


# The U1-70K and SRSF1 interaction is modulated by phosphorylation during the early stages of spliceosome assembly

Trent Paul<sup>1</sup> | Pengcheng Zhang<sup>2</sup> | Zihan Zhang<sup>1</sup> | Talia Fargason<sup>1</sup> |  
Naiduwadura Ivon Upekala De Silva<sup>1</sup> | Erin Powell<sup>1</sup> | Ethan Ekpenyong<sup>1</sup> |  
Shariq Jamal<sup>1</sup> | Yanbao Yu<sup>3</sup> | Peter Prevelige<sup>4</sup> | Rui Lu<sup>2</sup> | Jun Zhang<sup>1</sup> 

<sup>1</sup>Department of Chemistry, University of Alabama at Birmingham, Birmingham, Alabama, USA

<sup>2</sup>Department of Medicine, Division of Hematology/Oncology, University of Alabama at Birmingham, Birmingham, Alabama, USA

<sup>3</sup>Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware, USA

<sup>4</sup>Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA

## Correspondence

Jun Zhang, Department of Chemistry, University of Alabama at Birmingham, Birmingham, AL 35294, USA.  
Email: [zhanguab@uab.edu](mailto:zhanguab@uab.edu)

## Funding information

National Institutes of Health, Grant/Award Number: R35GM147091

**Review Editor:** Aitziber L. Cortajarena

## Abstract

In eukaryotes, pre-mRNA splicing is vital for RNA processing and orchestrated by the spliceosome, whose assembly starts with the interaction between U1-70K and SR proteins. Despite the significance of the U1-70K/SR interaction, the dynamic nature of the complex and the challenges in obtaining soluble U1-70K have impeded a comprehensive understanding of the interaction at the structural level for decades. We overcome the U1-70K solubility issues, enabling us to characterize the interaction between U1-70K and SRSF1, a representative SR protein. We unveil specific interactions: phosphorylated SRSF1 RS with U1-70K BAD1, and SRSF1 RRM1 with U1-70K RRM. The RS/BAD1 interaction plays a dominant role, whereas the interaction between the RRM domains further enhances the stability of the U1-70K/SRSF1 complex. The RRM interaction involves the C-terminal extension of U1-70K RRM and the conserved acid patches on SRSF1 RRM1 that is involved in SRSF1 phase separation. Our circular dichroism spectra reveal that BAD1 adapts an  $\alpha$ -helical conformation and RS is intrinsically disordered. Intriguingly, BAD1 undergoes a conformation switch from  $\alpha$ -helix to  $\beta$ -strand and random coil upon RS binding. In addition to the regulatory mechanism via SRSF1 phosphorylation, the U1-70K/SRSF1 interaction is also regulated by U1-70K BAD1 phosphorylation. We find that U1-70K phosphorylation inhibits the U1-70K and SRSF1 interaction. Our structural findings are validated through in vitro splicing assays and in-cell saturated domain scanning using the CRISPR method, providing new insights into the intricate regulatory mechanisms of pre-mRNA splicing.

## KEYWORDS

intrinsically disordered proteins, phosphorylation, SR proteins, SRSF1, U1-70K

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

## 1 | INTRODUCTION

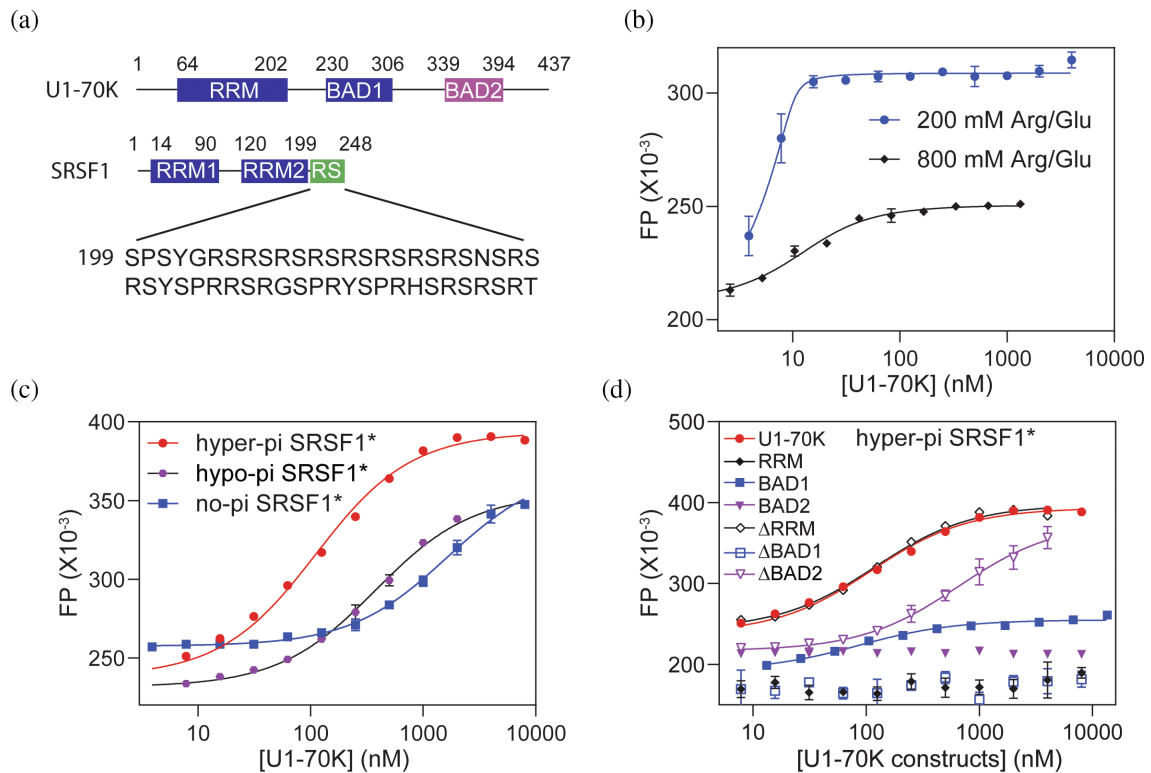
Most eukaryotic genes contain intervening sequences of exons and introns (Pan et al., 2008). To generate functional mRNA, introns need to be removed, and exons are joined together through a process known as pre-mRNA splicing. The pre-mRNA splicing occurs in the spliceosome, a dynamic RNA–protein complex that contains five small nuclear ribonucleoprotein particles (U1, U2, U4, U5, and U6), and hundreds of splicing factors (Staley & Guthrie, 1998; Will & Luhrmann, 1997). The spliceosome assembly commences with the recruitment of U1 and U2 complexes to the 5′ and 3′ splicing sites, respectively (Chabot & Steitz, 1987; Moore et al., 1993; Zillmann et al., 1987). These assembly steps rely on base pairing between U1/U2 RNA and conserved motifs around the splicing sites, such as the 5′ GU motif, and the 3′ AG motif preceded by a pyrimidine tract (Moore et al., 1993). However, base pairing is insufficient for splicing of most pre-mRNA, especially these undergoing alternative splicing. In addition to RNA base pairing, alternatively spliced pre-mRNA transcripts also possess regulatory RNA elements called splicing enhancers and silencers, located in both intronic and exonic regions. Splicing enhancers and silencers are recognized by splicing factors, to promote or reduce the usage of neighboring splicing sites, respectively (Anko, 2014; Long & Caceres, 2009; Manley & Krainer, 2010; Siebel et al., 1992; Siebel & Rio, 1990; Zhou & Fu, 2013).

Ser-Arg rich splicing factors (SR), which are the predominant splicing factors, recognize splicing enhancers and play a critical role in alternative splicing, a process regulating over 95% of the human protein genome (Anko, 2014; Long & Caceres, 2009; Manley & Krainer, 2010; Pan et al., 2008; Zhou & Fu, 2013). SR proteins contain one to two RNA-recognition motifs (RRM) and a featured arginine/serine region (RS). SR RRM domains specifically recognize splicing enhancers. The RS domains are subject to varying phosphorylation by SR protein kinases (SRPK) and Cyclin-dependent like kinases (Clk), facilitating protein–protein interactions (Colwill, Feng, et al., 1996; Colwill, Pawson, et al., 1996; Duncan et al., 1998; Gui, Lane, et al., 1994; Gui, Tronchere, et al., 1994; Hertel & Graveley, 2005; Wu & Maniatis, 1993). The binding of SR proteins to enhancers typically promotes inclusion of the corresponding exons (Cavaloc et al., 1999; Liu et al., 1998, 2000; Schaal & Maniatis, 1999; Tacke & Manley, 1995), although splicing inhibition is observed when SR proteins bind to introns for some RNA transcripts (Dembowski et al., 2012; Erkelenz et al., 2013).

SR proteins facilitate exon inclusion by interacting with the U1 component protein, U1-70K (also known as

snRNP70) in an RNA-independent way (Kohtz et al., 1994). U1-70K consists of RRM and two basic-acid dipeptide regions (BAD1 and BAD2) (Figure 1a). The U1-70K RRM tightly binds to a stem-loop region of U1 RNA and becomes a component protein of U1 snRNP (Kondo et al., 2015). The interaction between SR proteins and U1-70K has attracted considerable efforts to understand its binding mechanism. However, it remains elusive due to the challenges in obtaining soluble full-length SR proteins and U1-70K, the dynamic nature of the interaction, and complex phosphorylation background of these proteins. Controversy persists regarding which protein domains are essential for the SR/U1-70K interaction. Pull-down and yeast two-hybrid studies have suggested that phosphorylated SRSF1 RS is crucial for its interaction with U1-70K and other proteins (Hertel & Graveley, 2005; Wu & Maniatis, 1993). Nevertheless, another study proposed that the RRM domains of SRSF1 and U1-70K play a key role, while the SRSF1 RS is accessory (Cho et al., 2011). Still different from the aforementioned studies, a yeast-two hybrid study shows that the BAD1 domain of U1-70K is essential and sufficient for SRSF1 interaction (Cao & Garcia-Blanco, 1998). Most of these studies relied on crude cell extracts or indirect methods. Therefore, direct experimental evidence remains needed to elucidate the interaction between SR proteins and U1-70K (Zhou & Fu, 2013). Furthermore, the interaction between SR proteins and U1-70K is dynamic, a requirement for its biological function. After formation of B-complex, the U1 complex disassembles from the spliceosome before activation of splicing (Moore et al., 1993). Due to the transient nature of the interaction, the complex between SR proteins and U1-70K has never been observed in cryo-EM studies. Moreover, both SR proteins and U1-70K are heavily phosphorylated. SR proteins exist in un-, hypo-, and hyper-phosphorylated states. Previous studies have shown that hyper-phosphorylation of SR proteins is essential for U1-70K interaction (Kohtz et al., 1994). U1-70K is also a heavily phosphorylated protein (Woppmann et al., 1990). U1-70K in the cell has at least 13 different species according to the isoelectric points (Woppmann et al., 1990). When the dephosphorylation of U1-70K is abolished by thiophosphorylation, RNA splicing was inhibited (Tazi et al., 1993). It is still unknown how phosphorylation of U1-70K regulates its functions.

In this study, we have successfully addressed the long-standing challenge of U1-70K insolubility and low expression. Building on our recent success in solubilizing full-length SRSF1, we have a unique opportunity to unravel the mechanism underlying the U1-70K and SRSF1 interaction. Employing fluorescence polarization (FP) assays, we find that SRSF1 RRM1 interacts with



**FIGURE 1** BAD1 of U1-70K is essential for its interaction with phosphorylated SRSF1. (a) Domain architecture of U1-70K and SRSF1. BAD, basic acidic dipeptide region; RRM, RNA recognition motif; RS, Arg-Ser dipeptide repeats. The sequence of SRSF1 RS is displayed below. (b) FP binding profiles for U1-70K and U1 RNA in 20 mM Tris-HCl, pH 8.5, 0.1 mM TCEP, and 200 or 800 mM Arg/Glu. 10 nM of U1 RNA labeled with fluorescein at the 5' end was used. (c) FP binding assays for full-length U1-70K with SRSF1 in different phosphorylation states. Alexa Fluor 488 was labeled at T248C of SRSF1 (10 nM). The binding assays were performed in 200 mM Arg/Glu, 20 mM Tris-HCl, pH 8.5, and 0.1 mM TCEP. Hyper-pi SRSF1, Hypo-pi SRSF1, and no-pi SRSF1 indicate SRSF1 phosphorylated by Clk1, SRPK1, or non-phosphorylated, respectively. (d) FP binding profiles of hyper-phosphorylated SRSF1 with various U1-70K constructs. ΔRRM, ΔBAD1, and ΔBAD2 denote the U1-70K constructs without RRM, BAD1, or BAD2, respectively. \* indicates fluorescent labeled molecules. The standard deviation was estimated from three replicates.  $K_d$  values are listed in Table 1.

U1-70K RRM, and phosphorylated SRSF1 RS interacts with U1-70K BAD1, with U1-70K BAD2 being dispensable for SRSF1 interaction. The spectrum of hypo- to hyper-phosphorylation of SRSF1 RS enhances its binding to U1-70K progressively. These findings are further supported by our in vitro splicing assays, which show that exogenous BAD1 efficiently inhibits splicing, RRM has a mild effect, and BAD2 has no such inhibitory effect.

Using circular dichroism and FT infrared spectroscopies, we observe that upon RS binding, BAD1 undergoes a conformation transition from  $\alpha$ -helix to random coil and  $\beta$ -strand. Our NMR study and binding assays further unveil that the binding interfaces for SRSF1 RRM1 and U1-70K RRM are located on the conserved  $\alpha 1$  and  $\alpha 2$  edges of SRSF1 RRM1 and the C-terminal extension of U1-70K RRM. Notably, the  $\alpha 1$  and  $\alpha 2$  edges are sequestered by nonphosphorylated RS, as indicated by our NMR study. Therefore, SRSF1 phosphorylation plays a dual role in binding to U1-70K.

Interestingly, the SRSF1/U1-70K interaction is also modulated by the phosphorylation of U1-70K BAD1. The

phosphorylatable Arg-Ser region in BAD1 is responsible for the SRSF1 RS interaction, and the phosphorylation of this region reduces the RS/BAD1 binding by over 60-fold. To validate our structural biology findings, we perform a high-density CRISPR tiling scan to systematically investigate how mutations of U1-70K impact cell fitness. Our in-cell scanning results reveal that U1-70K RRM, especially its C-terminal extension, has a high impact on cell fitness, followed by BAD1 and BAD2 regions.

## 2 | RESULTS

### 2.1 | The U1-70K RRM and BAD1 domains are essential for SRSF1 interaction

It has been challenging to obtain sufficient amounts of soluble full-length U1-70K for biophysical and biochemical studies due to low expression, protein degradation, and insolubility. We have overcome these obstacles

**TABLE 1** Dissociation constants ( $K_d$ ) of U1-70K constructs to SRSF1 in different phosphorylation states.

Fluorophore labeled protein or RNA (10 nM)	Titratant	$K_d$ (nM)
U1 RNA	U1-70K	$0.20 \pm 0.28$ ( $7.0 \pm 2.3^a$ )
Hyper-pi SRSF1	U1-70K	$110 \pm 17$
Hypo-pi SRSF1	U1-70K	$370 \pm 40$
No-pi SRSF1	U1-70K	$1600 \pm 300$
Hyper-pi SRSF1	U1-70K $\Delta$ BAD1	Too weak to determine
Hyper-pi SRSF1	U1-70K $\Delta$ BAD2	$650 \pm 160$
Hyper-pi SRSF1	U1-70K $\Delta$ RRM	$110 \pm 14$
Hyper-pi SRSF1	U1-70K BAD1	$80 \pm 20$
Hyper-pi SRSF1	U1-70K BAD2	Too weak to determine
Hyper-pi SRSF1	U1-70K RRM	Too weak to determine

<sup>a</sup>The  $K_d$  value in a buffer containing 800 mM Arg/Glu.

through a combination of strategies, including the optimization of DNA codons, the development of new expressing strains, and the refinement of purification procedures. Detailed information can be found in Section 4. Given U1-70K's low solubility in the phosphate buffer (PBS), we dissolved the protein in a solution containing 20 mM Tris-HCl, pH 8.5, 200 mM Arg/Glu, and 0.1 mM TCEP. To ascertain that U1-70K purified by our protocol is in the native state, we measured its binding affinity to U1 RNA using fluorescence polarization assays (FP). We found that full-length U1-70K bound to U1 RNA with a sub-nanomolar affinity in the buffer with 200 mM Arg/Glu (Figure 1b, Table 1). As the RNA probe concentration was 10 nM, this binding affinity represents upper limits of  $K_d$  values. The real affinity could be higher (Figure 1b, Table 1). Even with the presence of 800 mM Arg/Glu, U1-70K maintains an impressively high binding affinity to U1 RNA, with a  $K_d$  of 5 nM (Figure 1b, Table 1).

As SRSF1 in the cell has different phosphorylation states, we sought to explore how these phosphorylation states impact its interaction with U1-70K. To this end, we prepared SRSF1 in three phosphorylation states: nonphosphorylated (referred to as no-pi SRSF1), hypophosphorylated (hypo-pi SRSF1, phosphorylated by SRPK1), and hyperphosphorylated (hyper-pi SRSF1, phosphorylated by Clk1). SRPK1 phosphorylates the arginine-serine repeats of SRSF1 and added 9–

13 phosphate groups (Figure S1a), while Clk1 phosphorylates both arginine-serine and serine-proline repeats and adds 18–22 phosphate groups (Figure S1b). These SRSF1 proteins were labeled with Alexa Fluor 488 at the C-terminal end (T248C). Full-length U1-70K and SRSF1 are soluble in the buffer with 200 mM Arg/Glu, in which we performed fluorescence binding assays. We found that U1-70K bound with hyper-pi SRSF1 with a  $K_d$  of  $110 \pm 17$  nM (Figure 1c, Table 1). Hypophosphorylated SRSF1 binds with U1-70K 3–4-fold weaker (Table 1), whereas nonphosphorylated SRSF1 binds to U1-70K at least 16-fold weaker (Table 1). These results suggest that phosphorylation of SRSF1 facilitates its interaction with U1-70K.

To identify the U1-70K domains responsible for SRSF1 binding, we prepared various U1-70K constructs and compared their binding affinities to hyperphosphorylated SRSF1 (Figure 1d, Table 1). We found that deletion of RRM or BAD2 had smaller impacts on SRSF1 binding than deletion of BAD1, which completely abolished the interaction between U1-70K and SRSF1. In addition, BAD1 alone interacts with SRSF1 with a binding affinity comparable with full-length U1-70K, whereas BAD2 or RRM alone cannot. These results suggest that BAD1 is essential for SRSF1 binding. It is noteworthy that the binding buffer used here contains 200 mM of Arg/Glu, which can inhibit weak interactions that may exist in PBS. Therefore, the roles of U1-70K RRM and BAD2 in SRSF1 interaction cannot be ruled out.

To confirm our FP results and to eliminate possible interference of Arg/Glu in protein interaction, we further performed pull-down assays in a buffer that has no Arg/Glu using purified His-SUMO-tagged U1-70K constructs and phosphorylated SRSF1 obtained from *E. coli* (Figure 2). Our pull-down assays confirmed the interaction between U1-70K and SRSF1 (Figure 2), in consistent with FP assays shown in Figure 1. Our pull-down assays showed that deletion of U1-70K RRM or BAD1 significantly reduced SRSF1 binding, whereas the deletion of BAD2 had a mild impact (Figure 2a), suggesting the essential role of RRM and BAD1 in SRSF1 binding. As SRSF1 has different phosphorylation states in the cell, and the cellular matrix and factors may impact protein-protein interactions, we also performed pull-down assays using the HeLa S3 cell lysate (Figure 2b). Our cell lysate pull-down assays were in line with those using recombinant SRSF1. Notably, our pull-down and FP assays show differences about roles of RRM and BAD1 in SRSF1 binding. These differences are partially attributed to the buffer difference. Our FP assays used a buffer containing 200 mM Arg/Glu to solubilize full-length SRSF1 and U1-70K. This buffer inhibits protein-protein interaction and therefore the role of the U1-70K RRM in SRSF1

binding is beyond detection of FP assays (more details are in the following section). Additionally, pull-down assays do not probe a real equilibrium state because dissociation occurs during washing steps. Therefore, pull-down assays are unable to detect binding processes with fast dissociation rates (Hulme & Trevethick, 2010; Jarmoskaite et al., 2020). Our FP assays show that BAD1 binds to SRSF1 with an affinity similar to full-length

U1-70K, whereas our pull-down assays indicate that BAD1 is much weaker in pulling down SRSF1. Therefore, it is likely the interaction mediated by BAD1 is kinetically unstable. In summary, our FP and pull-down assays suggest that both RRM and BAD1 of U1-70K are required to maintain a stable interaction with SRSF1.

We have demonstrated the essential roles of the U1-70K RRM and BAD1 in facilitating interaction with

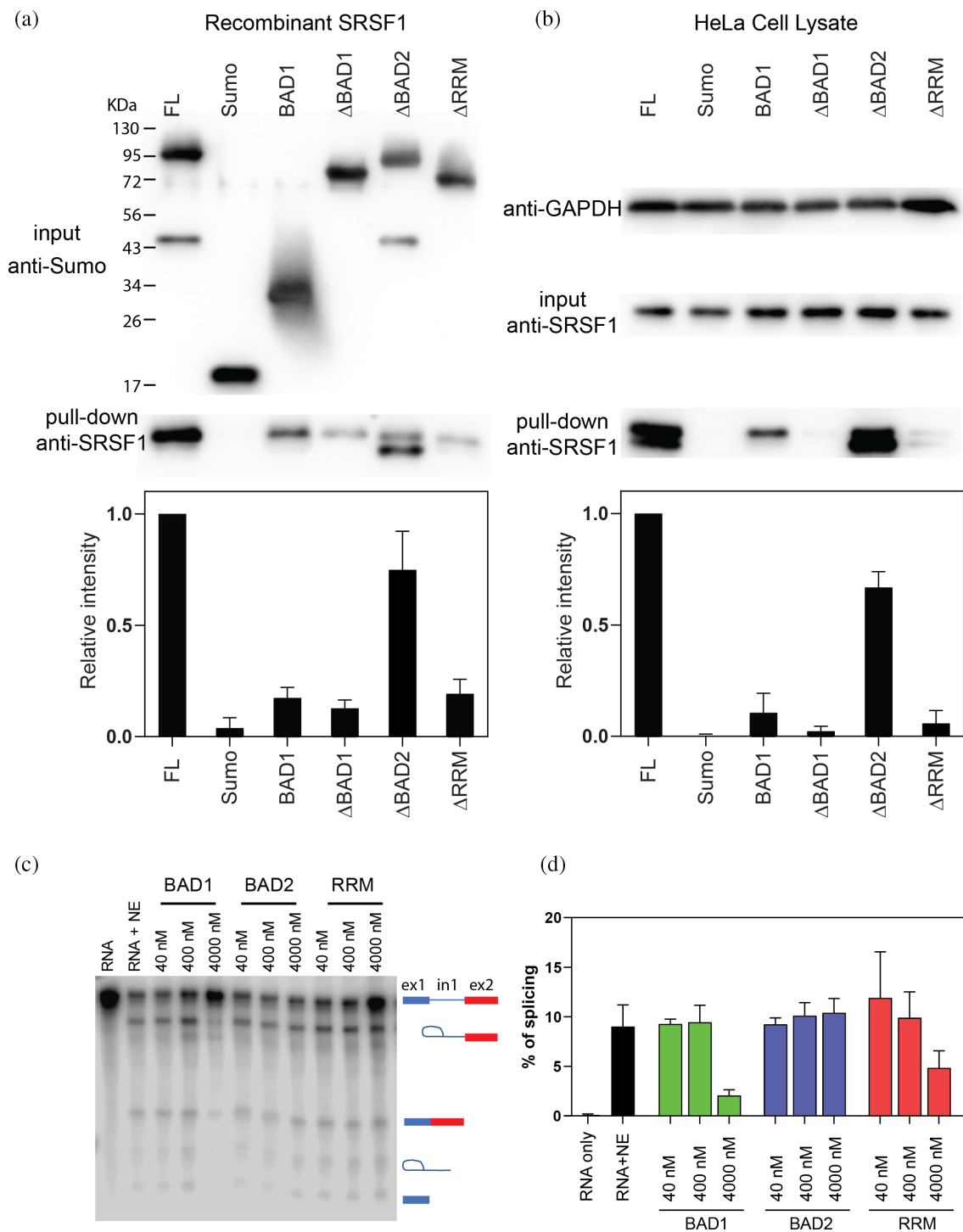


FIGURE 2 Legend on next page.



SRSF1, a critical step for the spliceosome assembly. Therefore, adding exogenous RRM or BAD1 will compete with endogenous U1-70K for SRSF1 interaction, which will consequently inhibit splicing. To validate this hypothesis, we conducted *in vitro* splicing assays on the  $\beta$ -globin gene, a well-known substrate dependent on SRSF1 (Cho et al., 2011; Das & Krainer, 2014; Krainer et al., 1984; Palhais et al., 2015; Zhang & Krainer, 2004; Zhang & Wu, 1996). In accordance with previous studies, we employed a miniature  $\beta$ -globin gene containing exon 1, intron 1, and exon 2 (Cho et al., 2011; Das & Krainer, 2014; Krainer et al., 1984; Palhais et al., 2015; Zhang & Krainer, 2004; Zhang & Wu, 1996). Our findings indicate that exogenous BAD1 and RRM of U1-70K effectively inhibit splicing of  $\beta$ -globin, whereas BAD2 does not exert a similar inhibitory effect (Figure 2c,d), which is consistent with our binding and pull-down assays.

## 2.2 | Map the domain–domain interactions between U1-70K and SRSF1

We have found that both U1-70K RRM and BAD1 are essential for formation of a stable complex with SRSF1. As both U1-70K and SRSF1 are multiple-domain proteins, we further wanted to map the direct interacting relationship between these domains, that is, which U1-70K domains interact with which SRSF1 domains (Figure 1a). We first wanted to determine which SRSF1 domains interact with U1-70K BAD1. To this end, we measured binding affinity of BAD1 to full-length SRSF1 and its individual domains (Figure 3a, Table 2). Since BAD1, phosphorylated SRSF1, and isolated SRSF1 domains are soluble in PBS, we conducted the following binding assays using PBS, which is closer to the intracellular buffer. In PBS, BAD1 interacts with full-length SRSF1 with a  $K_d$  of 12 nM (Figure 3a), which is 7-fold tighter than in 200 mM Arg/Glu (Figure 1d). BAD1

interacts with the phosphorylated RS tail of SRSF1 with a  $K_d$  of 1 nM, which is tighter than its interaction with full-length SRSF1. Our intramolecular study on SRSF1 reveals the intramolecular interaction between RS and the tandem SRSF1 (data to be published). Therefore, BAD1 can bind tighter with the isolated RS tail of SRSF1 because the competition from the intramolecular interaction is absent. The interaction between BAD1 and RRM1 or RRM2 are more than 200-fold weaker than phosphorylated RS. These results suggest that the SRSF1 region responsible for BAD1 interaction is the RS tail. To confirm this binding is not due to nonspecific interaction with the fluorophore, we performed a control assay between BAD1 and the Alexa Fluor 488 dye and found no binding between the fluorophore and BAD1 (Figure S2a).

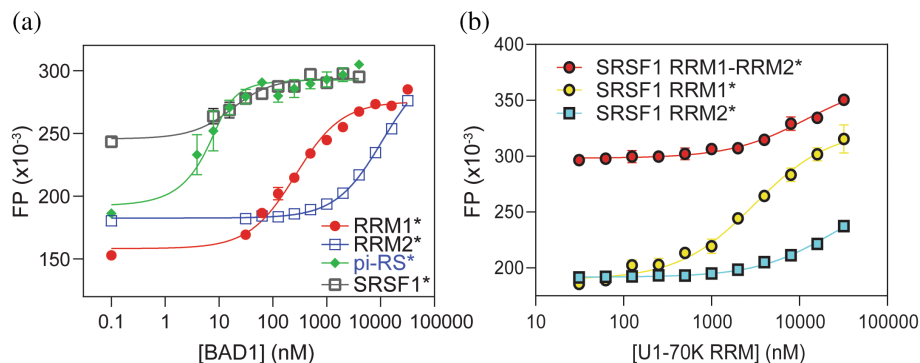
We further continued to determine which SRSF1 domain interacts with U1-70K RRM. We measured binding affinity of U1-70K RRM to SRSF1 RRM1, RRM2, or the RRM1–RRM2 tandem (Figure 3b). As these domains are highly soluble in PBS, the FP assays were performed in PBS. We found that U1-70K RRM binds to SRSF1 RRM1 with a  $K_d$  of 3  $\mu$ M but not with RRM2 or the RRM1–RRM2 tandem. We also found that the interaction between U1-70K RRM and SRSF1 RRM1 can be completely abolished by 200 mM Arg/Glu (Figure S2b). This is consistent with our FP binding assays that the RRM/RRM1 interaction does not contribute significantly to U1-70K/SRSF1 binding in the presence of 200 mM Arg/Glu. In summary, we concluded that BAD1 and RRM of U1-70K interact with phosphorylated RS and RRM1 of SRSF1, respectively.

## 2.3 | The phosphorylatable U1-70K BAD1 region is responsible for SRSF1 RS interaction

Based on the primary sequence feature, BAD1 can be divided into three regions, the N-terminal basic–acidic

**FIGURE 2** The U1-70K RRM and BAD1 domains are essential for its interaction with SRSF1. (a) His-Sumo tagged U1-70K constructs were immobilized on Ni-NTA beads as baits. Hyper-phosphorylated SRSF1 purified from *E. coli* was used as the prey. Sumo-specific and SRSF1-specific antibodies were used to detect the input U1-70K, and pull-down SRSF1, respectively. Pull-down band intensities were measured by ImageJ and normalized by input band intensities. BAD1, U1-70K BAD1 domain alone; FL, full-length U1-70K; Sumo, His-Sumo alone.  $\Delta$ BAD1,  $\Delta$ BAD2, and  $\Delta$ RRM denote the U1-70K constructs with corresponding domain deleted. The abbreviation is also used for panel b. (b) HeLa S3 cell lysate was used as the prey. GAPDH-specific antibody was used to ensure similar amounts of cells were used. All pull-down assays were performed three times, and the standard deviation was estimated from these replicates. (c) Exogenous U1-70K RRM and BAD1 inhibit *in vitro* splicing of  $\beta$ -globin. A representative image of a urea polyacrylamide gel used for *in vitro* splicing assays. A  $^{32}$ P labeled  $\beta$ -globin mini-gene, containing exon1 (ex1), intron 1 (in1), and exon 2 (ex2), was combined with nuclear extract (NE) for *in vitro* splicing assays. Schematic representations on the right illustrate the splicing product and intermediate species. Different concentrations of purified exogenous U1-70K BAD1, BAD2, and RRM were introduced to test their impacts on the splicing efficiency. (d) The percentage of spliced  $\beta$ -globin mRNA was quantified by measuring band intensities with ImageJ. To calculate the percentage of splicing, we determined the intensity ratio of the mature RNA product (comprising both ex1 and ex2) to the total input pre-mRNA. These *in vitro* splicing assays were repeated three times for error estimation, and the error bars represent the standard deviation.

**FIGURE 3** Map the domain–domain interaction between U1-70K and phosphorylated SRSF1 using FP binding assays. (a) FP binding assays for U1-70K BAD1 with SRSF1 constructs performed in PBS. (b) FP binding assays for U1-70K RRM with SRSF1 RRM1, RRM2, or the RRM1–RRM2 tandem. \* indicates Alexa Fluor 488 labeled molecules. 10 nM of fluorescein-labeled protein probes were used.  $K_d$  values are listed in Table 2.



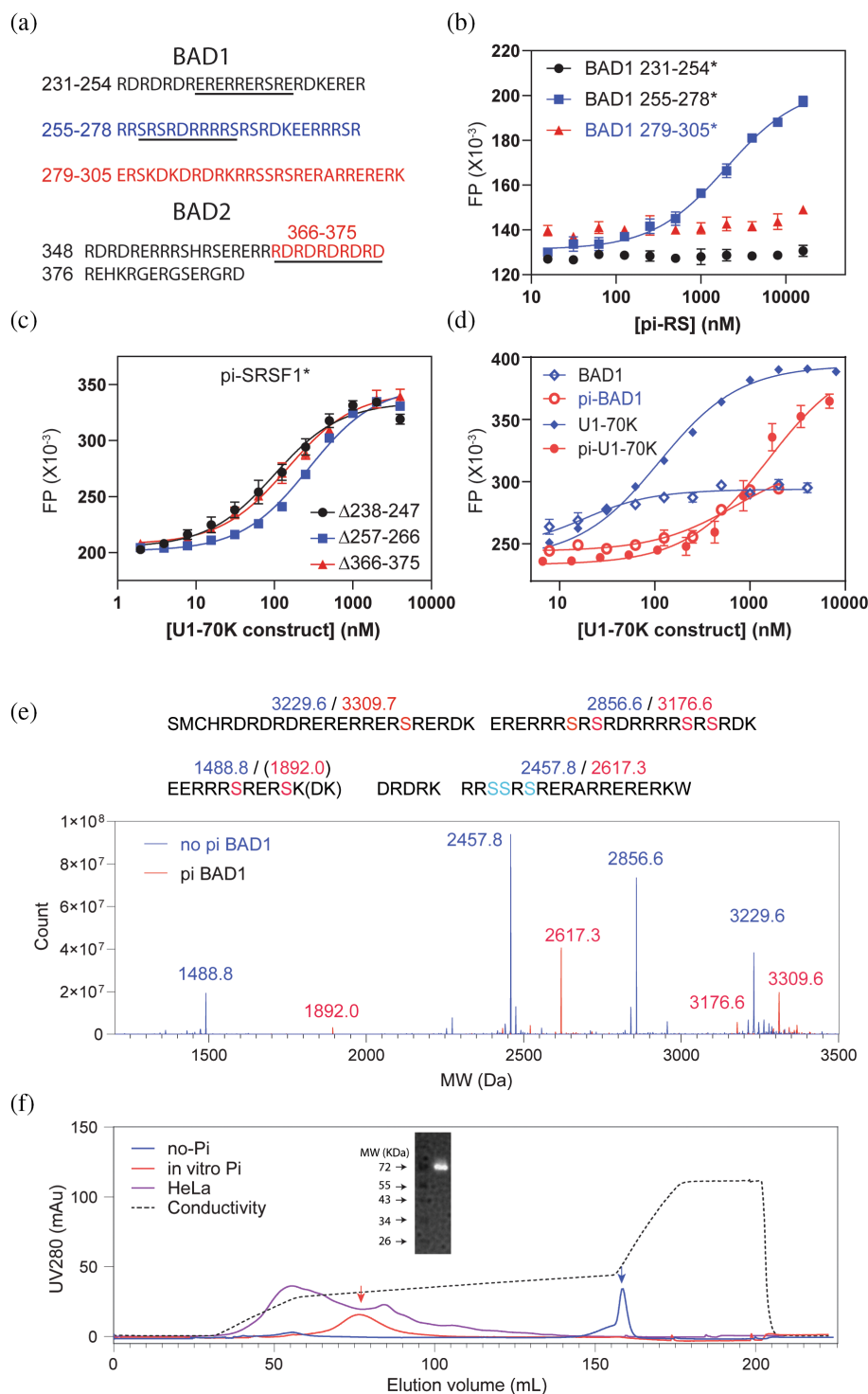
**TABLE 2** Dissociation constants ( $K_d$ ) of U1-70K domains to SRSF1 domains.

Titrants	Fluorophore labeled protein (10 nM)	$K_d$ (nM)
U1-70K BAD1	SRSF1 RRM1	240 ± 45
U1-70K BAD1	SRSF1 RRM2	10,000 ± 700
U1-70K BAD1	SRSF1 Pi-RS	1.0 ± 1.0
U1-70K BAD1	SRSF1	12 ± 6
U1-70K RRM	SRSF1 RRM1-RRM2	12,000 ± 4000
U1-70K RRM	SRSF1 RRM1	3000 ± 650
U1-70K RRM	SRSF1 RRM2	23,000 ± 6000

region (residues 231–254), the middle Arg–Ser rich regions (residues 255–278), and the C-terminal basic–acidic region (residues 279–306) (Figure 4a, Table 3). To determine which region is responsible for RS interaction, we performed FP assays using Alexa Fluor 488 labeled peptides (Figure 4b). We found that only the central region of BAD1, rich in Arg and Arg–Ser (residues 255–278), has detectable binding to SRSF1 RS (Figure 4b). To confirm our peptide binding results, we also performed binding assays using U1-70K constructs with deletions in BAD1 or BAD2 (Figure 4c). These three deletions were designed to probe roles of ER, RS, and DR repeats in SRSF1 binding. To ensure the binding results are comparable, the deleted peptide regions have the same length. Consistent with our peptide binding results, deletion of the RS had the most pronounced effect on SRSF1 binding (Figure 4c). Deletion of ER or DR repeats of BAD1 did not affect its binding to SRSF1 (Figure 4c). Notably, the RS region of BAD1 alone has a lower binding affinity to SRSF1, compared with the entire BAD1. This suggests that the RS region is the most important region, but other regions also contribute to the binding. This also explains our observation that deletion of the RS region cannot completely abolish BAD1’s binding to SRSF1 (Figure 4c).

RS repeats are potential substrates for SRPK and Clk kinase families. These two kinase families regulate functions of SR proteins and other splicing factors by phosphorylating arginine–serine dipeptide repeats. We wondered whether SRPK1 phosphorylated BAD1, and how this phosphorylation affected its interaction with SRSF1. We found that BAD1 was phosphorylated by SRPK1 in vitro, and the binding affinity of phosphorylated BAD1 to SRSF1, with a  $K_d$  of 700 nM, was 60-fold weaker than that of the nonphosphorylated protein. Using mass spectrometry, we found that SRPK1 adds 8 to 9 phosphate groups on BAD1, and further identified the phosphorylation sites (Figure 4e). We identified seven unambiguous phosphorylation sites (Figure 4e). Two of three serine residues in the RSSRS motif (residues 291–295) were phosphorylated.

U1-70K is a heavily phosphorylated protein. It is of interest to know whether U1-70K in the cell and the SRPK1-phosphorylated U1-70K in vitro have a similar pattern. Therefore, we prepared in vitro phosphorylated U1-70K by SRPK1 and found that SRPK1 adds in 8 to 9 phosphates groups to U1-70K (Figure S3), similar as SRPK1-phosphorylated BAD1. We also carried out a large-scale HeLa S3 cell culture and fractionated the cell lysate using an SP column. The in vitro phosphorylated U1-70K were eluted at the same salt concentration as the U1-70K from the HeLa S3 cell (Figure 4f). It is noteworthy that nonphosphorylated U1-70K was eluted from a very different salt concentration (Figure 4f). Therefore, it is likely that phosphorylated U1-70K in the cell may have a similar phosphorylation extent as the in vitro phosphorylated U1-70K by SRPK1. Similar to the impact of phosphorylation on BAD1, U1-70K phosphorylation exhibits a 14-fold weaker binding affinity to SRSF1 (Figure 4d,  $K_d$  of 1500 ± 250 nM vs. 110 ± 17 nM). Due to the low yield of U1-70K in the HeLa cell and difficulties in ionizing hyper-phosphorylated U1-70K, we could not determine the exact phosphorylation sites by mass spectrometry.



**FIGURE 4** The phosphorylatable Arg-Ser region of U1-70K BAD1 is responsible for SRSF1 binding, and its phosphorylation inhibits SRSF1 interaction. (a) Protein sequences of U1-70K BAD1 and BAD2. (b) FP binding assays for hyper phosphorylated SRSF1 RS (pi-RS) and BAD1 sub-regions. (c) FP binding assays for phosphorylated SRSF1 with U1-70K deletion constructs. The protein regions deleted are underlined in panel a. (d) FP binding assays for phosphorylated SRSF1 with non-phosphorylated and phosphorylated BAD1 and U1-70K. \* indicates the Alexa Fluor 488 labeled molecules.  $K_d$  values are listed in Table 3. (e) BAD1 of U1-70K is heavily phosphorylated by SRPK1. Mass spectra of non-phosphorylated (blue) and phosphorylated (red) BAD1. The masses of Lys-C cleaved peptides are listed in the unit of Da. (f) The cation chromatography (SP) profiles for non-phosphorylated (blue), SRPK1-phosphorylated U1-70K (red), and the HeLa S3 cell lysate (purple). The blue and red arrows indicate the elution position of non-phosphorylated and phosphorylated U1-70K, respectively. The inset western gel confirms that the HeLa S3 cell lysate fraction indicated by the red arrow contains U1-70K.

## 2.4 | Structural characterization of the interaction between U1-70K BAD1 and SRSF1 RS

We have shown that U1-70K BAD1 and phosphorylated RS play a dominant role in the U1-70K/SRSF1 interaction. Separately, these two protein domains have a high solubility in PBS (>1 mM). Our circular dichroism (CD) spectra show that BAD1 contains both  $\alpha$ -helical

structure and random coil, and hyper phosphorylated RS (pi-RS) is mainly disordered (Figure 5a, dotted lines). When these two domains were mixed in a concentration higher than 100 nM, phase separation was observed (Figure 5b). We collected circular spectra for the BAD1 and pi-RS using a two-chamber cuvette. Upon mixing the two proteins, we observed that the content of helical structure decreased dramatically (Figure 5a, black and purple lines). Notably, the decrease in signal is not due to



**TABLE 3** Dissociation constants ( $K_d$ ) of BAD1 regions to hyper phosphorylated SRSF1.

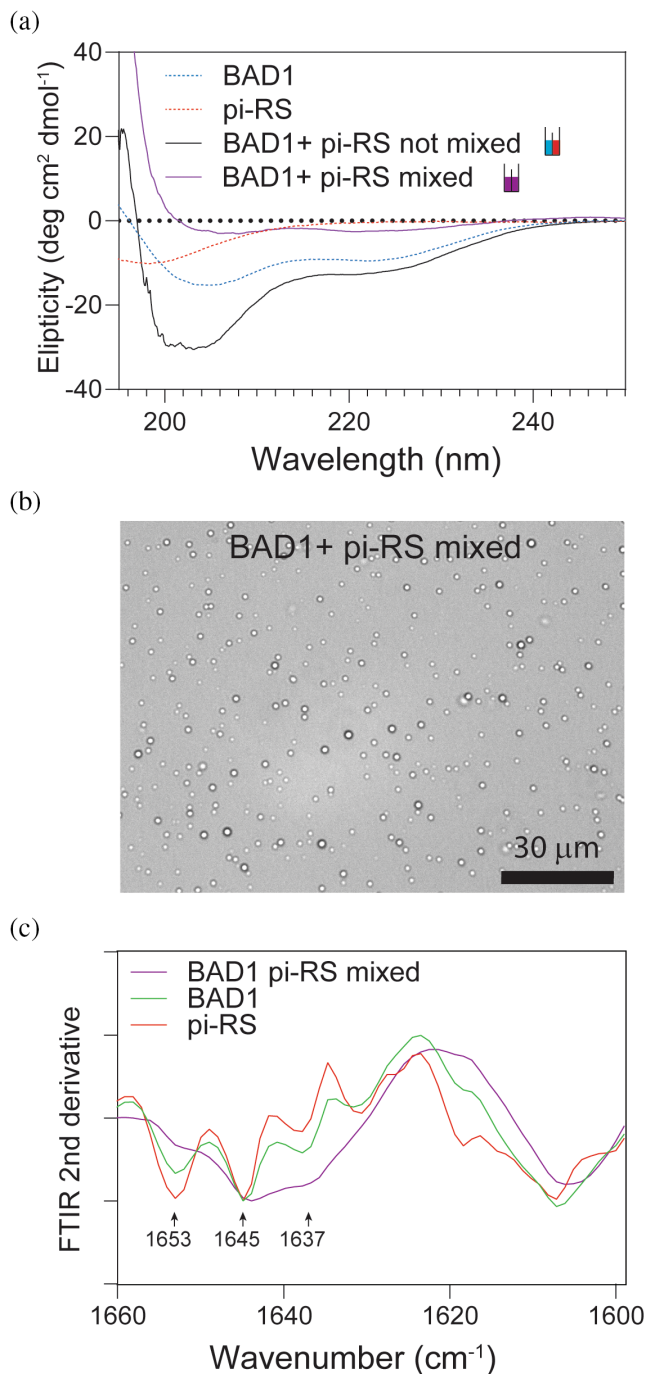
Titriments	Fluorophore labeled protein (10 nM)	$K_d$ (nM)
SRSF1 pi-RS	U1-70K BAD1 231–254	Too weak to determine
SRSF1 pi-RS	U1-70K BAD1 255–278	$2000 \pm 350$
SRSF1 pi-RS	U1-70K BAD1 279–305	Too weak to determine
U1-70K $\Delta$ 238–247	pi-SRSF1	$90 \pm 12$
U1-70K $\Delta$ 257–266	pi-SRSF1	$270 \pm 20$
U1-70K $\Delta$ 366–375	pi-SRSF1	$140 \pm 11$
U1-70K BAD1	pi-SRSF1	$12 \pm 6$
U1-70K pi-BAD1	pi-SRSF1	$700 \pm 150$

low photon transmission, as the working voltage of the photon multiplier tube was still lower than 600 V. We attribute this signal decrease to the cancellation of CD signals between the random coil and beta strand, as these two secondary structures exhibit opposite CD signals. Phase separation makes it challenging for NMR or X-ray characterization. However, phase separated sample is suitable for FT infrared spectra (FTIR), as protein droplets can be conveniently deposited onto the detector window. In addition, protein aggregation or phase separation does not interfere with infrared analysis. Our infrared spectrum for the BAD1 and RS mixture (Figure 5c) reveals that  $\beta$ -strands ( $1637\text{ cm}^{-1}$ ) and the random coil ( $1645\text{ cm}^{-1}$ ) dominate the spectrum. Only small amount of  $\alpha$ -helix ( $1653\text{ cm}^{-1}$ ) can be observed. Our CD and infrared analysis suggest that BAD1 undergoes a conformation switch from  $\alpha$ -helices to  $\beta$ -strands and random coil upon binding with RS.

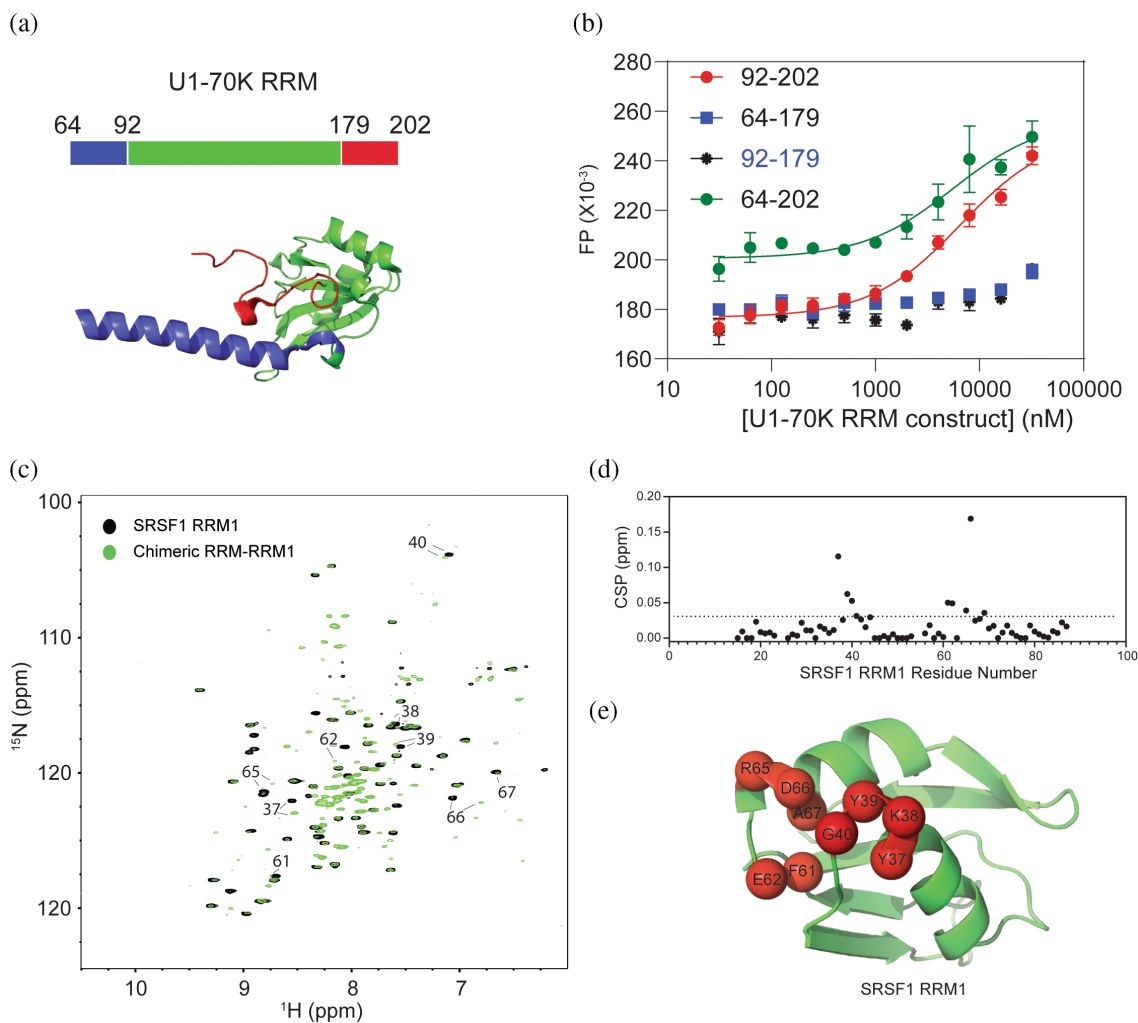
## 2.5 | SRSF1 RRM1 interacts with U1-70K RRM through a conserved site across the SR protein family

We have shown that U1-70K RRM interacts with SRSF1 RRM1. We further embarked on determining the interfaces for this interaction. U1-70K RRM has an N- and a C-terminal flanking regions (Figure 6a, Table 4). We found that the N-terminal flanking region of RRM is dispensable for SRSF1 RRM1 binding, while the C-terminal flanking region is essential (Figure 6b).

Due to the weak binding between RRM and RRM1, it is challenging to identify the SRSF1 RRM1 residues responsible for RRM interaction. To overcome this



**FIGURE 5** U1-70K BAD1 undergoes a conformation transition from  $\alpha$ -helix to  $\beta$ -strand and random coil upon interacting with hyper phosphorylated SRSF1 RS (pi-RS). (a) Circular dichroism spectra of individual U1-70K BAD1 (blue dotted line), individual hyper phosphorylated SRSF1 RS (pi-RS, red dotted line), before (black line) and after (purple) mixing in a two-chamber cuvette. (b) Phase separation of the BAD1/pi-RS mixture (500 nM for each protein). The image was recorded by a Cytation5 imaging system with a bright field microscope. (c) The second derivative of FT infrared spectra of BAD1, pi-RS, and mixture of the two proteins. Black arrows indicate the wavenumber for  $\alpha$ -helix, random coil, and  $\beta$ -strands, respectively.



**FIGURE 6** Identification of U1-70K RRM regions responsible for SRSF1 RRM1 interaction. (a) Domain architecture (top) and structure (bottom) of U1-70K RRM (PDB ID: 4PKD). (b) RRM1 interacts with U1-70K RRM that comprises the flanking regions (residues 64–202), but not with the RRM core (residues 64–179) or RRM without the C-terminal flank region (residues 92–179). RRM1 was labeled with Alexa Fluor 488, and assays were performed in PBS.  $K_d$  values are listed in Table 4. (c)  $^{15}\text{N}$ - $^1\text{H}$  HSQC of chimeric protein of U1-70K RRM/SRSF1 RRM1 (blue, RRM-RRM1) and SRSF1 RRM1 (black). The SRSF1 RRM1 residues demonstrating significant chemical shift perturbations (CSP) are labeled. (d) CSP analysis of RRM1. The dotted line indicates 2-fold of CSP RMSD. (e) The residues with significant CSP values (residues whose CSP are higher than the dotted line in panel d) are mapped onto the SRSF1 RRM1 structure (PDB ID: 1X4A).

**TABLE 4** Dissociation constants ( $K_d$ ) of U1-70K RRM constructs to SRSF1 RRM1.

Titants	Fluorophore labeled protein (10 nM)	$K_d$ (nM)
U1-70K RRM 64–202	SRSF1 RRM1	$6000 \pm 1400$
U1-70K RRM 64–179	SRSF1 RRM1	Too weak to determine
U1-70K RRM 92–202	SRSF1 RRM1	$7000 \pm 1100$
U1-70K RRM 92–179	SRSF1 RRM1	Too weak to determine

problem, we created a chimeric protein with the two domains connected by a flexible Gly-Ser linker. Comparing the  $^{15}\text{N}$  HSQC spectrum of the chimeric protein with RRM1 alone allowed us to identify the RRM1 residues involved in protein–protein interactions. As shown by Figure 6c, majority of RRM1 residues overlap with its counterpart in the chimeric protein. However, several residues exhibit obvious chemical shift perturbations (Figure 6d). Interestingly, all these residues are cluster around edges of two alpha-helices (Figure 6e). Our previous study has shown that these sites are conserved across

the SR proteins and involved in intramolecular interaction with the RS domain of SRSF1.

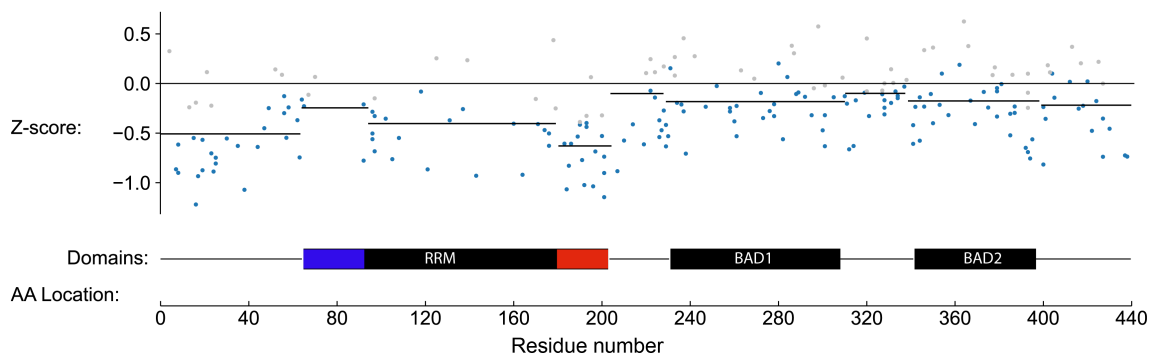
## 2.6 | In-cell identification of essential domains for U1-70K

Splicing factors like U1-70K are recognized as essential genes. To explore the roles of their various domains in cell fitness, we used a high-density CRISPR tiling scan approach for saturation mutagenesis screening of U1-70K coding exons. Our sgRNA library comprised 292 sgRNAs targeting U1-70K coding exons and 400 negative controls (non-targeting or targeting intergenic regions). We transduced this library into Cas9-expressing K562 cells and cultured them for 12 days. After culture, we noted a significant depletion of U1-70K sgRNAs (Figure 7). Mapping these sgRNAs to U1-70K amino acid positions revealed that the RRM domain, particularly its C-terminal flank, was most depleted, indicating its crucial role in cell fitness (Figure 7). Given that U1-70K RRM is essential for binding with U1 RNA, mutations unsurprisingly damage this key function. However, the C-terminal flank (red bar

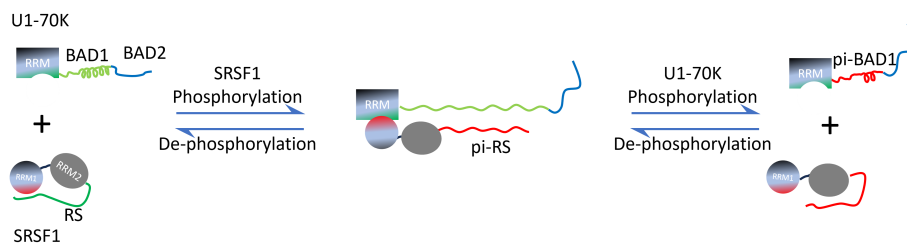
in Figure 7) plays a secondary role in binding to U1 RNA. Therefore, this region has additional crucial functions beyond RNA binding. Notably, the BAD1 and BAD2 domains also contribute to cell fitness, although to a lesser extent than the RRM domain, underscoring their varied importance in maintaining cell fitness. Interestingly, perturbations in BAD1 have a more pronounced impact on cell fitness compared to BAD2, consistent with our pull-down, SRSF1 binding, and *in vitro* splicing assays (Figure 2).

## 3 | DISCUSSION

This study reveals that U1-70K/SRSF1 interaction involves the direct contact of the C-terminal extension of U1-70K with the electronegative patch of SRSF1 RRM1, and SRSF1 RS with U1-70K BAD1 (Figure 8). In addition to SRSF1, SRSF2, and SRSF5 have also been reported to interact with U1-70K (Li et al., 2022). We found that the RRM1 domain and the RS domain of SRSF1 are crucial for its interaction with U1-70K, whereas the RRM2 domain of SRSF1 plays an inhibitory role. Notably, the



**FIGURE 7** CRISPR tiling scan across U1-70K domains in human K562 cells. Each dot represents the amino acid position (x-axis) and the Z-score of  $\log_2$ -fold change in sgRNA abundance between day 12 and day 0 post library infection and selection (y-axis). Blue dots highlight sgRNAs with significant differential representation. The black lines indicate the average Z-score of each domain.



**FIGURE 8** Schematic of the U1-70K/SRSF1 interaction and its modulation by phosphorylation. SRSF1 phosphorylation facilitates the interaction, whereas U1-70K phosphorylation inhibits the interaction. The RRM domains are shown in shapes, and disordered protein regions (BAD1, BAD2, and SRSF1 RS) are shown by lines. Electropositive and electronegative regions responsible for the U1-70K/SRSF1 interaction are shown in green and red, respectively. The nonphosphorylated RS interacts with the RRM1 electronegative patch responsible for interacting with U1-70K RRM. This self-inhibition is removed by phosphorylation of RS.

RRM1 and the RS domains are shared among all the 12 SR proteins, whereas RRM2 (also known as homologue RRM) is only present in SRSF1, 4, 5, 6, and 9. This suggests that the role of RRM1 in U1-70K binding we find is applicable to other SR proteins. Indeed, the SRSF1 RRM1 epitope (Figure 6e) for U1-70K interaction is conserved across the SR protein family. In addition to RRM1, we find that SRSF1 RS is critical for its tight binding to U1-70K. The RS domain size of SR protein family ranges from 50 aa to 300 aa. Therefore, it is possible that the contribution of RS domains to U1-70K interaction could vary across the SR protein family. This implies that SR proteins demonstrate not only diverse RNA-binding specificity, but also binding affinities to U1-70K.

It has long been known that phosphorylation plays a pivotal role in regulating the functions of SR proteins (Colwill, Feng, et al., 1996; Colwill, Pawson, et al., 1996; Duncan et al., 1998; Gui, Lane, et al., 1994; Gui, Tronchere, et al., 1994; Hertel & Graveley, 2005; Wu & Maniatis, 1993). For example, partial phosphorylation of the RS domain is required for the importing of SR proteins, and hyperphosphorylation is important for recruiting the U1 complex (Maertens et al., 2014). Our study here shows that hypo- to hyper-phosphorylation of SRSF1 RS progressively enhances its interaction with U1-70K. Phosphorylation reverses the charge property of SRSF1 RS and facilitates its interaction with electropositive BAD1. In addition, RS phosphorylation also regulates the interaction between SRSF1 RRM1 and U1-70K RRM (Figure 8). Our previous study has shown the intramolecular interaction between nonphosphorylated SRSF1 RS and the electronegative patch of RRM1 (Fargason et al., 2023). It is noteworthy that this electronegative patch largely overlaps with the RRM1 epitope for U1-70K interaction. Therefore, phosphorylation of SRSF1 RS removes the auto-inhibition and facilitates U1-70K interaction. In summary, phosphorylation of RS plays dual roles by fueling the RS/BAD1 interaction and removing the auto-inhibition on RRM1.

It has been controversial which U1-70K domains are essential for its interaction with SR proteins (Cao & Garcia-Blanco, 1998; Cho et al., 2011; Hertel & Graveley, 2005; Wu & Maniatis, 1993). Here, we have obtained purified full-length U1-70K and SRSF1, which allows us to perform fluorescence polarization binding assays and pull-down assays to investigate the roles of U1-70K domains in binding to SRSF1. We find that both RRM and BAD1 are essential for SRSF1 binding. We further confirm the importance of these two domains using *in vitro* splicing assays and evaluate their impacts on cell fitness with a high-density CRISPR tiling scan method. It is noteworthy that the C-terminal extension of U1-70K RRM is critical for cell viability, which is consistent with

its role in mediating interaction with SRSF1 RRM1. Perturbations on BAD1 and BAD2 also impact cell fitness, albeit to a lesser extent. Since the whole BAD1 contributes to binding with SRSF1 RS, perturbations on individual sites only have limited impacts. The CRISPR tiling scan can introduce mutations or deletions at target sites. Intrinsically disordered proteins are more tolerant to these perturbations, since these perturbations cannot perturb protein folding and structure. In addition, the BAD1 sequence contains repetitive RS and RR motifs. Mutation or deletion of one repeat can be readily compensated by neighboring repeats. This may explain the relatively lower impact of BAD1 and BAD2 compared with the C-terminal extension of U1-70K RRM.

We find the SRSF1/U1-70K interaction is also regulated by the phosphorylation of U1-70K. It has long been known that U1-70K is heavily phosphorylated in the cell (Woppmann et al., 1990). At least 13 species differing in the isoelectric point have been identified by 2D electrophoresis (Woppmann et al., 1990). A previous study has found that inhibiting the de-phosphorylation of U1-70K abolish RNA splicing (Tazi et al., 1993). In this study, we find that the majority of U1-70K purified from HeLa S3 cells is densely phosphorylated. Its SP chromatography profile is identical to the SRPK1-phosphorylated U1-70K purified recombinantly. This suggests that the phosphorylation pattern in HeLa S3 may be similar to the *in vitro* phosphorylated U1-70K. We find that the phosphorylation of BAD1 weakens its interaction with SRSF1, which is consistent with the previous finding that dephosphorylation is required for splicing (Tazi et al., 1993). In addition, most of the serine residues in U1-70K are clustered in BAD1. Therefore, it is likely that BAD1 is the region responsible for the heavy phosphorylation of U1-70K, and this phosphorylation represents an important mechanism to regulate U1-70K function. It still needs further research to determine the trigger for the phosphorylation and dephosphorylation of U1-70K in the cell.

The interaction between SRSF1 and U1-70K is distinct from ordinary protein–protein interactions in that it depends on disordered protein regions. A similar finding has been reported for the intrinsically disordered protein regions of histone H1, which interacts with prothymosin-alpha with pico-molar affinity (Borgia et al., 2018). Unlike the interaction between structured proteins, binding events mediated by disordered proteins have higher associate and dissociate rate constants, enabling a fast response to on/off signals (Borgia et al., 2018). This is critical to the functions of the dynamic complex formed between SRSF1 and U1-70K, which needs to disassemble before entering the B\* stage. In addition to mediating protein–protein interactions, SRSF1 RS and U1-70K BAD1 are also responsible for phase separation



(Fargason et al., 2023) (and unpublished data). Phase separation underpins the formation of nuclear speckles, which are closely related to RNA splicing. Therefore, our findings here hint that the SRSF1/U1-70K interaction can potentially modulate phase separation and distribution of these two proteins in the cell.

## 4 | MATERIALS AND METHODS

### 4.1 | U1-70K expression and purification

The DNA encoding human U1-70K was sub-cloned into pSMT3. To increase protein yield, U1-70K was co-expressed with DnaK and DnaJ that were sub-cloned in pBB535 (Addgene Plasmid # 27392) along with pRARE2 (Addgene plasmid # 26242). Transformed BL21star cells were cultured in TB media at 37°C to reach an OD<sub>600</sub> of 1.0, and 0.5 mM IPTG was added to induce protein expression. Cells were further cultured 16 h at 25°C. The cells were harvested by centrifugation (4000 RCF, 15 min at 4°C). The cell pellet was re-suspended in 20 mM Tris-HCl, pH 7.5, 25 mM imidazole, 0.1 mM TCEP, and 6M guanidinium HCl. The sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin, washed, and eluted in 200 mM Arg/Glu 20 mM MES pH 6.5, 500 mM imidazole, 0.2 mM TCEP, and 8M urea. The sample was further purified by a 5-mL HP SP column and exchanged into 800 mM Arg/Glu, 20 mM Tris-HCl pH 8.5, 0.1 mM TCEP, and 0.02% NaN<sub>3</sub> by centrifugal filtration.

The pSMT3 plasmid encoding U1-70K RRM (residues 64–202) was prepared by mutagenesis PCR. BL21-CodonPlus (DE3) cells were used for protein expression in LB media. Transformed cells were cultured at 37°C to reach an OD<sub>600</sub> of 0.6, and 0.5 mM IPTG was added to induce protein expression. Cells were further cultured 16 h at 22°C. The cells were harvested by centrifugation (4000 RCF, 15 min). The cell pellet was re-suspended in 25 mM HEPES, pH 7.5, 25 mM imidazole, 0.1 mM TCEP, 1 mM PMSF, 1 mg/mL lysozyme, and 1M NaCl. Cells were lysed by sonication, followed by centrifugation at 23,710 RCF for 40 min at 4°C to remove cell debris. The supernatant was applied to 5 mL of HisPur Ni-NTA resin, and eluted with 25 mM HEPES, pH 7.5, 0.5M NaCl, 0.2 mM TCEP, and 250 mM imidazole. The N-terminal SUMO-tag was removed by overnight cleavage at 4°C with 1 µg/mL Ulp1. The cleaved sample was purified by a 5-mL HiTrap Heparin column and polished by a HiLoad 16/60 Superdex75 column equilibrated with 25 mM HEPES, pH 7.5, 500 mM NaCl, and 0.2 mM TCEP.

### 4.2 | Protein purification of SRSF1 and its constructs

The DNA encoding human SRSF1 was sub-cloned into pSMT3 using BamH I and Hind III. The SRSF1 1–196 (ΔRS), SRSF1 14–90 (RRM1), and SRSF1 121–196 (RRM2) were prepared using mutagenesis PCR. All these mutants maintain the folded structure according to NMR spectra and bind with SRSF1 cognate RNA UCAGAGGA (De Silva et al., 2022; Fargason et al., 2023). All proteins were expressed by BL21-CodonPlus (DE3) cells in LB media or minimal media supplemented with proper isotopes for NMR experiments. Hyperphosphorylated SRSF1 was prepared by co-transformation of BL21-CodonPlus (DE3) cells using pSMT3/SRSF1 and CDC2-like kinase 1 (Clk1) cloned in pETDuet-1. Cells were cultured at 37°C to reach an OD<sub>600</sub> of 0.6 for LB or 0.8 for minimal media, and 0.5 mM IPTG was added to induce protein expression. Cells were further cultured 16 h at 22°C. The cells were harvested by centrifugation (4000 RCF, 15 min at 4°C).

#### 4.2.1 | Full-length SRSF1 (residues 1–248)

The cell pellet was re-suspended in 20 mM HEPES, pH 7.5, 150 mM Arg/Glu, 2M NaCl, 25 mM imidazole, 0.2 mM TCEP supplemented with 1 mM PMSF, 1 mg/mL lysozyme, 1 tablet of Pierce protease inhibitor, and 1 mM NaVO<sub>4</sub> for the hyperphosphorylated construct. After three freeze-thaw cycles, the sample was sonicated and centrifuged at 23,710g for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and then eluted with 60 mL of 20 mM MES pH 6.5, 300 mM imidazole, 600 mM Arg/Glu, and 0.2 mM TCEP. The eluted sample was cleaved with 2 µg/mL Ulp1 for 2 h at 37°C. The nonphosphorylated SRSF1 was then further purified by a 5-mL HiTrap Heparin column. The hyperphosphorylated SRSF1 was further purified by a 5-mL Cytiva Fast Flow Q column. The eluted samples from the ion exchange step were further purified by a HiLoad 16/60 Superdex 75 pg size exclusion column equilibrated with 800 mM Arg/Glu, pH 6.5, 1 mM TCEP, and 0.02% NaN<sub>3</sub>.

#### 4.2.2 | SRSF1 ΔRS (residues 1–196)

The cell pellet was resuspended in 20 mM Tris-HCl, pH 7.5, 2M NaCl, 25 mM imidazole, 0.2 mM TCEP, 1 mM PMSF, 0.5 mg/mL lysozyme, and 1 tablet of protease inhibitor. After three freeze-thaw cycles, the sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge.

The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and washed with 200 mL of 20 mM Tris-HCl, pH 7.5, 2M NaCl, 25 mM imidazole, and 0.2 mM TCEP. The sample was then eluted with 30 mL of 20 mM 2-Morpholinoethanesulfonic acid sodium salt (MES), pH 6.5, 500 mM imidazole, 500 mM Arg/Glu, and 0.2 mM TCEP. The eluted sample was cleaved with 2 µg/mL Ulp1 for 2 h at 37°C and diluted threefold with a buffer A of 20 mM MES, pH 6.0, 100 mM Arg/Glu, and 0.1 mM TCEP before being loaded onto a 5-mL HiTrap Heparin column. The sample was eluted over a gradient with a buffer B of 20 mM MES, pH 6.0, 100 mM Arg/Glu, 0.1 mM TCEP, 2M NaCl, and 0.02% NaN<sub>3</sub>. The protein was eluted around 50% buffer B. Fractions containing the target proteins were pooled, concentrated, and loaded onto a HiLoad 16/600 Superdex 75 pg size exclusion column equilibrated with 200 mM Arg/Glu, 0.2M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1 mM TCEP, and 0.02% NaN<sub>3</sub>.

#### 4.2.3 | RRM1 (residues 14–90)

The cell pellet was resuspended in a lysis buffer containing 25 mM HEPES, pH 8.5, 1M NaCl, 25 mM imidazole, 1 mM TCEP, 1 mg/mL lysozyme, 1 tablet of Pierce protease inhibitor, 1 mM PMSF, 0.02% NaN<sub>3</sub>. After three freeze-thaw cycles, the sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and washed with 100 mL of a loading buffer (25 mM HEPES, pH 8.5, 1M NaCl, 25 mM imidazole, 1 mM TCEP, 0.02% NaN<sub>3</sub>), followed by 100 mL of a high salt wash buffer (5M NaCl, 25 mL imidazole, 1 mM TCEP, and 0.02% NaN<sub>3</sub>), and finally an additional 100 mL of loading buffer. On-column cleavage was performed in 25 mL loading buffer supplemented with 2 µg/mL ULP1, 1 mM PMSF, 1 protease inhibitor tablet through inversion for 2 h. Protein was eluted using an elution buffer of 20 mM Tris pH 8.5, 500 mM imidazole 1 mM TCEP, 0.02% NaN<sub>3</sub>. The protein was concentrated and loaded onto a HiLoad 16/600 Superdex 75 pg size exclusion column equilibrated with 200 mM Arg/Glu, 0.2M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1 mM TCEP, and 0.02% NaN<sub>3</sub>.

#### 4.2.4 | RRM2 (residues 121–196)

The cell pellet was resuspended in a lysis buffer (25 mM HEPES, pH 8.5, 1M NaCl, 25 mM imidazole, 1 mM TCEP, 1 mg/mL lysozyme, 1 Pierce protease inhibitor tablet per 30 mL, 1 mM PMSF, 0.02% NaN<sub>3</sub>). After three freeze-thaw cycles, the sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter

Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and washed with 100 mL of a loading buffer (25 mM HEPES, pH 8.5, 1M NaCl, 25 mM imidazole, 1 mM TCEP, 0.02% NaN<sub>3</sub>), followed by 100 mL of a high salt wash buffer (5 M NaCl, 25 mL imidazole, 1 mM TCEP, and 0.02% NaN<sub>3</sub>) and finally an additional 100 mL of loading buffer. The protein was eluted using an elution buffer of 20 mM Tris pH 8.5, 500 mM imidazole 1 mM TCEP, 0.02% NaN<sub>3</sub>. The protein was further purified using a 5-mL Cytiva Fast Flow Q column on a gradient between a buffer A of 50 mM Arg pH 8.5, 1 mM and a buffer B of 50 mM Arg/Glu pH 8.5, 2M NaCl, 1 mM TCEP.

#### 4.2.5 | RS tail (residues 199–248)

The plasmid encoding SRSF1 RS tail (residues 199–248) was prepared by mutagenesis PCR, and BL21-CodonPlus (DE3) cells were used for protein expression in LB media. Transformed cells were cultured at 37°C to reach an OD600 of 0.6, and 0.5 mM IPTG was added to induce protein expression. Cells were further cultured 16 h at 22°C. The cells were harvested by centrifugation (4000 RCF, 15 min). The cell pellet was resuspended in 20 mM Tris-HCl, pH 7.5, 25 mM imidazole, 0.1 mM TCEP, and 6M guanidinium HCl. The sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and eluted with imidazole. The protein was purified using a mono S column. SRSF1 (199–248) was phosphorylated by incubation with SRPK1 and/or Clk1 (1 to 20 kinase to protein ratio) in 50 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, 2 mM ATP, 150 mM NaCl, 0.2 mM TCEP at 30°C overnight. The protein was purified using a mono Q column.

Hyper-phosphorylated SRSF1 was obtained by co-expression of SRSF1 with the CLK1 kinase as detailed by our previous research (Fargason et al., 2023). Hypo-phosphorylated SRSF1 was prepared by incubating 100 µM of SRSF1 with 5 µM of SRPK1 for 16 h at 30°C in 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 5 mM ATP, 0.2 mM TCEP, and 400 mM Arg/Glu. Hypo-phosphorylated SRSF1 was purified using a heparin column using pH to elute off the column using the same protocol as the unphosphorylated SRSF1.

### 4.3 | Fluorescence polarization binding assays

Fluorescence polarization assays were carried out using 10 nM of Alexa Fluor 488-labeled protein or RNA with

the binding partner concentrations ranging from 8000  $\mu\text{M}$  to 1 nM by 2-fold serial dilutions in 200 mM Arg/Glu, 20 mM Tris-HCl pH 8.5, 0.1 mM TCEP. The U1 RNA used was GGA GAA CCA UGA UCA CGA AGG UGG UUU UCC with fluorescein labeled at the 5' end. 50  $\mu\text{L}$  of samples were mixed in black 96-well plates (Costar). The fluorescence polarization data were gathered at 25°C using a BioTek Cytation5 imaging reader with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The binding affinities were determined using the quadratic equation below, where the fitting parameters  $F_{\text{min}}$ ,  $F_{\text{max}}$ , and  $K_{\text{d}}$  were the fluorescence polarization baseline, plateau value for the complex, and dissociation constant, respectively. The plateau value,  $F_{\text{max}}$ , depends on the size of the complex formed. Therefore, the same probe binding to different partners can yield different  $F_{\text{max}}$  values.  $[\text{P}_{\text{T}}]$  was the total titrant protein concentration, and  $[\text{L}_{\text{T}}]$  was the total fluorescently-labeled protein or RNA concentration (10 nM). Errors of  $K_{\text{d}}$  were calculated based on three independent measurements.

$$F_{\text{p}} = F_{\text{min}} + (F_{\text{max}} - F_{\text{min}}) \frac{([\text{P}_{\text{T}}] + [\text{L}_{\text{T}}] + K_{\text{d}}) - \left( \left( [\text{P}_{\text{T}}] + [\text{L}_{\text{T}}] + K_{\text{d}} \right)^2 - 4[\text{P}_{\text{T}}][\text{L}_{\text{T}}] \right)^{0.5}}{2[\text{L}_{\text{T}}]} \quad (1)$$

#### 4.4 | Pulldown and western blots

Purified 6His-SUMO-U1-70K (5  $\mu\text{g}$ ) or its various constructs were immobilized on 20  $\mu\text{L}$  of HisPur™ Ni-NTA Resin. The HeLa S3 lysate (prepared from 5 million of cells) was suspended in 1 mL of 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1% Triton X-100, protease inhibitors, PMSF, and phosphatase inhibitors. The Ni-NTA resin was incubated with the cell lysate at 4°C for 30 min. Samples were washed four times with 1 mL of 10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, protease inhibitor cocktail, PMSF, and 200 mM R/E. Proteins were eluted with 500 mM imidazole, 20 mM Tris-HCl, pH 8, 150 mM NaCl, 0.02% Tween, and PMSF. The samples were resolved on SDS-PAGE. Western blot transfer was performed at 400 mA for 1 h. SUMO Polyclonal Antibody (Catalog # 200-401-428) was used to detect U1-70K constructs, and anti-SF2 antibody (ab38017) was used to detect SRSF1. HRP-conjugated Affinipure Goat Anti-Rabbit IgG (H + L) was used as the secondary antibody. The images were collected using a SynGene imager, and band intensities were quantified by ImageJ.

#### 4.5 | Phosphorylation of U1-70K BAD1 and mass spectrometry

BAD1 (50  $\mu\text{M}$ ) was phosphorylated by SRPK1 (1  $\mu\text{M}$ ) for 1 h at 30°C in 50 mM Tris-HCl pH 7.5, 10 mM  $\text{MgCl}_2$ , 2 mM ATP, 150 mM NaCl, and 0.2 mM TCEP. Phosphorylated BAD1 was purified using a mono-S column. Phosphorylated BAD1 was exchanged into 500 mM guanidium HCl, 100 mM ammonium bicarbonate, pH 8.0, and 0.2 mM TCEP. BAD1 was further mixed with the Lys-C protease into a ratio of 20:1 and digested at 37°C for 24 h. Digested samples were diluted into 0.1% formic acid and separated on a 2.5  $\times$  150 mm A BEH SEC column (Waters) using 0.1% formic acid at a flow rate of 70  $\mu\text{L}/\text{min}$ . The elute of the column was connected inline to the ESI source of a Waters Synapt G2-S(i). Data was collected under MassLynx (Waters) in positive ion, resolution mode and the spectra were deisotoped using the MaxEnt3 module within MassLynx and the mass versus intensity plots.

#### 4.6 | FT infrared spectroscopy

U1-70K BAD1 (residues 229–306) and hyperphosphorylated SRSF1 RS (residues 199–248) were buffer exchanged into 5 mM MES pH 6.33, 10 mM NaCl, 0.1 mM TCEP, and concentrated to 1 mM. Samples were freeze-dried overnight, resuspended in 1 mL of  $\text{D}_2\text{O}$ , incubated at room temperature for 1 h, and re-freeze dried overnight to remove exchangeable hydrogens. The protein sample volume was reduced to 1/10th the original volume to increase the concentration using  $\text{D}_2\text{O}$ . Samples were scanned 1024 times on a Bruker ALPHA FT-IR spectrometer.

#### 4.7 | Circular dichroism

Proteins were exchanged into 100 mM NaF, 0.1 mM TCEP, 20 mM K phosphate, pH 7.5. CD spectra were collected on a Jasco J-815 CD spectrometer with a 1-mm light path cuvette. Spectral regions from 260 to 180 nm were collected and repeated three times.

#### 4.8 | In vitro splicing

The splicing efficiency was measured using the model mini gene  $\beta$ -globin as reported in the previous study (Krainer et al., 1984). The mini-gene was transcribed into RNA using the T7 polymerase and labeled by  $[\alpha\text{-}^{32}\text{P}]$

UTP (3000 Ci/mmol, 10 mCi/mL, 250 uCi, Cat. No: BLU007H250UC). After transcription, the RNA sample was cleaned by an Invitrogen Centri-Spin 40 column to remove the majority of unintegrated NTP. Then the RNA was capped by the vaccinia capping system (NEB M2080S) and purified by 6% urea PAGE gel. The purified RNA was subject to splicing assays using the protocol as detailed by Krainer et al. (1984). Briefly, *in vitro* splicing reactions (25  $\mu$ L) contained 1 mM ATP, 20 mM creatine phosphate, 3.2 U RNase inhibitor, 3.2 mM MgCl<sub>2</sub>, 2.6% PVA, 20 mM HEPES-KOH pH 7.4, 5000 CPM [<sup>32</sup>P]-labeled pre-mRNA, and 45 ng of HeLa S3 nuclear extract (NE). Reactions were incubated at 30°C for 4 h. Subsequently, 32  $\mu$ L of 1 mg/mL proteinase K stock was added to each sample, and the solution was incubated at 37°C for 15 min. Then, 6  $\mu$ L of 3M sodium acetate and 200  $\mu$ L of stop solution (300 mM sodium acetate, 0.1% SDS, 62.5  $\mu$ g/mL tRNA) were added. Additionally, 200  $\mu$ L of Phenol/Tris saturated solution (Catalog number: 327125000) was added to each sample. The samples were vortexed for 1 min, then spun at 14,000 rpm for 5 min. Next, 180  $\mu$ L of the sample was decanted, and 600  $\mu$ L of ice-cold ethanol was added. The samples were incubated at -20°C overnight. The samples were then spun at 14,000 rpm for 10 min at 4°C, and the supernatant was removed. RNA pellets were dissolved in formamide/EDTA stop dye (formamide with 0.1% bromophenol blue, 0.1% xylene cyanol, and 2 mM EDTA) and heated at 95°C for 3 min. The RNA samples were analyzed on a 6% urea PAGE gel at 200 V for 45 min. Subsequently, the gel was exposed to a phosphor-imaging screen overnight then imaged using an Amersham Typhoon5 (Part#29187191). The band intensities were quantified by ImageJ and three independent splicing assays were performed for estimation of error.

#### 4.9 | CRISPR tiling scan and data analysis

CRISPR tiling scan was performed as previously described (Zhou et al., 2023). Briefly, U1-70K protein tiling CRISPR library consists of 292 all possible sgRNAs targeting exons of human U1-70K and 400 negative control guides that are either non-targeting ( $n = 200$ ) or targeting intergenic regions ( $n = 200$ ). sgRNA oligos were synthesized at Genscript and cloned into LRG2.1\_Neo (Addgene, 125593) vector by Gibson Assembly. Lentiviruses of U1-70K protein tiling CRISPR library were packaged in HEK293FT cells and were used to infect Cas9-expressing human K562 cells at less than 0.5 MOI (multiplicity of infection) to ensure that each cell contained no more than one sgRNA. The transduced cells were selected by neomycin for 4 days, followed by cell

collection both at this point and 12 days following continued culture. Genomic DNA was isolated, and sgRNA sequences were amplified by two round PCR reactions using Nextera primers as described before (Zhou et al., 2023), followed by deep sequencing on the NovaSeq PE150 platform. After sequencing, sgRNA counts were extracted and mapped to sgRNA sequences by MAGeCK *count* command (Li et al., 2014). sgRNA counts were then normalized by DESeq2 (Love et al., 2014). Protein domains were predicted by ProTiler (He et al., 2019).

#### AUTHOR CONTRIBUTIONS

**Trent Paul:** Conceptualization; investigation; methodology; writing – original draft; validation. **Pengcheng Zhang:** Investigation; methodology. **Zihan Zhang:** Investigation. **Talia Fargason:** Investigation. **Naiduwadura Ivon Upekala De Silva:** Investigation. **Erin Powell:** Investigation. **Ethan Ekpenyong:** Investigation. **Shariq Jamal:** Investigation. **Yanbao Yu:** Investigation; methodology. **Peter Prevelige:** Methodology; investigation. **Rui Lu:** Methodology; investigation. **Jun Zhang:** Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; validation; methodology; formal analysis; project administration; supervision; resources.

#### ACKNOWLEDGMENTS

We want to thank Dr. Ronald Shin, the manager of UAB Central Alabama High-Field NMR facility. This work was supported by the United States National Institutes of Health (R35GM147091 to Jun Zhang).

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

#### ORCID

Jun Zhang  <https://orcid.org/0000-0002-5842-7424>

#### REFERENCES

- Anko ML. Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin Cell Dev Biol.* 2014;32: 11–21.
- Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature.* 2018;555:61–6.
- Cao W, Garcia-Blanco MA. A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem.* 1998;273:20629–35.
- Cavaloc Y, Bourgeois CF, Kister L, Stevenin J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA.* 1999;5:468–83.
- Chabot B, Steitz JA. Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. *Mol Cell Biol.* 1987;7:281–93.



- Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, et al. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A*. 2011;108:8233–8.
- Colwill K, Feng LL, Yeakley JM, Gish GD, Cáceres JF, Pawson T, et al. SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J Biol Chem*. 1996;271:24569–75.
- Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, et al. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J*. 1996;15:265–75.
- Das S, Krainer AR. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res*. 2014;12:1195–204.
- De Silva NIU, Fargason T, Zhang Z, Wang T, Zhang J. Inter-domain flexibility of human Ser/Arg-rich splicing factor 1 allows variable spacer length in cognate RNA's bipartite motifs. *Biochemistry*. 2022;61:2922–32.
- Dembowski JA, An P, Scoulos-Hanson M, Yeo G, Han J, Fu XD, et al. Alternative splicing of a novel inducible exon diversifies the CASK guanylate kinase domain. *J Nucleic Acids*. 2012; 2012:816237.
- Duncan PI, Stojdl DF, Marius RM, Scheit KH, Bell JC. The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. *Exp Cell Res*. 1998;241:300–8.
- Erkelenz S, Mueller WF, Evans MS, Busch A, Schöneweis K, Hertel KJ, et al. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA*. 2013;19:96–102.
- Fargason T, de Silva NIU, Powell E, Zhang Z, Paul T, Shariq J, et al. Peptides that mimic RS repeats modulate phase separation of SRSF1, revealing a reliance on combined stacking and electrostatic interactions. *Elife*. 2023;12:e84412.
- Gui JF, Lane WS, Fu XD. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*. 1994; 369:678–82.
- Gui JF, Tronchere H, Chandler SD, Fu XD. Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A*. 1994;91: 10824–8.
- He W, Zhang L, Villarreal OD, Fu R, Bedford E, Dou J, et al. De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. *Nat Commun*. 2019;10:4541.
- Hertel KJ, Graveley BR. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem Sci*. 2005; 30:115–8.
- Hulme EC, Trevethick MA. Ligand binding assays at equilibrium: validation and interpretation. *Br J Pharmacol*. 2010;161: 1219–37.
- Jarmoskaite I, AlSadhan I, Vaidyanathan PP, Herschlag D. How to measure and evaluate binding affinities. *Elife*. 2020;9:e57264.
- Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann R, Garcia-Blanco MA, et al. Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*. 1994; 368:119–24.
- Kondo Y, Oubridge C, van Roon AM, Nagai K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife*. 2015; 4:e04986.
- Krainer AR, Maniatis T, Ruskin B, Green MR. Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell*. 1984;36:993–1005.
- Li Q, Jiang Z, Ren S, Guo H, Song Z, Chen S, et al. SRSF5-mediated alternative splicing of M gene is essential for influenza A virus replication: a host-directed target against influenza virus. *Adv Sci (Weinh)*. 2022;9:e2203088.
- Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15:554.
- Liu HX, Chew SL, Cartegni L, Zhang MQ, Krainer AR. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol*. 2000;20:1063–71.
- Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*. 1998;12:1998–2012.
- Long JC, Cáceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*. 2009;417:15–27.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Öztöp I, et al. Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci U S A*. 2014;111:2728–33.
- Manley JL, Krainer AR. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev*. 2010;24:1073–4.
- Moore MJ, Query CC, Sharp PA. Splicing of precursors to mRNA by the spliceosome, vol. 55. New York: Cold Spring Harbor Laboratory Press; 1993.
- Palhais B, Præstegaard VS, Sabaratnam R, Doktor TK, Lutz S, Burda P, et al. Splice-shifting oligonucleotide (SSO) mediated blocking of an exonic splicing enhancer (ESE) created by the prevalent c.903+469T>C MTRR mutation corrects splicing and restores enzyme activity in patient cells. *Nucleic Acids Res*. 2015;43:4627–39.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40:1413–5.
- Schaal TD, Maniatis T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol*. 1999;19:1705–19.
- Siebel CW, Fresco LD, Rio DC. The mechanism of somatic inhibition of drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev*. 1992;6:1386–401.
- Siebel CW, Rio DC. Regulated splicing of the drosophila P transposable element third intron in vitro: somatic repression. *Science*. 1990;248:1200–8.
- Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*. 1998;92:315–26.
- Tacke R, Manley JL. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J*. 1995;14:3540–51.
- Tazi J, Kornstädt U, Rossi F, Jeanteur P, Cathala G, Brunel C, et al. Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature*. 1993;363:283–6.

- Will CL, Luhrmann R. Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol.* 1997;9:320–8.
- Woppmann A, Patschinsky T, Bringmann P, Godt F, Luhrmann R. Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res.* 1990;18:4427–38.
- Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell.* 1993;75:1061–70.
- Zhang Z, Krainer AR. Involvement of SR proteins in mRNA surveillance. *Mol Cell.* 2004;16:597–607.
- Zhang WJ, Wu JY. Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol Cell Biol.* 1996;16:5400–8.
- Zhou Z, Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma.* 2013;122:191–207.
- Zhou X, Zhang P, Aryal S, Zhang L, Lu R. UTX loss alters therapeutic responses in KMT2A-rearranged acute myeloid leukemia. *Leukemia.* 2023;37:226–30.
- Zillmann M, Rose SD, Berget SM. U1 small nuclear ribonucleoproteins are required early during spliceosome assembly. *Mol Cell Biol.* 1987;7:2877–83.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Paul T, Zhang P, Zhang Z, Fargason T, De Silva NIU, Powell E, et al. The U1-70K and SRSF1 interaction is modulated by phosphorylation during the early stages of spliceosome assembly. *Protein Science.* 2024;33(8): e5117. <https://doi.org/10.1002/pro.5117>