopy and X-ray diffraction studies revealed the gels to be composed of homogenous cross- β polymers (Kato et al. 2012).

Two features of these LC domain polymers distinguished them from prototypic amyloid fibers. First, the fibers polymerized from LC domains were found to be labile to dilution. Pathogenic, prion-like amyloids are not only stable to dilution but also resistant to melting by detergents and chaotropic reagents. Despite sharing similar morphologies, as visualized by electron microscopy, the cross- β fibers formed from LC sequences must be different in a fundamental way from pathogenic amyloid fibers—they are labile to depolymerization. Second, X-ray diffraction analyses of amyloid fibers have yielded sheet-to-sheet distances within protomers ranging from 6 to 11 Å (Greenwald and Riek 2010). For reasons yet to be determined, polymeric fibers formed from the LC domains of RNA binding proteins are uniform in yielding intra-protomer, sheet-to-sheet distances of 10 Å (Kato et al. 2012).

The phenomenon of gelation, or what we now term “hydrogel formation,” was evolved into a simple microscopic assay. Droplets of concentrated solutions of purified LC domains (0.5 μ l) are spotted onto glass cover slips at the bottom of a chamber slide. Gelation is allowed to take place at room temperature over 24–48 h. We have routinely used mCherry:LC fusion proteins, such that the resulting hydrogel droplet is red in color. Droplets can then be challenged with test proteins. Soluble GFP was observed to penetrate hydrogel droplets freely, but was

M. Kato and S.L. McKnight

rapidly washed away when buffer was applied to the chamber slide wells. In contrast, GFP linked to any of a number of LC domains was retained by hydrogel droplets formed from mCherry:FUS or mCherry:hnRNPA2. Briefly put, hydrogel binding was evolved into a rapid and simple binding assay (Han et al. 2012; Kato et al. 2012).

We initially used correlative mutagenesis to investigate the biologic utility of cross- β polymers. It has long been known that the LC domains of the FUS, EWS, and TAF15 proteins can be translocated onto DNA binding domains as causative mutations driving a variety of forms of human cancer (Arvand and Denny 2001; Guipaud et al. 2006; Lessnick and Ladanyi 2012). In the context of these fusion proteins, the LC sequences have been shown to function as potent activation domains. Systematic mutagenesis studies on the LC domains of FUS and TAF15 have shown parallel effects on polymerization and transcriptional activation (Kwon et al. 2013). Mutations that severely impede cross- β polymerization, as deduced by hydrogel trapping, correlatively inhibit the ability of the LC domains of FUS and TAF15 to activate transcription. Mutations that affect polymerization to a moderate degree impinge transcriptional activation moderately, and mutations that do not affect polymerization do not impede transcription. These experiments and observations are consistent with the interpretation that the LC domains of FUS and TAF15 must be capable of polymerization to perform their role in transcriptional activation, yet do not directly indicate the existence of LC domain polymers in cells.

LC DOMAIN FIBERS BIND REGULATORY PROTEINS THAT ARE THEMSELVES DISORDERED

Among the list of cellular proteins precipitated by b-isox microcrystals are two categories of proteins long studied in the biogenesis of messenger RNA. The largest subunit of RNA polymerase II was precipitated by b-isox microcrystals, as were a number of RNA binding proteins involved in alternative pre-mRNA splicing.

In the former case, it was observed that the only form of the largest subunit of RNA polymerase II that was precipitated was the isoform fully devoid of phosphorylation of the C-terminal domain (CTD). The CTD of mammalian RNA polymerase II consists of 52 repeats of the heptapeptide sequence YSPTSPS (Corden et al. 1985). During the transcription cycle, cyclin-dependent kinase enzymes phosphorylate various serine and threonine residues within the CTD repeats (Heidemann et al. 2013). These phosphorylation events release promoter-bound RNA polymerase II such that it can transition into the elongation phase (Egloff et al. 2012).

The fact that unphosphorylated RNA polymerase II was precipitated by b-isox microcrystals, but that all of the phosphorylated forms were left in the cell lysate, offered two ideas. First, these data indicated that the CTD might be the determinant for interaction between RNA polymerase II and b-isox crystals (given that the precipitation of RNA polymerase II was strictly gated by CTD phosphorylation). Second, if correct, the observation suggested that, if phosphorylated, the CTD might be unable to bind b-isox crystals. Simple biochemical and molecular biological experiments confirmed both of these predictions (Kwon et al. 2013).

The CTD of RNA polymerase II fits the description of LC sequences. The 362 residues of the CTD consist solely of four amino acids: serine, threonine, tyrosine, and proline. However, unlike the LC domains of FUS, TAF15, EWS, and the hnRNPs, the CTD of RNA polymerase II does not self-polymerize. Before these studies, Tom Cech and colleagues had reported binding of RNA polymerase II to the FUS protein (Schwartz et al. 2012). Following the lead offered by the Cech studies, we tested for CTD binding to various hydrogel droplets, including those generated by the LC domains of FUS, EWS, TAF15, and several hnRNPs. When linked to GFP, the CTD of RNA polymerase II was found to bind to hydrogel droplets formed from the LC domains FUS, EWS, and TAF15, but not hnRNPs. The observed binding was fully reversed on exposure to the cyclin-dependent

kinase enzymes known to phosphorylate the CTD (Heidemann et al. 2013). We have speculated that the phosphorylation-sensitive binding of the CTD to polymeric fibers formed from the LC domains of FUS, EWS, and TAF15 may represent a biochemical reconstitution of CTD regulation in living cells (Kwon et al. 2013).

The second category of LC sequences that fail to polymerize, yet may be part of the broader paradigm of cell organization that we are studying, is represented by the repetitive serine: arginine (SR) domains associated with RNA binding proteins that regulate pre-mRNA splicing. These SR domains were first discovered by Gall and Roth in the early 1990s (Roth et al. 1990, 1991). SR-containing proteins abide in nuclear speckles implicated in pre-mRNA splicing (Lamond and Ladanyi 2003) and are regulated by a cyclin-like kinase (CLK) enzyme dedicated to the phosphorylation of serine residues within SR domains (Colwill et al. 1996; Duncan et al. 1998; Menegay et al. 2000; Aubol et al. 2013).

When fused to GFP, SR domains do not facilitate polymerization. Such fusion proteins, however, do bind avidly to hydrogel droplets formed from the LC domains of hnRNPs (Kwon et al. 2014). This binding is liberated by the administration of the CLK enzyme and ATP. Mutation of all serine residues within the SR domain of the SRSF2 alternative splicing protein to glycine resulted in a protein that still bound to hnRNP hydrogels, but could not be liberated by the CLK enzymes and ATP. When expressed in living cells, the phosphorylation-immune domain (glycine:arginine [GR] repeats replacing SR repeats) led to the formation of nucleoli-proximal speckles that were resistant to melting via overexpression of the CLK enzyme. We interpret the effect of the GR derivative of SRSF2 to be plugging up nuclear speckles at their site of origin (nucleoli) and thereby impeding the flow of information from gene to messenger RNA to protein.

Studies of the CTD of RNA polymerase II and the SR domains of alternative splicing factors provide evidence of a secondary level of organization of LC domains. The primary level of organization is interpreted to be homotypic

Cross- β Polymerization of LC Sequence Domains

cross- β polymerization of LC domains typified by the FUS, EWS, TAF15, and hnRNPs. Once polymerized, these fibers facilitate lateral interaction with repeated sequences typified by the CTD of RNA polymerase II or the SR domains of splicing factors. Such lateral interactions have been visualized by fluorescence microscopy (Kwon et al. 2013) and are dynamically reversible by the protein kinase enzymes known to regulate CTD and SR function.

POLY-DIPEPTIDES ENCODED BY THE REPEAT EXPANSION IN *C9orf72* CLOG INFORMATION FLOW

Elegant studies conducted over the past five years have shown that expansion of a hexanucleotide repeat sequence in the first intron of the human *C9orf72* gene leads to upwards of 40% of heritable forms of ALS (DeJesus-Hernandez et al. 2011; Renton et al. 2011). Unaffected people contain 10–20 repeats of the hexanucleotide sequence GGGGCC in the *C9orf72* gene. The repeat number expands to upwards of 1000 copies in affected individuals.

Three possible mechanisms have been considered as to how this repeat expansion may lead to disease. First, it might disrupt production of the appropriate protein encoded by the *C9orf72* gene (DeJesus-Hernandez et al. 2011; Renton et al. 2011). Second, sense or antisense transcripts of the repeat, both of which are transcribed in affected patients, might poison cells (Lagier-Tourenne et al. 2013; Mizielinska et al. 2013; Zu et al. 2013; Haeusler et al. 2014). Finally, aberrant translation of these transcripts in the form of simple sequence poly-dipeptides, which are indeed produced in diseased tissue by non-ATG-mediated translation (Zu et al. 2011), might be toxic (Ash et al. 2013; Donnelly et al. 2013; Mori et al. 2013).

Among these three possibilities, we have focused on the poly-dipeptides encoded by the repeat expansion of *C9orf72* as a potential mechanism of disease pathogenesis. The sense strand of the hexanucleotide repeat can encode poly-GA, poly-GP, and poly-GR dipeptides, whereas the antisense strand can encode poly-PA, poly-PG, and poly-PR dipeptides. Noticing

M. Kato and S.L. McKnight

the similarity between the poly-GR dipeptide and our mutated derivative of the SR domain of the SRSF2 splicing factor, wherein serine residues were changed to glycine, we wondered whether the former poly-dipeptide might bind to fibrous polymers and impede the pathway of information transfer from gene to message to protein.

To this end, we synthesized synthetic peptides containing 20 repeats of either the GR dipeptide or the PR dipeptide. These were incubated at varying doses with cultured mammalian cells, leading to the observation that both of the poly-dipeptides are toxic at levels from 1 to 10 μ M (Kwon et al. 2014). The PR poly-dipeptide was considerably more toxic than the GR poly-dipeptide because of a 20 $\times$ longer half-life in tissue culture medium. Studies of RNA biogenesis in cells exposed to the poly-PR toxin gave evidence of impediments in both pre-mRNA splicing and ribosomal RNA biogenesis. Of particular note was an unusual pattern of aberrant splicing of the mRNA encoding a potassium channel designated excitatory amino acid transporter 2 (EAAT2). Upon exposure to the PR poly-dipeptide, cultured mammalian cells caused skipping of exon 9 and the aberrant addition of 1008 nucleotides of intronic sequence downstream from the splice donor site of exon 7. Precisely this same pattern of aberrant EAAT2 splicing has been reported for the brain tissue of ALS patients carrying the *C9orf72* repeat expansion (Lin et al. 1998). We consider the latter observation to constitute compelling evidence favoring disease pathogenicity enacted by expression of the PR and GR poly-dipeptides in patients carrying a repeat expansion in the heritable *C9orf72* form of ALS.

A MOLECULAR FOOTPRINTING METHOD TO PROBE FOR LC DOMAIN POLYMERS IN NUCLEI

Two types of correlative mutagenesis experiments have provided evidence that LC domain polymers might have biologic function. Mutations that impede polymerization of the LC domain of FUS correlatively affect its ability to partition into stress granules (Kato et al. 2012).

Likewise, mutations of either the FUS or TAF15 LC domain correlatively affect both polymerization and transcriptional activation (Kwon et al. 2013).

In efforts to develop a more direct method of asking whether LC domains polymerize in living cells, we turned to *N*-acetylimidazole (NAI). This chemical directly acetylates lysine, arginine, asparagine, serine, threonine, and tyrosine residues (Riordan et al. 1965; Timasheff and Gorbunoff 1967). Using glutathione-S transferase (GST) as a test case, we treated the enzyme with NAI in both its folded and denatured state. The latter protein was labeled with a heavy amino acid isotope such that peptides derived from the folded and denatured protein could be separated and analyzed by SILAC mass spectrometry. These methods gave clear evidence that the NAI chemical could only modify the folded GST enzyme on its surface, but could liberally modify the denatured protein. The difference maps of modification yielded a molecular footprint of the folded state of GST (Xiang et al. 2015).

This same method was then used to develop a footprint of polymeric fibers formed from the LC domain of hnRNPA2. The footprint of recombinant hnRNPA2 polymers was then compared with native hnRNPA2 protein in freshly isolated nuclei prepared from mammalian cells. With the exception of a single amino acid, tyrosine 324, a clear match was observed between the two footprints (Xiang et al. 2015).

The aforementioned tyrosine residue was protected from NAI acetylation in recombinant polymers of the LC domain of hnRNPA2, yet was sensitive to modification in the native, nuclear form of the protein. Noticing that Y324 was closely flanked by proline residues, we wondered whether the isomeric state of the peptide bonds of these proline residues might affect NAI sensitivity. This idea was prompted by knowledge that the cellular enzyme that facilitates *cis-trans* isomerization of the peptide bonds of proline residues, peptidyl-prolyl *cis-trans* isomerase (PPIA), is tightly associated with RNA binding proteins (Lauranzano et al. 2015). When the LC domain of hnRNPA2 was co-expressed in bacterial cells with the PPIA enzyme,

then purified, polymerized, and exposed to NAI, the resulting footprint perfectly matched that of native, nuclear hnRNPA2 (Xiang et al. 2015). In other words, tyrosine 324, which was protected in fibers formed from recombinant protein expressed alone, became NAI sensitive on co-expression with PPIA. This experimental approach gives direct experimental evidence of hnRNPA2 polymers of the native protein in freshly prepared nuclei, causing us to have some degree of confidence in our claim favoring the biologic utility of LC domain polymerization.

A BREWING CONTROVERSY

Since the brilliant discoveries of Reed Wickner, biologists have known that LC domains can form highly stable, prion-like polymers (Wickner 1994; see also Wickner 2016). It was Wickner's studies that resolved the riddle of non-Mendelian inheritance in the context of the PSI⁺ strain of yeast. This phenotype was traced to the yeast Sup35 protein, which encodes a translation termination factor. Wickner discovered that when Sup35 exists in an aggregated, prion-like state, it causes the PSI⁺ phenotype, manifested by partial loss of function and nonsense suppression. That Sup35 prions can be passed from cell to cell fully explained the non-Mendelian genetics of the PSI⁺ phenotype. It has long been recognized that the prion-forming domain of Sup35 is an LC sequence.

LC domains appear abundantly throughout the proteomes of eukaryotic cells—from yeast to humans. These sequences are found within nearly every DNA and RNA regulatory protein, yet the biologic utility of LC domains has been enigmatic for decades. In the late 1980s, the McKnight laboratory discovered one of the most potent and acclaimed transcriptional activation domains ever used or studied—the acidic activation domain of the herpes simplex virus VP16 protein (Triezenberg et al. 1988a,b). Despite decades of use, it remains unclear how the disordered activation domain of VP16 actually works.

Until several years ago, thinking in the field about LC domains diverged. On one end of

Cross- β Polymerization of LC Sequence Domains

the spectrum, certain researchers argued that LC domains served to form irreversible, prion-like polymers. These rock-solid polymers were proposed to confer phenotypic variability to organisms such as yeast (Newby and Lindquist 2013). Prion-like polymers might serve an evolutionary benefit for the organism, considering, for example, that if just one in a million yeast cells contained one such prion-like polymer, the cell might survive threatening conditions. The biological value of these sequences cannot be ascertained in any way other than within an evolutionary context. We will call this the “north pole” of LC sequence studies.

The “north pole” concept for the utility of LC domains cannot be easily studied experimentally—a million-year trial of cell fitness for evolution is impossible. However, an exciting line of research that may circumvent this conundrum has arisen from the discovery that the permanence of prions formed by LC sequences might correspond to the permanence of memory (Si et al. 2010; Majumdar et al. 2012). An LC domain associated with a protein that can facilitate cytoplasmic extension of poly-A tails on mRNA, termed cytoplasmic polyadenylation element binding (CPEB), was found to form permanent, sodium dodecyl sulfate (SDS)-resistant polymers when expressed in yeast. Further experimentation has provided evidence that this same LC domain can adopt a prion-like conformation in neurons. The irreversibility of oligomerization, allowing deposition of a permanently insoluble RNA:protein complex at specific synaptic structures, has been postulated to correspond to the permanence of memory.

If the permanence of prion-like, irreversible aggregates of LC domains can be described as the “north pole” of this field, then the “south pole” is occupied by researchers who contend that LC domains never form any structure at all. In this line of thought, which is now gaining a larger audience, LC domains act in a random-coil, totally unfolded state. The experimental approach of these researchers is based on the finding that when LC domains are mixed under certain experimental conditions, they phase-separate in a manner not unlike oil droplets

M. Kato and S.L. McKnight

partitioning out of water. This, it is contended, occurs *sans* molecular structure (Altmeyer et al. 2015; Elbaum-Garfinkle et al. 2015; Lin et al. 2015; Molliex et al. 2015; Nott et al. 2015; Patel et al. 2015). By use of “fuzzy,” “molten,” or “slithering” interactions, unfolded LC domains are believed to associate with proper specificity to control the formation of membrane-free intracellular puncta, including nuclear speckles, stress granules, P bodies, and neuronal granules.

Here, we describe an alternative hypothesis that exists between these two poles—a “temperate zone,” so to speak. We suggest that certain LC domains can form labile, reversible cross- β polymers, and that these labile polymers are fundamental to the organization of the aforementioned intracellular puncta.

There are thousands of LC domains in the proteome of even the simplest eukaryotic cell (such as baker’s yeast). If LC domains can form labile polymers, some interesting questions can be posed considering future research in the field:

1. What are the rules that govern the interaction between LC domain polymers, as viewed from either a homotypic or heterotypic perspective?
2. Why do LC domain polymers formed from a variety of different RNA regulatory proteins, including FUS, EWS, TAF15, and multiple hnRNPs, all display X-ray diffraction rings at 10 Å (when cross- β polymers studied over the past five decades display diffraction rings ranging from 6 to 11 Å)?
3. What are the posttranslational events that balance polymerization?
4. What is the precise molecular structure of an LC domain polymer?
5. How do missense mutations causative of neurodegenerative disease tip the balance of polymerization to depolymerization toward the aggregated state?
6. How does the iterative repeat structure of the CTD of RNA polymerase II, or the SR repeats

of splicing factors, bind to LC domain polymers?

7. Why do the CTD repeats of RNA polymerase II bind to LC polymers formed from FUS, EWS, and TAF15, but not to hnRNP polymers?
8. How does phosphorylation of either the CTD or SR domains interfere with binding to their cognate polymers?

It is reasonable to conclude that neither our group nor other researchers in the field currently understand how LC domains actually work in living cells—that is the bad news. However, the good news is that many new ideas are circulating in the field, and this enigmatic area of the life sciences is experiencing newfound interest. From the outset of our studies on LC domains, our approach has been built on what is clearly an experimental artifact—the nonbiologic crystallization of the b-isox chemical. Given this unconventional start, we offer a closing caveat to readers. Our “temperate zone” concept of labile LC domain polymers is built on the most surprising and quirky of beginnings!

ACKNOWLEDGMENTS

This work is supported by a grant from the National Institutes of Health (U01GM107623) and unrestricted funds provided to S.L.M. by an anonymous donor.

REFERENCES

*Reference is also in this collection.

- Altmeyer M, Neelsen KJ, Teloni F, Pozdnyakova I, Pellegrino S, Grofte M, Rask MBD, Streicher W, Jungmichel S, Nielsen ML, et al. 2015. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun* 6: 8088.
- Arvand A, Denny CT. 2001. Biology of EWS/ETS fusions in Ewing’s family tumors. *Oncogene* 20: 5747–5754.
- Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW III, Rademakers R, et al. 2013. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77: 639–646.
- Aubol BE, Plocinik RM, Hagopian JC, Ma CT, McGlone ML, Bandyopadhyay R, Fu XD, Adams JA. 2013. Parti-

- tioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *J Mol Biol* **425**: 2894–2909.
- Ballatore C, Lee VM, Trojanowski JQ. 2007. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci* **8**: 663–672.
- Benilova I, Karran E, De Strooper B. 2012. The toxic A β oligomer and Alzheimer's disease: An emperor in need of clothes. *Nat Neurosci* **15**: 349–357.
- Bentmann E, Haass C, Dormann D. 2013. Stress granules in neurodegeneration—Lessons learnt from TAR DNA binding protein of 43 kDa and fused in sarcoma. *FEBS J* **280**: 4348–4370.
- Blum ES, Schwendeman AR, Shaham S. 2013. PolyQ disease: Misfiring of a developmental cell death program? *Trends Cell Biol* **23**: 168–174.
- Chen S, Ferrone FA, Wetzell R. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci* **99**: 11884–11889.
- Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI. 1996. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J* **15**: 265–275.
- Corden JL, Cadena DL, Ahearn JM Jr, Dahmus ME. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci* **82**: 7934–7938.
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, et al. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**: 245–256.
- Donnelly CJ, Zhang PW, Pham JT, Haeusler AR, Mistry NA, Vidensky S, Daley EL, Poth EM, Hoover B, Fines DM, et al. 2013. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**: 415–428.
- Duncan PI, Stojdl DF, Marius RM, Scheit KH, Bell JC. 1998. The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. *Exp Cell Res* **241**: 300–308.
- Egloff S, Dienstbier M, Murphy S. 2012. Updating the RNA polymerase CTD code: Adding gene-specific layers. *Trends Genet* **28**: 333–341.
- Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CCH, Eckmann CR, Myong S, Brangwynne CP. 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci* **112**: 7189–7194.
- Greenwald J, Riek R. 2010. Biology of amyloid: Structure, function, and regulation. *Structure* **18**: 1244–1260.
- Guipaud O, Guillonnet F, Labas V, Praseuth D, Rossier J, Lopez B, Bertrand P. 2006. An in vitro enzymatic assay coupled to proteomics analysis reveals a new DNA processing activity for Ewing sarcoma and TAF(II)68 proteins. *Proteomics* **6**: 5962–5972.
- Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, et al. 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**: 195–200.
- Han TW, Kato M, Xie S, Wu L, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL. 2012. Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell* **149**: 768–779.
- Hardy J, Selkoe DJ. 2002. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**: 353–356.
- Heidemann M, Hintermair C, Voss K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* **1829**: 55–62.
- Johnson BS, Snead D, Lee JJ, McCaffery JM, Shorter J, Gitler AD. 2009. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J Biol Chem* **284**: 20329–20339.
- Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, et al. 2012. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**: 753–767.
- Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A, et al. 2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**: 467–473.
- Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. 2013. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**: 1049–1060.
- Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL. 2014. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**: 1139–1145.
- Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR, Jiang J, Watt AJ, Chun S, Katz M, et al. 2013. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci* **110**: E4530–E4539.
- Lamond AI, Spector DL. 2003. Nuclear speckles: A model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**: 605–612.
- Lauranzano E, Pozzi S, Pasetto L, Stucchi R, Massignan T, Paoletta K, Mombrini M, Nardo G, Lunetta C, Corbo M, et al. 2015. Peptidylprolyl isomerase A governs TARDBP function and assembly in heterogeneous nuclear ribonucleoprotein complexes. *Brain* **138**: 974–991.
- Lessnick SL, Ladanyi M. 2012. Molecular pathogenesis of Ewing sarcoma: New therapeutic and transcriptional targets. *Annu Rev Pathol* **7**: 145–159.
- Lin CL, Bristol LA, Jin L, Dykes-Hoberg M, Crawford T, Clawson L, Rothstein JD. 1998. Aberrant RNA processing in a neurodegenerative disease: The cause for absent EAAT2, a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron* **20**: 589–602.
- Lin Y, Protter DSW, Rosen MK, Parker R. 2015. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol Cell* **60**: 208–219.
- Majumdar A, Cesario WC, White-Grindley E, Jiang H, Ren F, Khan MR, Li L, Choi EM, Kannan K, Guo F, et al. 2012. Critical role of amyloid-like oligomers of *Drosophila* Orb2 in the persistence of memory. *Cell* **148**: 515–529.

M. Kato and S.L. McKnight

- Menegay HJ, Myers MP, Moeslein FM, Landreth GE. 2000. Biochemical characterization and localization of the dual specificity kinase CLK1. *J Cell Sci* **113**: 3241–3253.
- Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, Isaacs AM. 2013. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* **126**: 845–857.
- Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj Anderson P, Kim Hong J, Mittag T, Taylor JP. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**: 123–133.
- Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretschmar HA, Cruts M, Van Broeckhoven C, et al. 2013. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science* **339**: 1335–1338.
- Newby GA, Lindquist S. 2013. Blessings in disguise: Biological benefits of prion-like mechanisms. *Trends Cell Biol* **23**: 251–259.
- Nomura T, Watanabe S, Kaneko K, Yamanaka K, Nukina N, Furukawa Y. 2014. Intracellular aggregation of mutant FUS/TLS as a molecular pathomechanism of amyotrophic lateral sclerosis. *J Biol Chem* **289**: 1192–1202.
- Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowitz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, et al. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* **57**: 936–947.
- Patel A, Lee Hyun O, Jawerth L, Maharana S, Jahn M, Hein MY, Stoyanov S, Mahamid J, et al. 2015. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**: 1066–1077.
- Perutz MF, Johnson T, Suzuki M, Finch JT. 1994. Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci* **91**: 5355–5358.
- Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, et al. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**: 257–268.
- Riordan JF, Wacker WEC, Vallee BL. 1965. N-acetylimidazole: A reagent for determination of “free” tyrosyl residues of proteins. *Biochemistry* **4**: 1758–1765.
- Roth MB, Murphy C, Gall JG. 1990. A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J Cell Biol* **111**: 2217–2223.
- Roth MB, Zahler AM, Stolk JA. 1991. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol* **115**: 587–596.
- Sadek H, Hannack B, Choe E, Wang J, Latif S, Garry MG, Garry DJ, Longgood J, Frantz DE, Olson EN, et al. 2008. Cardiogenic small molecules that enhance myocardial repair by stem cells. *Proc Natl Acad Sci* **105**: 6063–6068.
- Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR. 2012. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev* **26**: 2690–2695.
- Si K, Choi YB, White-Grindley E, Majumdar A, Kandel ER. 2010. Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell* **140**: 421–435.
- Timasheff SN, Gorbunoff MJ. 1967. Conformation of proteins. *Ann Rev Biochem* **36**: 13–54.
- Triebenberger SJ, Kingsbury RC, McKnight SL. 1988a. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* **2**: 718–729.
- Triebenberger SJ, LaMarco KL, McKnight SL. 1988b. Evidence of DNA: Protein interactions that mediate HSV-1 immediate early gene activation by VP16. *Genes Dev* **2**: 730–742.
- Wickner RB. 1994. Evidence for a prion analog in *S. cerevisiae*: The [URE3] non-Mendelian genetic elements as an altered URE2 protein. *Science* **264**: 566–569.
- * Wickner RB. 2016. Yeast and fungal prions. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a023531.
- Xiang S, Kato M, Wu Leeju C, Lin Y, Ding M, Zhang Y, Yu Y, McKnight SL. 2015. The LC domain of hnRNP A2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**: 829–839.
- Zu T, Gibbons B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, et al. 2011. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci* **108**: 260–265.
- Zu T, Liu Y, Banez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH, et al. 2013. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci* **110**: E4968–E4977.

Cold Spring Harbor Perspectives in Biology

Cross- β Polymerization of Low Complexity Sequence Domains

Masato Kato and Steven L. McKnight

Cold Spring Harb Perspect Biol 2017; doi: 10.1101/cshperspect.a023598 originally published online November 11, 2016

Subject Collection [Prion Biology](#)

Genetic PrP Prion Diseases

Mee-Ohk Kim, Leonel T. Takada, Katherine Wong, et al.

Neurodegenerative Disease Transmission and Transgenesis in Mice

Brittany N. Dugger, Daniel P. Perl and George A. Carlson

Toward the Atomic Structure of PrP^{Sc}

Jose A. Rodriguez, Lin Jiang and David S. Eisenberg

Bioassays and Inactivation of Prions

Kurt Giles, Amanda L. Woerman, David B. Berry, et al.

Functional Prions in the Brain

Joseph B. Rayman and Eric R. Kandel

The Amyloid Phenomenon and Its Links with Human Disease

Christopher M. Dobson

Tau Positron Emission Tomography Imaging

Hartmuth C. Kolb and José Ignacio Andrés

Prion-Like Polymerization in Immunity and Inflammation

Xin Cai, Hui Xu and Zhijian J. Chen

Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases

Michael G. Erkkinen, Mee-Ohk Kim and Michael D. Geschwind

Prion Properties of SOD1 in Amyotrophic Lateral Sclerosis and Potential Therapy

Caroline Sibilla and Anne Bertolotti

Mapping Neurodegenerative Disease Onset and Progression

William W. Seeley

Erratum: Functional Prions in the Brain

Joseph B. Rayman and Eric R. Kandel

Pathology of Neurodegenerative Diseases

Brittany N. Dugger and Dennis W. Dickson

TIA-1 Is a Functional Prion-Like Protein

Joseph B. Rayman and Eric R. Kandel

Molecular Genetics of Neurodegenerative Dementias

Flora I. Hinz and Daniel H. Geschwind

Cross- β Polymerization of Low Complexity Sequence Domains

Masato Kato and Steven L. McKnight

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

**All Modifications and
Oligo Types Synthesized**
Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?
Your wish is our command.