Cell

Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains

Ilmin Kwon,[1,5] Masato Kato,[1,5] Siheng Xiang,[1] Leeju Wu,[1] Pano Theodoropoulos,[1] Hamid Mirzaei,[1] Tina Han,[1,3] Shanhai Xie,[1,4] Jeffry L. Corden,[2] and Steven L. McKnight[1,*]

[1]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390-9152, USA

[2]Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[3]Present address: Department of Physiology, Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA

[4]Present address: Peloton Therapeutics, Dallas, TX 75235, USA

[5]These authors contributed equally to this work

*Correspondence: steven.mcknight@utsouthwestern.edu

http://dx.doi.org/10.1016/j.cell.2013.10.033

SUMMARY

The low-complexity (LC) domains of the products of the fused in sarcoma (FUS), Ewings sarcoma (EWS), and TAF15 genes are translocated onto a variety of different DNA-binding domains and thereby assist in driving the formation of cancerous cells. In the context of the translocated fusion proteins, these LC sequences function as transcriptional activation domains. Here, we show that polymeric fibers formed from these LC domains directly bind the C-terminal domain (CTD) of RNA polymerase II in a manner reversible by phosphorylation of the iterated, heptad repeats of the CTD. Mutational analysis indicates that the degree of binding between the CTD and the LC domain polymers correlates with the strength of transcriptional activation. These studies offer a simple means of conceptualizing how RNA polymerase II is recruited to active genes in its unphosphorylated state and released for elongation following phosphorylation of the CTD.

INTRODUCTION

Numerous forms of cancer result from translocation events wherein the amino terminal, low-complexity (LC) domains of any of three related RNA-binding proteins become fused to a variety of different DNA-binding domains (Arvand and Denny, 2001; Guipaud et al., 2006; Lessnick and Ladanyi, 2012). The relevant RNA-binding proteins include the products of the fused in sarcoma (FUS), Ewings sarcoma (EWS), and TAF15 genes. Collectively, these three RNA-binding proteins are referred to as FET (FUS/EWS/TAF15) proteins (Andersson et al., 2008). The proteins encoded by these varied translocation products are causative of transformation (Bertolotti et al., 1999; Crozat et al., 1993; Ichikawa et al., 1999; Rabbitts et al., 1993; Zinszner et al., 1994). They further display dual dependency upon both the DNA-binding domain, which can be represented by members of the homeobox, zinc finger, ETS, or leucine zipper families of DNA-binding domains, as well as the LC domains donated by FUS, EWS, or TAF15. The DNA-binding domains are understood to direct the cancer-causing fusion proteins to appropriate batteries of genes adequate to facilitate cell growth or survival. By contrast, the LC sequences donated by members of the FET family are understood to function as transcriptional activation domains.

Over the past several decades the fields of biochemistry, biophysics, and molecular biology have achieved a concrete understanding of how DNA-binding domains function. The atomic structures of many such domains have been resolved by either X-ray crystallography or NMR spectroscopy, often in complex with their specific DNA substrates. By contrast, far less is known about the manner in which transcriptional activation domains operate at a mechanistic level. Prototypic activation domains are comprised of low-complexity sequences that exist in an intrinsically disordered, random coil conformation (Huntley and Golding, 2002; Uversky, 2002). The observation that evolutionary pressure of cancer cell formation has repeatedly led to the selection of the LC domains of FET proteins as the fusion partner to a variety of different DNA-binding domains strongly hints that these particular LC sequences may be exceptionally potent transcriptional activation domains.

The LC domain of FUS donated to the translocation product causative of cancer has a highly skewed distribution of amino acids. Of the 220 residues within the FUS LC domain, 84% are comprised of only four amino acids—glycine, serine, glutamine, and tyrosine. The domain contains zero representatives of glutamic acid, lysine, arginine, cysteine, histidine, valine, leucine, isoleucine, tryptophan, or phenylalanine. By having an amino acid composition dominated by only a four letter code (G, S, Q, and Y), the LC domain of FUS would appear to be more like nucleic acids than typical proteins that fold into their ultimate, three-dimensional shape by use of a much wider reliance on all 20 types of amino acid residues. When incubated at high concentrations, the LC domain of FUS has been found to polymerize

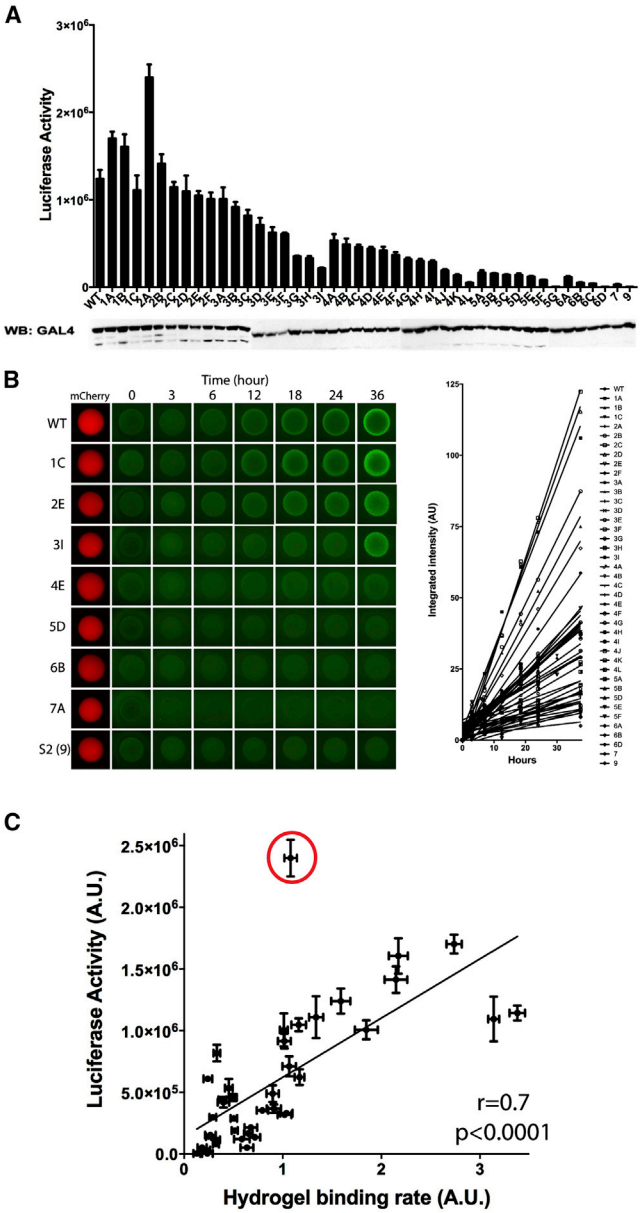

Figure 1. Correlation Between Transcriptional Activation and Hydrogel Binding of Native and Mutated Derivatives of the LC Domain of FUS

(A) The LC domain of FUS was linked to the DNA-binding domain of GAL4 and assayed for activation of a GAL4 reporter gene in transiently transfected U20S cells. Activity of the native LC sequence was compared with 43 variants wherein different number of tyrosine residues were randomly mutated to serine (see text). Identities of specific tyrosine-to-serine changes in each mutant are shown in Table S1. Expression levels for all test protein were assayed by western blotting as shown below histograms.

(B) The GFP-linked LC domains of FUS carrying the same mutations as (A) were exposed to mCherry:FUS hydrogels (left) and initial binding rates were measured (right).

(C) A correlation plot between the transactivation activity and hydrogel-binding rate of the individual LC mutants. Note that there is but one significant outlier indicated by a red circle. This is the "2A mutant" and described in more detail in the text and Figure 5.

into uniform, amyloid-like fibers (Han et al., 2012; Kato et al., 2012). Although morphologically similar to pathogenic amyloid aggregates, polymeric fibers formed from the LC domain of FUS are labile to depolymerization, raising the possibility that reversible LC polymerization may engender functional utility. Given two measurable features of the LC domain of FET proteins (transcriptional activation capacity and polymerization propensity), we hereby describe experiments that test the correlative relationship between the two.

RESULTS

The FUS LC domain contains 27 repeats of the triplet sequence [G/S]Y[G/S], and derivatives carrying 5, 9, 15, and 27 tyrosine-to-serine mutations in randomly chosen triplets showed progressively diminished capabilities of polymerization and stress granule recruitment (Kato et al., 2012). In order to initiate studies of the transcriptional activation capacity of the FUS LC domain, it was linked to the GAL4 DNA-binding domain and assayed by transient transfection for activation of firefly luciferase activity driven by a GAL4-dependent reporter gene. Forty-three mutated derivatives were prepared wherein individual tyrosine residues of the triplet sequence within the LC domain were randomly mutated to serine. Three mutants randomly changed a single tyrosine, 6 changed two tyrosines, 9 changed three tyrosines, 12 changed four tyrosines, 7 changed five tyrosines, 4 changed six tyrosines, 1 changed seven tyrosines, and 1 changed nine tyrosines (Table S1). As shown in Figure 1A, concordance was observed between the number of tyrosine residues changed to serine and the degree of impediment upon of transcriptional activation. With a single exception, variants carrying either one or two tyrosine-to-serine changes exhibited transcriptional activation capacity indistinguishable from the GAL4:wild-type fusion protein. The exceptional variant (2A) was considerably more active than the wild-type LC domain of FUS. Variants carrying three tyrosine-to-serine changes tended to be slightly less active than those carrying zero, one or two mutations. The trend in loss of activity was observed to track with variants carrying four, five, or six random tyrosine-to-serine mutations, and mutants carrying seven or nine tyrosine-to-serine mutations revealed no detectable capacity to activate expression of the GAL4-responsive luciferase target gene.

A simple assay for the propensity of LC domains to polymerized into amyloid-like fibers has been described previously (Kato et al., 2012). Briefly, upon prolonged incubation of a hybrid protein linking the LC domain of FUS to mCherry, the protein polymerizes into a hydrogel-like state. Microscopic gel droplets can be formed in chamber slides, exposed to soluble test protein, then scored for trapping of the test protein via confocal microscopy. The molecular basis of trapping has been deduced, by TIRF microscopy, to result from copolymerization of the GFP-tagged test protein into existing polymeric fibers in the mCherry:FUS hydrogel droplets (Kato et al., 2012). Forty out of the 43 variants described above were fused to GFP and tested for their ability to be trapped by hydrogel droplets composed of mCherry linked to the wild-type LC domain of FUS. As shown in Figure 1B, variants carrying one or two tyrosine-to-serine changes were trapped by the mCherry:FUS hydrogel droplets in a manner