UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Characterization of mRNA splicing factors : ASF/SF2 and Ire1p

Permalink

https://escholarship.org/uc/item/5qz3j405

Author

Chakrabarti, Sutapa

Publication Date

2007

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Characterization of mRNA splicing factors: ASF/SF2 and Ire1p

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Chemistry

by

Sutapa Chakrabarti

Committee in charge:

Professor Gourisankar Ghosh, Chair
Professor Joseph A. Adams
Professor Simpson Joseph
Professor Maho Niwa
Professor Joseph P. Noel
Professor Emmanuel Theodorakis

2007

Copyright

Sutapa Chakrabarti, 2007

All Rights Reserved.

The dissertation of Sutapa Chakrabarti is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2007

TABLE OF CONTENTS

Signature Page……... iii

Table of Contents……... iv

List of Figures…… vii

List of Tables……. x

Acknowledgements…… xi

Vita and Publications……. xiii

Abstract…….. xiv

I. Introduction…… 1

A. Diverse mechanisms of mRNA splicing……... 2

B. Pre-mRNA splicing and the role of SR proteins……... 3

C. Regulation of SR protein function and sub-cellular localization.. 7

D. SR protein specific kinases……... 10

E. The protein kinase fold……. 12

F. Docking interactions in protein kinases mediate substrate specificity…….. 15

G. Focus of study……... 16

II. Materials and Methods……... 19

A. Preparation of recombinant DNA……. 20

B. Mutagenesis of ASF/SF2, Clk/Sty and SRPK1…… 22

C. Expression and purification of recombinant proteins……... 24

D. Co-immunoprecipitation assays of SRPK1 and ASF/SF2…… 28

E. In vitro kinase assays, inhibition and start-delay-trap assays…... 29

F. Phosphatase assays…… 30

G. Indirect immunofluorescence assays…… 30

H. Electrophoretic mobility shift assays…… 31

III. Recognition of the SR protein ASF/SF2 by the SR Protein Kinase, SRPK1…… 34
A. Introduction……... 35
B. Results……... 41
1. A docking motif on ASF/SF2 is responsible for the high affinity interaction with SRPK1……... 41
2. The docking motif of ASF/SF2 plays a role in dictating its mode of phosphorylation by SRPK1……..... 44
3. The docking motif of ASF/SF2 is important for its localization to nuclear speckles……..... 46
4. The docking motif ASF/SF2 limits its phosphorylation by SRPK1……... 48
5. Docking motif mutations of ASF/SF2 bypass the Clk/Sty function for its release from speckles…… 50
6. Specific residues within the RRM2 domain of ASF/SF2 contribute to its high affinity interaction with SRPK1…….. 55
7. Mutation of ASF/SF2 contact residues alters its steady-state localization but not its phosphorylation…… 56
C. Discussion……. 61

IV. Recognition of Specific RNA Sequence Elements by ASF/SF2 …….. 68
A. Introduction……... 69
B. Results……... 75
1. The phospho-serine residue in the P-2 pocket of SRPK1 is not a priming site for hypophosphorylation of ASF/SF2….. 75
2. Recognition of 5’ splice site and exonic splicing enhancer sequences by ASF/SF2…….. 77
3. Both RRM domains and the linker between the two RRMs are essential for ESE recognition of ASF/SF2…….. 84

4. Specific residues in the linker region contribute to the RNA binding affinity and specificity of ASF/SF2…… 89
C. Discussion……. 89

V. Expression, Purification and Crystallization of an Active Fragment of the Yeast Kinase-Endoribonuclease, Ire1p…… 94
A. Introduction……... 95
B. Materials and Methods…….. 101
1. Preparation of recombinant DNA……. 101
2. Mutagenesis of Ire1p……..... 102
3. Expression and purification of Ire1p……. 103
4. Limited proteolysis of Ire1p…….. 107
5. In vitro kinase assay…….. 107
6. Crystallization of Ire1p…….. 108
C. Results……... 108
1. Identification of stable, well-behaved constructs of Ire1p… 108
2. Crystallization of Ire1p…….. 116
3. Modification of protein for crystallization…… 121
4. Oligomeric State of Ire1p…….. 125
D. Discussion……. 127

VI. Discussion…….. 130
A. Identification of an unusual docking element in SR proteins…... 131
B. Understanding the interaction between SR proteins and RNA sequence elements…….. 133
C. Regulation of splicing factors by phosphorylation……... 135

References…….. 137

LIST OF FIGURES

Figure 1.1.	Domain organization of SR and SR-like proteins……..…...	5
Figure 1.2.	Role of SR proteins in spliceosome assembly……..	6
Figure 1.3.	Effect of phosphorylation on localization of SR proteins	9
Figure 1.4.	Domain organization of SR protein kinases……	11
Figure 1.5.	The protein kinase fold……	14
Figure 1.6.	Docking interactions mediated by protein kinases……	17
Figure 3.1.	Sequence of the RS domains of ASF/SF2 and Npl3p…….....	36
Figure 3.2.	Identification of a docking groove in SRPK1....…….....	38
Figure 3.3.	ASF/SF2 makes a 3-point contact with the kinase core of SRPK1..	40
Figure 3.4.	Identification of a docking motif in ASF/SF2…….……..	42
Figure 3.5.	The docking motif of ASF/SF2 dictates its mechanism of phosphorylation…….	45
Figure 3.6.	The RS domain of ASF/SF2 does not affect its mode of phosphorylation…….....	47
Figure 3.7.	The docking motif of ASF/SF2 is essential for its recruitment to nuclear speckles…….	49
Figure 3.8.	The docking motif of ASF/SF2 restricts its phosphorylation to the RS1 motif…….....	51
Figure 3.9.	Phosphorylation of the RS2 motif of ASF/SF2 by Clk/Sty releases it from nuclear speckles…….	52

Figure 3.10. SRPK1 and Clk/Sty regulate ASF/SF2 through differential phosphorylation of the RS domain…… 54

Figure 3.11. The contact residues of ASF/SF2 are responsible for its high affinity interaction with SRPK1…… 57

Figure 3.12. Stable association of ASF/SF2 with SRPK1 is essential for its proper subcellular localization……. 58

Figure 3.13. ASF/SF2 contact mutants can be hypophosphorylated by SRPK1 but are not hyperphosphorylated efficiently…….. 60

Figure 3.14. Multiple sequence alignment of select SR proteins…….. 63

Figure 3.15: Proposed model for regulation of ASF/SF2 by SRPK1 and Clk/Sty 67

Figure 4.1: Primed phosphorylation of SR proteins by SRPKs……... 70

Figure 4.2: ESE-dependent recruitment of U1snRNP to the 5' splice site…….. 72

Figure 4.3: A phospho-serine of ASF/SF2 binds the P-2 pocket of SRPK1….. 74

Figure 4.4: Mutation of specific residues within the inter-RRM linker of ASF/SF2 partially affects its subcellular localization……... 76

Figure 4.5: ASF/SF2(S116A/S119A) shows a time-dependent defect in its subcellular localization…….. 78

Figure 4.6: Mutation of residues in the inter-RRM linker of ASF/SF2 does not affect phosphorylation of its RS domain……..... 79

Figure 4.7: RNA binding properties of different forms of ASF/SF2……... 81

Figure 4.8: ASF/SF2 binds 5' splice site sequences with very low affinity…… 82

Figure 4.9 Purification of ASF/SF2(ΔRS)…….. 83

Figure 4.10: ASF/SF2 binds exonic splicing enhancer sequences with low affinity and high specificity..... 85
Figure 4.11: The individual RRM domains of ASF/SF2 do not bind ESEs…….. 86
Figure 4.12: The inter-RRM linker of ASF/SF2 enhances ESE recognition……. 88
Figure 4.13: Specific residues within the inter-RRM linker of ASF/SF2 contribute to its affinity and specificity of ESE recognition..... 90
Figure 4.14: Genetic complementation assay to test the effect of ASF/SF2 mutants on cell viability…… 93
Figure 5.1: The Unfolded Protein Response…… 97
Figure 5.2: Expression and purification of the cytoplasmic domain of Ire1p…. 110
Figure 5.3: Sequence alignment of Ire1 kinase-nucleases across species……... 111
Figure 5.4: Sequence and structural alignment of the kinase domain of Ire1p with the kinase cores of GSK3β, CDK2, PKR and PKA…….. 112
Figure 5.5: Purification of His-Ire1p(646-1115)……. 113
Figure 5.6: Identification and purification of a stable kinase core of Ire1p……. 115
Figure 5.7: Design of Ire1p constructs for crystallization……... 117
Figure 5.8: Crystallization of His-Ire1p(606-1115):ATP/$MgCl_2$…… 118
Figure 5.9: Crystals of Ire1p obtained from different constructs and conditions 120
Figure 5.10: Purification and crystallization of Ire1p(606-1115)…….. 122
Figure 5.11: Purification and crystallization of His-Ire1p(606-1115) mutants…. 124
Figure 5.12: Oligomeric states of the different constructs of Ire1p……... 126
Figure 6.1: Sequence alignment of individual RRM domains of SR proteins… 134

LIST OF TABLES

Table 2.1 Composition of Buffers used for Minimal Media ……..…..... 27

ACKNOWLEDGMENTS

I would like to thank my dissertation advisor, Dr. Gourisankar Ghosh, for his support and encouragement and for sharing his passion and enthusiasm for science. Working with him has been an enjoyable and interesting experience. My project in the G. Ghosh lab led me to work very closely with a group of people, making the process that much more exciting for me. I would like to thank Jacky Ngo for his significant contribution to my project and his company through all the late nights of grant writing. Thanks to Brad Nolen, Randall Lukasiewicz and Suhyung Cho for their insightful comments and contribution to the overall project. Special thanks to Suhyung for teaching me the basics of RNA biochemistry and her infinite patience with me throughout this process. Kayla Giang, Nhat Huynh and Jean Yeh were especially helpful in producing a lot of the material that I used in experiments described later.

My project has been a very collaborative one, giving me a chance to work with members of several other labs. I would like to thank Dr. X-D. Fu and people in his lab for providing me with material essential for my work. Special thanks to J.H. Ding for sharing his expertise on immunofluorescence assays and fluorescence microscopy. The work carried out in Dr. J.A. Adams' lab was very informative and especially useful to my project. I would also like to thank Dr. M. Niwa and Aditi Chawla for involving me in an exciting field of research and Dr. S. Joseph and his lab for help with RNA biochemistry.

Anu Moorthy, Tom Huxford, Rashmi Talwar, Debin Huang and Sebastien Vallee have provided guidance and helpful suggestions all along the way. I also thank Anu for the immense time and effort she spent in training me, which I now realize is a huge task.

I am lucky to have many wonderful friends in graduate school. Amanda has encouraged, applauded and criticized me as she deemed appropriate. Through her, I have come to realize that one doesn't have to be related to be part of a family. Don has been a great friend all throughout and a constant reminder of the 'good old days' in Urey Hall. Erika never fails to brighten up my gloomy days with her sunny nature and numerous anecdotes. Moni and Soumen have provided me with home-style comfort at all hours of the day. Jessica and Anthony, along with Vivien, have cheered me on to the end and I greatly appreciate their confidence in me. Olga offered scientific criticism, technical expertise and her honest input whenever I asked for it. I would also like to thank my friends outside of San Diego - Bipro, Somu, Nupur, Megha, and my cousin Sudakshina who shared my excitement when things worked well for me in lab and more importantly, listened to my endless complaints when they didn't. A special thanks to Nupur for proof-reading my entire thesis, a task that I would never undertake voluntarily.

For once, I would like to thank my sister, Sulagna, who was my first science teacher and has been my biggest cheer-leader, in a way that only sisters can be. Finally, although I can't thank my mum and dad (Drs. Asima and Syamalendu Chakrabarti) enough for everything, I would like to mention that their unconditional belief in me has been my biggest source of sustenance in graduate school.

Chapter III, in part, is reprint of material as it appears in *Molecular Cell*, 2005, **20**, 77-89, Ngo, J. C., Chakrabarti, S., Ding, J. H., Velazques-Dones, A., Nolen, B., Aubol, B. E., Adams, J. A., Fu, X. D. and Ghosh, G. The dissertation author was the primary researcher and author of this publication.

VITA

1999	B.Sc(Hons), St. Stephen's College, Delhi University, India
2001	M.Sc, Indian Institute of Technology Bombay, India
2007	Ph.D., University of California, San Diego

PUBLICATIONS

Ngo, J. C., Chakrabarti, S., Ding, J. H., Velazquez-Dones, A., Nolen, B., Aubol, B. E., Adams, J. A., Fu, X. D., and Ghosh, G. (2005). Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. Mol Cell *20*, 77-89.

Aubol, B. E., Chakrabarti, S., Ngo, J., Shaffer, J., Nolen, B., Fu, X. D., Ghosh, G., and Adams, J. A. (2003). Processive phosphorylation of alternative splicing factor/splicing factor 2. Proc Natl Acad Sci U S A *100*, 12601-12606.

Nolen, B., Ngo, J., Chakrabarti, S., Vu, D., Adams, J. A., and Ghosh, G. (2003). Nucleotide-induced conformational changes in the Saccharomyces cerevisiae SR protein kinase, Sky1p, revealed by X-ray crystallography. Biochemistry *42*, 9575-9585.

ABSTRACT OF THE DISSERTATION

Characterization of mRNA splicing factors: ASF/SF2 and Ire1p

by

Sutapa Chakrabarti

Doctor of Philosophy in Chemistry

University of California, San Diego, 2007

Professor Gourisankar Ghosh, Chair

Mechanisms of mRNA splicing greatly vary in complexity, ranging from simple reactions mediated by a two-component system to highly complex ones involving a macromolecular assembly called spliceosome. The former example of mRNA splicing is observed in the unfolded protein response (UPR) pathway, where intron excision is mediated by the kinase-endonuclease Ire1p and exon ligation is brought about by t-RNA ligase. Complex splicing reactions mediated by the spliceosome are more common and involve core ribonucleoprotein components and several accessory protein factors. The SR proteins are one such family of accessory factors. They are characterized by the presence of at least one N-terminal RNA recognition motif (RRM) and a C-terminal domain rich in arginine-serine dipeptides (RS domain), which can be extensively phosphorylated by SRPK and Clk/Sty kinases.

The crystal structure of SRPK1 bound to an SR-protein derived peptide led to the identification of a docking motif on a prototype SR protein, ASF/SF2. The docking

motif serves to restrict phosphorylation of ASF/SF2 by SRPK1, thus generating a hypophosphorylated species which localizes to nuclear speckles. Subsequent hyperphosphorylation by Clk/Sty leads to its release from the speckles. We prove that docking motif is an important regulatory element for phosphorylation and subcellular localization of ASF/SF2. The X-ray crystal structure of the SRPK1:ASF/SF2 complex suggests that interactions between the kinase core of SRPK1 and the RRM2 domain of ASF/SF2 contribute to their high affinity interaction. We show that mutation of contact residues in ASF/SF2 affects its nuclear localization, but does not perturb its phosphorylation by SRPK1. We suggest that the regulation of ASF/SF2 by SRPK1 extends beyond that of a substrate by its kinase, and that SRPK1 acts as a chaperone for ASF/SF2. We also investigated the RNA-binding property of ASF/SF2 and found that, like other auxiliary splicing factors, ASF/SF2 binds specific sequence elements in mRNA with low affinity. We also show that in addition to both RRM domains, the inter-RRM linker contributes significantly to the affinity and specificity of RNA binding by ASF/SF2.

Our interest in mRNA splicing and protein kinases led us to study Ire1p, which is a sensor and effector of UPR in yeast. Upon induction of UPR, Ire1p participates in a non-conventional splicing reaction to upregulate the transcription factor Hac1p, which induces UPR-responsive genes. We wanted to understand how the kinase and nuclease domains of Ire1p functionally interact with each other to orchestrate UPR, using X-ray crystallography as a tool. We have successfully purified and crystallized fragments of Ire1p, which encompass its kinase and nuclease domains. Attempts to determine the structure of the kinase-nuclease domain of Ire1p are currently underway.

I: Introduction

A. Diverse mechanisms of mRNA splicing

The gene expression program of a cell is regulated at a number of stages such as transcription, mRNA splicing, mRNA export, translation and post-translational modification. Although the mRNA splicing reaction is identical in every case (removal of introns and ligation of exons), the mechanisms and components involved in mediating this reaction can be very diverse. The simplest example of mRNA splicing is that of splicing of the transcription factor Hac1p, which occurs upon induction of the unfolded protein response (UPR) in yeast. Splicing of *HAC1* mRNA involves only two catalytic components – the kinase-ribonuclease Ire1p which cleaves the intron in the pre-mRNA and Rlg1p (t-RNA ligase) which seals together the free exons [1, 2]. In contrast to the mechanism described above, splicing of most other pre-mRNA transcripts in eukaryotes is highly complex and involves a multi-step reaction catalyzed by several protein and RNA components which are organized into a macromolecular assembly called the spliceosome [3]. The level of complexity in spliceosome-mediated splicing also differs among eukaryotes. Splicing in yeast is relatively simple compared to that in higher eukaryotes in a number of ways – whereas more than 95% of humans genes are interrupted by introns (average of 8 introns per gene), only 5% of the genes in yeast are interrupted by an intron (single intron per gene) [4, 5]. Furthermore, while the mammalian splicing reactions involve a host of auxiliary factors such as SR proteins to facilitate the reaction, yeast lack the functional equivalents of most of these factors, thus simplifying their splicing mechanism and spliceosome composition.

B. Pre-mRNA splicing and the role of SR proteins

Pre-mRNA splicing in higher eukaryotes can be divided into two categories; constitutive splicing, where every exon in the pre-mRNA is included in the processed mRNA, and alternative splicing where only select exons are included in the final product [6]. At least 74% of human multi-exon genes are alternatively spliced, suggesting the importance of their contribution in generating the complexity of the human proteome [7]. It has been estimated that 15% of point mutations that cause human genetic diseases affect splicing [8]. Both constitutive and alternative splicing are mediated by the spliceosome, which is a dynamic complex of several U snRNPs (U1, U2, U4, U5 and U6) and auxiliary factors such as hnRNPs and SR proteins [9]. Transient protein-protein and protein-RNA interactions within the spliceosome, together with post-translational modification of the proteins therein, add a level of complexity to the assembly and catalytic function of the spliceosome.

SR proteins are a family of conserved, essential splicing factors [10]. They are characterized by the presence of one or more N-terminal RNA recognition motifs (RRMs) and a C-terminal domain rich in arginine-serine (RS) dipeptides, from which they derive their name. The RRM domains, as the name suggests, are involved in sequence-specific RNA binding while the RS domain is believed to function as a protein-protein/protein-RNA interaction unit [11]. The RS domain can be interchanged among different SR proteins without any loss of function, suggesting that this domain acts as independent functional module [12]. Ten SR proteins have been identified in humans till date – the prototype alternate splicing factor/splicing factor 2 (ASF/SF2), SC35, SRp75, SRp55, SRp54, SRp40, SRp38, SRp30c, SRp20 and 9G8 [13]. Although evolutionarily

conserved in higher eukaryotes, the budding yeast *Saccharomyces cerevisiae* does not possess any prototype SR proteins. However, yeast has two SR-like proteins, Npl3p and Gbp2p, which contain RRM domains and one or more RS dipeptides interspersed in an arginine-glycine-glycine (RGG) rich domain [14, 15]. Figure 1.1 shows a schematic representation of select SR proteins from humans and yeast.

Several studies have established the involvement of SR proteins in constitutive and alternative splicing [10, 16-18]. SR proteins recognize positive splicing signals within the exons of mRNA, called exonic splicing enhancers (ESEs), through their RRM domains and mediate protein-protein interactions with components of the U snRNPs through their RS domains, thus facilitating spliceosome assembly [11, 19-21]. Specifically, enhancer-bound SR proteins recruit U1 snRNP to the 5' splice site and U2AF65 to the poly-pyrimidine tract within the intron, leading to the formation of the 'E' complex of the spliceosome [22, 23]. Furthermore, SR proteins are involved in recruitment of U2 snRNP and transition of the spliceosome from the 'A' to the 'B' complex, where the U4/U6.U5 tri-snRNP is recruited [24, 25]. The roles of SR proteins in different steps of spliceosome assembly have been summarized in Figure 1.2. SR proteins also directly recognize 5' splice site sequences of certain pre-mRNAs and RS domains of SR proteins have been shown to be sufficient for spliceosome assembly and catalysis [26, 27] . In addition to recruitment of spliceosome machinery, SR proteins serve to suppress negative splicing signals, called exonic splicing silencers (ESSs) which are recognized by the hnRNP class of proteins, and thus promote exon inclusion [28, 29]. This function of SR proteins is especially important for their roles in alternative splicing [30]. Studies demonstrate that while SR proteins are functionally redundant in

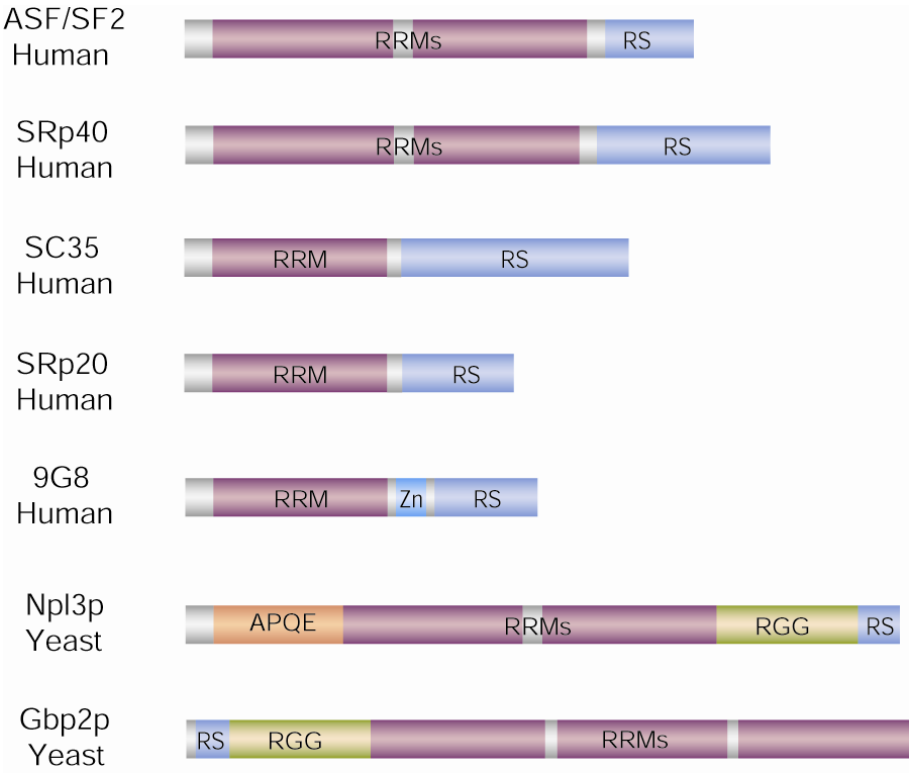


Figure 1.1. Domain organization of SR and SR-like proteins. Human SR proteins contain at least one RRM followed by an RS domain. The SR protein 9G8 is unusual since it has a zinc knuckle between its RRM and RS domain. The yeast SR-like proteins also contain RRM domains and one or more RS dipeptides which are dispersed in their RGG domains.

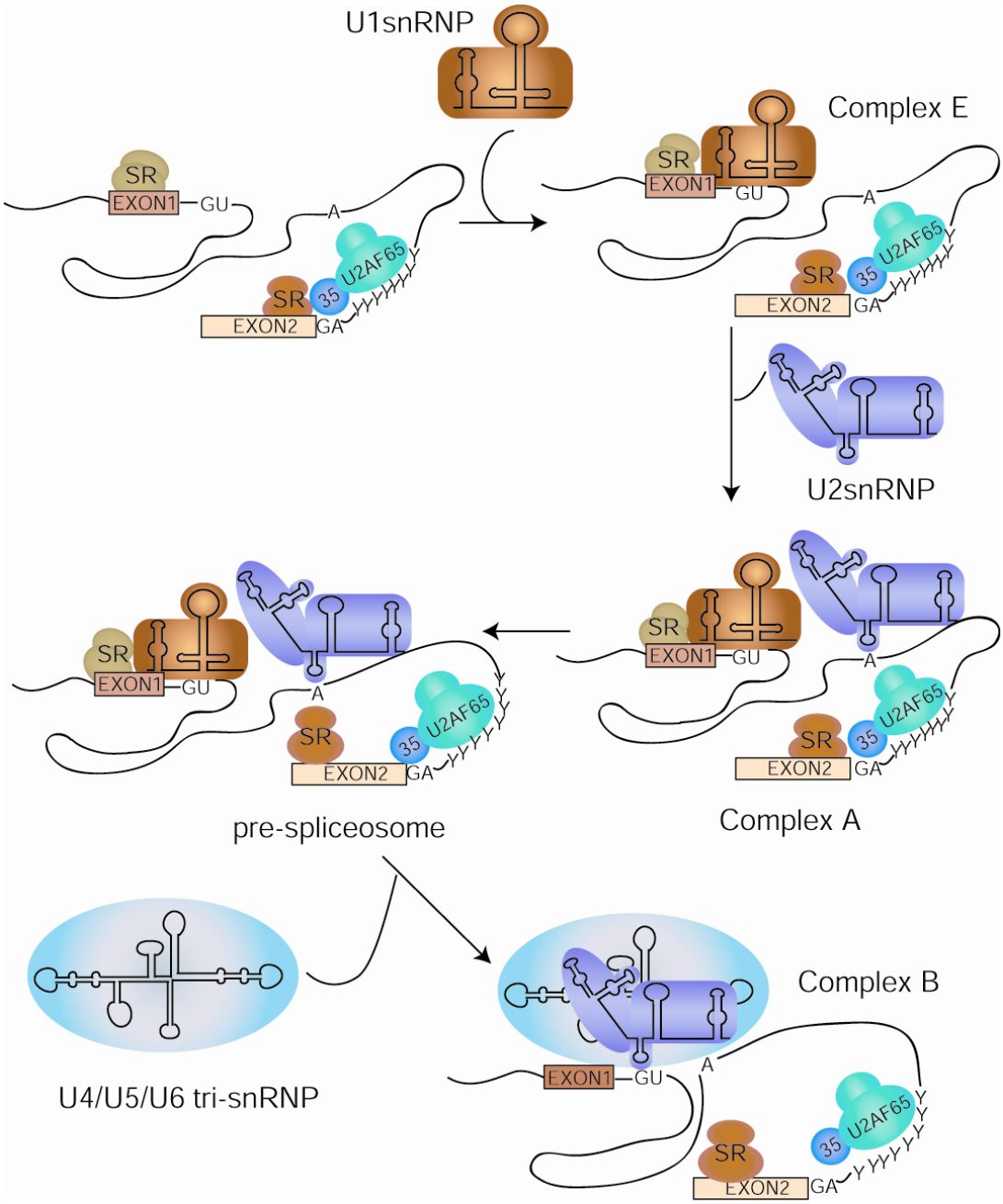


Figure 1.2. Role of SR proteins in spliceosome assembly. SR proteins are involved in every step of early spliceosome assembly. They recruit U1snRNP to the 5' splice and U2AF heterodimer to the poly-Y tract. SR proteins also contact the branch-point mRNA and function in the recruitment of U2 snRNP and U4/U5/U6 trisnRNP to the spliceosome. (Adapted from an article by H. Shen and M.R. Green, [123])

constitutive splicing, their roles in alternative splicing are not interchangeable [31]. Gene knockout experiments with SC35 show that T-cell development in mice is affected due to a defect in alternative splicing of a receptor tyrosine phosphatase, CD45 [32]. Similarly, ASF/SF2 knockout mice display a defect in postnatal heart development due to misregulation in splicing of the Ca^{2+}/Calmodulin dependent kinase II delta (CAMKIIδ) transcript [33]. Constitutive exons contain multiple elements responsive to different SR proteins, all of which facilitate spliceosome recruitment. This allows one SR protein to functionally compensate for the absence of another, thus ensuring that splicing of essential genes is not compromised. On the other hand, alternatively spliced exons often contain splicing signals specific for a particular SR protein and are thus more responsive to variation in levels of SR proteins.

C. Regulation of SR protein function and sub-cellular localization

Although the first SR protein, ASF/SF2, was identified by its role in 5' splice site cleavage and lariat formation, the composite SR protein family was identified by its immuno-reactivity against a phospho-specific antibody, mAb104 [34, 35]. SR proteins are now known to be phosphorylated at the serine residues in their RS domains, and can exist in hypo- or hyper-phosphorylated forms depending on the extent of phosphorylation [36]. Phosphorylation modulates protein-protein and protein-RNA interactions mediated by SR proteins, thus serving as a mode of regulating their function in spliceosome assembly and catalysis [37, 38]. Cycles of phosphorylation and dephosphorylation of SR proteins have also been shown to be essential for pre-mRNA splicing [39-41]. Additionally, presence of both phosphatases and phosphatase inhibitors modulate splicing

at different stages of spliceosome assembly and pre-mRNA cleavage [42, 43]. Together, these results establish that the correct phosphorylation state of SR proteins is critical for its function as a splicing factor. Cellular activities of SR proteins, other than pre-mRNA splicing, are also modulated by RS domain phosphorylation. ASF/SF2 and 9G8 act as adaptors for mRNA export only in their hypophosphorylated forms [44, 45]. Also, ASF/SF2 associates with spliced mRNA and activates translation when present in a hypophosphorylated form [46].

In addition to their cellular functions, the subcellular localization of SR proteins is also dictated by phosphorylation of their RS domains [47, 48]. All SR proteins localize to the nucleus in their steady state to form dense inter-chromatin granules, which are referred to as nuclear speckles [36]. However, some SR proteins are permanently retained in the nucleus while others shuttle between the nucleus and the cytoplasm [49]. SC35 and ASF/SF2 are examples of nuclear-resident and nucleo-cytoplasmic shuttling SR proteins respectively. Extensive studies on localization of SR proteins have established that phosphorylation of the RS domain is essential for both nuclear import and nucleo-cytoplasmic shuttling of SR proteins [50, 51]. Interestingly, SR proteins have been shown to associate with their specific importin, Transportin-SR, in a phosphorylation dependent manner [52, 53]. Furthermore, the extent of phosphorylation of the RS domain affects subcellular trafficking of SR proteins, thus adding another level of complexity to the regulation of localization [54]. While hyperphosphorylation of SR proteins by nuclear kinases releases them from the speckles, partial dephosphorylation is believed to be essential for nuclear export of shuttling proteins [44, 45, 55-57]. Figure 1.3 summarizes the effect of RS domain phosphorylation on localization of SR proteins.

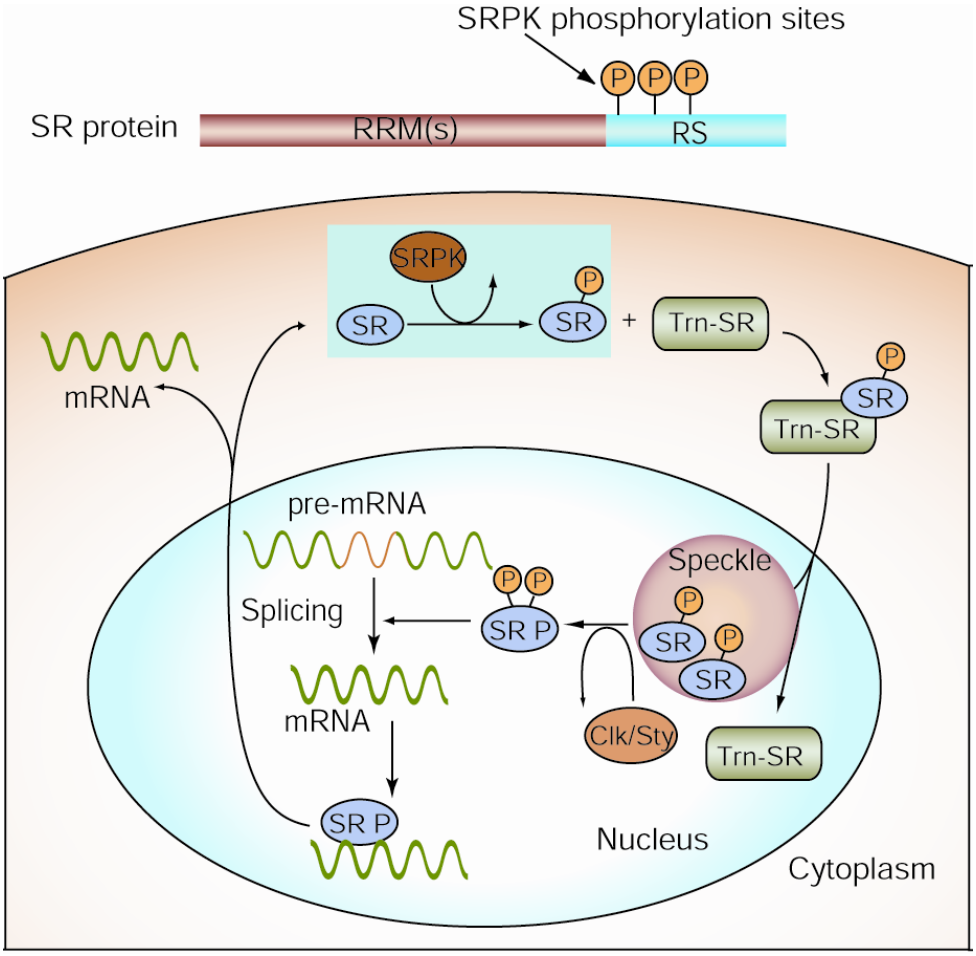


Figure 1.3. Effect of RS domain phosphorylation on subcellular localization of SR proteins. Hypophosphorylation of SR proteins in the cytoplasm by SRPK leads to its import into nuclear speckles, in a Transportin-SR (Trn-SR) dependent manner. Hyperphosphorylation of SR proteins in the speckles by Clk/Sty causes them to be released from the speckles and recruited to sites of splicing. Specific SR proteins are then dephosphorylated, which facilitates their interaction with mRNA and leads to their export along with the spliced mRNA.

D. SR protein specific kinases

SR proteins are phosphorylated at their RS domains predominantly by two families of protein kinases – the SR protein kinases (SRPKs) and the Cdc2-like-kinase/Serine-Threonine-Tyrosine (Clk/Sty) family of kinases [36]. SRPKs are serine-specific kinases and are phylogenetically classified as members of the CMGC family of serine/threonine kinases. They are evolutionarily conserved among higher eukaryotes. Humans contain two distinct family members, SRPK1 and SRPK2, and SRPK1a which is an alternatively spliced form of SRPK1 [58-61]. While SRPK1 is fairly ubiquitous, SRPK1a and SRPK2 are expressed in a tissue-specific manner. In contrast to the multiple SRPKs in higher eukaryotes, yeast contains only one SRPK, known as Sky1p (SRPK in yeast) (Figure 1.4) [62]. SRPK1, the first member of the SRPK family to be identified, was isolated by its ability to phosphorylate the SR protein SC35 [58]. Subsequent biochemical and genetic experiments led to the identification of SRPKs in other species [62, 63]. All SRPKs have unusual domain architecture in that their kinase domain is bifurcated by an insert, called the spacer domain, which varies in length and primary sequence among species. Although the spacer domain is not required for catalytic activity of the kinase, it has been shown to play a substantial role in regulating subcellular localization and maintaining structural stability of SRPK1 [64, 65]. SRPKs are predominantly cytoplasmic and translocate to the nucleus very briefly in a cell-cycle dependent manner [64]. They are constitutively active protein kinases and do not need activation loop phosphorylation or binding of regulatory factors to induce catalytic activity. The X-ray crystal structures of Sky1p and SRPK1 show that both kinases possess unique structural elements that maintain them in their constitutively active forms

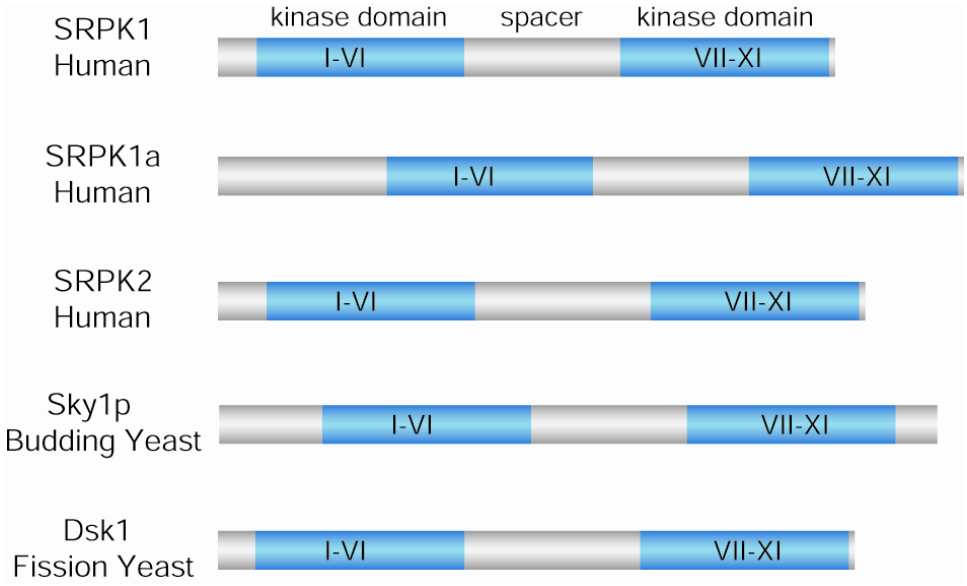


Figure 1.4. Domain organization of SR protein kinases. The kinase domains of all SRPKs across species are bifurcated by the spacer insert. SRPKs also have non-homologous N- and C-terminal extensions which may vary considerably in length.

[65, 66]. In absence of regulation of catalytic activity, it appears that SRPKs are regulated by cellular compartmentalization. Phosphorylation of SR proteins in the cytoplasm by SRPKs is critical for their nuclear import and cellular functions, thus rendering this family of kinases a potent regulator of SR proteins [36].

Contrary to SRPKs, Clk/Sty belongs to a family of dual-specificity kinases and phosphorylates serine, threonine and tyrosine residues [67]. In addition to a functional kinase domain, Clk/Sty contains an RS domain which is believed to mediate protein-protein interactions with the RS domains of SR proteins [56]. It is a nuclear kinase and specifically co-localizes to nuclear speckles with SR proteins [56, 67]. It has been shown that hyper-phosphorylation of SR proteins by Clk/Sty leads to their dispersion from nuclear speckles and recruitment to sites of active transcription and splicing [68-70]. Additionally, hyper-phosphorylation of SR proteins by Clk/Sty leads to a reduction in their in vitro splicing activity [71, 72]. It appears that SRPK and Clk/Sty kinases act in a concerted manner to achieve optimal phosphorylation of SR proteins and thus serve as their spatial and temporal regulators.

E. The protein kinase fold

The human kinome comprises of 518 protein kinases and makes up 1.7% of the entire genome [73]. Protein kinases regulate nearly every aspect of cellular function including gene expression, metabolism, cell proliferation and apoptosis to name a few. Although they are involved in very diverse cellular functions, all protein kinases essentially share a common structural scaffold which has evolved to mediate a phospho-transfer reaction using adenosine triphosphate (ATP) as a cofactor [74]. The crystal

structure of the cAMP-dependent protein kinase (PKA) was the first solved for a protein kinase and defined the kinase fold (Figure 1.5) [75]. The conserved kinase domain consists of two parts- the N-terminal small lobe which contains mainly β-strands and the C-terminal large lobe which is largely α-helical. Although differences in primary structure exist within the kinase core, certain features such as the glycine-rich loop, invariant lysine-glutamate ion pair, the catalytic loop and the activation segment are very well conserved among all protein kinases (Figure 1.5) [76]. These features have enormous functional significance and are involved in nucleotide-binding or catalysis.

Since kinases are involved in almost every cellular process, their stringent regulation is essential for proper functioning of a cell. There are two common mechanisms for regulation of protein kinases – phosphorylation of one or more residues at the activation loop (located within the activation segment) of the kinase and binding of a regulatory factor to the protein kinase [77, 78]. The activation segment of a kinase is defined as the region between the two conserved tripeptide motifs DFG (N-terminus) and APE (C-terminus) and includes the magnesium binding loop, the activation loop and the P+1 loop [78]. Phosphorylation of the activation loop induces a large conformational change which can switch a kinase from its inactive to active state (eg, PKA and the insulin receptor kinase IRK) [79, 80]. Binding of a regulatory factor can also induce large conformation changes in protein kinases and switch them from an inactive to an active state. A well-studied example of this mode of regulation is the cyclin dependent kinase (CDK) family, where binding of a regulatory cyclin to a CDK allows it to adopt an active conformation [81, 82]. The CDK2-cyclinA complex is an interesting example of

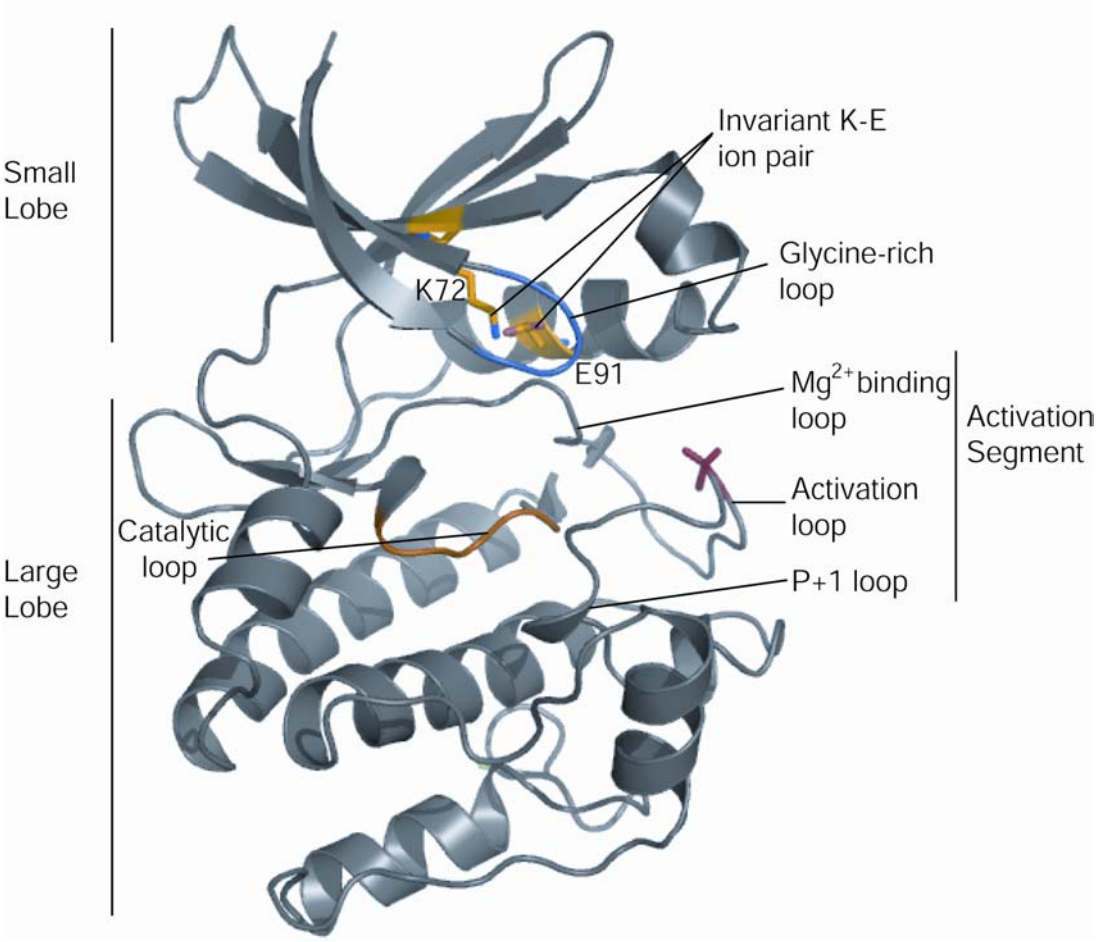


Figure 1.5. The protein kinase fold. Ribbon diagram of the protein kinase core as determined from the X-ray crystal structure of PKA (PDB ID: 1ATP). The small lobe, large lobe and other conserved kinase core features have been indicated.

regulation of protein kinases since it requires cyclinA binding as well as phosphorylation of its activation loop to adopt a fully active conformation [83, 84].

F. Docking interactions in protein kinases mediate substrate specificity

Substrate recognition by protein kinases can be divided into two components: interactions at the active site of the kinase and interactions at sites distal to the active site, referred to as docking interactions [85, 86]. Although most protein kinases display a preference for the target sequence in their substrates, this preference is not stringent and often, not enough to confer substrate specificity in vivo. Thus, distal docking interactions between a kinase and its substrate provide the specificity in recognition that is required in vivo [87]. Docking interactions also play roles in recruiting activators, inhibitors and scaffolding proteins to kinases and may induce allosteric changes within a kinase, leading to its activation [85, 88]. Analysis of the human kinome reveals that the majority of the serine/threonine kinases do not contain a recognizable docking module. This suggests that there exists a great diversity in the sequence and structural motifs that serine/threonine kinases employ for their docking interactions.

Docking interactions have been extensively studied in the MAP kinase family of serine/threonine kinases [89]. X-ray crystallographic studies coupled with H-D exchange/ mass-spectrometric techniques have identified two distinct docking elements in this kinase family [90-92]. The first is a hydrophobic groove formed between strands β7 and β8 and the αD-αE loop of the kinase, which recognizes the DEJL docking motif (consensus sequence: R/K-X_{2-4}-ϕ-X- ϕ) in substrates. This docking groove is utilized by the kinases p38α, ERK2 and Fus3 to interact with their substrates [90, 93, 94]. The

second docking element in MAP kinases is located very close to the active site of the kinase and is a groove formed by parts of helices αF and αG and the MAP kinase insert (a structural element unique to CMGC kinases). This docking groove recognizes a DEF motif (consensus sequence F-X-F) in substrates. This docking groove is utilized in the ERK signaling pathway for distinct substrate recognition events [95, 96]. Occasionally, substrates of MAP kinases contain both the DEJL and DEF docking motifs, thus facilitating their recruitment to multiple kinases [94-96].

Apart from the docking interactions described above, other regions of a kinase can be utilized for substrate docking. Interaction between the RNA-dependent protein kinase PKR and its substrate eIF2α is mediated predominantly by the non-canonical helix αG of PKR [97]. Another interesting example of extensive docking interactions between a kinase and its substrate is that of the yeast kinase Fus3 binding to the Ste5 scaffold [98]. Ste5 binds to Fus3 in a bipartite manner - the C-terminus of the binding peptide contains a DEJL motif that interacts with the large lobe of the kinase, while the N-terminus arranges itself into a single β strand that interacts with β strands in the small lobe of Fus3 to form a seven-strand β sandwich. All in all, it is clear that protein kinases use a myriad of different interaction surfaces to recognize their cognate substrates with high specificity in vivo (Figure 1.6).

G. Focus of study

The SR protein, ASF/SF2, interacts with the SR protein kinase, SRPK1, with an unusually high affinity. Also, ASF/SF2 is rapidly, but partially, phosphorylated at its RS domain in a processive manner. Our aim was to determine the molecular basis of this

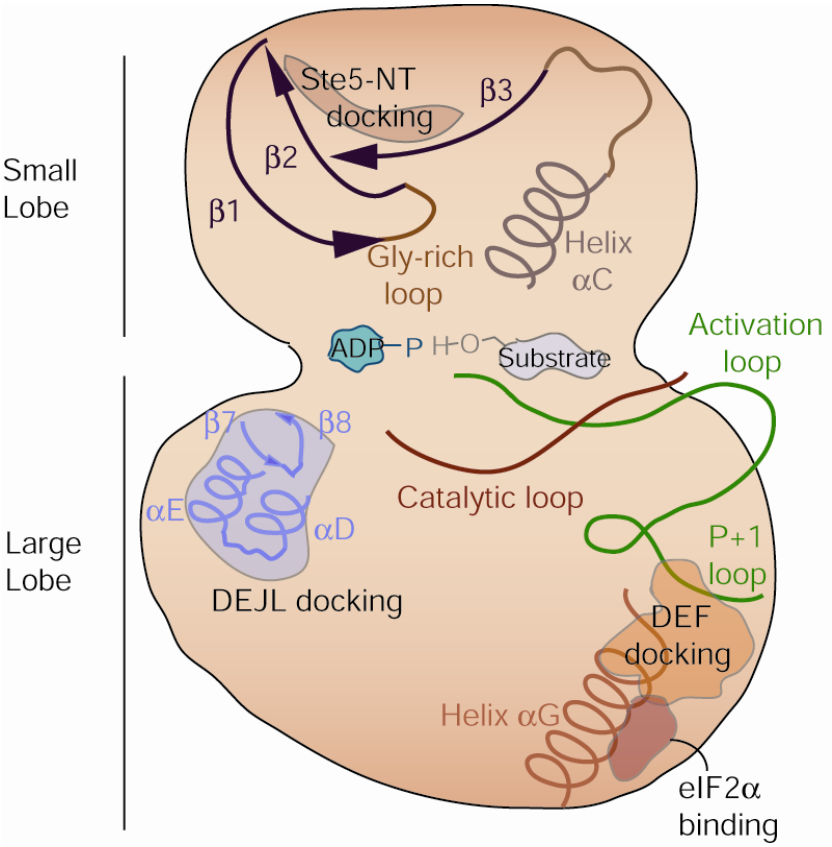


Figure 1.6. Docking interactions mediated by protein kinases. While recognition of the target residue on the substrate always occurs at the same active site (denoted by Substrate-OH and ATP), kinases use a variety of structural elements to mediate docking interactions with their substrates. Specific secondary structural elements utilized for docking of DEJL and DEF motifs to MAP kinases, eIF2α to PKR and the N-terminus of Ste5 docking peptide to Fus3 are indicated on the kinase scaffold. (Adapted from a review by Z. Shi and N.G. Ahn, [85])

unique phosphorylation of ASF/SF2 by SRPK1 and to test if it had any relevance in regulation of the function of ASF/SF2 in vivo. We used the X-ray crystal structures of SRPK1 bound to an SR-like peptide and ASF/SF2 as a basis for our hypotheses on regulation of ASF/SF2. Additionally, we wished to understand recognition of exonic splicing enhancers by SR proteins at a molecular level and thus investigated the RNA-binding function of ASF/SF2 and the regulation of this event.

Our combined interest in X-ray crystallographic studies on proteins kinases and molecular mechanisms of pre-mRNA splicing led us to undertake structural studies on the yeast kinase-endoribonuclease, Ire1p. We attempted to determine the X-ray crystal structure of a constructs of Ire1p that encompasses its kinase and ribonuclease domains with a hope to understand regulation of the novel splicing activity of this protein by its own kinase domain.

II: Materials and Methods

A. Preparation of recombinant DNA

The plasmids pET19b ASF/SF2(FL) and pET14bGST ASF/SF2(FL) harboring the cDNA of the full-length human ASF/SF2 (residues 1-248) were provided by Dr. Xiang-Dong Fu (Cellular and Molecular Medicine, University of California, SanDiego).

pGEX4T2 ASF/SF2(ΔRS2/1-219), ASF/SF2(ΔRS/1-196) and ASF/SF2(ΔCT/1-179): The inserts for these expression plasmid were generated by PCR amplification of the relevant cDNA segment from pET19b ASF/SF2(FL). A common 5' primer was used for all the constructs:

ASF(1)NTBamH1: 5' CGTGGATCCATGTCGGGAGGTGGT 3'

The 3' primers for all these constructs carried a NotI restriction site.

ASF(219)CTNotI: 5' CCGTTAAAGCGGCCGCCTAGCTTTTGCTACGGCT 3'

ASF(196)CTNotI : 5' CCGTTAAAGCGGCCGCCTACCCATCAACTTTAAC 3'

ASF(179)CTNotI: 5' GGACCAAAAGCGGCCGCTTACTTAGTGTTATCCAG 3'

pGEX4T2 ASF/SF2(ΔDock): The insert for this expression plasmid was generated by PCR amplification using pBabeHygro ASF/SF2(Δ183-196) (a gift from Dr. X-D Fu) as a template and the primers ASF(1)NTBamH1 and ASF(248)CTNotI.

ASF(248)CTNotI: 5' AGGACCAAAAGCGGCCGCTTATGTACGAGAGCGAGA 3'

The resulting DNA fragments were gel purified, double digested with the restriction enzymes BamHI and NotI (New England Biolabs) and then gel purified again. This fragment was then ligated into the vector pGEX4T2 (GE Healthcare) which was

previously double digested by the same restriction enzymes. The ligated products were transformed into *E. coli* JM109 cells and colonies thus obtained were screened for the presence of ASF/SF2 cDNAs by restriction analyses and sequencing.

pRCβActin-Flag ASF/SF2(wt), ASF/SF2(ΔRS2) and ASF/SF2(ΔDock): The inserts for the mammalian expression plasmids were generated by PCR amplification using their corresponding GST-expression clones as templates. A common 5' primer was used:

ASF(1)NTNdeI: 5' GAACTACATATGTCGGGAGGTGGTGTG 3'

The 3' primers used here were the same as those used to generate the GST-expression clones – ASF(248)CTNotI and ASF(219/ΔRS2)CTNotI.

Following PCR amplification, the subcloning procedure was identical to that above, except that the enzymes used for double digestion were NdeI and NotI and the vector used was pRCβActin-Flag (a gift from Dr. Yi Chen).

pET24dTEV ASF/SF2(ΔRS/12-196), ASF/SF2(R1/12-90), ASF/SF2(R2/118-196), ASF/SF2(R1L/12-121) and ASF/SF2(LR2/90-196): The inserts for these expression plasmids were generated by PCR amplification using pGEX4T2 ASF/SF2(ΔRS2) as a template. All 5' primers had a BamHI restriction site, while all 3' primers had a NotI restriction site.

ASF(12)NTBamHI: 5' CGCGGATCCATGGGGAACAACGATTGC 3'

ASF(90)NTBamHI: 5' CGCGGATCCATGCGAAGCGGCCGTGGAACA 3'

ASF(118)NTBamHI: 5’ CGCGGATCCATGCGGTCTGAAAACAGAG 3’

ASF(196)CTNotI: sequence listed above

ASF(90)CTNotI: 5’ CCGTTAAAGCCGCCTATCGAGGAAACTCCAC 3’

ASF(121)CTNotI: 5’ CCGTTAAAGCGGCCGCCTAGTTTTCAGACGGCT 3’

Following PCR amplification, the subcloning procedure was identical to that above, except that the enzymes used for double digestion were BamHI and NotI and the vector used was pET24dTEV (a gift from Dr. Greg Van Duyne, University of Pennsylvania).

The expression plasmid for His-SRPK1ΔNS was a gift from Dr. J. C. Ngo.

B. Mutagenesis of ASF/SF2, Clk/Sty and SRPK1

Mutagenesis of ASF/SF2 was carried out according to manufacturer’s protocol for the Stratagene Quikchange Mutagenesis Kit. Presence of the mutation was verified in every clone by sequencing.

pGEX4T2 ASF/SF2(ΔRS R191A/K193A) and ASF/SF2(ΔRS2 R191A/K193A): pGEX 4T2 ASF/SF2(ΔRS) and ASF/SF2(ΔRS2) were used as templates for the PCR reactions and the following pair of primers were used in both cases:

ASF_RK_AA1: 5’ ACTGCCTACATCGCGGTTGCAGTTGATGGGCCC 3’

ASF_RK_AA2: 5’ GGGCCCATCAACTGCAACCGCGATGTAGGCAGT 3’

pRCβActin-Flag ASF/SF2(R191A/K193A) was sub-cloned from pET14bGST ASF/SF2 (FL R191A/K193A) (generated as above) using primers ASF(1)NTNdeI and ASF(248) CTNotI.

pRCβActin-Flag ASF/SF2(S4A): The composite mutations S205A/S207A/S209A/S211A (S4A) were generated in two parts: first residues S205 and S207 of pRC βActin-Flag ASF/SF2(wt) were mutated to alanines using the primers ASF_S205207AA1 and ASF_S205207AA1. Once the mutant pRCβActin-Flag ASF/SF2(S205A/S207A) was obtained, it was used as a template for another round of mutagenesis using primers ASF S209/211A-1 and ASF-S209/211A-2. The sequences of all the primers used in making these mutations have been listed below.

ASF_S205207AA1:5'CCGAGCTATGGCCGAGCTCGCGCGCGTTCTCGCTCTCGTAGC 3'

ASF_S205207AA2: 5' GCTACGAGAGCGAGAACGCGCGCGAGCTCGGCCATAGCTCGG 3'

ASF-S209/211A-1: 5' GGAAGATCTCGATCTCGAGCGCGTGCACGTAGCAGAAGCCGT3'

ASF-S209/211A-2: 5' ACGGCTTCTGCTACGTGCACGCGCTCGAGATCGAGATCTTCC 3'

pGEX4T2 ASF/SF2(S4A) was sub-cloned from pRCβActin-Flag ASF/SF2(S4A) using primers ASF(1)NTBamH1 and ASF(248) CTNotI.

The clones pRCβActin-Flag ASF/SF2 (W134A), (Q135A) and (R154A) were gifts from Kayla Giang and Nhat Huynh.

pRCβActin-Flag ASF/SF2 (S116A), (S119A) and (S116A/S119A) were subcloned from pIND4 ASF/SF2 (S116A), (S119A) and (S116A/S119A) (gifts from Kayla Giang and Nhat Huynh) respectively, using primers ASF(1)NTNdeI and ASF(248) CTNotI.

pIND4 ASF/SF2 (E120A): pIND4 ASF/SF2 (a gift from Dr. X-D Fu) was used as a template for the mutagenesis reaction and the primers used were ASF_E120A_1 and ASF_E120A_2.

ASF_E120A_1: 5' GCCAGGCGGGCTGCAAACAGAGTGGTTG 3'

ASF_E120A_2: 5' CAACCACTCTGTTTGCAGCCCGCCTGGC 3'

pRCβActin-Flag (E120A) was subcloned from pIND4 ASF/SF2 (E120A) as above.

pCDNA3-HA Clk/Sty(FL) was a gift from Drs. X-D Fu and J. H. Ding.

pCDNA3-HA Clk/Sty (K190R): pCDNA3HA Clk/Sty(FL) was used as a template in the mutagenesis reaction, along with the primers listed below.

Clk_K190R_1: 5' CATGTAGCAGTAAGAATAGTTAAAAAT 3'

Clk_K190R_2: 5' ATTTTTAACTATTCTTACTGCTACATG 3'

pCMV-myc SRPK1(FL) was a gift from Drs. X-D Fu and X. Y. Zhong

pCMV-myc SRPK1(K109R): pCMV-myc SRPK1(FL) was used as a template in the mutagenesis reaction and primers used in this reaction were a gift from Dr. J.C. Ngo.

C. Expression and purification of recombinant proteins

GST-ASF/SF2: *E. coli* BL21(DE3) cells transformed with the appropriate GST-expression plasmid were grown in 2L LB media (containing 100μg/mL Ampicillin) at 37°C to an OD_{600} of 0.2 and induced with 0.1mM IPTG overnight at 25°C. The cells were harvested by centrifugation. The cell pellet was then resuspended in 120mL of lysis buffer containing 20mM Tris pH 8.5, 50mM NaCl, 10% glycerol, 1mM PMSF and 0.1X

Protease Inhibitor Cocktail (Sigma-Aldrich). The cells were lysed by sonication and clarified by centrifugation at high speed (13,500 – 14,000 r.p.m) for 45 minutes. The resultant supernatant was loaded onto a gravity Q-sepharose Fast Flow column at 4°C. The column was washed with 100mL lysis buffer. Presence of the protein in the flow through and wash was determined by SDS-PAGE analysis. The flow through and wash of the Q-column were pooled, filtered through a 0.8μ syringe filter, the concentration of NaCl was increased to 500mM, and then loaded onto a pre-equilibrated GST column. The GST column was washed with GST binding buffer (20mM Tris pH 8.5, 500mM NaCl, 10% glycerol and 1mM DTT). The protein was then eluted off the column by GST elution buffer (GST binding buffer supplemented with 10mM reduced glutathione). The purity of the elutions was checked by SDS-PAGE analysis. The desired elutions were pooled, and dialyzed into buffer containing 20 mM Tris pH 8.5, 500 mM NaCl, 10% glycerol and 1mM DTT. Following dialysis, the protein was concentrated using Centriprep-30 (Millipore) and loaded onto a 16/60 Superdex 200 size-exclusion column, pre-equilibrated with buffer containing 20mM Tris pH 8.5, 500mM NaCl, 5% glycerol and 1mM DTT, for a final purification step. The peak fractions were pooled, concentrated to approximately 2mg/mL and flash frozen in small aliquots. A typical yield for this protein preparation is 1mg/1L of LB culture.

His-ASF/SF2(ΔRS): *E. coli* BL21(DE3) cells transformed with pET24dTEV ASF/SF2(ΔRS) were grown in 5mL LB media (containing 15μg/mL Kanamycin) at 37°C for 8 hours. The starter culture was centrifuged at high speed to obtain a pellet, which was then washed and resuspended in 5mL of minimal media (composition and

preparation described below). This was used as inoculum for 1L minimal media, which was grown at 37°C to an $O.D_{600}$ of 0.4 and induced with 0.1 mM IPTG for 16 hours at 25°C. The cells were harvested as above and lysed in lysis buffer containing 20mM Tris pH 7.5, 250mM NaCl, 50mM urea, 5mM imidazole, 10% glycerol, 1mM PMSF and 0.1X PIC. The cell lysate was clarified by centrifugation (as above) and filtration through a 0.8μ syringe filter. The clarified lysate was then diluted to 2X of its initial volume with dilution buffer (20mM Tris pH 7.5, 250mM NaCl, 5mM imidazole, 10% glycerol, 1mM PMSF) and was loaded onto a Ni^{2+}-NTA Agarose column (Invitrogen) at 4°C. The column was washed with 5mM and 40mM imidazole buffers at 4°C and the protein was eluted in 200mM imidazole buffer at 25°C (composition of buffers used for Ni^{2+}-NTA column were the same as the dilution buffer, except for variations in amounts of imidazole). The eluted protein was assayed for purity and RNA contamination by SDS-PAGE and agarose gel electrophoresis respectively. The elutions that contained minimal RNA contaminations were pooled and dialyzed against a total of 3L of dialysis buffer containing 20mM Tris pH 7.5, 200mM NaCl, 5% glycerol, 1mM DTT for 3 hours at 25°C. Following dialysis, the protein was concentrated using Centriprep-10 (Millipore) and loaded onto a 16/60 Superdex 75 size-exclusion column, pre-equilibrated with buffer containing 20mM Tris pH 7.5, 200mM NaCl, 5% glycerol and 1mM DTT, for a final purification step. The peak fractions were pooled, concentrated to approximately 2mg/mL and flash frozen in small aliquots. A typical yield for this preparation is 20mg/1L of minimal media culture.

His-ASF/SF2(ΔRS S116A/S119A) was a gift from Dr. S. Cho.

Composition of Minimal Media: To prepare the final minimal media, the indicated volumes of Buffer 1, amino acid buffers A, B and C, nucleotide base buffer and 20 mL of 100X Eagle vitamin solution (Gibco/BRL) were mixed together, brought up to a total volume of 2L with autoclaved water and adjusted to pH 7.4. The composition and volume of each of the buffers is listed in Table 2.1. All buffers were separately sterile filtered.

Table 2.1: Composition of Buffers used for Minimal Media

Buffer	Volume	Composition
Buffer1	1L	7.5mM $(NH_4)_2SO_4$, 8.5mM NaCl, 55mM KH_2PO_4, 100mM K_2HPO_4, 1mM $MgSO_4$, 20mM glucose, 1mg/L $CaCl_2$, 1μg/L $ZnSO_4$, 1μg/L $MnCl_2$, 1mg/L $FeSO_4$, 1μg/L MoO_3, 20mg/L d-biotin, 1mg/L thiamine and 1μg/L $CuCl_2$ - adjusted to pH 7.4
Buffer 2	-	137mM NaCl, 2.5mM KCl, 10mM Na_2HPO_4 and 1.76mM KH_2PO_4 – adjusted to pH 7.0
Amino acid buffer A	50mL	400mg leucine, 300mg each of alanine, glutamate, lysine, glutamine and valine, 200mg each of aspartate, glycine, histidine, isoleucine, asparagine, proline, threonine and methionine and 100mg cysteine, dissolved in Buffer 2
Amino acid buffer B	25mL	200mg phenylalanine and 100mg each of tryptophan and tyrosine, dissolved in Buffer 2 with drop-wise addition of 1 M NaOH
Amino acid buffer C	20mL	4g each of arginine and serine dissolved in Buffer 2
Nucleotide base buffer	35mL	100mg each of adenine, guanine, cytosine, thymine and uracil, dissolved in Buffer 2 with drop-wise addition of 1 M NaOH

His- ASF/SF2(R1), ASF/SF2(R2), ASF/SF2(R1L) and ASF/SF2(LR2): *E. coli* BL21(DE3) cells transformed with the appropriate His-expression plasmid were grown in 2L LB media (containing 15μg/mL Kanamycin) at 37°C to an OD_{600} of 0.3 and induced with 0.1mM IPTG overnight at 25°C. The cells were harvested by centrifugation and the proteins were purified in a manner identical to His-ASF/SF2(ΔRS). The purity and yield of the proteins varied marginally with each preparation.

D. Co-immunoprecipitation assays of SRPK1 and ASF/SF2

Co-IP of His-SRPK1ΔNS and GST-ASF/SF2: Complexes of purified His-SRPK1ΔNS and GST-ASF/SF2 were formed by incubating equimolar amounts (corresponding to ~10μg) of each protein at 25°C for 1 hour. The complex was then diluted up to 0.1mg/mL in binding buffer (20mM Tris 8.5, 200mM NaCl, 5% glycerol, 50mg/mL, 1mM DTT, 0.5% TritonX-100). GST-ASF/SF2 was immunoprecipitated overnight at 4°C using 20ng of α-GST antibody (Santa Cruz) immobilized on Protein A/G agarose (Upstate, NY). Non-specifically bound protein was washed off using a total of 4mL of wash buffer (binding buffer without any BSA or DTT) and the fraction of SRPK1 bound to ASF-SF2 was visualized by SDS-PAGE and Western blotting using α-His antibody (Penta His, Qiagen).

Co-IP of myc-SRPK1(K109R) and Flag-ASF/SF2: Human embryonic kidney (HEK) 293 cells were co-transfected with pCMV-myc SRPK1(K109R) and pRCβActin-Flag ASF/SF2 using Lipofectamine 2000 (Invitrogen) according to manufacturer's

protocol. Cells were harvested 48 hours post transfection in 100μL lysis buffer (20mM Tris pH 7.5, 200mM NaCl, 1% Triton-X 100, 2mM DTT, 5mM p-nitro phenyl phosphate, 2mM Na_2VO_4, 1mM PMSF, 1X protease inhibitor cocktail) and approximately 500μg of cell extract was used for each reaction. Extracts were diluted up to 2μg/μL in lysis buffer and pre-cleared with Protein G agarose (Upstate, NY) for 1 hour at 4°C. Myc-SRPK1(K109R) was immuno-precipitated overnight at 4°C with 125ng of α-SRPK1 antibody (BD Biosciences) immobilized on Protein G agarose. Non-specifically bound protein was washed off using a total of 4mL of wash buffer (10mM Tris 7.5, 150mM NaCl, 0.2% TritonX-100). Immunoprecipitated SRPK1 and the fraction of ASF/SF2 bound to it were visualized by SDS-PAGE and Western blotting using α-myc (Cell Signaling) and α-Flag (M2, Sigma) antibodies respectively.

E. In vitro kinase assays, inhibition and start-delay-trap assays

In vitro kinase assay: Equimolar amounts (1-2μM each) of His-SRPK1 and GST-ASF/SF2 were incubated in 1X kinase buffer (50 mM Tris-Cl pH 7.5, 10 mM MgCl2, 1 mg/ml BSA and 1 mM DTT) for 5 minutes at 25°C. The reaction was initiated using 200 μM ATP and trace amounts of γ^{32}P-ATP and was allowed to proceed at 25°C. Equal volume aliquots were removed for analysis by SDS-PAGE at the indicated times. Phosphorylated proteins were detected by SDS-PAGE and autoradiography and their relative amounts were quantified using the software ImageQuant (Molecular Dynamics).

Phosphorylation inhibition assay: 1μM of SRPK1 was pre-incubated with an equimolar amount of ASF/SF2(ΔRS2) and increasing amounts of ASF/SF2 inhibitor-like proteins (1X - 70X) in 1X kinase buffer at 25°C for 30 minutes. 200 μM ATP and trace amounts of γ^{32}P-ATP were added to the reaction cocktails after incubation to initiate the reaction. The reactions were allowed to proceed for 30 minutes and analyzed by autoradiography.

Start-delay-trap assays: These assays were performed in a manner identical to the in vitro kinase assay, except 10mM inhibitor peptide (YRTRDAPRERAPTR – a gift from Dr. J.A. Adams) was added at the indicated time after initiation of the reaction. The reactions were analyzed as above.

F. Phosphatase assays

Flag-ASF/SF2 constructs were singly transfected in HeLa cells or co-transfected with HA-Clk/Sty constructs as described in Chapter III. Cells were harvested either 24 or 48 hours post transfection (indicated in Chapter III) and protein was extracted in 50μL RIPA buffer. 30 μg of cell lysate was treated with 0.01 units of CIP (NEB) in presence of 10X NEB buffer 3 for 1 hour at 37°C. Reactions were resolved by SDS-PAGE and visualized by Western Blotting using α-Flag antibody.

G. Indirect immunofluorescence assays

HeLa cells were plated onto poly-lysine coated glass coverslips in 6-well plates, 12-24 hours prior to transfection. Flag-ASF/SF2 constructs were either singly transfected or co-transfected with HA-Clk/Sty using Lipofectamine 2000 (manufacturer's protocol). Cells were fixed for immunofluorescence assays 24-48 hours post transfection, unless specifically mentioned in the text. Cells were fixed using 4% paraformaldehyde for 30 minutes, followed by permeabilization using 0.1% TritonX-100 for 10 minutes. In order to prevent non-specific antibody staining, cells were blocked with 1% BSA and washed with 0.1% BSA after primary and secondary antibody staining. All the assay solutions were made up in 1X phosphate-buffered saline (PBS). ASF/SF2 was detected by α-Flag antibody (1:1000, diluted in blocking solution) followed by staining using FITC-conjugated anti-mouse secondary antibody. In order to stain cellular nuclei, coverslips were mounted in mounting media containing DAPI (Vectashield, Vector Laboratories). Coverslips were analyzed under a Zeiss Axiophot microscope and images acquired with a Hamamatsu ORCA-ER digital camera with Improvision OpenLab 3.1.5 software.

H. Electrophoretic mobility shift assays

The RNA aptamers opt_5'ss and Ron_ESE (sequences given below) were purchased from Dharmacon Research, deprotected according to manufacturer's protocol and purified by urea-PAGE. The sequences of the RNA aptamers are given below:

opt_5'ss: 5' - GCAAGGUAAGUAAGGUAAGU - 3'

Ron_ESE: 5'- AGGCGGAGGAAGC - 3'

The RNA aptamer S_ESE was provided by Dr. S. Cho. S_ESE is a 30 nucleotide long RNA aptamer which contains the following ESE sequence: 5'- GAGACGA - 3'

5' end labeling of RNA aptamer: 50pmoles of purified RNA aptamer was incubated with 20units of T4 polynucleotide kinase (T4 PNK, New England Biolabs) and 40μCi γ^{32}P-ATP in presence of T4 PNK buffer at 37°C for 2 hours. The reaction was quenched by addition of 0.5mM EDTA, following which the labeled RNA aptamer was purified away from excess γ^{32}P-ATP using a G-25 Sephadex spin column. The radioactive count (and concentration) of the labeled RNA aptamer was determined by liquid scintillation counting.

In vitro transcription of SV40_5'ss: The RNA aptamer SV40_5'ss was generated by in vitro transcription from a partially double-stranded DNA template, the bottom strand of which had an overhang corresponding to the complementary sequence of SV40_5'ss (in addition to the binding site for T7 RNA polymerase on both strands). The RNA aptamer was internally labeled by adding α^{32}P-UTP (40μCi) to the transcription reaction. The transcription reaction contained 40mM Tris pH 8.1, 40mM $MgCl_2$, 100mg/mL PEG8000, 2mM spermine, 5mM DTT and 0.01% TritonX-100 in addition to 1μmole template DNA, 4mM nucleotide tri-phosphate mix (A,U, G and C) and T7 RNA polymerase. The reaction was incubated at 37°C for 7 hours, followed by DNaseI treatment (37°C/30 minutes) to remove the template. The reaction was then quenched by addition of 50mM EDTA, following which labeled RNA was extracted by chloroform

extraction and ethanol precipitation. The RNA aptamer was finally purified by urea-PAGE prior to being used in gel-shift assays. The sequence of the SV40_5'ss aptamer is as follows:

5' – GCUAAGGUAAAUAUAGG – 3'

Gel shift assay: 10,000cpm (corresponding to approximately 10fmoles) of labeled RNA was incubated with increasing amounts of ASF/SF2 in presence of 1X binding buffer (10mM Tris 7.5, 150mM NaCl, 10% glycerol, 0.1% NP40, 0.05μg/μL BSA, 10mM $MgCl_2$, 1mM DTT) and 5μg yeast t-RNA at 25°C for 30 minutes. The corresponding unlabeled RNA served as a specific inhibitor while a 12 nucleotide long fragment of U1snRNA served as a non-specific inhibitor. The inhibitors were added in the binding reaction along with the labeled RNA, wherever indicated in Chapter IV. The protein:RNA complexes were resolved on a 6% native polyacrylamide gel (made up in 1X Tris-Borate-EDTA) that had been pre-run in 1X TBE at 25°C for 1 hour. The gel was run at a constant voltage of 200V at 25°C for 1.5 hours, dried and analyzed by autoradiography.

III: Recognition of the SR protein ASF/SF2 by the SR Protein Kinase, SRPK1

A. Introduction

The SR protein ASF/SF2 is a modular protein containing two N-terminal RRMs followed by a short linker which connects it to the C-terminal RS domain. The RS domain of ASF/SF2 is bifurcated in that it contains a tract of eight RS dipeptides (RS1 motif) followed by short interrupted stretches of RS dipeptides (RS2 motif) (Figure 3.1A). ASF/SF2 is extensively phosphorylated at its C-terminal RS domain by members of the SRPK and Clk/Sty family of kinases [99]. Recently, a human SRPK, SRPK1, has been shown to interact with ASF/SF2 with high affinity and phosphorylate approximately half the sites in its RS domain in a processive manner in vitro, where once the kinase locks onto the substrate, the resultant complex does not dissociate until phosphorylation is complete [100]. Unlike SRPK1, Clk/Sty is a nuclear kinase and thus appears to regulate ASF/SF2 in the nucleus through phosphorylation [68, 69, 72, 101, 102]. It is thought to phosphorylate significantly more residues in the RS domain of ASF/SF2 [56, 102]. Previous studies have shown a precise requirement for distinct phosphorylated states of ASF/SF2 for its different biological functions and its own nuclear import. In the nucleus, ASF/SF2 and most other splicing factors are concentrated in distinct nuclear bodies called nuclear speckles, the formation of which is dependent on the degree of phosphorylation of these proteins [51, 55, 56, 68, 69, 103]. Consistent with these observations, the transient overexpression of Clk/Sty has been shown to cause release of ASF/SF2 from nuclear speckles.

Recently the X-ray crystal structure of an active form of SRPK1, SRPKΔNS, was solved in complex with an SR-protein derived peptide and ADP in our laboratory (Figure 3.2). The 9mer peptide in the ternary complex ($RR\underline{R}^{1}E^{2}\underline{R}^{3}S^{4}P^{5}T^{6}\underline{R}^{7}$) was derived from

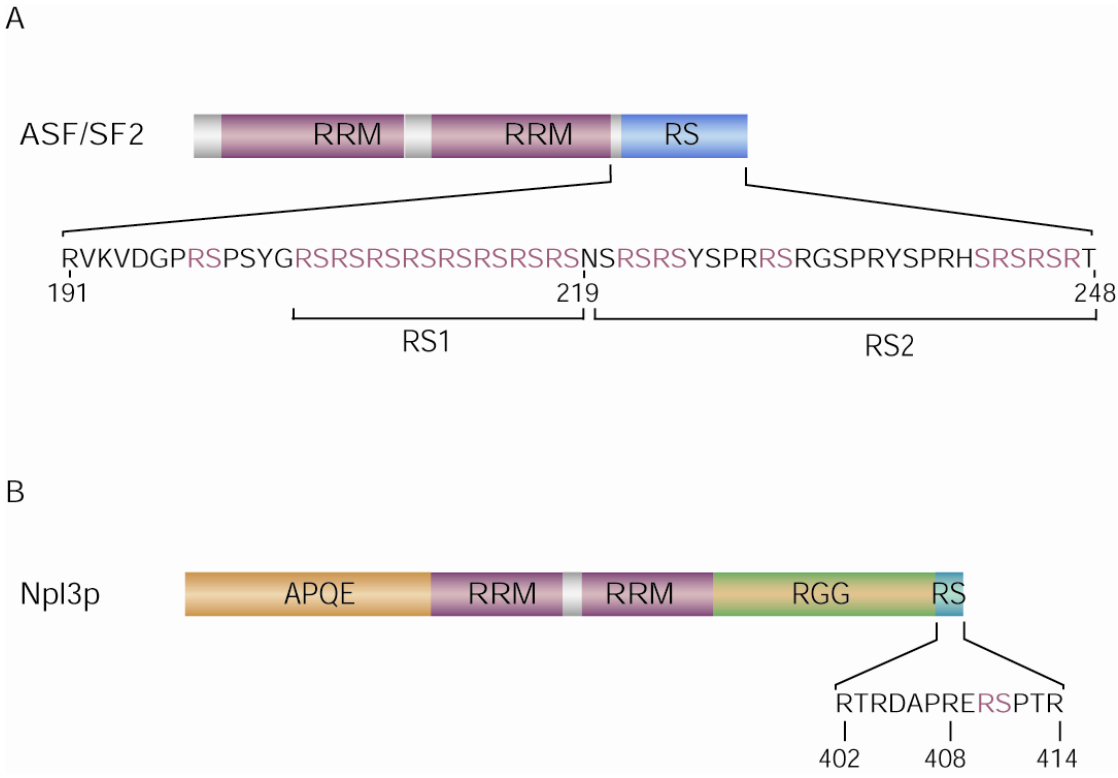


Figure 3.1. Sequence of the RS domains of ASF/SF2 and Npl3p. A. The C-terminal RS domain of ASF/SF2 can further be divided into the RS1 and RS2 motifs. The RS1 motif consists of a tract of contiguous RS dipeptides, whereas the RS2 motif contains dispersed RS dipeptides B. The RS domain of Npl3p consists of a single RS dipeptide which is located at the extreme C-terminus of the protein.

the single phosphorylation site on the yeast SR-like protein Npl3p (residues 408 – 414, with S411 being the P site serine) (Figure 3.1B). It was modified by adding two arginines at the N-terminus to enhance accessibility of the peptide to the crystal solvent channel. Although the peptide was initially designed to bind to the substrate-binding groove of SRPK1ΔNS, it surprisingly docks onto a distal site analogous to the DEF motif binding site in the kinase ERK2 [91]. The peptide recognizes a deep groove in SRPK1ΔNS formed by the MAP kinase insert, the loop connecting helices αF and αG, and the helix αG of the kinase (Figure 3.2A). Particularly, basic residues of the substrate peptide make strong electrostatic interactions with acidic residues within this groove of the kinase (Figure 3.2B). Mutation of contact residues (D548, D564, E571 and K615) within this groove significantly reduces binding of the kinase to ASF/SF2 (Figure 3.2C). The mutant kinase also lacks the ability to phosphorylate its substrate processively, further emphasizing the importance of this groove in substrate recognition (Figure 3.2D). We refer to this groove of the kinase as the docking groove. Since the substrate peptide fits into the docking groove of the kinase with near perfect complementarity, it was thought that this peptide mimics a docking motif on ASF/SF2.

It was previously unknown as to why ASF/SF2 binds SRPK1 with high affinity and is phosphorylated at only half of its available serines. The biological significance of such restricted phosphorylation was also unknown. Additionally, how SRPK1 and Clk/Sty coordinate their activities to regulate SR proteins was not clear. We have identified a docking motif in ASF/SF2 and show that it is critical for restricted phosphorylation of ASF/SF2 by SRPK1 both in vivo and in vitro. Our results suggest a mechanism for regulation of ASF/SF2 localization by the two kinases.

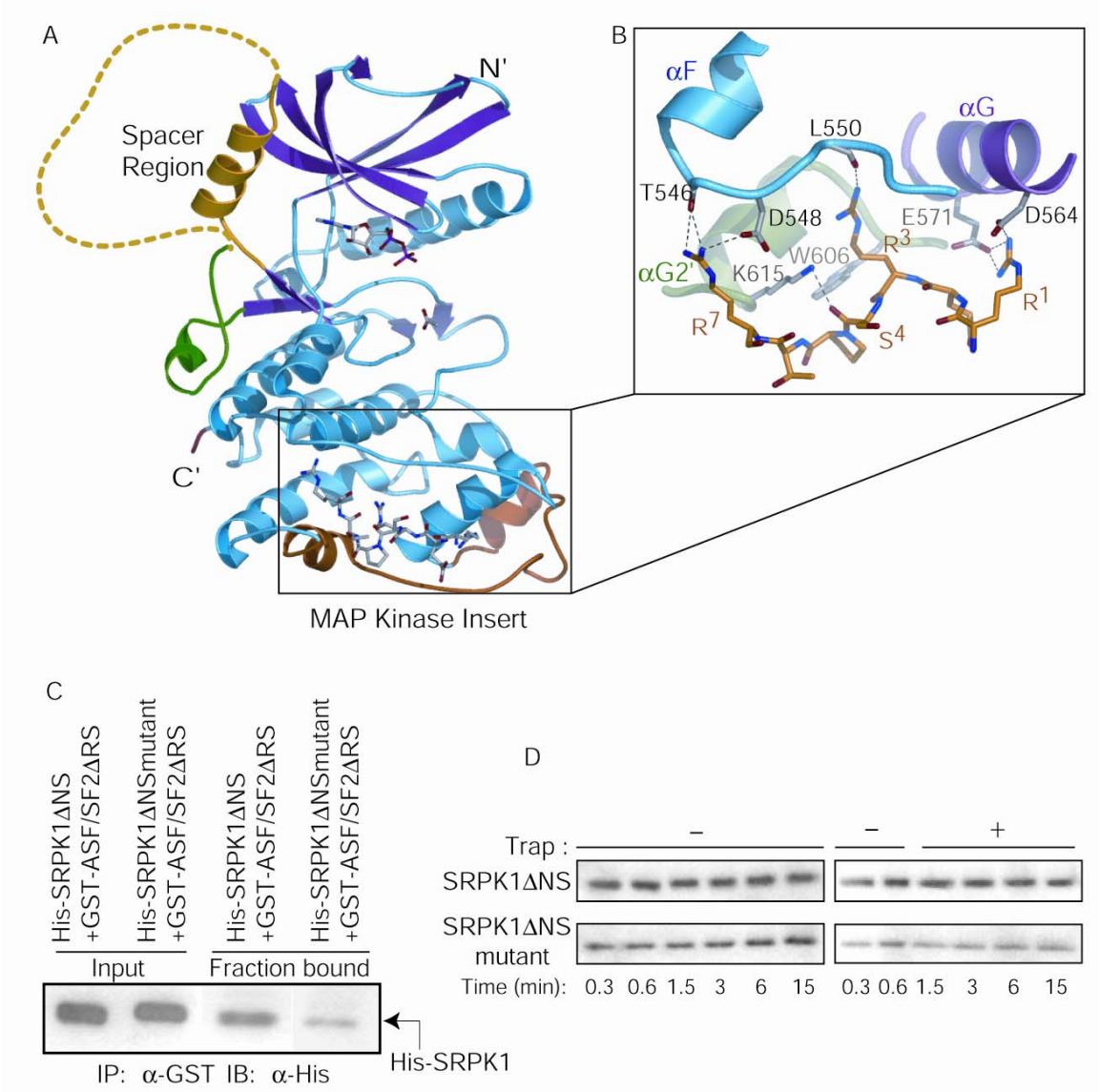


Figure 3.2. Identification of a docking groove in SRPK1. A. Overall structure of SRPK1/ peptide/ADP complex. The peptide binds to a docking groove in the large lobe of the kinase, made up of helix αG, the loop connecting helices αF and αG and the MAP kinase insert. B. Details of the interactions between basic residues of the peptide and residues from the docking groove of the kinase. C. Co-precipitation assays of wild type and the docking groove mutant of SRPK1 with ASF/SF2(ΔRS). Mutation of key docking groove residues considerably weakens the affinity of SRPK1 for ASF/SF2. D. Start-delay-trap assays with wild-type and mutant SRPK1 using ASF/SF2(ΔRS2) as a substrate. In contrast to the wild type kinase, the docking groove mutant of SRPK1 does not phosphorylate ASF/SF2 in a processive manner.

In order to further investigate the interaction between SRPK1 and ASF/SF2, our laboratory has purified and crystallized a fragment of ASF/SF2 (ASF/SF2 ΔRRM1/RS2) in complex with SRPK1. As expected, the structure of the complex revealed additional interactions between SRPK1 and ASF/SF2. ASF/SF2 has the canonical β1-α1-β2-β3-α2-β4 RRM fold, with one additional β strand (βN) between α2 and β4 [104]. ASF/SF2 makes a 3-point interaction with the small and large lobes of SRPK1 using residues in its RRM2 domain (Figure 3.3A). One such interaction involves the SWQDLKD motif which is conserved in all SR proteins containing two RRMs [105]. This motif lies at the N-terminus of helix α1 and contacts the tip of the glycine-rich loop of SRPK1. In particular, W134 and Q135 of ASF/SF2 appear to be stacked against H90 and W88 of SRPK1 (Figure 3.3B). Additionally, Q135 also contacts the backbone of the glycine-rich loop of SRPK1. The other interaction between SRPK1 and ASF/SF2 involves the large lobe of the kinase and a residue from the β turn between strands β2 and β3 of ASF/SF2. The residue R154 protrudes into a pocket formed by helices αD and αF of SRPK1 and is predominantly stabilized by interactions with the side-chains of residues E543 and Y549 (Figure 3.3C). In addition to the contact points described above, the non-canonical β strand βN allows for the formation of another hairpin loop (between βN and β4) which engages both the small and the large lobe of SRPK1 (Figure 3.3D). In this study, we investigate the role of these contact residues in the regulation of ASF/SF2. We focus our attention on residues W134, Q135 and R154 and show that these residues are crucial for interaction of ASF/SF2 with SRPK1 in vivo. We also show that perturbation of the SRPK1-interaction surface affects the steady-state localization of ASF/SF2. Together

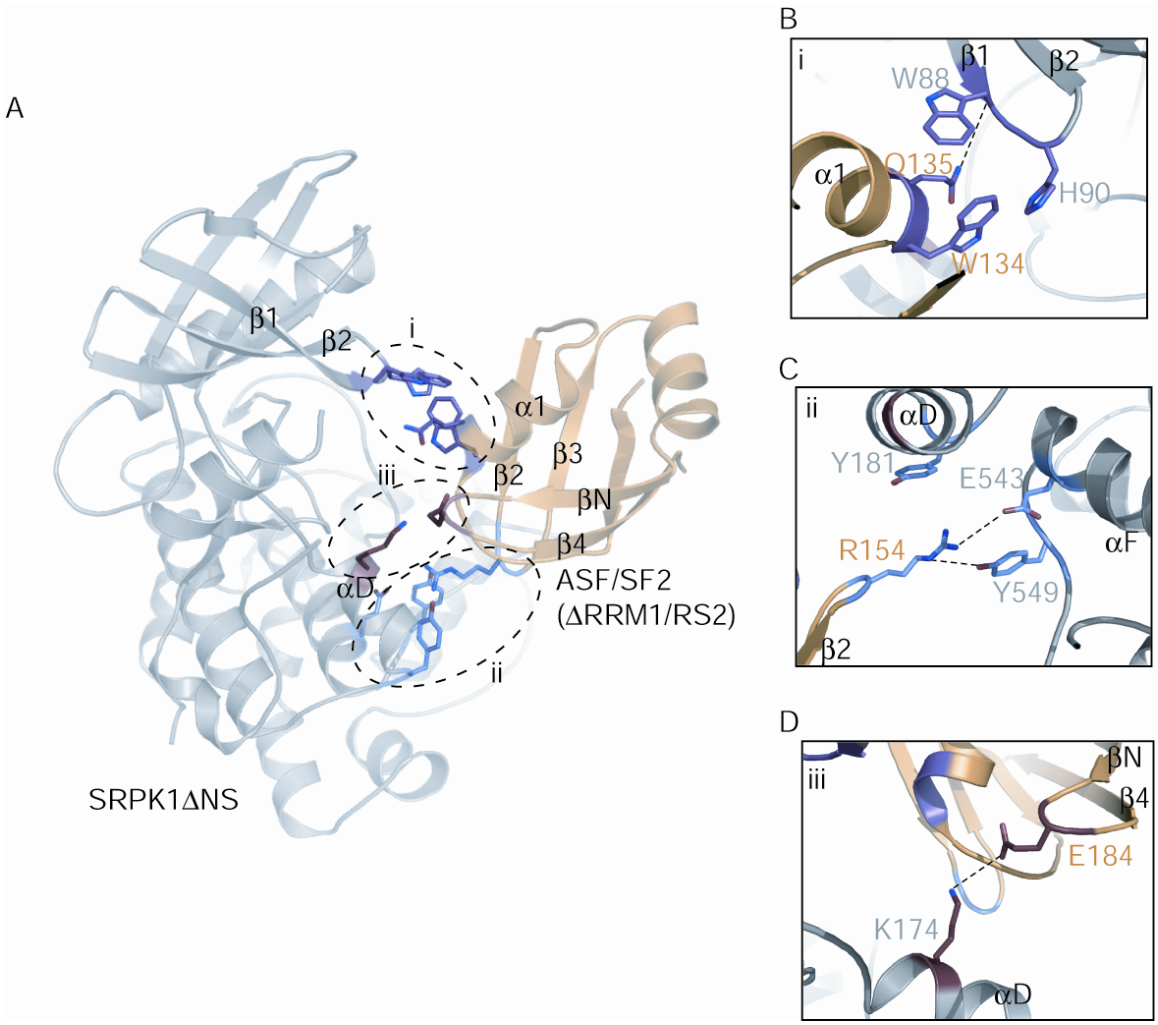


Figure 3.3. ASF/SF2 makes a 3-point contact with the kinase core of SRPK1. A. Overall structure of the ternary complex of SRPK1:ASF/SF2:AMP-PNP. Details of each interaction are shown in the insets B. Interactions between residues from the conserved SWQDLKD (W134 and Q135) motif of ASF/SF2 and the glycine-rich loop of SRPK1. C. A conserved basic residue (R154) in the β-turn between β-strands 2 and 3 projects into a pocket formed by helices αD and αF of SRPK1. D. The β-hairpin formed by strands β4and βN contacts both the small and the large lobes of SRPK1.

our study proposes a model of stringent kinase-dependent regulation of the SR protein, ASF/SF2.

B. Results

1. A docking motif on ASF/SF2 is responsible for the high affinity interaction with SRPK1

In order to identify the region/regions of ASF/SF2 responsible for stable binding to SRPK1, we have made systematic deletion constructs of ASF/SF2 and qualitatively estimated their affinity for SRPK1ΔNS (Figure 3.4A). Since SRPK1ΔNS is identical to SRPK1(FL) with respect to its activity and ability to interact with ASF/SF2, we used this truncated protein for all our in vitro biochemical studies and henceforth refer to it as SRPK1. We observe that deletion of a part or the entire RS domain of ASF/SF2 does not reduce affinity of the truncated proteins, ASF/SF2(ΔRS2) and ASF/SF2(ΔRS), for SRPK1 (Figure 3.4B). The slight reduction in the amount of SRPK1 precipitated by ASF/SF2(ΔRS2) as compared to ASF/SF2(ΔRS) could be due to the presence of contaminating RNA non-specifically bound to the RS domain of ASF/SF2 during protein purification. However, deletion of 17 more residues from the C-terminus in addition to the entire RS domain, significantly reduces binding of the resultant construct, ASF/SF2(ΔCT), to SRPK1 (Figure 3.4B, lanes 2, 3 and 4). We conclude that this 17 amino acid stretch of ASF/SF2 is critical for its interaction with SRPK1. To verify that this amino acid stretch is indeed critical for binding SRPK1, we performed phosphorylation-inhibition assays where we tested the ability of ASF/SF2(ΔRS) and ASF/SF2(ΔCT) to inhibit phosphorylation of ASF/SF2(ΔRS2) by SRPK1ΔNS1 (Figure

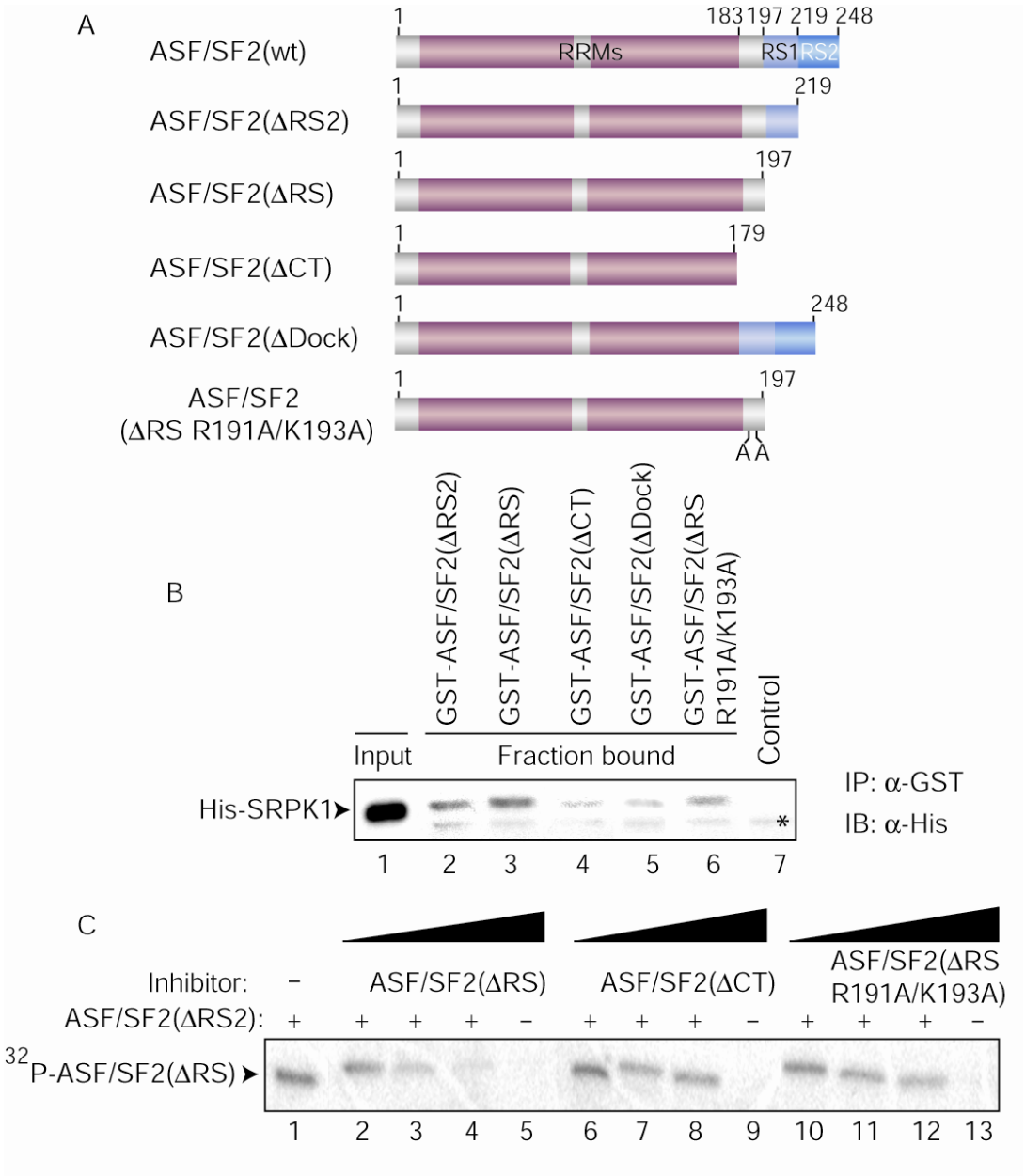


Figure 3.4. Identification of a docking motif in ASF/SF2. A. Constructs of ASF/SF2 generated to delineate the region responsible for interaction with SRPK1. B. Co-precipitation assay of SRPK1 with the different constructs of ASF/SF2 shown above. ASF/SF2-(ΔCT), (ΔDock) and (ΔRS R191A/K193A) do not bind SRPK1 efficiently. C. Phosphorylation inhibition assay using ASF/SF2(ΔRS2) as a substrate and ASF/SF2-(ΔRS), (ΔCT) and (ΔRS R191A/K193A) as potential inhibitors. ASF/SF2(ΔRS) blocks phosphorylation effectively as opposed to ASF/SF2-(ΔCT) and (ΔRS R191A/K193A).

3.4C). We found that ASF/SF2(ΔRS) is capable of inhibiting phosphorylation when present in amounts as low as 15- fold molar excess of substrate. Presence of 70-fold excess of ASF/SF2(ΔRS) abolishes phosphorylation by SRPK1ΔNS1. In contrast, ASF/SF2(ΔCT) does not function as a potent inhibitor even at its highest concentration (Figure 3.4C, compare lanes 4 and 8). These results suggest that segment(s) present within the first 196 residues (encompassing the RRMs of ASF/SF2), particularly the amino acid stretch preceding the RS domain is critical for binding SRPK1.

In order to determine if the amino acid stretch between the RRMs and the RS domain functions as a docking motif for SRPK1, we designed an internal deletion construct that removed 14 residues (aa 184-197) connecting the second RRM and the RS1 motif of ASF/SF2. We tested the ability of this deletion construct, ASF/SF2 (ΔDock), to interact with the kinase. Again, we observe a significant decrease in binding affinity of this truncated protein for the kinase (Figure 3.4B, compare lanes 2, 3 and 5). The structure of the SRPK1/peptide/ADP complex shows extensive hydrogen bonding and hydrophobic interactions between the kinase and the peptide. These interactions can only be mediated by long basic side chains such as arginine and lysine. The complex structure therefore suggests that the docking groove on SRPK1 might recognize a consensus motif of R-X-R/K-X-X-X-R on its substrate. The amino acid stretch between residues 183-198 contains a sequence, $\underline{R}^{191}V\underline{K}VDGP\underline{R}^{198}$, that conforms to the docking motif sequence proposed above. In particular, two of the three basic residues involved in contacting the kinase are conserved and the third one is shifted one position towards the C-terminus. We mutated amino acids R191 and K193 to alanines and assayed for the ability of this mutant protein to bind SRPK1 as before. The mutant, ASF/SF2(ΔRS

R191A/K193A) is defective in binding SRPK1 and does not inhibit phosphorylation of ASF/SF2(ΔRS2) (Figure 3.4B, lane 6 and Figure 3.4C, lane 12). Overall, our observations suggest that the amino acid stretch between the RRMs and the RS domain (residues 184-197) encompasses a docking motif on ASF/SF2.

2. The docking motif of ASF/SF2 plays a role in dictating its mode of phosphorylation by SRPK1

We proceeded to determine the mechanisms of phosphorylation of ASF/SF2(ΔDock) and ASF/SF2(ΔRS2 R191A/K193A) as compared to that of ASF/SF2(ΔRS2) using a start-delay-trap assay. The start-delay-trap assay performed here is a modified form of the start-trap assay described in [100]. In this assay, phosphorylation of the substrate by the kinase is initiated by addition of ATP and the trap (peptide inhibitor) is then added after a short time lag. We observed that phosphorylation of ASF/SF2(ΔRS2) is not inhibited by the trap peptide whereas phosphorylation of both ASF/SF2(ΔRS2 R191A/K193A) and ASF/SF2(ΔDock) is inhibited under identical conditions (Figure 3.5, right panels). We have not determined the K_d for complexes of all the studied ASF/SF2 constructs with SRPK1 and therefore simply suggest that the docking motif plays an important role in governing the mechanism of phosphorylation of ASF/SF2. In all, these experiments allow us to suggest that in addition to the RRM-mediated contact of ASF/SF2 with SRPK1, the docking motif - docking groove interactions between ASF/SF2 and SRPK1 enable the kinase to lock onto the substrate and also allow proper positioning of the RS domain, thus facilitating processive phosphorylation.

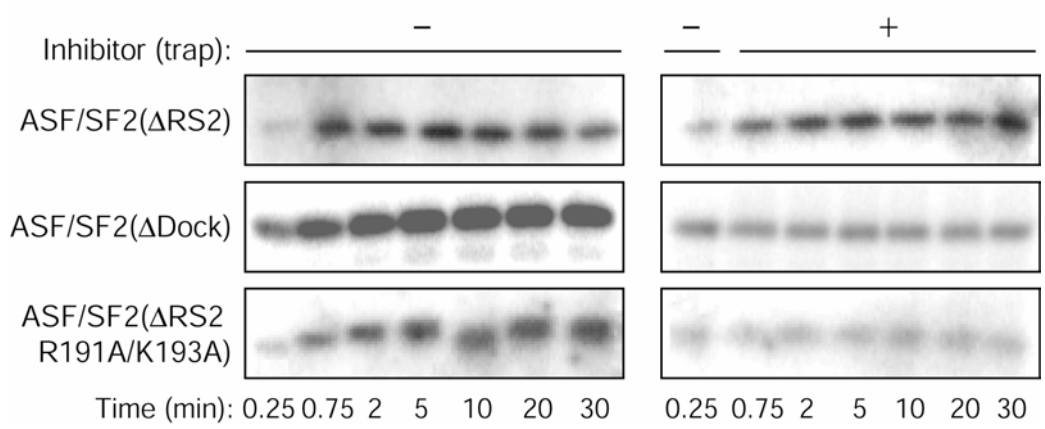


Figure 3.5. The docking motif of ASF/SF2 dictates its mechanism of phosphorylation. Start-delay-trap assays with ASF/SF2-(wt), (ΔDock) and (ΔRS2 R191A/K193A) show that unlike ASF/SF2(wt), both ASF/SF2-(ΔDock) and (ΔRS2 R191A/K193A) are inhibited by the trap peptide.

The above experiments led us to speculate that in addition to the docking motif, an intact RS1 motif might also be essential for regulation of phosphorylation of ASF/SF2. To investigate the role of the RS1 motif in phosphorylation of ASF/SF2, we mutated the first four serines in the RS1 motif of full-length ASF/SF2 (S205, S207, S209 and S211) to alanines. We refer to this mutant as ASF/SF2(S4A). In vitro phosphorylation assays using this mutant protein as a substrate shows that it is phosphorylated as efficiently as ASF/SF2(ΔRS2) and incorporates approximately half the number of phosphates compared to ASF/SF2(ΔRS2) (Figures 3.6A and 3.6B). This implies that ASF/SF2(S4A) is phosphorylated only at the four available serines within RS1 and no compensatory phosphorylation occurs elsewhere in the RS domain in vitro. To further elucidate any mechanistic differences between the mutant and wild type ASF/SF2 proteins, we repeated the start -delay- trap assay with these proteins, using the same inhibitor peptide as before. We observe that although there is the expected decrease in phosphate incorporation in ASF/SF2(S4A), there is no inhibition on addition of the trap peptide (Figure 3.6C). In view of our current results, we infer that while the docking motif of ASF/SF2 plays an important role, the RS domain itself does not affect the mechanism of phosphorylation.

3. The docking motif of ASF/SF2 is important for its localization to nuclear speckles

To explore if the docking motif of ASF/SF2 plays a role in vivo, we performed localization assays on wild type ASF/SF2 and different ASF/SF2 docking motif mutants. Since localization of ASF/SF2 depends on its phosphorylation, we hypothesized that any defect in the phosphorylation mechanism of the docking motif mutant in vivo would translate into a defect in localization of the mutant proteins. Wild type and mutant Flag-

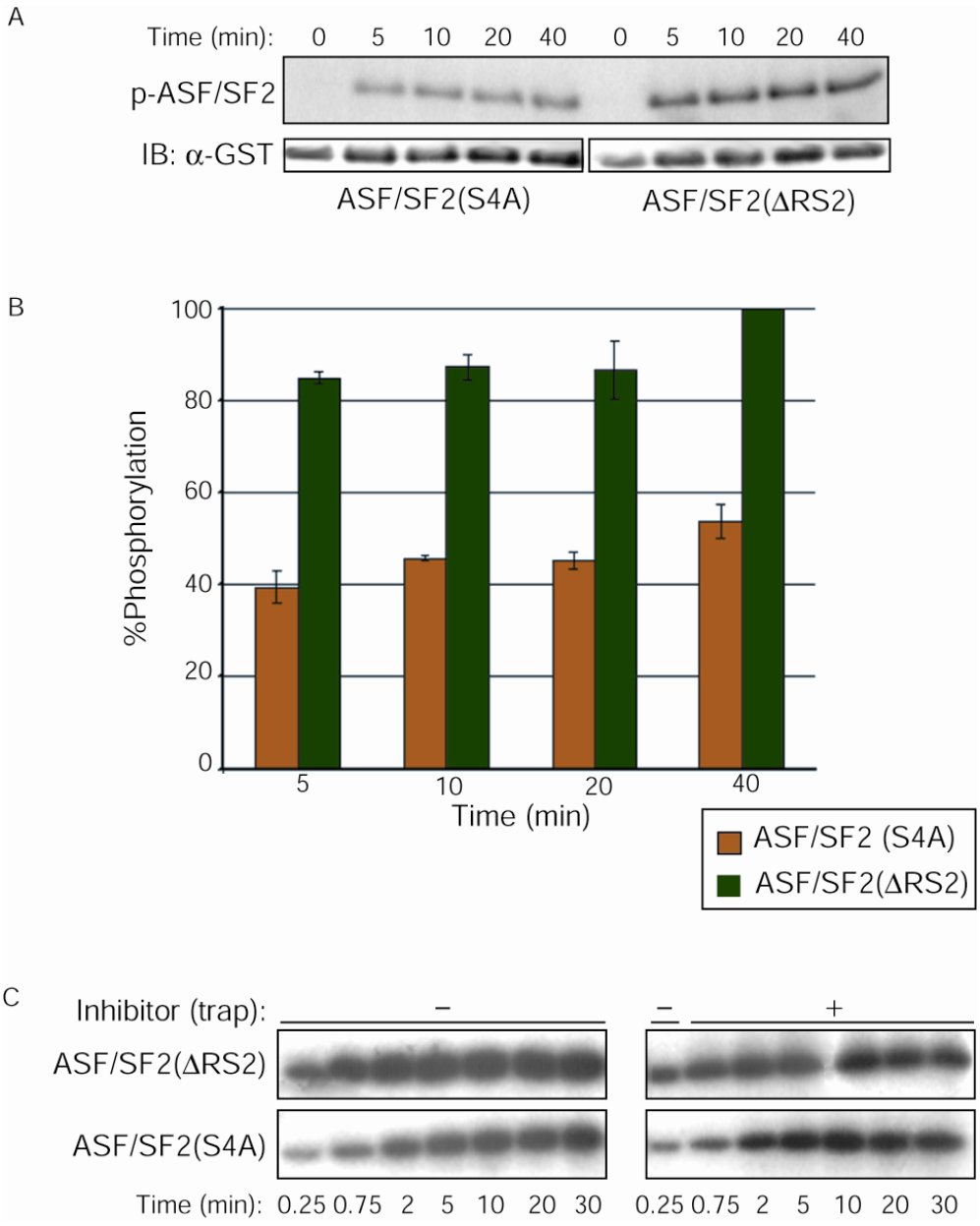


Figure 3.6. The RS domain of ASF/SF2 does not affect its mode of phosphorylation. A. In vitro kinase assays with SRPK1 using ASF/SF2(S4A) and ASF/SF2(ΔRS2)as substrates. B. Quantification of amounts of phosphate incorporated in each substrate in the assay above. ASF/SF2(S4A) incorporates half the number of phosphates compared to ASF/SF2(ΔRS2). Efficiencies of phosphorylation of both substrates are comparable. Error bars represent the standard deviation from three independent experiments. C. Start-delay-trap assays with ASF/SF2(ΔRS2) and ASF/SF2(S4A). Neither of the two proteins is inhibited by the trap peptide.

ASF/SF2 proteins were transiently over-expressed in HeLa cells and subcellular localization was assayed by indirect immunofluorescence using α-Flag antibody. We observed that ASF/SF2(wt) as well as all the docking motif mutants translocate to the nucleus, and as expected, transiently expressed ASF/SF2(wt) is detected predominantly in nuclear speckles (Figure 3.7, first panel). ASF/SF2(S4A) is also mostly deposited in nuclear speckles (> 50% of the transfected cells show speckles) (Figure 3.7, second panel). In contrast, ASF/SF2(ΔDock) is mostly diffused throughout the nucleus (only 12% of transfected cells show nuclear speckles compared to >75% for wild type) (Figure 3.7, third panel). The defect of ASF/SF2(R191A/K193A) in speckle formation is intermediate (35% of transfected cells show nuclear speckles) (Figure 3.7, last panel). Since phosphorylation of the RS domain is sufficient for nuclear import, it is not surprising that these mutants localize to the nucleus. However, it seems that phosphorylation of every serine within the RS1 motif is not essential for speckle formation.

4. The docking motif of ASF/SF2 limits its phosphorylation by SRPK1

The inability of the docking motif mutants to localize to nuclear speckles could be due to a defect in interaction of the mutant proteins with some other components of the nuclear speckles. Alternatively, the phosphorylation states of the mutant proteins may differ significantly from that of the wild type protein, leading to aberrant nuclear localization. In order to elucidate differences in the phosphorylation states of wild type and docking motif mutant proteins, we performed an in vitro phosphorylation assay using identical amounts of ASF/SF2(wt) and ASF/SF2(ΔDock). We observe that in vitro

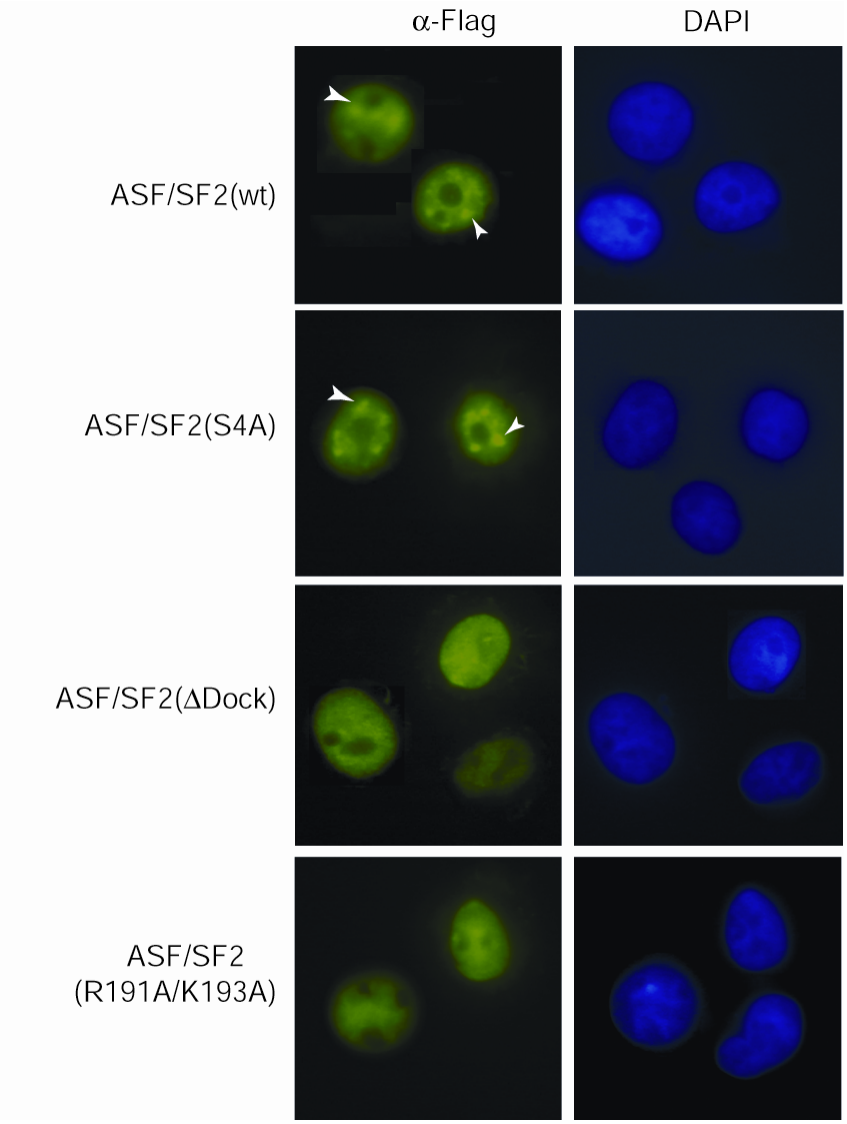


Figure 3.7. The docking motif of ASF/SF2 is essential for its recruitment to nuclear speckles. Indirect immunofluorescence assays of Hela cells transfected with Flag-ASF/SF2(wt), (ΔDock) and (R191A/K193A) show that ASF/SF2(wt) is in nuclear speckles whereas ASF/SF2(ΔDock) and ASF/SF2(R191A/K193A) are diffused throughout the nucleus. Arrows indicate nuclear speckles.

ASF/SF2(ΔDock) is phosphorylated as efficiently as ASF/SF2(wt) and incorporates higher amounts of phosphate than ASF/SF2(wt) (Figures 3.8A and 3.8B). Our experiments suggest that significantly more serines in the RS domain are phosphorylated by SRPK1 once the docking motif of ASF/SF2 is deleted. We have shown previously that ASF/SF2(ΔDock) does not localize to nuclear speckles. The fact that this molecule escapes restricted phosphorylation by SRPK1 suggests that hyperphosphorylation might be the reason why ASF/SF2(ΔDock) fails to localize to the nuclear speckles.

5. Docking motif mutations of ASF/SF2 bypass the Clk/Sty function for its release from speckles

Previous experiments have shown that over-expression of Clk/Sty along with ASF/SF2(wt) causes a release of ASF/SF2 from nuclear speckles [68-70]. We reason that Clk/Sty phosphorylates serines within the RS2 motif of ASF/SF2, an event essential for its dissociation from speckles. If this hypothesis is correct, then ASF/SF2(ΔRS2) would not be released from speckles even upon co-expression with Clk/Sty. Both Flag-ASF/SF2(wt) and ASF/SF2(ΔRS2), when transiently over-expressed in HeLa cells, localize to nuclear speckles, as determined by indirect immunofluorescence (Figure 3.9, left panels). We have compared the localization of ASF/SF2(wt) and ASF/SF2(ΔRS2) in the presence of over-expressed Clk/Sty. Whereas ASF/SF2(wt) is diffused throughout the nucleus in presence of Clk/Sty, ASF/SF2(ΔRS2) is still retained in the speckles (Figure 3.9, middle top and bottom panels respectively). Over-expression of a catalytically inactive mutant of Clk/Sty (catalytic lysine mutated to arginine), Clk/Sty(K190R), does not have any effect on ASF/SF2 localization, confirming that

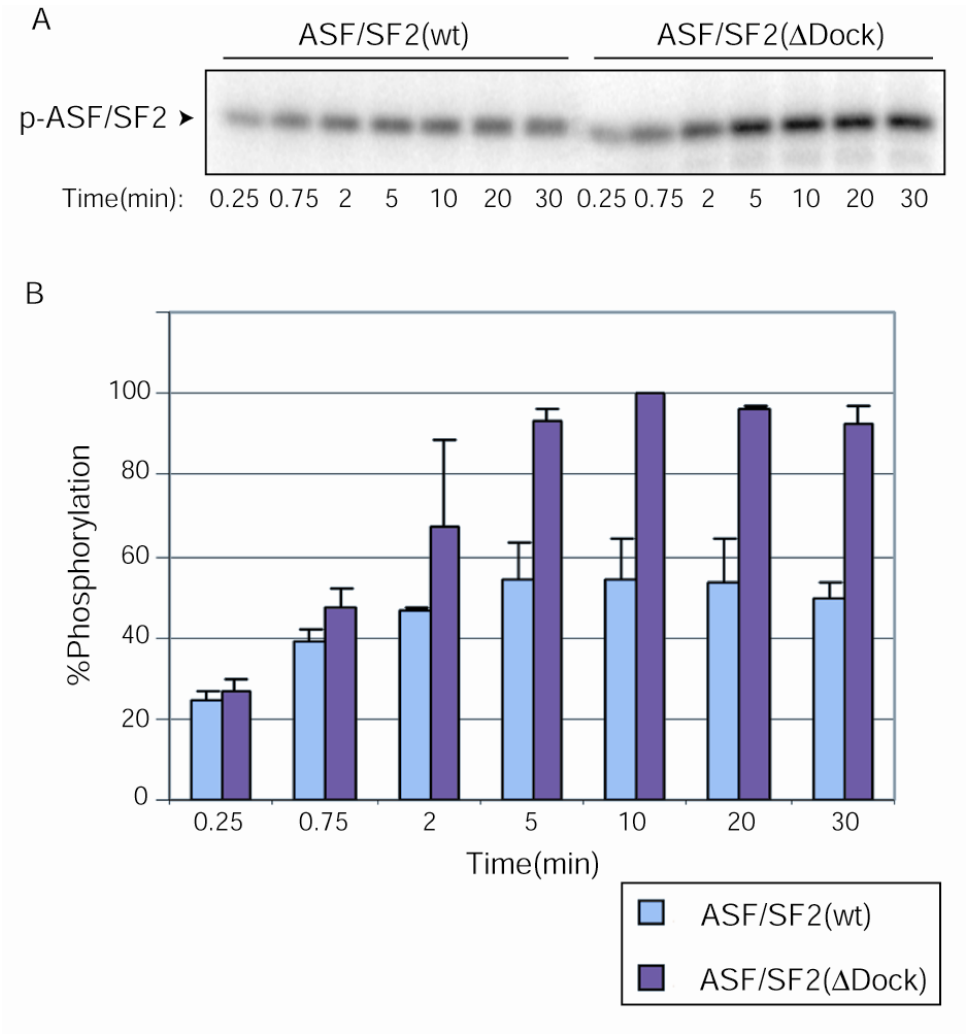


Figure 3.8. The docking motif of ASF/SF2 restricts its phosphorylation to the RS1 motif. A. In vitro kinase assay with SRPK1, using ASF/SF2(wt) and ASF/SF2(ΔDock) as substrates. B. Quantification of amounts of phosphate incorporated in each substrate in the assay above. ASF/SF2(ΔDock) incorporates approximately twice the number of phosphates compared to ASF/SF2(wt). The efficiency of phosphorylation of both substrates is comparable. Error bars represent the standard deviation from three independent experiments.

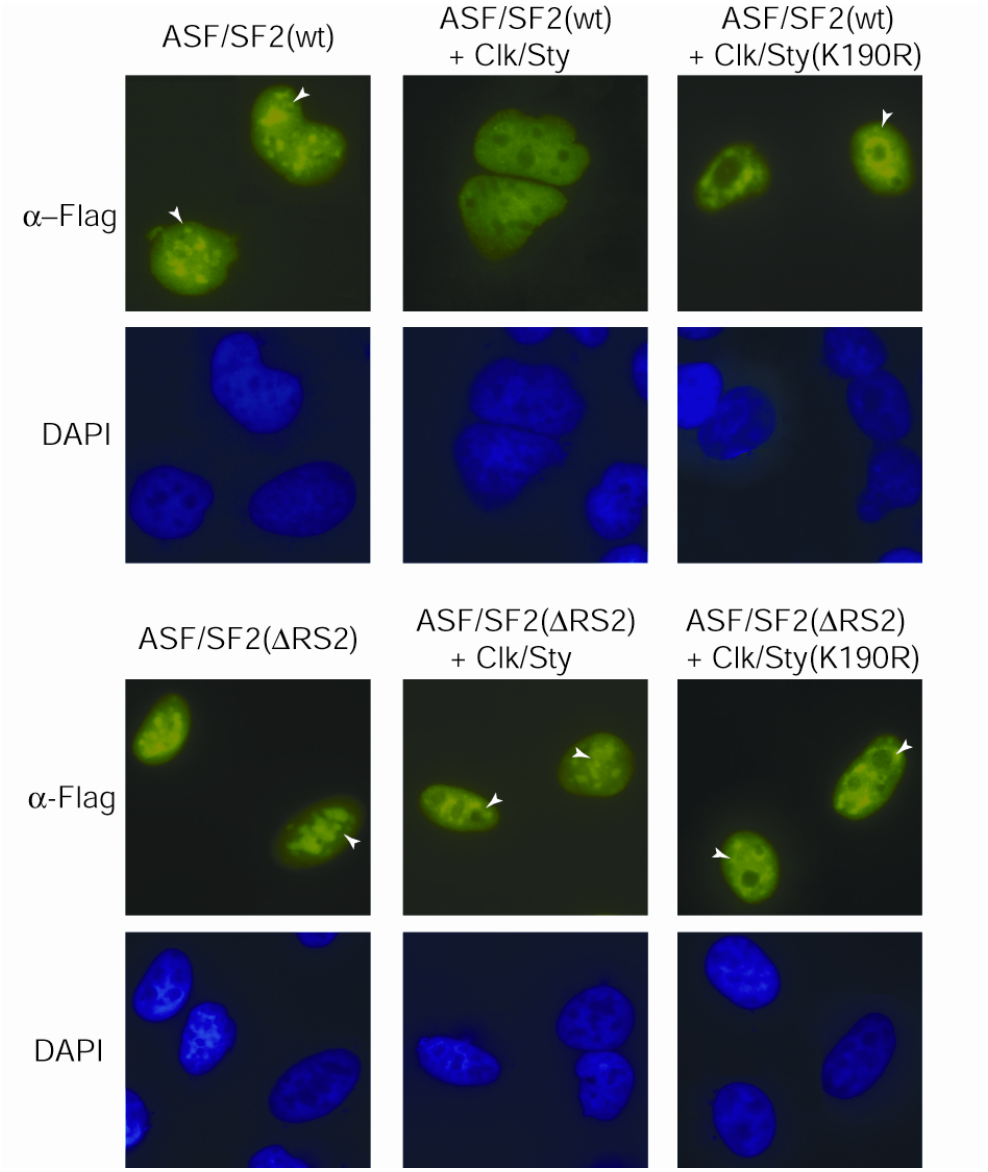


Figure 3.9. Phosphorylation of the RS2 motif of ASF/SF2 by Clk/Sty releases it from nuclear speckles. Indirect immunofluorescence assays of HeLa cells transfected with Flag- ASF/SF2(wt) and ASF/SF2(ΔRS2) show that both proteins are deposited to nuclear speckles in their steady state. Overexpression of active Clk/Sty causes dissolution of speckles in case of ASF/SF2(wt) but has no such effect on ASF/SF2(ΔRS2). Overexpression of ASF/SF2 with a catalytically inactive form of Clk/Sty has no effect on speckle formation in either case. Arrows indicate nuclear speckles.

breakdown of the nuclear speckles is a direct consequence of hyperphosphorylation (Figure 3.9, right panels). To further determine if the effect on subnuclear localization of ASF/SF2 is due to different phosphorylated states, we sought to compare the mobility of ASF/SF2(wt) and ASF/SF2(ΔRS2), obtained from HeLa cells co-transfected with either wild type Clk/Sty or Clk/Sty(K190R) in addition to ASF/SF2, by SDS/PAGE (Figure 3.10A). We observe that mobility of ASF/SF2 increases upon treatment with Calf Intestinal Phosphatase (CIP), confirming that the different slow migrating species of ASF/SF2 resolved by SDS-PAGE are phosphorylated products. Upon cotransfection of ASF/SF2(wt) with Clk/Sty, we observe two species, the more prominent of which is the slowest migrating species (Figure 3.10A, lane 1). The faster migrating species corresponds to one obtained upon cotransfection of ASF/SF2(wt) with Clk/Sty(K190R) (Figure 3.10A, lanes 1 and 3). We conclude that the two species of ASF/SF2(wt) obtained upon co-transfection with Clk/Sty, correspond to hypo- and hyperphosphorylated forms of ASF/SF2(wt). The species obtained upon CIP treatment of the extract corresponds to unphosphorylated ASF/SF2(wt) (Figure 3.10A, lane 2). ASF/SF2(ΔRS2) exists as only one slow migrating species, corresponding to one phosphorylation state, irrespective of the presence or absence of an active form of Clk/Sty (Figure 3.10A, lanes 5 and 7). CIP treatment of extracts containing ASF/SF2(ΔRS2) also yield an unphosphorylated form of this protein (Figure 3.10A, lanes 6 and 8). These observations strongly suggest that Clk/Sty phosphorylates the RS2 motif of ASF/SF2, yielding a hyperphosphorylated species. Moreover, these experiments demonstrate that the hypophosphorylated form of ASF/SF2 is deposited into speckles and that Clk/Sty regulates the release of ASF/SF2 from speckles by converting the

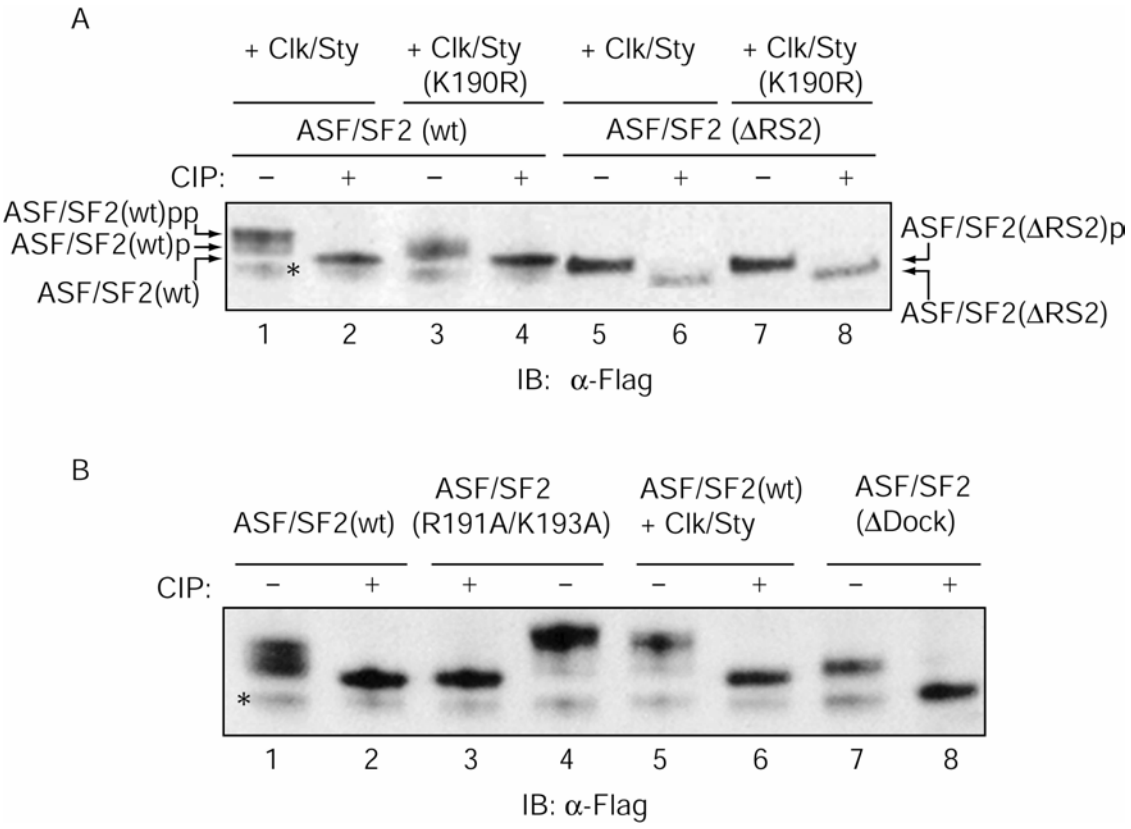


Figure 3.10. SRPK1 and Clk/Sty regulate ASF/SF2 through differential phosphorylation of the RS domain. A. SDS-PAGE analysis of HeLa cell lysates of Flag- ASF/SF2(wt) and ASF/SF2(ΔRS2) co-transfected with Clk/Sty. Co-transfection of ASF/SF2(wt) with active Clk/Sty yields a hyperphosphorylated species in addition to the hypo-phosphorylated one whereas ASF/SF2(ΔRS2) exists only in one phosphorylated state irrespective of the presence of active Clk/Sty. B. SDS-PAGE analysis of HeLa cell lysates of Flag-ASF/SF2(wt) (in presence or absence of Clk/Sty), ASF/SF2 (R191A/K193A) and ASF/SF2(ΔDock). The mobility of ASF/SF2(R191A/K193A) is similar to that of hyperphosphorylated ASF/SF2(wt), indicating enhanced phosphorylation. ASF/SF2(wt) exists in multiple phosphorylation states, whereas ASF/SF2(R191A/K193A) and ASF/SF2(ΔDock) are present predominantly in a single hyperphosphorylated form. * indicates a non-specific band detected by the α-Flag antibody.

hypophosphorylated form of ASF/SF2 into a hyperphosphorylated form by phosphorylation of the RS2 motif.

In order to reaffirm the role of the docking motif of ASF/SF2 in restricted phosphorylation by SRPK1, we compared phosphorylation states of ASF/SF2(wt), ASF/SF2(R191A/K193A) and ASF/SF2(ΔDock) obtained from HeLa cells transiently expressing these proteins (Figure 3.10B). We observed that ASF/SF2(ΔDock) and ASF/SF2(R191A/K193A) exist predominantly as one slower migrating species whereas ASF/SF2(wt) exists in at least two different forms (Figure 3.10B, lanes 7, 4 and 1). The mobility shift observed in ASF/SF2(R191A/K193A), as compared to ASF/SF2(wt), corresponds to that observed upon cotransfection of ASF/SF2(wt) with active Clk/Sty, indicating that ASF/SF2(R191A/K193A) is constitutively hyperphosphorylated (Figure 3.10B, compare lanes 1, 4 and 5). Since ASF/SF2(ΔDock) is smaller in size than ASF/SF2(wt), we cannot correlate its migration on SDS-PAGE to its phosphorylation state in comparison to ASF/SF2(wt). However, the existence of ASF/SF2 (R191A/K193A) as a single hyperphosphorylated species suggests that the single species of ASF/SF2(ΔDock) represents its hyperphosphorylated state. This experiment confirms the role of the docking motif in restricting phosphorylation of ASF/SF2 by SRPK1 to the RS1 motif alone, thus allowing Clk/Sty to hyperphosphorylate ASF/SF2 when required.

6. Specific residues within the RRM2 domain of ASF/SF2 contribute to its high affinity interaction with SRPK1

We have shown above that in addition to the docking motif, the RRMs of ASF/SF2 play a role in its high affinity interaction with SRPK1. The X-ray crystal

structure of the SRPK1:ASF/SF2 complex indicates specific interactions between residues in the RRM2 domain of ASF/SF2 with the small and large lobe of SRPK1 (Figure 3.3B and 3.3C). We mutated these contact residues (W134, Q135 and R154) of ASF/SF2 to alanines and compared the ability of these mutants to bind SRPK1 with that of wild type ASF/SF2 (Figure 3.11). Flag-ASF/SF2 constructs were cotransfected with a catalytically inactive mutant of SRPK1, SRPK1 K109R, into HEK-293 cells, followed by immunoprecipitation with α-SRPK1 antibody. Use of an inactive form of SRPK1 in co-immunoprecipitation assays isolated the ASF/SF2:SRPK1 binding event from all other steps in the phosphorylation reaction and enabled us to capture a complex which is likely to be of a transient nature due to rapid turnover of the substrate. We observe that mutation of residues W134, Q135 and R154 to alanines abolish the binding of ASF/SF2 to SRPK1, thus denoting their role in this interaction (Figure 3.11, compare lanes 9, 10, 11 and 12).

7. Mutation of ASF/SF2 contact residues alters its steady-state localization but not its phosphorylation

Having identified the residues of ASF/SF2 that are important for its interaction with SRPK1, we wanted to test their role in the sub-cellular localization of ASF/SF2. It has been firmly established that phosphorylation of the RS domain of ASF/SF2 by SRPK1 is essential for its nuclear localization, but there is no evidence to show if high affinity interaction with SRPK1 is also necessary. We studied the sub-cellular localization of the ASF/SF2 mutants by transiently over-expressing Flag-tagged constructs in HeLa cells as before (Figure 3.12). We found that whereas ASF/SF2(wt) is

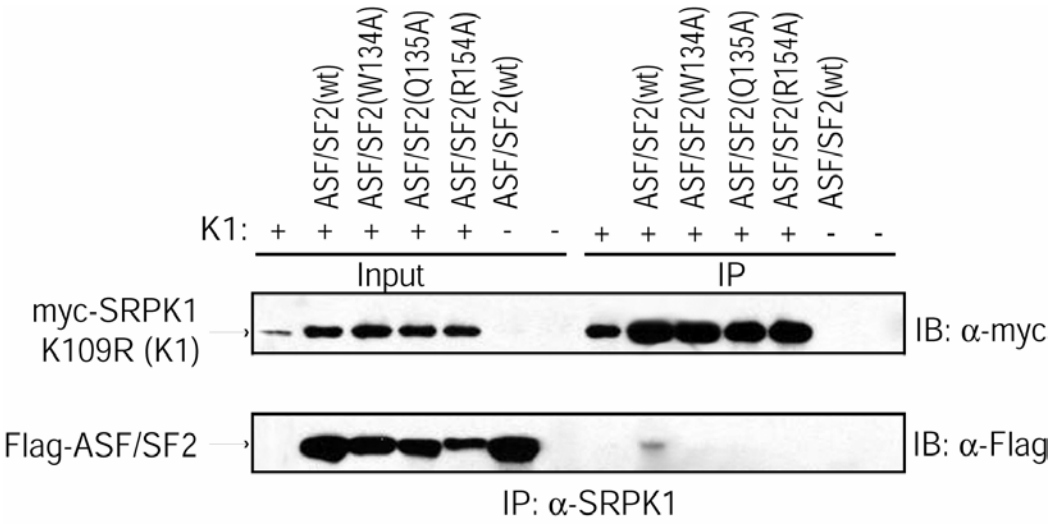


Figure 3.11. The contact residues of ASF/SF2 are responsible for its high affinity interaction with SRPK1. Key contact residues of ASF/SF2 (W134, Q135 and R154) were identified from the X-ray crystal structure of the SRPK1:ASF/SF2 complex. Co-immunoprecipitation assays of ASF/SF2 with a catalytically inactive form of SRPK1 show that binding of ASF/SF2 to SRPK1 is completely abolished upon mutation of the contact residues.

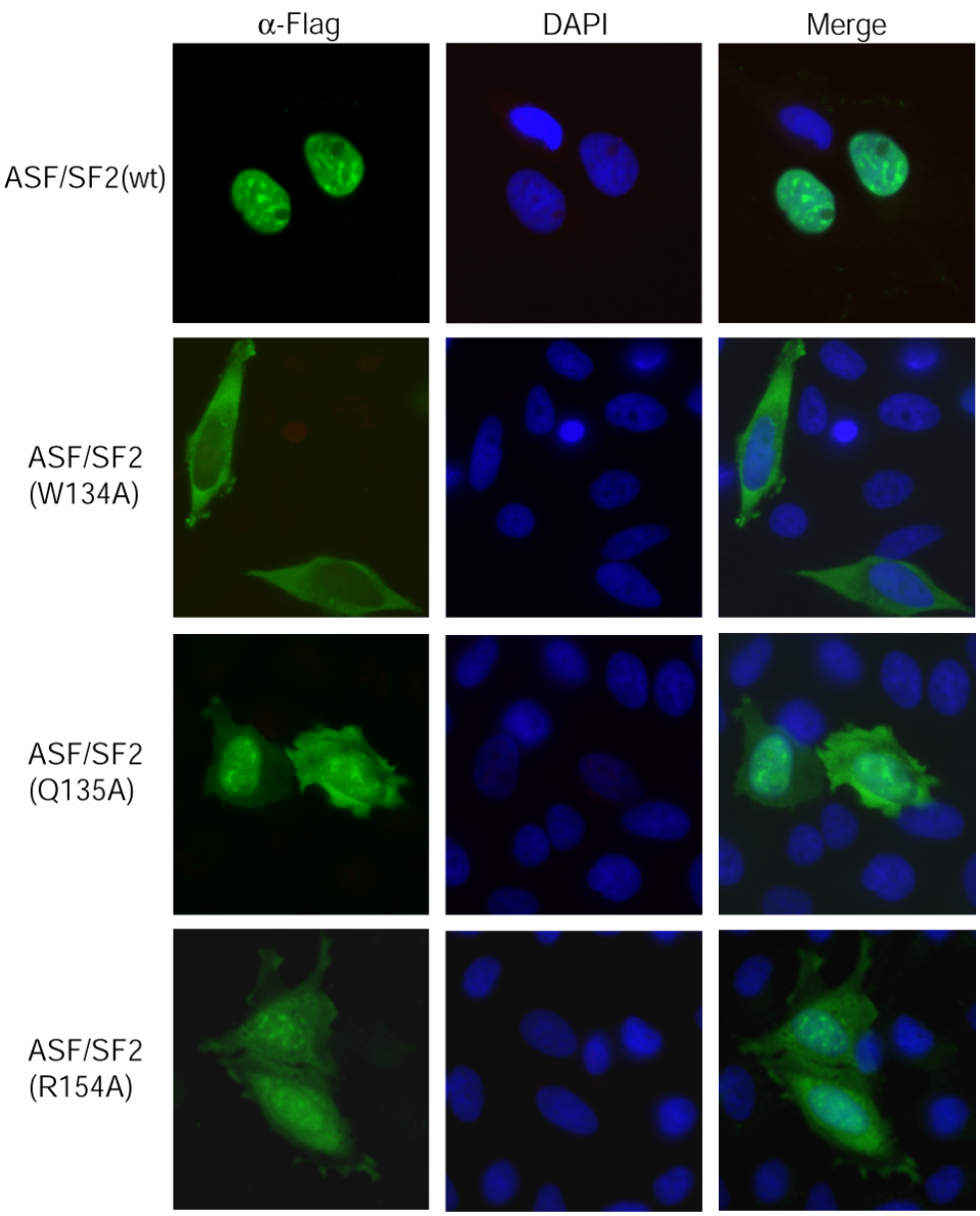


Figure 3.12. Stable association of ASF/SF2 with SRPK1 is essential for its proper subcellular localization. Indirect immunofluorescence assays of HeLa cells transfected with Flag- ASF/SF2-(wt), (W134A), (Q135A) and (R154A) show that ASF/SF2(wt) localizes to nuclear speckles whereas ASF/SF2-(W134A), (Q135A) and (R154A) are distributed both in the nucleus and the cytoplasm.

in the nuclear speckles as expected, ASF/SF2(W134A) and ASF/SF2(R154A) localize to the cytoplasm to a great extent (~50% of the transfected cells show ASF/SF2 either exclusively in the cytoplasm or distributed throughout both the nucleus and cytoplasm) (Figure 3.12, first, second and fourth panels). ASF/SF2(Q135A) is present predominantly in the nucleus, with only 10% of the transfected cells showing cytoplasmic/nucleocytoplasmic localization of the protein (Figure 3.12, third panel).

To test if the aberrant cellular localization of ASF/SF2 contact mutants is due to a defect in their phosphorylation, we sought to determine the phosphorylation states of these mutant proteins by comparing their mobility on SDS-PAGE with that of ASF/SF2(wt) as described earlier (Figure 3.13). CIP treatment of the cell lysates confirmed that the different mobility shifts of ASF/SF2 observed in SDS-PAGE are due to differential phosphorylation of the protein. We observed that 24 hours post transfection all ASF/SF2 mutants are phosphorylated as efficiently as the wild type protein (Figure 3.13A, lanes 1, 3, 5 and 7). However, 48 hours post transfection, a slow-migrating species, corresponding to the hyperphosphorylated form of ASF/SF2, accumulates in ASF/SF2(wt) and ASF/SF2(Q135A) and is absent in ASF/SF2(W134A) and ASF/SF2 (R154A) (Figure 3.13B, compare lanes 1, 3, 5 and 7). This species is absent in the two mutants that partially localize to the cytoplasm since hyperphosphorylation of ASF/SF2 occurs only when it is present in nuclear speckles and is mediated by the nuclear kinase Clk/Sty. This suggests that the defect in localization observed in the ASF/SF2 contact mutants is not due to inefficient phosphorylation by SRPK1, but rather a consequence of loss of binding with SRPK1.

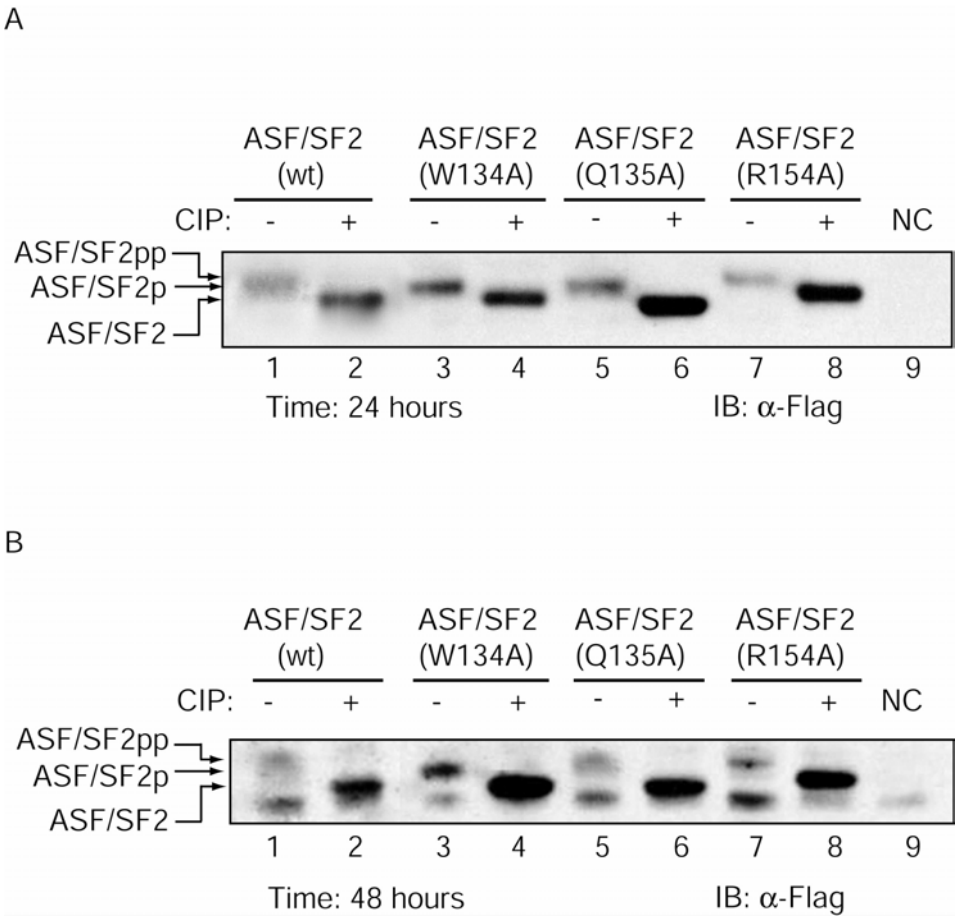


Figure 3.13. ASF/SF2 contact mutants can be hypophosphorylated by SRPK1 but are not hyperphosphorylated efficiently. A. SDS-PAGE analysis of lysates of HeLa cells transiently expressing ASF/SF2- (wt), (W134A), (Q135A) and (R154A) for 24 hours. All ASF/SF2 contact mutants are hypophosphorylated in a manner similar to ASF/SF2(wt). B. SDS-PAGE analysis of lysates of HeLa cells transiently expressing ASF/SF2- (wt), (W134A), (Q135A) and (R154A) for 48 hours. ASF/SF2- (wt) and (Q135A) get hyperphosphorylated over time, whereas ASF/SF2-(W134A) and (R154A) remain exclusively in the hypophosphorylated form.

C. Discussion

Our studies based on the X-ray crystal structures of SRPK1 bound to its peptide and SR protein substrates provide insight into the various distal recognition elements present in substrates of protein kinases. Docking interactions between a kinase and its interacting partners (substrates, activators) have been extensively characterized in MAP kinases [90, 91, 94, 106]. Additionally, several other kinases have been shown to rely upon distal docking interactions to achieve substrate specificity [86], [87]. Since the consensus sequence of the phosphorylation site for a particular kinase is often fairly degenerate, the docking interactions are a key element in regulating specificity and preventing aberrant phosphorylation by proteins kinases. The X-ray crystal structure of a complex of SRPK1 and a 9mer peptide derived from Npl3p led to the identification of a docking groove in SRPK1. This docking groove is comprised of helix αG, the loop connecting helices αF and αG and the MAP kinase insert. The MAP kinase insert is a common feature of all kinases in the CMGC family (of which SRPK1 is a member) and together with helix αG has been shown to form a docking groove in Cdk2 and GSK3β [107-109].

The study outlined here describes the identification of a novel docking motif in SR proteins. We suggest that the ideal docking motif sequence for an SRPK1 substrate is R-X-R/K-X-X-X-R and have shown that the basic residues within this sequence are crucial for docking interactions. We have not tested the importance of the intervening valine residues (V192 and V194) within the ASF/SF2 docking motif in SRPK1 recognition. Molecular dynamics simulations with SRPK1 and ASF/SF2 show that these valines do not interact with the docking groove of SRPK1 in any way and are in fact,

detrimental to substrate docking (J. Gullinsgrad, personal communication). The X-ray crystal structure of the SRPK1:ASF/SF2 complex shows that part of the RS1 motif of ASF/SF2 binds to the docking groove of SRPK1. This indicates that electrostatic interactions between basic residues of the docking motif and acidic residues of the docking groove are the major determinants for this binding event. This also suggests that there are multiple docking events occurring in course of phosphorylation of ASF/SF2 by SRPK1. To enable cycles of docking and release, the docking motif of the substrate should be such that it can bind with moderate affinity and yet be released fairly easily. Presence of basic and hydrophobic residues in the docking motif of ASF/SF2 may modulate its affinity in order to achieve precise phosphorylation of the RS domain. Subsequent to the identification of the ASF/SF2 docking motif, multiple docking motifs have been identified in RGG domain of the yeast SR-like protein Npl3p [110]. The docking motif of Npl3p conforms to the sequence requirement mentioned above and contains basic residues (arginines) interspersed with hydrophobic residues (phenylalanines). It has been proposed that the presence of multiple docking motifs increases their local concentration, thus resulting in effective recognition of Npl3p by the docking groove of the kinase Sky1p. Docking motif sequences preceding the RS domain are also conserved in other human SR proteins such as Srp40, SRp75 and SRp55, suggesting that docking interactions may be a common mode of recognition of all SR proteins by SRPKs (Figure 3.14).

It was previously shown that SRPK1 interacts with ASF/SF2 with high affinity (50nM) and phosphorylates it in a processive manner. We show here that direct interactions between the docking groove of SRPK1 and the docking motif of ASF/SF2

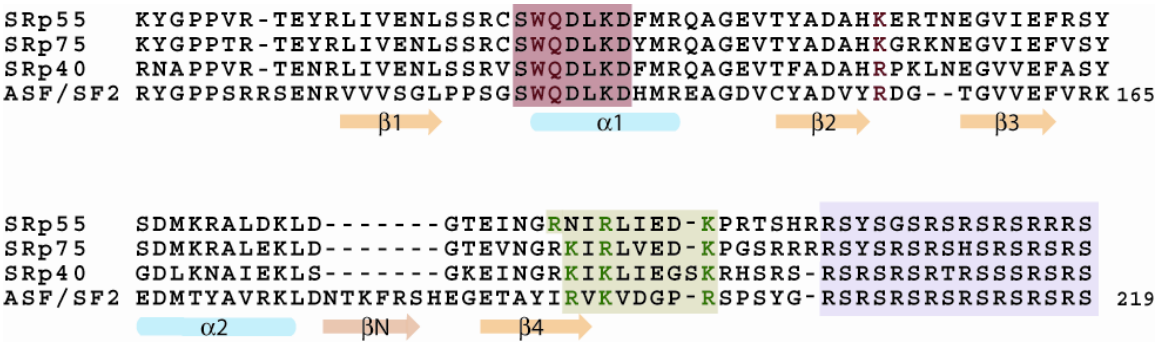


Figure 3.14. Multiple sequence alignment of select SR proteins. The RRM2 domains and part of the RS domains of the SR proteins SRp55, SRp75 and SRp40 and ASF/SF2 were used to generate a multiple sequence alignment of these SR proteins. Secondary structural elements, based on the X-ray crystal structure of ASF/SF2 in the SRPK1:ASF/SF2 complex, have been aligned with the multiple sequence alignment. Beta sheets are denoted by yellow arrows and alpha helices by cyan cylinders. The SWQDLKD motif is highlighted in red, the docking motif in green and the RS domain in mauve. Conserved contact residues are colored red and basic residues within the docking motif are colored green.

contribute to the binding affinity of the two proteins, which governs the mode of phosphorylation of ASF/SF2. Even more intriguing is the fact that the docking groove-docking motif interactions restrict the number of serines in ASF/SF2 that can be phosphorylated by SRPK1. We have shown that typically only serines within the RS1 motif of ASF/SF2 are phosphorylated by SRPK1 and upon deletion of the docking motif, significantly more serines in the RS domain are phosphorylated. We show conclusively that distinct phosphorylation states of ASF/SF2 determine its subcellular and subnuclear localization. In vivo, SRPK1 phosphorylates ASF/SF2 to generate a hypophosphorylated species which is deposited into nuclear speckles. Release of ASF/SF2 from these speckles requires its hyperphosphorylation by Clk/Sty. Nuclear speckles are thought to be sites of storage for splicing factors, from where they are recruited to the sites of active transcription (and consequently splicing) by the action of Clk/Sty [68, 69]. Since splicing factors have to be released from nuclear speckles in order to participate in splicing, it is likely that the hyperphosphorylation and release event is stringently regulated in cells by specific signals to activate Clk/Sty.

In addition to the docking motif, we have shown that the RRM domains of ASF/SF2 contribute to its high affinity interaction with SRPK1. The X-ray crystal structure of the SRPK1:ASF/SF2 complex indicates that the RRM2 domain of ASF/SF2 mediates a 3-point contact with the small and large lobes of SRPK1. The amino acids of ASF/SF2 involved in this 3-point contact are conserved in all SR proteins containing 2 RRMs (Figure 3.14). We show that disruption of any one these contact residues in ASF/SF2 abolishes its interaction with SRPK1. Furthermore, we show that although hypophosphorylation by SRPK1 is not affected, these mutants fail to localize to the

nucleus as efficiently as wild type ASF/SF2. However, hyperphosphorylation of the localization-defective mutants is affected, once again proving that hyperphosphorylation of ASF/SF2 is a strictly nuclear event. Although it is not surprising that the ASF/SF2 contact mutants are phosphorylated by SRPK1, the defect observed in nuclear localization is unexpected. Previous studies have deemed phosphorylation of the RS domain to be sufficient for nuclear import of SR proteins [51]. Our observations suggest that stable interaction with SRPK1 along with phosphorylation is essential for nuclear import of ASF/SF2. We suggest that SRPK1 binds ASF/SF2 and keeps it in a natively folded phosphorylated form until it interacts with the SR protein-specific imortin, Transportin-SR. SR proteins are abundant in cells, and being highly charged, are prone to aggregation in solution. Also, Transportin-SR has been shown to interact exclusively with phosphorylated RS domains of SR proteins [52]. Thus we speculate that SRPK1 acts as a chaperone for ASF/SF2 preventing its aggregation and dephosphorylation by cellular phosphatases prior to nuclear import.

We can now begin to understand the sequence of phosphorylation events of ASF/SF2 and its relation to biological function. Our studies on the recognition of ASF/SF2 by SRPK1 allow us to propose a model for regulation of ASF/SF2. We have show the basis and necessity for high affinity interaction of SRPK1 and ASF/SF2. ASF/SF2 binds SRPK1 with high affinity in the cytoplasm through its RRM2 domain and docking motif, leading to phosphorylation of its RS1 motif. This hypophosphorylated ASF/SF2 is then imported to the nuclear speckles in a Transportin-SR dependent manner, where it can be hyperphosphorylated by Clk/Sty and recruited to sites of splicing. It has been shown earlier that dephosphorylation of ASF/SF2 is required for splicing activity

[39, 42, 111]. Moreover, nuclear export of mRNA mediated by ASF/SF2 requires a hypophosphorylated form of ASF/SF2. It is likely that at some point during spliceosome assembly or catalysis, ASF/SF2 undergoes dephosphorylation, which converts the hyperphosphorylated state generated in the nucleus back to a hypo- or unphosphorylated state. Recently, PP1 and PP2A family phosphatases have been implicated in pre-mRNA splicing, by specifically acting upon components of U2 and U5 snRNPs [112]. However no specific phosphatases for mammalian SR proteins have been identified as yet. Figure 3.15 summarizes our model for regulation of ASF/SF2 by SRPK1 and Clk/Sty.

The text of this chapter is, in part, reprint of material as it appears in *Molecular Cell*, 2005, **20**, 77-89, Ngo, J. C., Chakrabarti, S., Ding, J. H., Velazques-Dones, A., Nolen, B., Aubol, B. E., Adams, J. A., Fu, X. D. and Ghosh, G. The dissertation author was the primary researcher and author of this publication.

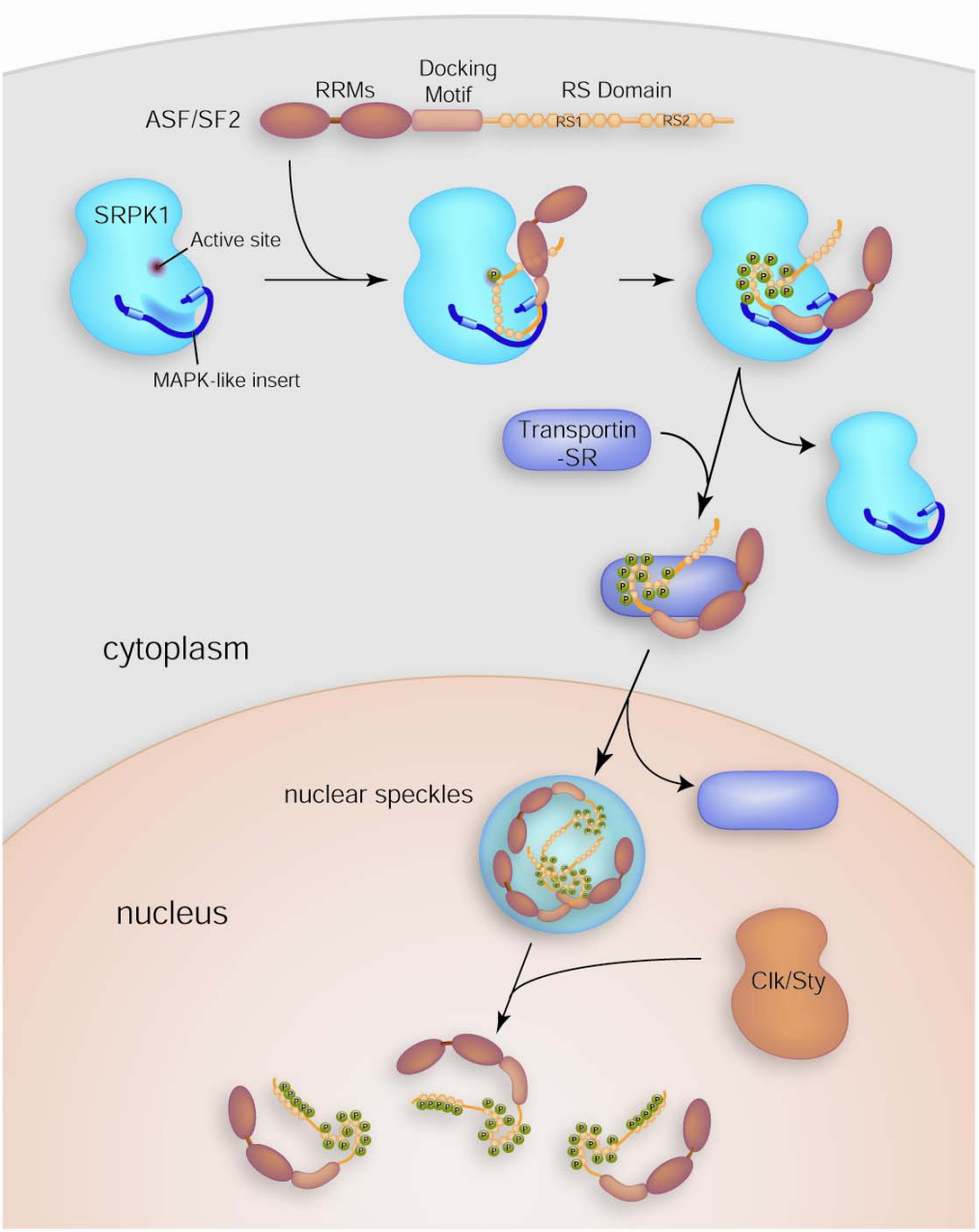


Figure 3.15. Proposed model for regulation of ASF/SF2 by SRPK1 and Clk/Sty. SRPK1 binds ASF/SF2 with high affinity in the cytoplasm and phosphorylates its RS1 motif. This hypophosphorylated ASF/SF2 interacts with Transportin-SR and is deposited in nuclear speckles. Hyperphosphorylation of ASF/SF2 in the speckles by Clk/Sty releases it from the speckles and recruits it to sites of splicing.

IV: Recognition of Specific RNA Sequence Elements by ASF/SF2

A. Introduction

ASF/SF2 is a prototype member of the SR protein family and consists of two N-terminal RRM domains and a C-terminal RS domain. Phosphorylation of the RS domain of ASF/SF2 by SRPK and Clk/Sty kinases regulates its subcellular localization and cellular functions such as mRNA export, constitutive and alternative splicing and mRNA translation [36, 37, 44, 45, 48, 71, 113, 114]. Structural studies on human and yeast SRPKs have led to the identification of conserved structural elements, which dictate the mechanism of phosphorylation of SR proteins by these kinases. One such structural feature is the basic 'P-2' pocket (comprised of two conserved arginine residues, R568 and R571 in Sky1p) which is capable of binding a negatively charged molecule. Molecular dynamics simulations and biochemical studies on Sky1p suggest that this basic pocket binds a negatively charged group located at the P-2 position of the substrate (E409 in Npl3p, phospho-serine in ASF/SF2) and places the P-site serine (S411 in Npl3p, next serine in the contiguous RS tract of ASF/SF2) in the active site of the kinase (Figure 4.1) [66]. These studies propose a model for primed sequential phosphorylation of the RS domain of SR proteins, where every phospho-serine, when bound to the P-2 pocket of the kinase, acts as a priming site for phosphorylation of the next available serine (Figure 4.1). However this model does not explain how the first serine in the RS domain gets phosphorylated, since ASF/SF2 does not have an acidic residue at the P-2 position for any serine within the RS domain. It is possible that a particular serine residue in ASF/SF2 is phosphorylated by a different kinase and this phospho-serine act as the priming phosphorylation site for the RS domain. It has been reported that ASF/SF2 can

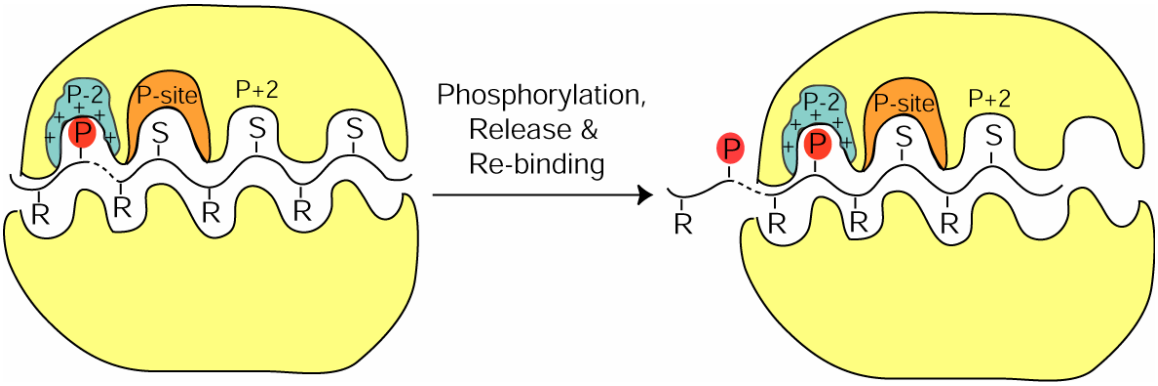


Figure 4.1. Primed phosphorylation of SR proteins by SRPKs. A negatively charged group from the SR protein (phospho-serine or acidic amino acid, highlighted in red) binds the basic P-2 pocket of the kinase (highlighted in blue) and positions the P-site serine of the substrate in the active site of the kinase (highlighted in orange). Once the P-site serine is phosphorylated, it can slide over and occupy the P-2 pocket of the kinase, placing the next available serine in the active site of the kinase which leads to sequential phosphorylation of the RS domain.

be phosphorylated by proteins other than SRPKs and Clk/Sty, such as DNA topoisomerase I and Akt [115, 116].

Unlike its phosphorylation and sub-cellular localization, the RNA binding function of ASF/SF2 is not clearly understood as yet. Previous studies have shown that ASF/SF2 is capable of binding 5' splice site sequences in the introns of pre-mRNA and directly influencing splice site selection [26, 117]. More recent studies show that ASF/SF2 recognizes exonic splicing enhancer sequences (ESEs) within the exons of pre-mRNA and indirectly influence splice site selection [118]. It is now widely believed that the RRM domains of ASF/SF2 bind the ESE of mRNA, while the RS domain is free to interact with core components of the spliceosome thus facilitating recruitment of the spliceosome to splice sites (Figure 4.2) [21]. In support of this model, ASF/SF2 and U1-70K, a component of U1snRNP, were shown to interact through their RS domains [40, 119]. ESE-dependent splice site recognition by the spliceosome is particularly important in case of alternative splicing or when introns contain weak splice sites. Several ESE sequences have been identified by in vitro and in vivo SELEX experiments and computational studies [120, 121]. ESE sequences are degenerate and may be present in multiple copies in an exon, leading to difficulties in identifying a pattern of ESE recognition by SR proteins. Contrary to previous studies, reports by Shen *et al.* show that the RS domains of SR proteins contact the splicing signals (ESE, branchpoint of pre-mRNA and splice sites) to promote spliceosome assembly [122, 123]. Additionally, studies show that the RS domain of SR and non-SR protein, fused to an unrelated mRNA binding protein, is sufficient to mediate splicing in select cases [27, 124, 125]. Several groups have attempted to address the role of a particular SR protein (and specific

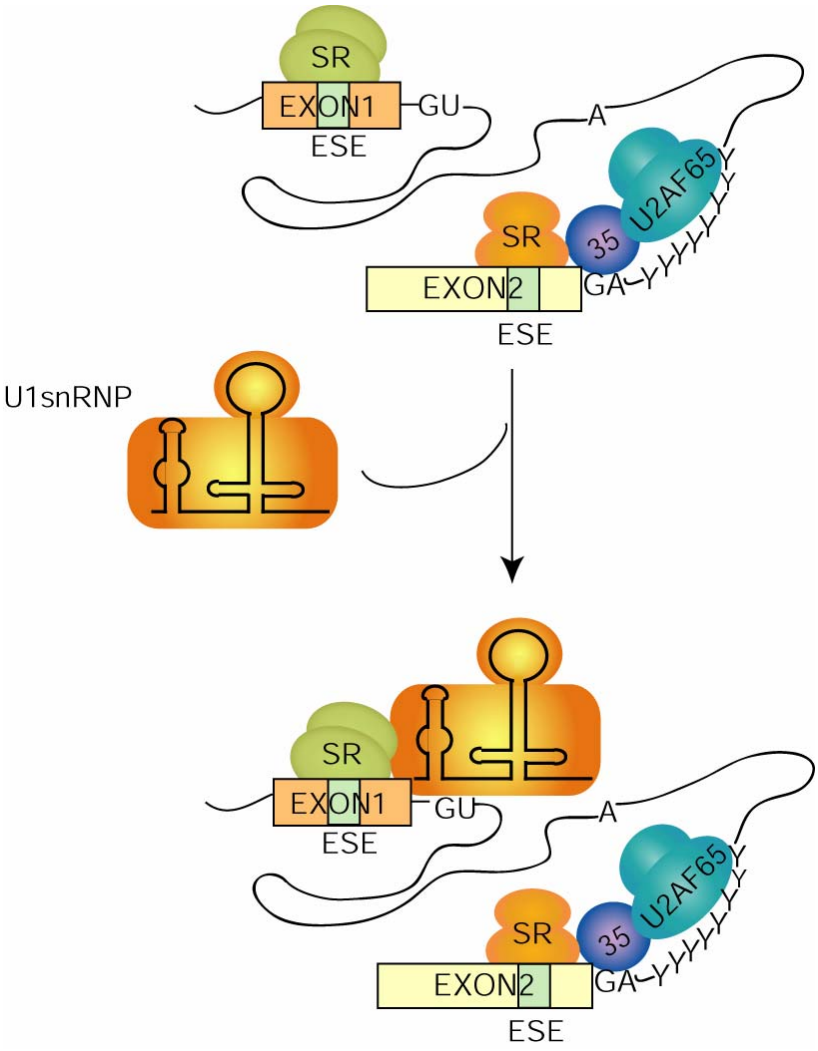


Figure 4.2. ESE-dependent recruitment of U1snRNP to the 5' splice site. SR proteins specifically bind ESE-sequences within pre-mRNA through their RRM domains and simultaneously interact with components of U1snRNP through their RS domain, thus facilitating recruitment of U1snRNP to the neighboring 5' splice site.

domains therein) in splicing by performing in vitro splicing assays using HeLa extracts and recombinant SR proteins [37, 39]. The functional redundancy of SR proteins in constitutive splicing limits the information that can be gleaned from these assays. Furthermore, SR proteins are highly charged and prone to aggregation in solution, leading to problems in purification of the recombinant proteins. Hence, biochemical information on the mechanism of RNA recognition by SR proteins has not been forthcoming.

Here we discuss how an unexpected observation from the X-ray crystal structure of the SRPK1:ASF/SF2 complex led us to investigate the RNA binding property of ASF/SF2 in detail. The electron density map of the SRPK1:ASF/SF2 complex shows strong density occupying the P-2 pocket of the kinase. The structural model of the complex was refined considerably with a phospho-serine residue fit to this electron density (Figure 4.3). The phospho-serine appears to be part of the inter-RRM linker (S116 or S119) of ASF/SF2 from a symmetry-related molecule. Most of the linker region is disordered in the structure of the SRPK1:ASF/SF2 complex, suggesting that it is a highly flexible region and could easily adjust its conformation so as to place the phospho-serine in the P-2 pocket of SRPK1. Surprisingly, we found that these serine residues do not act as priming phosphorylation sites for the RS domain but play a significant role in modulating ESE recognition by ASF/SF2. Our studies reveal a novel regulatory element for RNA binding by this SR protein and form a basis for further investigation of the role of RRM-mediated RNA recognition in pre-mRNA splicing.

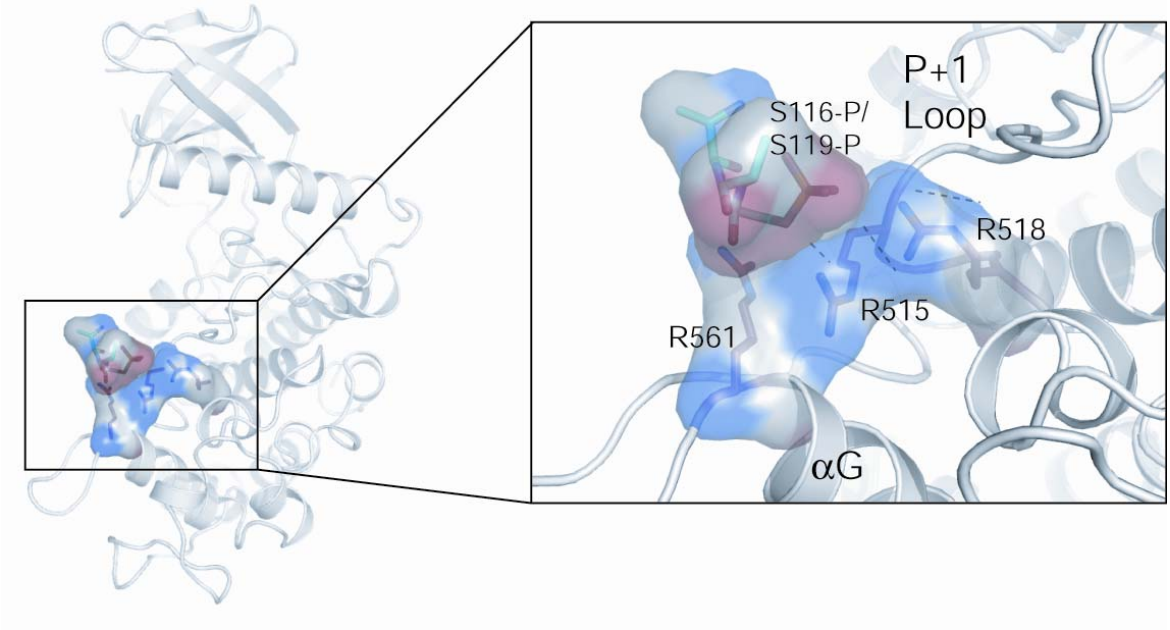


Figure 4.3. A phospho-serine of ASF/SF2 binds the P-2 pocket of SRPK1. The X-ray crystal structure of the SRPK1:ASF/SF2 complex highlighting electrostatic interactions between a phospho-serine residue (red) from the linker between the two RRM domains of ASF/SF2 (either S116-P or S119-P) and arginine residues (blue) from the P-2 pocket of SRPK1.

B. Results

1. The phospho-serine residue in the P-2 pocket of SRPK1 is not a priming site for hypophosphorylation of ASF/SF2

We first attempted to determine if the phospho-serine residue in the P-2 pocket of SRPK1 behaves as a primed phosphorylation site for sequential phosphorylation of the RS domain of ASF/SF2. To this end, we generated Flag-tagged constructs of ASF/SF2 where either or both of the potentially phosphorylated serine residues (S116 and S119) within the linker region were mutated to alanines and assayed their subcellular localization. We reasoned that if priming phosphorylation at either of the two serines is essential for phosphorylation of the RS domain, then mutation of these residues would abolish RS domain phosphorylation and affect localization of ASF/SF2 to nuclear speckles. Indirect immunofluorescence assays on HeLa cells transiently expressing ASF/SF2 constructs for 24 hours show that ASF/SF2(S116A) and ASF/SF2(S119A) localize to nuclear speckles in a manner identical to the wild type protein (Figure 4.4, first, second and third panels). However, ASF/SF2(S116A/S119A) shows a marginal defect in localization, where the protein is diffused throughout the nucleus and cytoplasm in approximately 25% of the transfected cells (Figure 4.4, fourth panel). Since the defect observed was relatively small, we hypothesized that phosphorylation of the RS domain in ASF/SF2 (S116A/S119A) was not abolished, but probably impeded due to mutation of the two serine residues. To test this hypothesis, we performed localization studies on HeLa cells transiently expressing ASF/SF2(S116A/S119A) for 12, 18, 24 and 48 hours. We expected a greater defect in nuclear localization at the early time points due to impaired phosphorylation of the RS domain, and a reversal of this defect within 48 hours.

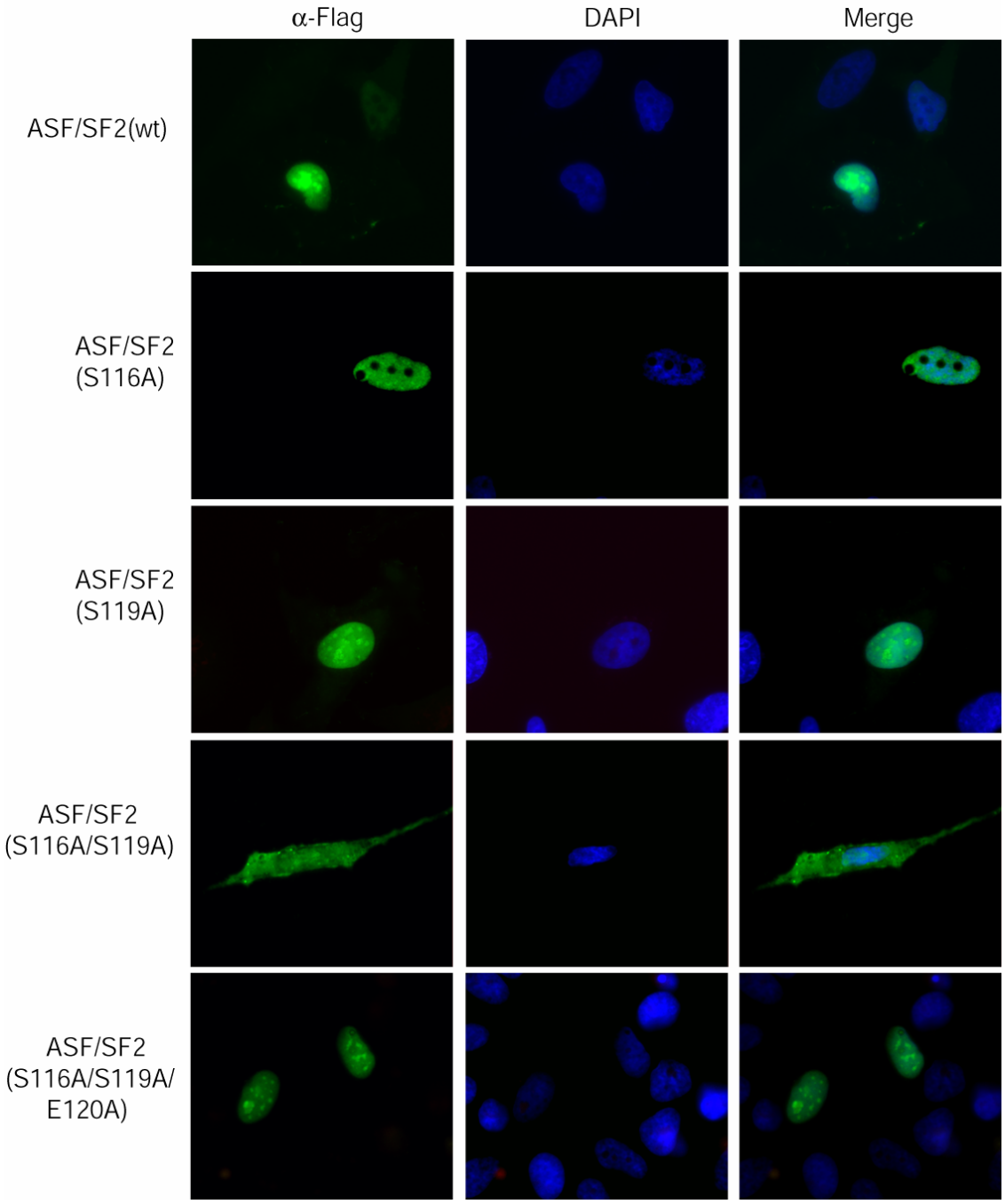


Figure 4.4. Mutation of specific residues within the inter-RRM linker of ASF/SF2 partially affects its subcellular localization. Indirect immunofluorescence assays of HeLa cells transfected with Flag- ASF/SF2-(wt), (S116A), (S119A), (S116A/S119A) and (S116A/S119A/E120A) show that while ASF/SF2(S116A/S119A) is partly diffused throughout the nucleus and cytoplasm, all the other mutants localizes to nuclear speckles in a manner identical to ASF/SF2(wt).

Surprisingly, we find that ASF/SF2(S116A/S119A) shows a defect in localization only 24 hours post transfection and localizes to nuclear speckles like the wild type protein at all other time points (Figure 4.5).

The localization pattern of ASF/SF2(S116A/S119A) led us to investigate if, in absence of the two serines, any other residue within the linker region could mimic a priming phosphorylation site and thus negate the effect of the mutations. We scanned the sequence of the linker region for any residue that could occupy the P-2 pocket of SRPK1 and found a conserved glutamate (E120) that could fulfill this role. However, mutation of this residue along with the two serines does not affect the localization of the resultant protein, ASF/SF2(S116A/S119A/E120A), in any way (Figure 4.4, fifth panel). To further confirm that the phospho-serine residue observed in the P-2 pocket of SRPK1 does not regulate phosphorylation of the RS domain of ASF/SF2, we sought to determine the phosphorylation states of the ASF/SF2 serine mutants generated above. We compared the mobility of different ASF/SF2 proteins, transiently expressed in HeLa cells, by SDS-PAGE and found that ASF/SF2- S116A, S119A and S116A/S119A are phosphorylated in a manner identical to wild type ASF/SF2 (Figure 4.6). Our results indicate that the observed phospho-serine residue in the P-2 pocket of SRPK1 does not play a role in priming phosphorylation, but may participate in regulating some other functional aspect of ASF/SF2.

2. Recognition of 5' splice site and exonic splicing enhancer sequences by ASF/SF2

Since the serine residues in the linker between the two RRM domains do not play a role in regulation of phosphorylation, we turned our attention to focus on the RNA

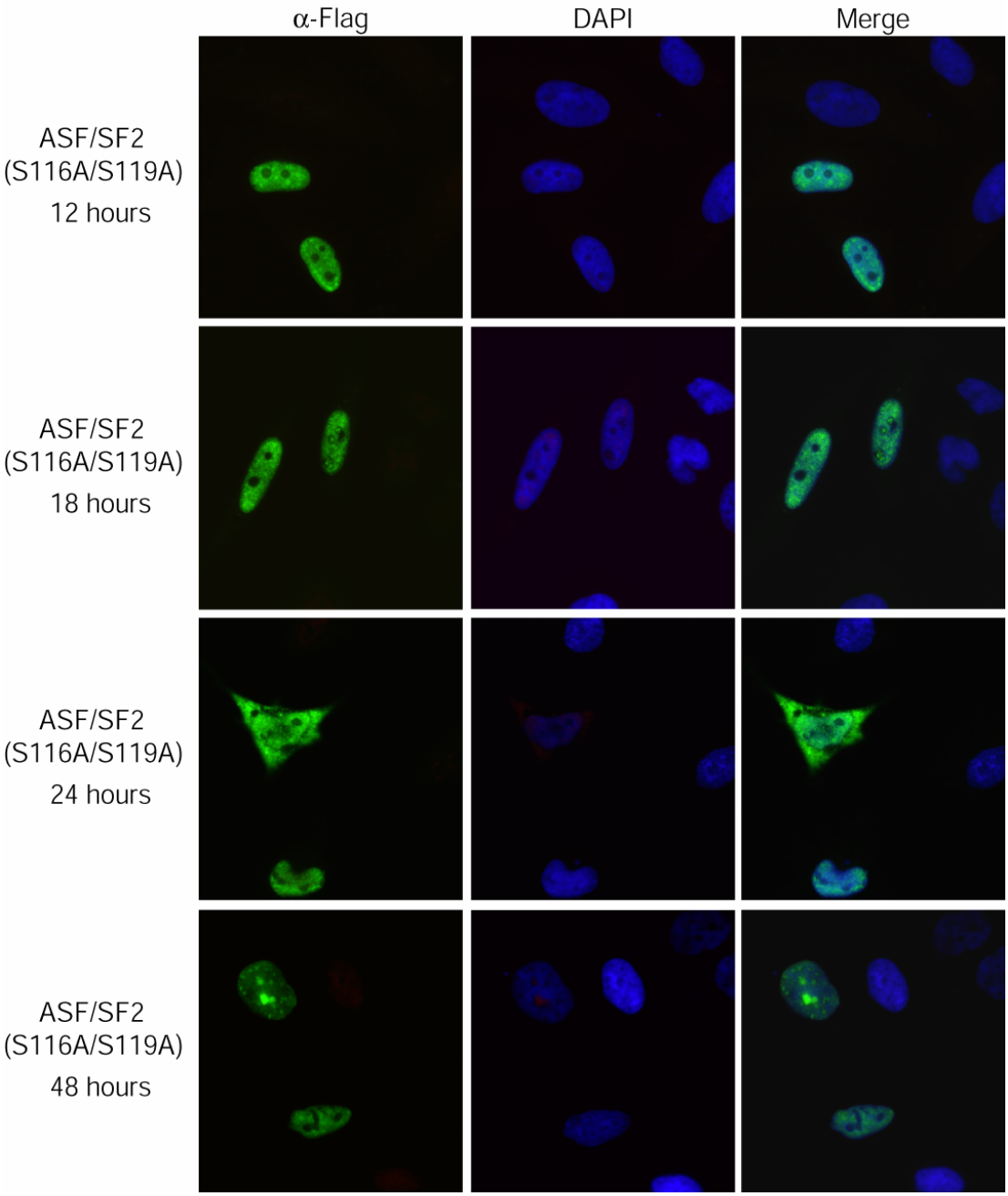


Figure 4.5. ASF/SF2(S116A/S119A) shows a time-dependent defect in its subcellular localization. Indirect immunofluorescence assays of HeLa cells transiently expressing Flag-ASF/SF2(S116A/S119A) for the indicated times show a localization defect only 24 hours post-transfection, while at all other time points the protein localizes to nuclear speckles in a manner identical to ASF/SF2(wt).

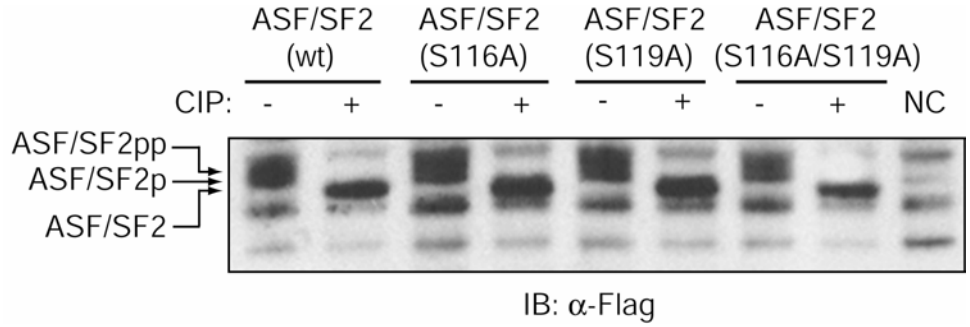


Figure 4.6. Mutation of residues in the inter-RRM linker of ASF/SF2 does not affect phosphorylation of its RS domain. SDS-PAGE analysis of lysates of HeLa cells transiently expressing ASF/SF2- (wt), (S116A), (S119A) and (S116A/S119A). All ASF/SF2 linker mutants are phosphorylated in a manner similar to ASF/SF2(wt).

binding properties of ASF/SF2. Our first step was to identify the minimal construct of ASF/SF2 that is well-behaved in solution and retains its RNA binding property. The RNA binding properties of ASF/SF2(FL) and ASF/SF2(ΔRS) have been previously compared in our laboratory (Figure 4.7). It was observed that ASF/SF2(FL) tends to aggregate in solution, and is often not suitable for in vitro RNA binding assays. Hypophosphorylated ASF/SF2(FL) (ASF/SF2 phosphorylated by SRPK1), shows some RNA binding affinity which is weaker than that exhibited by ASF/SF2(ΔRS). In light of these observations, we used ASF/SF2(ΔRS) for our biochemical studies on RNA binding.

Our preliminary electrophoretic mobility shift assays (EMSA) with GST-ASF/SF2(ΔRS) and a 17-mer RNA aptamer bearing the 5' splice site sequence from the SV40 small t intron (SV40_5'ss) showed that ASF/SF2 binds to 5' splice site RNA with very low affinity (Figure 4.8A). EMSAs performed with GST-ASF/SF2(ΔRS) and an RNA aptamer containing the optimal 5' splice site sequence (opt_5'ss) identified by Zuo *et al* show a similar trend of splice site recognition (Figure 4.8B).

Since ASF/SF2 has been implicated in ESE-dependent splice site recognition, we also tested its ability to bind an ESE containing RNA aptamer. Our first step was to purify large amounts of soluble ASF/SF2(ΔRS) that was free of cellular RNA contamination. We expressed His-ASF/SF2(ΔRS) in *E. coli* BL21*(DE3)* grown in minimal (M9) media and purified it by Ni^{2+}-NTA affinity and size-exclusion chromatography. Size-exclusion chromatography removed all the non-specific RNA associated with the protein (Figure 4.9). We performed EMSAs of ASF/SF2(ΔRS) with a SELEX-derived ESE aptamer (S_ESE) and an aptamer containing the ESE sequence

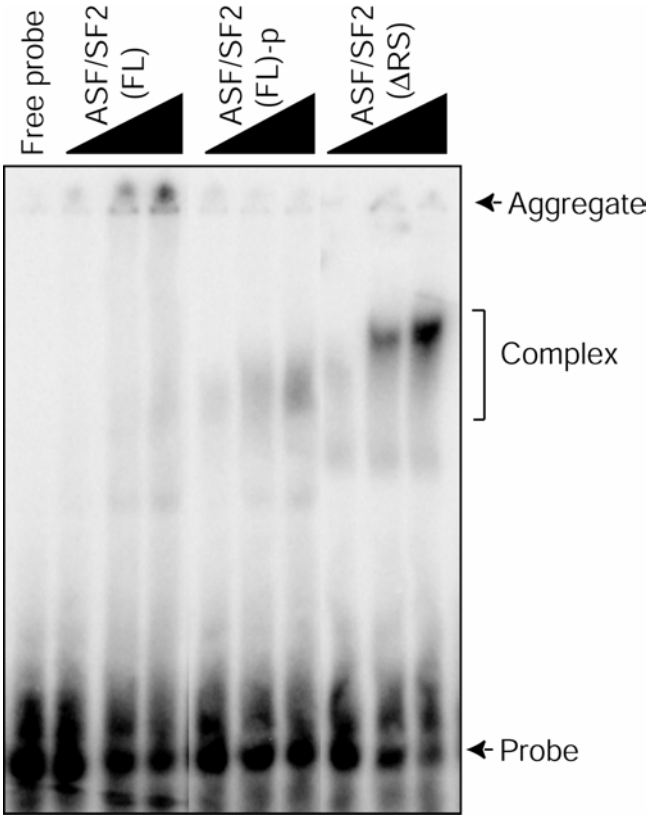


Figure 4.7. RNA binding properties of different forms of ASF/SF2. EMSA of ASF/SF2(FL), ASF/SF2(FL) phosphorylated by SRPK1 (ASF/SF2(FL)-p) and ASF/SF2(ΔRS) with a specific RNA aptamer. ASF/SF2(FL) tends to aggregate in solution while ASF/SF2(FL)-p shows weak binding to the RNA aptamer. ASF/SF2(ΔRS) shows significant binding to the RNA aptamer. (Figure credit: S. Cho)

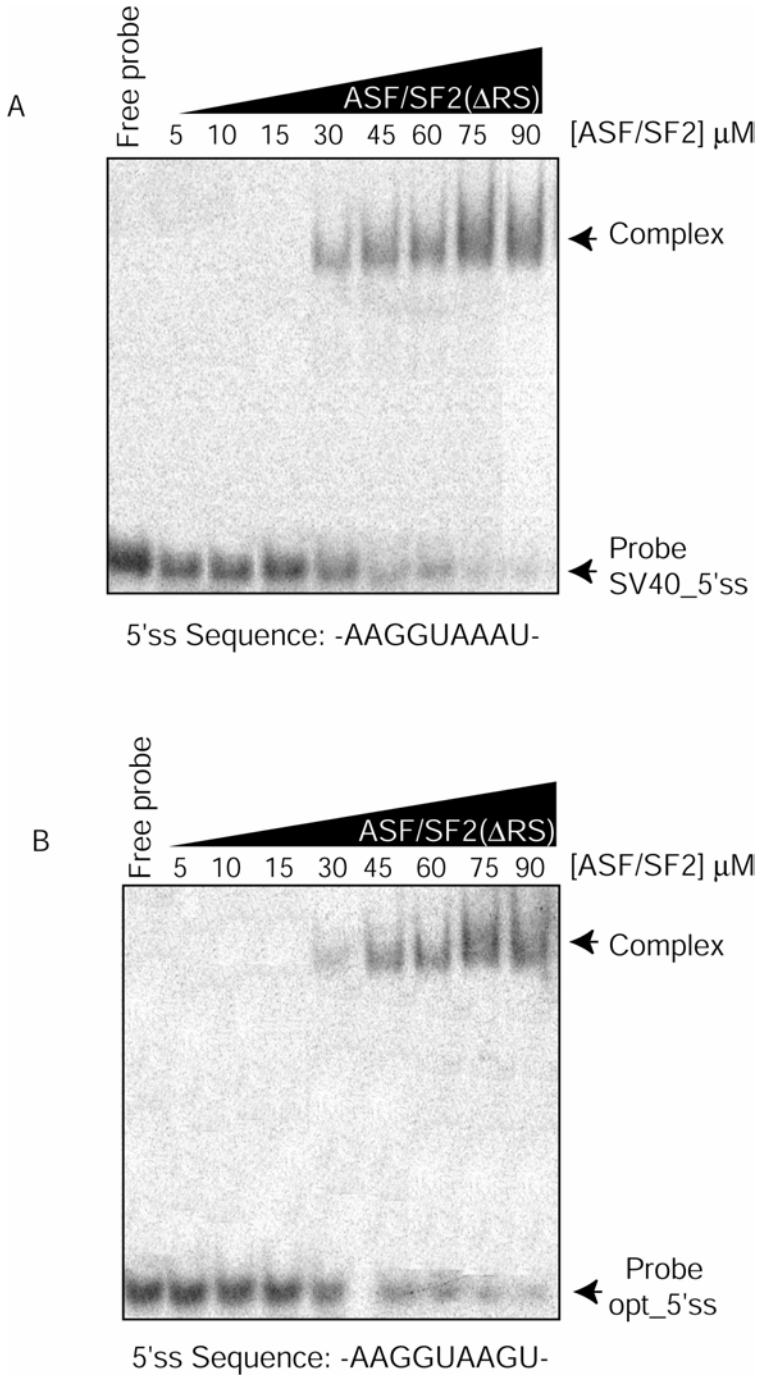


Figure 4.8. ASF/SF2 binds 5' splice site sequences with very low affinity. A. EMSA of ASF/SF2(ΔRS) with an RNA aptamer containing the 5' splice site sequence of the SV40 small t intron. B. EMSA of ASF/SF2(ΔRS) with an RNA aptamer containing an optimal 5' splice site sequence as described by Zuo et al. The RNA binding affinity of ASF/SF2(ΔRS) to both 5' splice site containing aptamers is in the high micro-molar range.

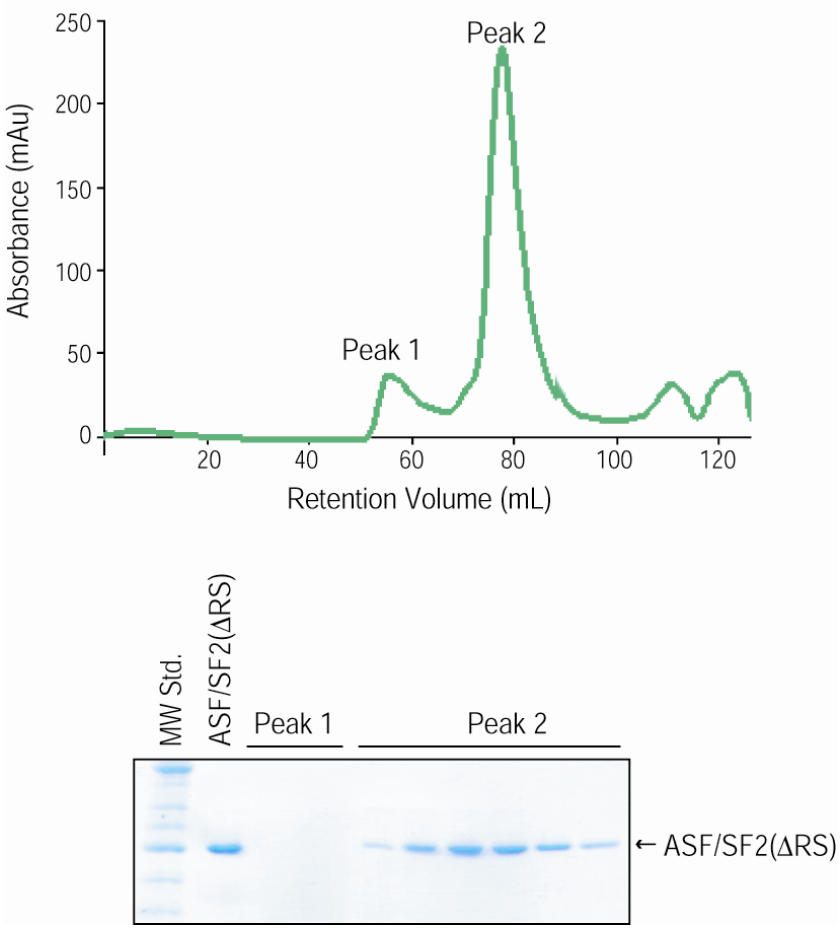


Figure 4.9. Purification of ASF/SF2(ΔRS). Size-exclusion profile and corresponding SDS-PAGE analysis of purified ASF/SF2(ΔRS). Size-exclusion chromatography separates the non-specifically bound cellular RNA (Peak 1, no protein detected on SDS-PAGE) from ASF/SF2(ΔRS) (Peak 2)

from the Ron protooncogene (Ron_ESE) (Figures 4.10A and 4.10B) [120, 121]. We found that although ASF/SF2 binds ESE RNA aptamers with a much higher affinity than 5' splice site sequences, the binding affinity is still relatively moderate (approximately 8μM). However, despite the low affinity interaction, recognition of ESE sequences by ASF/SF2 is fairly specific (Figure 4.10). Since the ASF/SF2:ESE interaction is stronger and more specific that that of ASF/SF2:5' splice site, we used the ESE-containing aptamer (Ron_ESE) to study the RNA binding function of ASF/SF2 in greater detail.

3. Both RRM domains and the linker between the two RRMs are essential for ESE recognition of ASF/SF2

Previous semi-quantitative studies have shown that the individual RRM domains of ASF/SF2 are not sufficient for its optimal RNA binding and splicing activity in vitro [117, 126]. Additionally, it was suggested that the two RRM domains have to be present within the same polypeptide to have significant RNA binding activity. Since the affinity of ASF/SF2(ΔRS) for RNA was far greater than the cumulative affinity of the individual RRM domains, it seemed that an additional recognition element was required for RNA binding by ASF/SF2. In order to analyze the RNA binding function of ASF/SF2 in greater detail, we generated a series of constructs that encompass each individual RRM domain, without and with the adjacent linker region (Figure 4.11A). All these proteins were expressed in *E. coli* BL21(DE3) cells and purified in their soluble form by Ni^{2+}-NTA affinity and size-exclusion chromatography. As expected, EMSAs with ASF/SF2 fragments that contain a single RRM domain but lack the linker, ASF/SF2(R1) and ASF/SF2(R2), showed that neither of the two RRMs bound the ESE aptamer with an

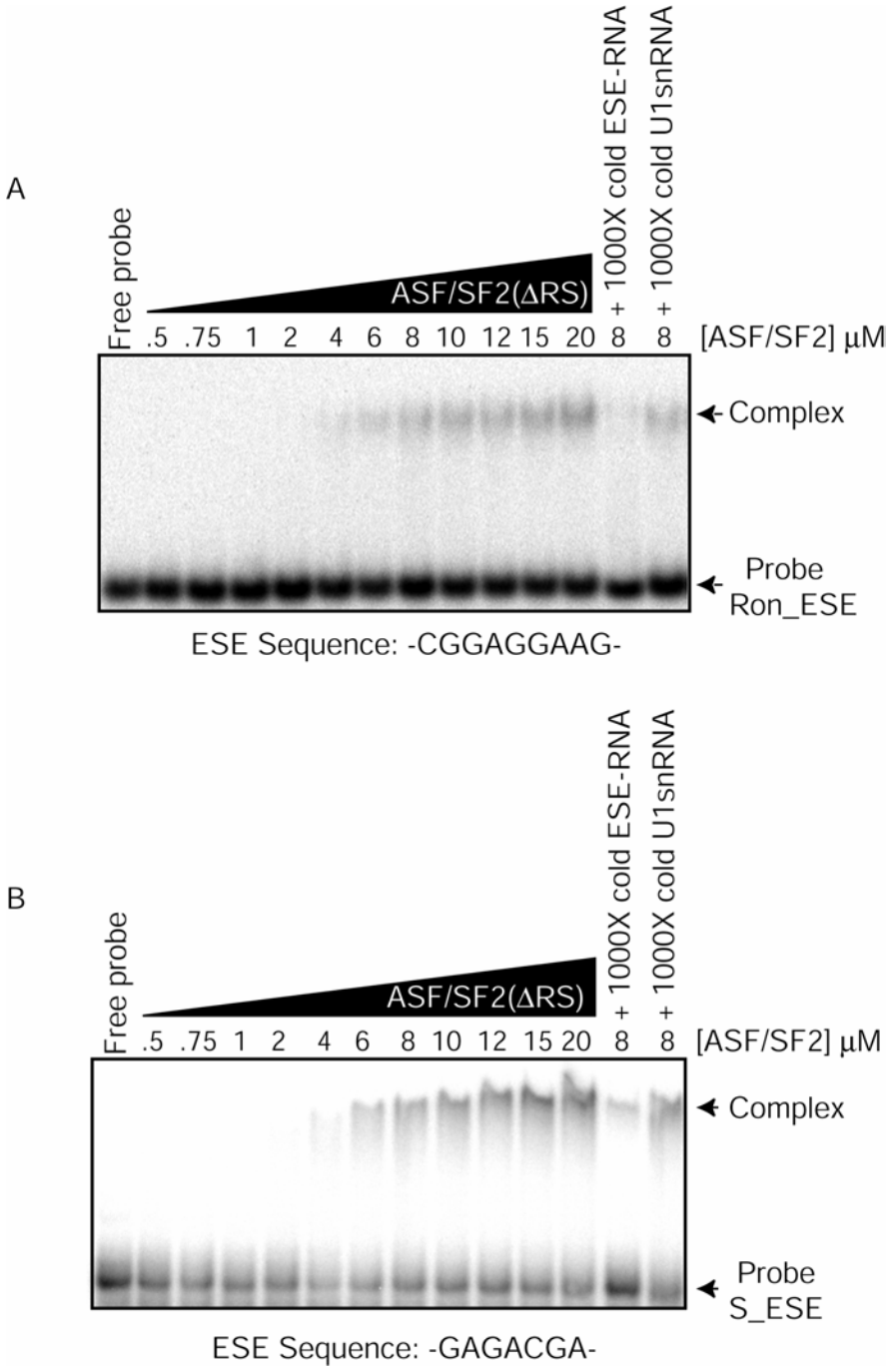


Figure 4.10. ASF/SF2 binds exonic splicing enhancer sequences with moderate affinity and high specificity. A. EMSA of ASF/SF2(ΔRS) with an RNA aptamer containing an ESE from the Ron protooncogene. B. EMSA of ASF/SF2(ΔRS) with an RNA aptamer containing a SELEX-derived ESE. The corresponding unlabeled RNA aptamers served as specific competitors of RNA binding, while a 12-mer RNA aptamer containing part of the U1snRNA was used as a non-specific competitor. The binding affinity of ASF/SF2(ΔRS) to ESE sequences is approximately 8μM (estimated as $B_{max}/2$).

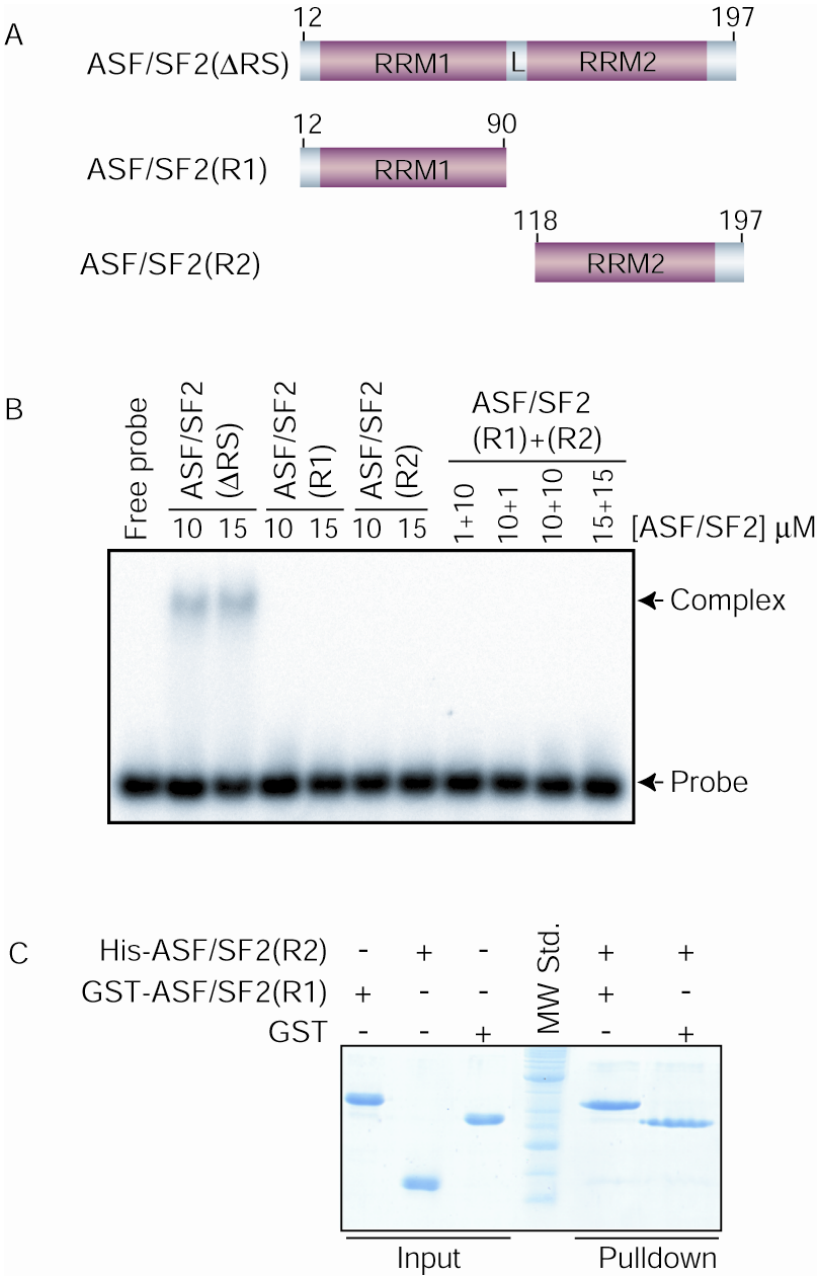


Figure 4.11. The individual RRM domains of ASF/SF2 do not bind ESEs. A. Constructs of ASF/SF2 generated to determine the role of each RRM domain in ESE recognition. B. EMSA of the Ron_ESE aptamer with ASF/SF2-(ΔRS), (R1) and (R2). ASF/SF2-(R1) and (R2) do not bind RNA independently or when mixed together in solution. The concentrations of ASF/SF2-(R1) and (R2) used in each mixture are indicated. C. GST pulldown assay of GST-ASF/SF2(R1) with ASF/SF2(R2). ASF/SF2-(R1) and (R2) do not interact with each other in solution.

affinity comparable to ASF/SF2(ΔRS) (Figure 4.11B). Furthermore, RNA binding by ASF/SF2 could not be recapitulated by simply mixing together ASF/SF2(R1) and ASF/SF2(R2) in an equi-molar ratio in presence of the ESE aptamer. We wanted to eliminate the possibility that, when mixed in high concentrations, the two RRMs could interact with each other to occlude their RNA binding surfaces and abolish an additive binding event. To this end, we performed GST-pulldown assays with ASF/SF2(R2) and GST-ASF/SF2(R1) to verify that there was no interaction between the two RRM domains in vitro (Figure 4.11C). Together, our results indicate that either both RRM domains have to be tethered together by a linker to present the right conformation for RNA binding (suggested by Shamoo et al. [127]), or the linker region itself contributes significantly to the RNA binding affinity of ASF/SF2.

To analyze the role of the linker domain in ESE recognition, we performed EMSAs with ASF/SF2 fragments that contain the linker in addition to individual RRM domains (ASF/SF2- R1L and LR2) as above (Figure 4.12). We observe that although ASF/SF2(R1L) does not bind the ESE RNA, ASF/SF2(LR2) shows some weak binding which is enhanced upon addition of an equimolar amount of ASF/SF2(R1) (Figure 4.12B). This indicates that the linker region, along with the two RRMs, contributes to the RNA binding affinity of ASF/SF2 and does not merely function as a tether for the two RRM domains. Overall, our results clearly show that apart from the RS domain, every other region of ASF/SF2 is essential for its ESE recognition.

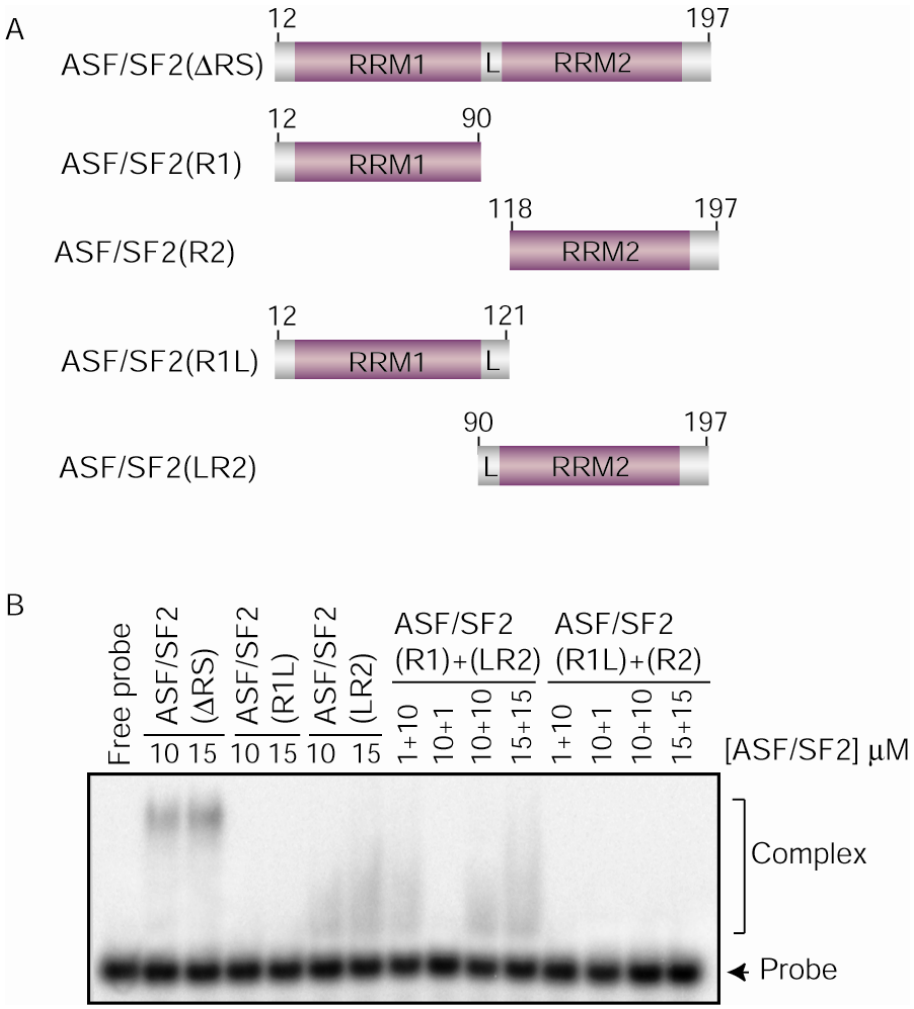


Figure 4.12. The inter-RRM linker of ASF/SF2 enhances ESE recognition. A. Constructs of ASF/SF2 generated to delineate the region of the protein responsible for recognition of ESE-RNA. B. EMSA of the Ron_ESE aptamer with ASF/SF2-(ΔRS), (R1L) and (LR2). ASF/SF2-(R1L) does not bind RNA independently or when mixed together with ASF/SF2(R2) in solution. ASF/SF2(LR2) shows weak binding with the ESE-RNA, which is enhanced upon addition of ASF/SF2(R1). Proteins were mixed together such that all 3 components of the composite RNA binding platform of ASF/SF2 (RRM1, linker and RRM2) are present in solution. Concentrations of individual components in the mixture are indicated.

4. Specific residues in the linker region contribute to the RNA binding affinity and specificity of ASF/SF2

Our studies on ASF/SF2 show that residues S116 and S119 located in the linker between the two RRM domains play a role in regulation of ASF/SF2, independent of RS domain phosphorylation. The role of the linker region in ESE recognition by ASF/SF2 was also established above. Next, we wanted to test specifically if residues S116 and S119 are predominantly responsible for the RNA binding affinity of the linker region. It is possible that these residues play a functional role in ESE recognition and their absence translates as a small defect in localization of ASF/SF2(S116A/S119A) in vivo. To test this hypothesis we compared the RNA binding properties of ASF/SF2(ΔRS) and ASF/SF2(ΔRS S116A/S119A). EMSAs performed with both proteins and the ESE aptamer showed that ASF/SF2(ΔRS S116A/S119A) is significantly impaired in its ability to bind RNA (Figure 4.13). Moreover, RNA recognition by the mutant protein is fairly non-specific since the ASF/SF2:ESE complex can be disrupted by addition of excess non-specific RNA. These observations indicate that residues S116 and S119 present in the linker region govern the RNA binding affinity and specificity of ASF/SF2 to a considerable extent.

C. Discussion

The studies described above bring forth several interesting features in the recognition of RNA sequence elements by ASF/SF2. We established that ASF/SF2 binds ESE-containing RNA aptamers with relatively low affinity but high specificity. Low affinity binding of specific RNA aptamers by RNA binding proteins is not

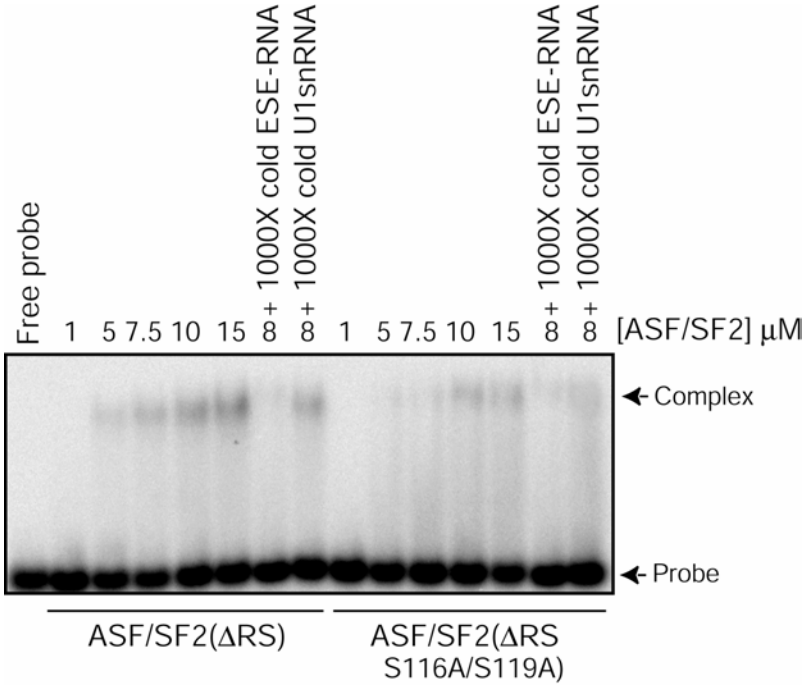


Figure 4.13. Specific residues within the inter-RRM linker of ASF/SF2 contribute to its affinity and specificity of ESE-RNA recognition. EMSA of ASF/SF(ΔRS) and ASF/SF2(ΔRS S116A/S119A) with the Ron_ESE aptamer. Unlabeled Ron_ESE RNA served as a specific competitor of RNA binding, while a 12-mer U1 snRNA aptamer was used as a non-specific competitor The affinity and specificity of ASF/SF2(ΔRS S116A/S119A) is drastically reduced in comparison to ASF/SF(ΔRS), as concluded from its weak binding to ESE-RNA and dissociation of the ASF/SF2:ESE complex by the non-specific competitor.

unprecedented. The SR protein, SRp20, recognizes a 4-nucleotide specific ESE with an affinity of 18μM [128]. In addition to SR proteins, another class of auxiliary splicing factors called hnRNPs bind their specific RNA sequences (exonic splicing silencers) with low affinity [129]. Since spliceosome assembly and catalysis is a dynamic process, it is likely that SR proteins and hnRNPs undergo multiple cycles of binding and release of RNA, which is facilitated by a low affinity interaction. However, increase of local concentration of SR proteins at the splice sites, presence of interacting partners and post-translational modifications of SR proteins can serve to stabilize their interaction with specific RNA sequence elements.

Furthermore, our studies have identified a novel regulatory element for the ESE-RNA recognition function of ASF/SF2. We show that the linker between the two RRM domains plays a significant role in affinity and specificity of RNA recognition by ASF/SF2 and also identify specific residues (S116 and S119) therein that are responsible for this function. The path leading up to this identification was somewhat serendipitous as these residues were initially thought to play a completely different role (that of priming phosphorylation) in ASF/SF2 regulation, as suggested by the crystal structure of the SRPK1:ASF/SF2 complex. Our cell-based assays conclusively show that residues within the linker region do not act as priming sites for phosphorylation of the RS domain of ASF/SF2. Additionally, in vitro kinase assays with wild type ASF/SF2 and ASF/SF2(S116A/S119A) show that there are no differences in phosphorylation of the two proteins (S. Cho, personal communication). Studies with certain multiple-RRM containing proteins such as poly-A binding protein (PABP), Drosophila Sex-lethal protein (Sxl) and the U2 auxiliary factor, U2AF65, have established a requirement for at

least two RRM domains in RNA binding [130-132]. However, the linker between the two RRM domains did not play a functional role in any of the proteins mentioned above. RNA recognition by ASF/SF2 is similar to these proteins in that both its RRM domains are essential for RNA binding but different since its linker is also involved in this process. The affinity and specificity in RNA binding conferred by residues S116 and S119 within the linker is remarkable, suggesting that they may mediate direct contacts with RNA. Phosphorylation of these serines by a nuclear kinase may diminish RNA binding of ASF/SF2 and thus serve as another mode of regulation of this SR protein. Further biochemical studies are required to determine if this pattern of RNA recognition is common to all SR proteins containing two RRM domains.

Genetic complementation studies with ASF/SF2 show that the RS domain of ASF/SF2 is dispensable for cell survival in mouse embryonic fibroblasts (MEFs) [133]. However, mutation of key RNA binding residues within the RRM1 domain (F56 and F58) abolishes the ability of ASF/SF2 to complement cell survival, indicating that the critical function of ASF/SF2 relies on its ability to bind RNA. The experimental setup for the genetic complementation assay developed by Lin et al. is outlined in Figure 4.14. We propose to use this complementation assay to determine the functional significance of the residues S116 and S119 in vivo. If these residues are indeed critical for RNA binding of ASF/SF2, then mutation to alanines would abolish the ability of the resultant protein to complement cell survival. This assay would confirm the role of the linker region of ASF/SF2 in RNA recognition in vivo.

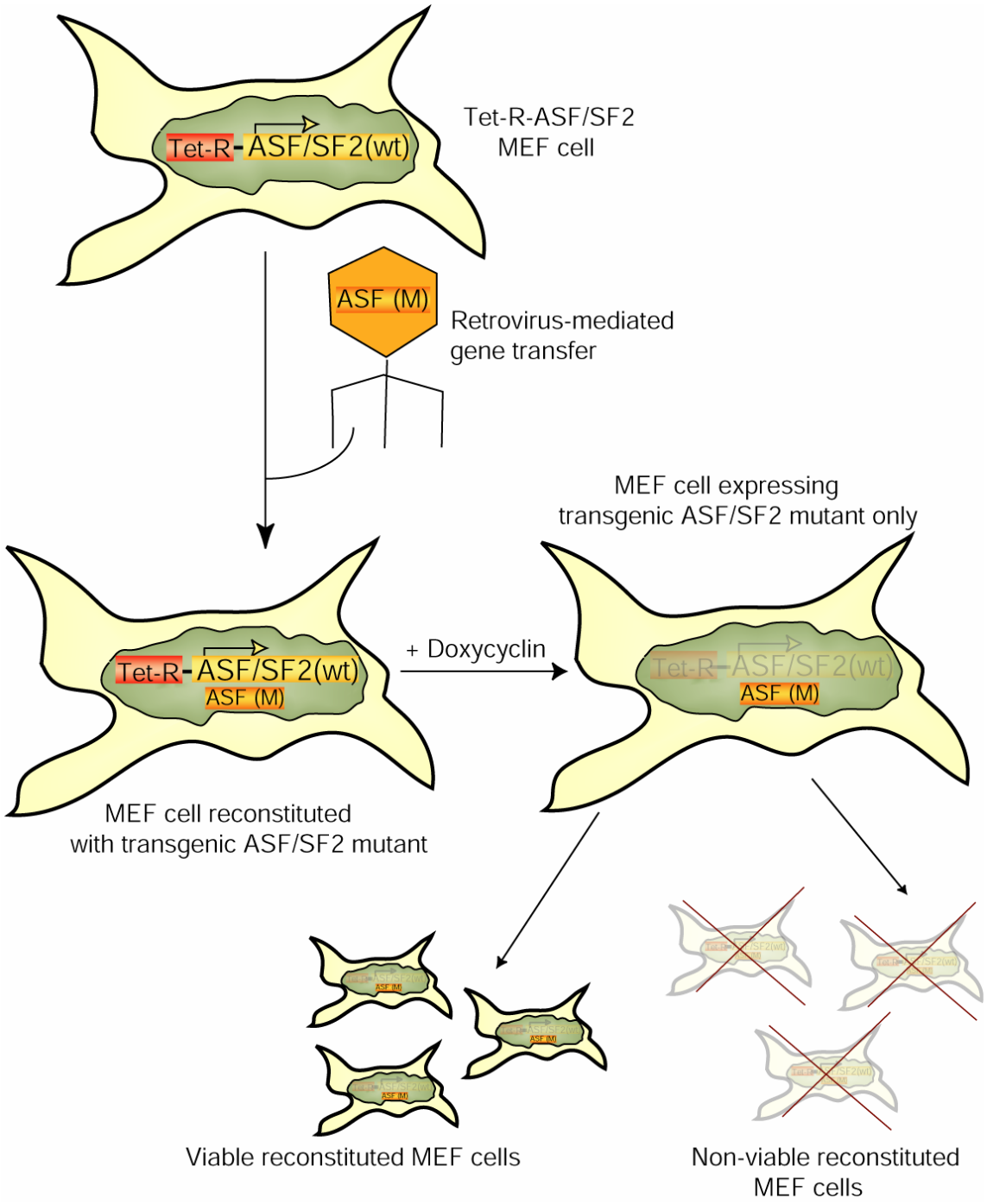


Figure 4.14. Genetic complementation assay to test the effect of ASF/SF2 mutants on cell viability. The desired ASF/SF2 mutant is transferred into mouse embryonic fibroblast (MEF) cells, which contain ASF/SF2(wt) under a tetracycline repressor, using the retrovirus mediated gene transfer system. Expression of ASF/SF2(wt) can be shut off by addition of a tetracycline analog (doxycyclin), thus allowing one to assay for the ability of the mutant ASF/SF2 to complement cell survival. This genetic complementation assay was developed by Lin et al [133].

V: Expression, Purification and Crystallization of an Active Fragment of the Yeast Kinase-Endoribonuclease, Ire1p

A. Introduction

Cells respond to a variety of external and internal stimuli by modifying their pattern of gene expression. This is directly mediated by several inducible factors which are activated in response to stimuli to induce or repress their target genes. There exist several other modes of regulation of gene expression at the level of mRNA processing and protein production. Pre-mRNA splicing, mRNA export and translation, protein folding and post-translational modifications are a few of the check-points in the process of gene expression which occur downstream of transcription. The endoplasmic reticulum (ER) is a cellular compartment responsible for the structural maturation of proteins entering the secretory pathway [134]. Proteins must be correctly folded and glycosylated before they can transit from the ER to the Golgi and other ER-distal compartments [135, 136]. A number of cellular proteins such as immunoglobulin binding protein (BiP), protein disulfide isomerase, and peptidyl-prolyl-cis-trans-isomerase are localized to the ER lumen and act as chaperones to promote proper folding and prevent aggregation of folding intermediates of proteins [137]. Accumulation of misfolded and aberrantly modified proteins in the ER lumen (a condition referred to as ER stress) leads to activation of a signal transduction pathway called the Unfolded Protein Response (UPR). UPR is predominantly a transcriptional program designed to upregulate genes involved in protein folding and alleviating ER stress [138, 139]. Experimentally, UPR can be induced by treatment of cells with drugs such as tunicamycin (blocks N-linked glycosylation of proteins), dithiothreitol (reduces disulfide bonds in proteins) and thapsigargin (depletes ER calcium stores). In yeast, UPR targets include genes encoding chaperones, oxido-reductases, ER-associated degradation components and proteins

functioning downstream of the secretory pathway [140]. These target genes contain a cis-acting promoter element (~22 base pairs) called the unfolded protein responsive element (UPRE) that is necessary and often sufficient to mediate a response to accumulation of unfolded proteins in the ER lumen [141, 142]. Together, the UPR induced gene products act to dispose or salvage the misfolded polypeptides in the ER lumen and increase the protein folding capacity of the ER [143].

In yeast, UPR is mediated by a single sensor, Ire1p. *IRE1* was originally identified as a gene required for inositol auxotrophy and later shown to be responsible for tunicamycin-dependent induction of the *KAR2* gene (encoding BiP) in yeast, thus linking it to the UPR pathway [144-146]. Although dispensable under normal conditions, *IRE1* was found to be essential for cell viability during ER stress [145, 146]. Ire1p is a single-span transmembrane protein with an N-terminal ER luminal domain, a kinase domain and a C-terminal endoribonuclease domain [2, 145, 146] (Figure 5.1A). The ER luminal domain of Ire1p is responsible for sensing the accumulation of unfolded proteins in the ER and transducing the signal to the cytosolic effector domains. How these unfolded proteins are recognized by Ire1p has been a subject of debate. Previous studies proposed a model where Ire1p is maintained in a monomeric inactive state by binding to BiP [147]. Upon accumulation of unfolded proteins in the ER lumen, BiP is titrated away from Ire1p, leading to dimerization and activation of Ire1p [148]. This model of Ire1p activation portrays BiP as the primary sensor of ER stress and has been questioned in light of a more recent study on the BiP/Ire1p complex and the crystal structure of the luminal domain of Ire1p [149, 150]. The report by Kimata et al. suggests that BiP is not an ER stress sensor, but instead functions to limit Ire1p activation to respond only to ER

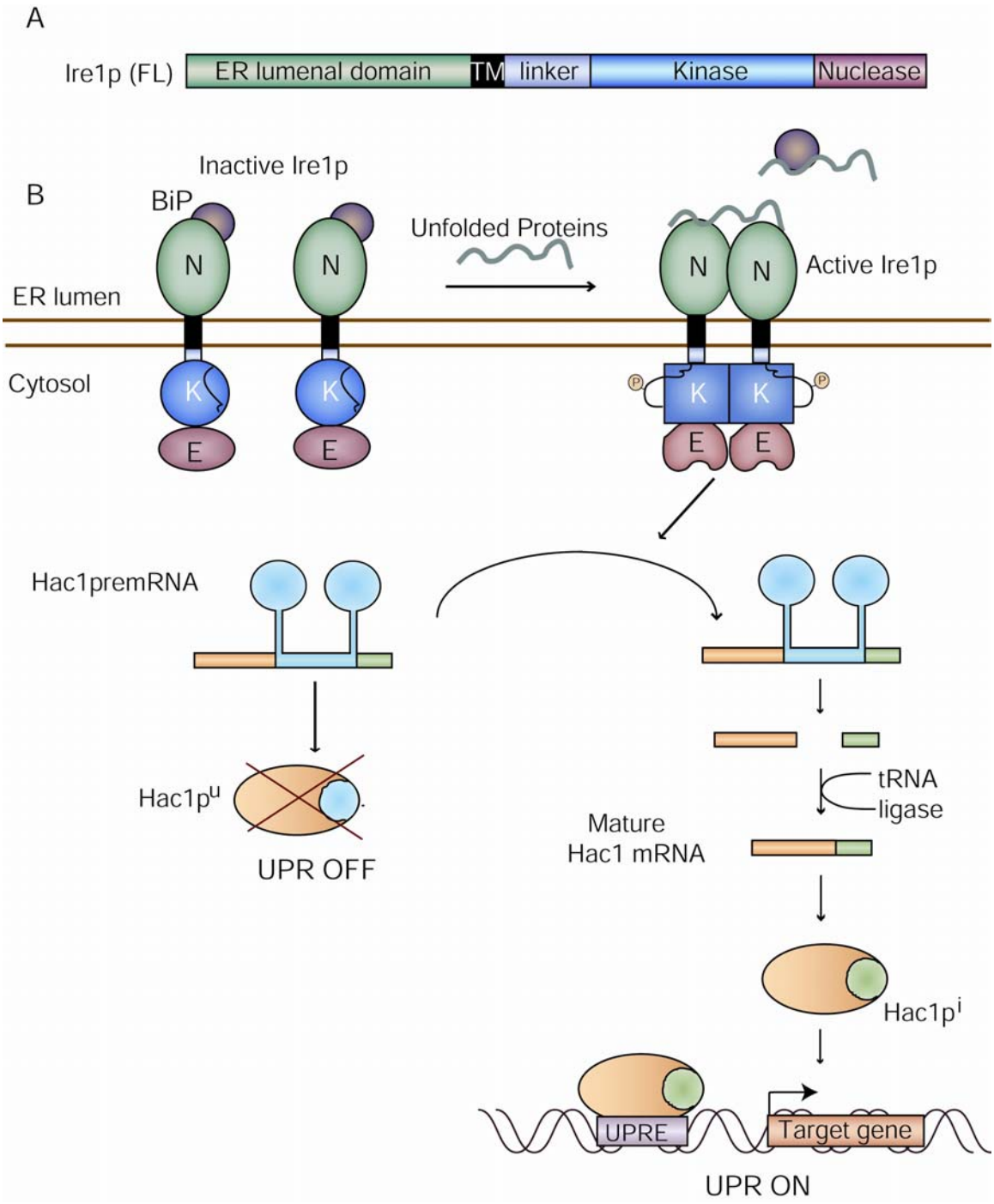


Figure 5.1. The Unfolded Protein Response. A. Domain organization of the yeast UPR senor, Ire1p. B. Schematic representation of the activation of Ire1p and downstream signaling events upon UPR induction.

stress. The structure suggests that the luminal domain of Ire1p is capable of sensing ER stress by directly binding unfolded proteins/peptides which leads to the formation of higher-order oligomers and Ire1p activation. Oligomerization of the N-terminal luminal domain mediates the same effect in the C-terminal cytosolic domains of Ire1p [151].

The cytosolic region of Ire1p comprises of a kinase domain, which is followed by an endoribonuclease domain. A linker of variable length and sequence connects the transmembrane region to the kinase domain. Ire1p is unusual in that it is the only membrane-bound serine-threonine kinase in yeast and undergoes dimerization and trans-autophosphorylation in a manner reminiscent of receptor tyrosine kinases [151, 152]. Sensing of unfolded proteins by the ER luminal domain of Ire1p leads to dimerization, autophosphorylation and activation of its kinase domain, which in turn activates its endoribonuclease domain [2]. The mechanism of signal transduction from the kinase to the nuclease domain of Ire1p is not clearly known, but is thought to be caused by a conformational change in the kinase domain upon activation and nucleotide (preferably ADP) binding [153]. A similar mode of dimerization dependent activation of the nuclease domain was found in RNase L, a protein closely related to Ire1, which functions in antiviral and antiproliferative mechanisms of interferon action [154-156]. RNase L is an unusual ribonuclease in that it has a naturally inactive kinase-like domain and binds 2'-5'oligoadenylate to stimulate its nuclease activity [157].

The only known substrate, to date, of the nuclease domain of Ire1p is the mRNA encoding the transcription factor Hac1p [158, 159]. Hac1p is a basic leucine zipper (b-ZIP) transcription factor responsible for upregulation of UPR target genes [158, 160]. The uninduced form of the *HAC1* mRNA, *HAC1*^{u} contains a non-classical intron at its 3'

end which interacts with the 5' untranslated region of the mRNA to block its own translation [161]. Upon UPR induction, the intron is removed by a 2-site cleavage reaction mediated by Ire1p. The two cleaved mRNA fragments are ligated by the tRNA ligase Rlg1p to generate a mature form of the *HAC1* mRNA, *HAC1*^{i} (induced) [1]. The chemistry of the overall splicing reaction bears more similarity to tRNA splicing than conventional spliceosome-mediated mRNA splicing [162]. The *HAC1*^{i} mRNA produced by this reaction is then translated to produce the functional form of Hac1p which translocates to the nucleus and activates transcription by binding to the UPRE sequence in the promoters of target genes. Figure 5.1B shows a schematic representation of the UPR pathway in yeast.

UPR in higher eukaryotes is more complex owing to the presence of three sensor proteins, each of which acts in a slightly different manner [163]. One of the sensors, IRE1, is conserved in all eukaryotes [164, 165]. The mammalian IRE1 family includes two transmembrane protein kinase-nucleases; IRE1α which is expressed ubiquitously in all cell types and IRE1β whose expression is largely limited to the intestinal epithelia [164]. Mammalian Ire1 cleaves the mRNA encoding the transcription factor XBP1 to generate an active form of the protein which upregulates UPR target genes [166, 167]. Additionally, IRE1 has been shown to cleave its own mRNA, 28*S* rRNAs and mRNAs encoding plasma membrane and other secreted proteins, thus adding another level of regulation and response to accumulation of unfolded proteins [168-170]. The other two UPR transducers in higher eukaryotes are the PKR-like ER kinase (PERK) and the membrane bound transcription factor ATF6. Both PERK and ATF6 contain ER luminal domains to sense the accumulation of unfolded proteins [171, 172]. In response to ER

stress, PERK acts by phosphorylating the alpha subunit of the eukaryotic initiation factor (eIF2α) to repress general translation while ATF6 is cleaved by specific proteases to release its cytosolic domain which then translocates to the nucleus to activate transcription [171-174].

The yeast Ire1p-mediated regulation of ER stress is a simple paradigm for UPR in eukaryotes and thus makes for a model study system. Although much work has been done at a molecular level on the sensing mechanism of the luminal domain of Ire1p, relatively little information is available about the transduction of signal from the ER sensors to the cytosolic effectors. Oligomerization, trans-autophosphorylation and nucleotide binding of Ire1p have been shown to be important in activating its nuclease domain, but the underlying mechanisms have not been clearly established.

The focus of this study was to determine the X-ray crystal structure of a fragment of Ire1p encompassing its kinase and nuclease domains. We wished to determine the mechanism of activation of the kinase and nuclease domains of Ire1p and understand how these two domains functionally interact with each other in order to orchestrate UPR, using X-ray crystallography as a tool. Additionally, the structure of Ire1p would be interesting from the standpoint of a protein kinase as it is the only serine-threonine kinase to be regulated by dimerization and trans-autophosphorylation and would also contribute to our increasing repertoire of protein kinase structures. Determining the X-ray crystal structure of a protein can be a challenging process and obtaining well-diffracting crystals is often the major obstacle. Crystallization can be deterred due to presence of impurities in the protein preparation, heterogeneity in post-translational modification or presence of flexible regions within the protein that interfere with crystal packing. Several measures

can be adopted to overcome these problems and facilitate crystallization of the target protein. In this chapter, we report the expression, purification and crystallization of a construct of Ire1p and describe our attempts to obtain well-diffracting crystals of the same.

B. Materials and Methods:

1. Preparation of recombinant DNA

The plasmid pGEX6P1 Ire1p(556-1115) harboring the cDNA of the full-length cytoplasmic domain of Ire1p was provided by Dr. Maho Niwa (Division of Biology, University of California, SanDiego).

pET15b Ire1p(646-1115), Ire1p(642-1115), Ire1p(606-1115), Ire1p(585-1115), Ire1p (570-1115), Ire1p(646-1077) and Ire1p (656-1001): The inserts for these expression plasmid were generated by PCR amplification of the relevant cDNA segment from pGEX6P1 Ire1p(556-1115). The 5' primers for all these constructs carried an Nde1 restriction site.

Ire1(646)NTNdeI: 5'GGAATTCCATATGAAAAAGGGCCGAAAATC3'

Ire1(642)NTNdeI: 5'GGAATTCCATATCAACAACAACACCAAAAGAGG3'

Ire1(606)NTNdeI: 5'GGAATTCCATATGAAGCAAGTTGTTTTTG3'

Ire1(585)NTNdeI: 5'GGAATTCCAATATGAAATCGCTAAATTGTCC3'

Ire1(570)NTNdeI: 5'GGAATTCCATATGAAAATTGGATTTATGC3'

Ire1(656)NTNdeI: 5'GGAATTCCATATGCGAAAATCACGCATTGC3'

The 3' primers for all the constructs had a BamHI restriction site. A common 3' primer was used for the first six constructs:

Ire(1115)CTBamHI: 5'CGGGGATCCTTATGAATACAAAAATTCAC3'

The following 3' primers were for the constructs Ire1p(646-1077) and Ire1p(656-1001):

Ire1(1001)CTBamHI: 5'CGGGGATCCTTATCTGTTTTCAATTTC3'

Ire1(1077)CTBamHI: 5'CGCGGATCCCTAATCGGGTACCGGCCC3'

The resulting DNA fragments were gel purified, double digested with the restriction enzymes NdeI and BamHI (New England Biolabs) and then gel purified again. These fragment was then ligated into the vector pET15b (Novagen) which was previously double digested by the same restriction enzymes. The ligated products were transformed into *E. coli* JM109 cells and colonies thus obtained were screened for the presence of Ire1p cDNAs by restriction analyses and sequencing.

pTYB1 Ire1p(606-1115): The insert for this plasmid was generated by PCR amplification using pet15bIre1p(606-1115) as a template and the primers Ire1(606)NTNdeI and Ire1(1115)CTXhoI, the sequence of which is given below:

Ire1(1115)CTXhoI: 5'CCGCTCGAGTGAATACAAAAATTCACG 3'

Following PCR amplification, the subcloning procedure was identical to that above, except that the enzymes used for double digestion were HindIII and XhoI and the vector used was pTYB1 (New England Biolabs).

2. Mutagenesis of Ire1p

Mutagenesis of Ire1p was carried out according to manufacturer's protocol for the Stratagene Quikchange Mutagenesis Kit. Presence of the mutation was verified in every clone by sequencing.

pET15b Ire1p(642-1115 D828A): pET15b Ire1p(642-1115) was used as a template for the PCR reaction and the primers used for mutagenesis were a gift from Dr. Maho Niwa. pET15b Ire1p(606-1115 D797N) and pET15b Ire1p(606-1115 S840D/S841D/T843D): pET15b Ire1p(606-1115) was used as a template for the PCR reaction and the primers used for mutagenesis in both cases were a gift from Dr. Maho Niwa.

3. Expression and purification of Ire1p

GST-Ire1p(556-1115): *E. coli* BL21(DE3)pLysS cells transformed with the plasmid pGEX6P2 Ire1p(556-1115) were grown in 2L LB media (containing 100μg/mL Ampicillin) at 37°C to an OD_{600} of 0.2 and induced with 0.1mM IPTG for 7 hours at 25°C. The cells were harvested by centrifugation and frozen at -80°C for 30 minutes to enable efficient lysis. The thawed cell pellet was then resuspended in 120mL of lysis buffer containing 20mM Tris pH 7.5, 50mM NaCl, 10% glycerol, 1mM PMSF and 0.1X Protease Inhibitor Cocktail (Sigma-Aldrich). The cells were lysed by sonication and clarified by centrifugation at high speed (13,500 – 14,000 r.p.m) for 45 minutes. The resultant supernatant was loaded onto a gravity Q-sepharose Fast Flow column at 4°C. The column was washed with 100mL lysis buffer. Presence of the protein in the flow through and wash was determined by SDS-PAGE analysis. The flow through and wash of the Q-column were pooled, filtered through a 0.8μ syringe filter, the concentration of NaCl was increased to 500mM, and then loaded onto a pre-equilibrated GST column (GE Healthcare). The GST column was washed with GST binding buffer (20mM Tris pH 7.5, 500mM NaCl, 10% glycerol and 1mM DTT). The protein was then eluted off the column by GST elution buffer (GST binding buffer supplemented with 10mM reduced

glutathione). The purity of the elutions was checked by SDS-PAGE analysis. The cleaner elutions were pooled, concentrated using Centriprep-30 and flash frozen.

His-Ire1p(646-1115): *E. coli* BL21(DE3) cells bearing the plasmid pET15b Ire1p(646-1115) were grown in 2L LB media (containing 100μg/mL Ampicillin) at 37°C to an OD_{600} of 0.8 and induced with 0.1mM IPTG for 16 hours at 25°C. The cells were harvested as above and lysed in lysis buffer containing 20mM Tris pH 7.5, 250mM NaCl, 5mM imidazole, 10% glycerol, 1mM PMSF and 0.1X PIC. The cell lysate was clarified by centrifugation (as above) and filtration through a 0.8μ syringe filter and was loaded onto a Ni^{2+}-NTA Agarose column at 4°C. The column was washed with 5mM and 40mM imidazole buffers and the protein was eluted in 200mM imidazole buffer (all buffers used for Ni^{2+} affinity chromatography contained 20mM Tris pH 7.5, 250mM NaCl, 10% glycerol and 1mM PMSF in addition to the indicated amounts of imidazole). The eluted protein was assayed for purity by SDS-PAGE and dialyzed against a total of 3L of buffer containing 20mM Tris pH 7.5, 200mM NaCl, 5% glycerol, 1mM DTT for 3 hours. An additional dialysis of 30 minutes against 1L of low salt dialysis buffer (20mM Tris pH 7.5, 100mM NaCl, 5% glycerol and 1mM DTT) was carried out to reduce the concentration of NaCl in the protein even further. After dialysis, the protein was diluted with an equal volume of dilution buffer (20mM Tris pH 7.5, 5% glycerol, 1mM DTT) and filtered through a 0.22μ syringe filter. The filtered protein was then loaded onto a HiTrap Q Sepharose HP column (GE Healthcare) at 25°C. The column was washed with 10mL low salt buffer (20mM Tris pH 7.5, 50mM NaCl, 5% glycerol, 1mM DTT) and the protein was eluted off the column in a 50mM – 750mM NaCl gradient (other buffer

components same as low salt buffer). Purity of the peak fractions was assayed by SDS-PAGE analysis. The peak fractions were then pooled, concentrated and loaded onto a 16/60 Superdex 200 size-exclusion column, pre-equilibrated with buffer containing 20mM Tris pH 7.5, 100mM NaCl, 5% glycerol and 1mM DTT, for a final purification step. The peak fractions were pooled, concentrated to approximately 10mg/mL and flash frozen in small aliquots. A typical yield for this protein preparation is 4mg/1L of LB culture.

His-Ire1p(646-1077): Purification of this protein was performed in a manner similar to that of His-Ire1p(646-1115), but the purity and yield of the preparation was inferior. The protein was concentrated to 1.8mg/mL and the yield was 1mg/1L of LB culture.

His-Ire1p(556-1115): Purification of this protein was attempted using the protocol for His-Ire1p(606-1115), but aborted after the first chromatographic step due to insufficient yield of the protein.

His-Ire1p(606-1115), His-Ire1p(585-1115), His-Ire1p(570-1115), His-Ire1p(656-1001), and His-Ire1p(606-1115 S840D/S841D/T843D): These proteins were purified in a manner identical to His-Ire1p(646-1115). The poly-His tag of His-Ire1p(606-1115 S840D/S841D/T843D) was removed in one preparation by incubation of the protein with thrombin (15cleavage units of Thrombin (Sigma)/1mg of purified protein) for 2 hours at 25°C prior to size-exclusion chromatography. Size-exclusion chromatography was then

performed as before. The purity and yield of the proteins varied marginally with each preparation.

His-Ire1p(642-1115 D828A) and His-Ire1p(606-1115 D797N): Purification of these proteins was essentially identical to that of His-Ire1p(606-1115) with variations in NaCl concentration in the different buffers. The lysis buffers for these preparations have 500mM NaCl instead of 250mM NaCl while the size-exclusion buffer had 150mM NaCl instead of 100mM NaCl. The final concentrations of the proteins were 2.0 - 3.0mg/mL and the yield per 1L of culture was 1.5 - 2mg.

Ire1p(606-1115): Ire1p(606-1115) was purified using the IMPACT-CN (Intein Mediated Purification with an Affinity Chitin-binding Tag) system (New England Biolabs) according to manufacturer's protocol with some variations (buffer composition and additional chromatography steps). Briefly, *E. coli* BL21(DE3) cells transformed with the plasmid pTYB1 Ire1p(606-1115) were grown as for His-Ire1p(646-1115). The cells were harvested and lysed in 120mL of lysis buffer containing 20mM Tris pH 7.5, 500mM NaCl, 10% glycerol, 0.1mM EDTA, 1mM PMSF and 0.1X PIC. The cell lysate was clarified by centrifugation and filtration and was loaded onto a chitin column (New England Biolabs) at 4°C. The column was washed once with 10 column volumes of lysis buffer and then flushed with 3 column volumes of cleavage buffer (lysis buffer supplemented with 40mM DTT) at 25°C. The chitin resin was then resuspended in cleavage buffer and incubated at 25°C for 16 hours. The protein was finally eluted by running 2 column volumes of lysis buffer through the chitin column. The elutions were

pooled and concentrated and size-exclusion chromatography was performed as described above for His-Ire1p(606-1115). The peak fractions were pooled, concentrated to approximately 9mg/mL and flash frozen in small aliquots. A typical yield for this protein preparation is 1mg/1L of LB culture.

4. Limited proteolysis of Ire1p

To determine a stable core fragment of Ire1p, limited proteolysis reactions were carried out with the protease trypsin followed by mass-spectrometric analysis of the generated fragments. 10μg of His-Ire1p(646-1077) was incubated with 0.1μg of trypsin (Sigma) in presence of 1X reaction buffer (25mM Tris pH 7.5, 50mM NaCl and 1mM DTT) at 25°C. Aliquots of this reaction were removed at regular intervals up to 40 minutes and quenched by boiling (95°C/2 minutes) with SDS-sample loading buffer. The products of the proteolysis reaction were analyzed by SDS-PAGE. For mass-spectrometric analysis, the reaction was scaled up to 5X of that for SDS-PAGE analysis and quenched with 0.1% formic acid. The molecular weight of the stable fragment generated was estimated by MALDI-TOF analyses at The Scripps Research Institute.

5. In vitro kinase assay

2μM of Ire1p was incubated with 200 μM ATP and 5μCi γ-^{32}P-ATP in presence of 1X kinase reaction buffer (25mM Tris pH 7.5, 50mM NaCl, 1mM DTT, 1mM $MgCl_2$ and 1mg/mL BSA) at 25°C. Equal volume aliquots of the reaction were removed at regular intervals up to 30 minutes and quenched by boiling (95°C/2 minutes) with SDS-sample loading buffer. The reactions were analyzed by SDS-PAGE and autoradiography.

6. Crystallization of Ire1p

Micro crystals of His-Ire1p(606-1115) were obtained in the presence of ATP by the micro-batch diffusion method using Hampton Crystal Screen 2 (Hampton Research). The drop contained 1μL of protein at a concentration of 6.0mg/mL, preincubated with 5-fold molar excess of ATP and 10-fold molar excess of $MgCl_2$, and 1μL of screening solution. The crystals were optimized to grow by hanging drop vapor diffusion in a solution of 1M sodium citrate and 2mM DTT. Crystals were eventually obtained with His-Ire1p(585-1115), His-Ire1p(570-1115), His-Ire1p(606-1115 S840D/S841D/T843D), Ire1p(606-1115 S840D/S841D/T843D) and His-Ire1p(606-1115 D797N). Crystals of His-Ire1p(606-1115) and His-Ire1p(585-1115) were also grown in presence of ADP and AMP-PNP as well as in their apo form. Ire1p crystals were also grown in 1.4M sodium malonate and 2mM DTT. A wide variety and concentration of cryo-protectants were used to cryo-protect Ire1p crystals prior to flash freezing in liquid nitrogen for data collection. Some of the cryo-protectants used are: ethylene glycol (20-30%), glycerol (15-30%), PEG 400 (25-30%), sodium malonate (2.0-2.5M) and sucrose (15-25%). The crystals were tested for diffraction at the UCSD X-ray facility.

C. Results

1. Identification of stable, well-behaved constructs of Ire1p

We first attempted to purify the full-length cytoplasmic domain of Ire1p. A GST-fusion construct of Ire1p(556-1115) was obtained from Dr. Maho Niwas's lab and purification was attempted by multiple chromatographic steps. Purification of the full-

length cytoplasmic domain of Ire1p was difficult as this protein was susceptible to proteolytic degradation and aggregation. Also, the bacterial chaperone DnaK was thought to be associated with the protein, further contributing to the impurity of the preparation (Figure 5.2A). The protein was estimated to be only 40% pure and its purity could not be improved by other chromatographic steps. However, this protein was found to be active in an in vitro kinase assay and a nuclease assay (A. C. and M. N). An attempt to purify a His-tagged version of Ire1p(556-1115) was not successful either, due to significant contamination by a 70kDa protein which was presumed to be DnaK (Figure 5.2B).

Since we were unable to purify the full-length cytoplasmic domain of Ire1p successfully, we decided to make a shorter construct of Ire1p for crystallization. A sequence alignment of the cytoplasmic domains of Ire1p across different species showed that although the kinase and nuclease domains of these proteins are well conserved, there is no significant homology among residues which are N-terminal to the kinase domain (Figure 5.3). This stretch corresponds to the linker between the transmembrane and the kinase domain of Ire1p and may be flexible. In yeast Ire1p, the linker is rich in charged residues (lysines and aspartates) which may contribute to the aberrant behavior of the full-length cytoplasmic protein. A structure based sequence alignment of Ire1p with other protein kinases whose 3-dimensional structures have been determined indicated that the first typical secondary structural element (β1) begins at residue 670 (Figure 5.4). We therefore designed a shorter construct of Ire1p starting at residue 646 (Figure 5.5A). This protein was successfully purified and was not prone to degradation or aggregation, unlike Ire1p(556-1115) (Figure 5.5B). The protein was active in an in vitro kinase assay,

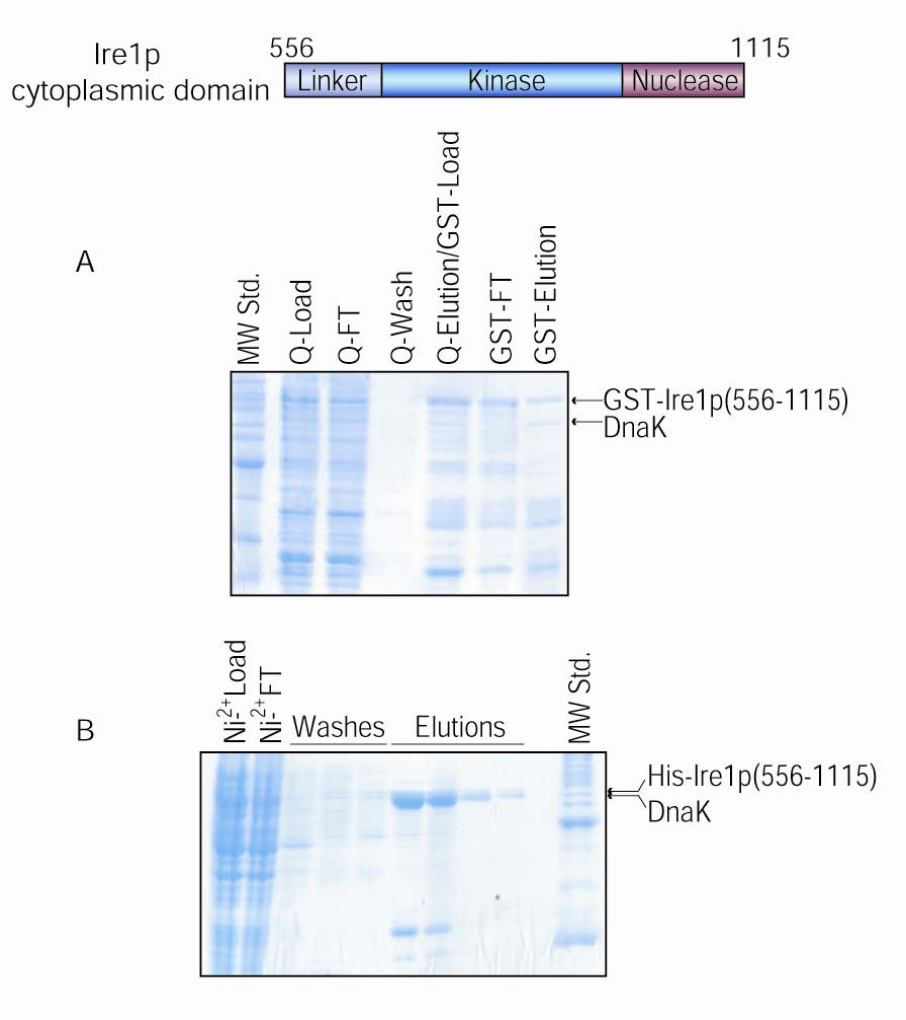


Figure 5.2. Expression and purification of the cytoplasmic domain of Ire1p. A. SDS-PAGE analysis of multiple chromatographic steps in the purification of GST-Ire1p(556-1115) B. SDS-PAGE analysis of purification of His-Ire1p (556-1115).

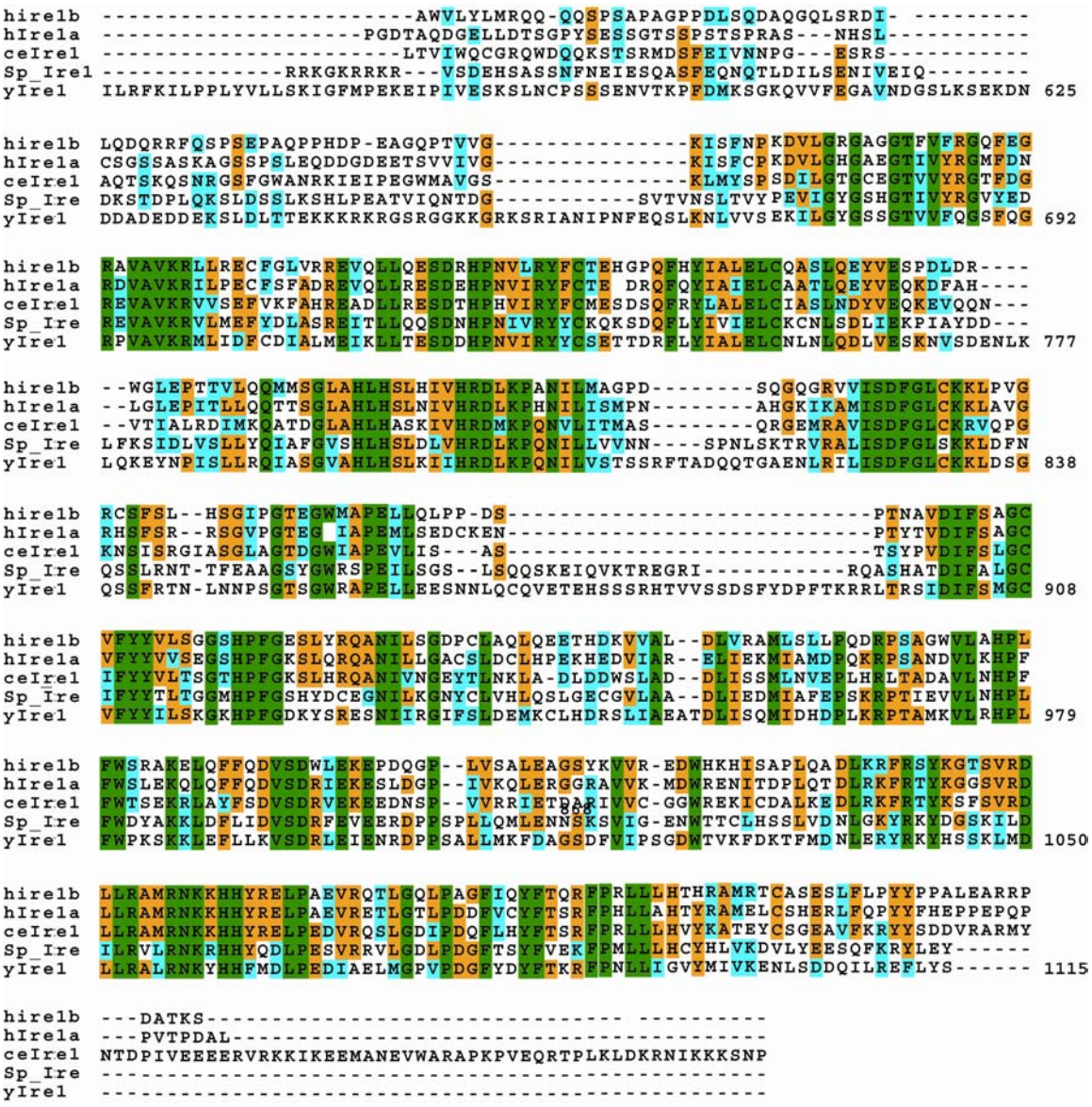


Figure 5.3. Sequence alignment of Ire1 kinase-nucleases across species. The cytosolic domain of the *S. cerevisiae* Ire1 *(*yIre1p) was aligned with the cytosolic domains of Ire1 from humans (hIre1α and hIre1β), *C. elegans* (ceIre1) and *S. pombe* (Sp_Ire). Identical residues are colored green, conserved residues cyan and similar residues orange.

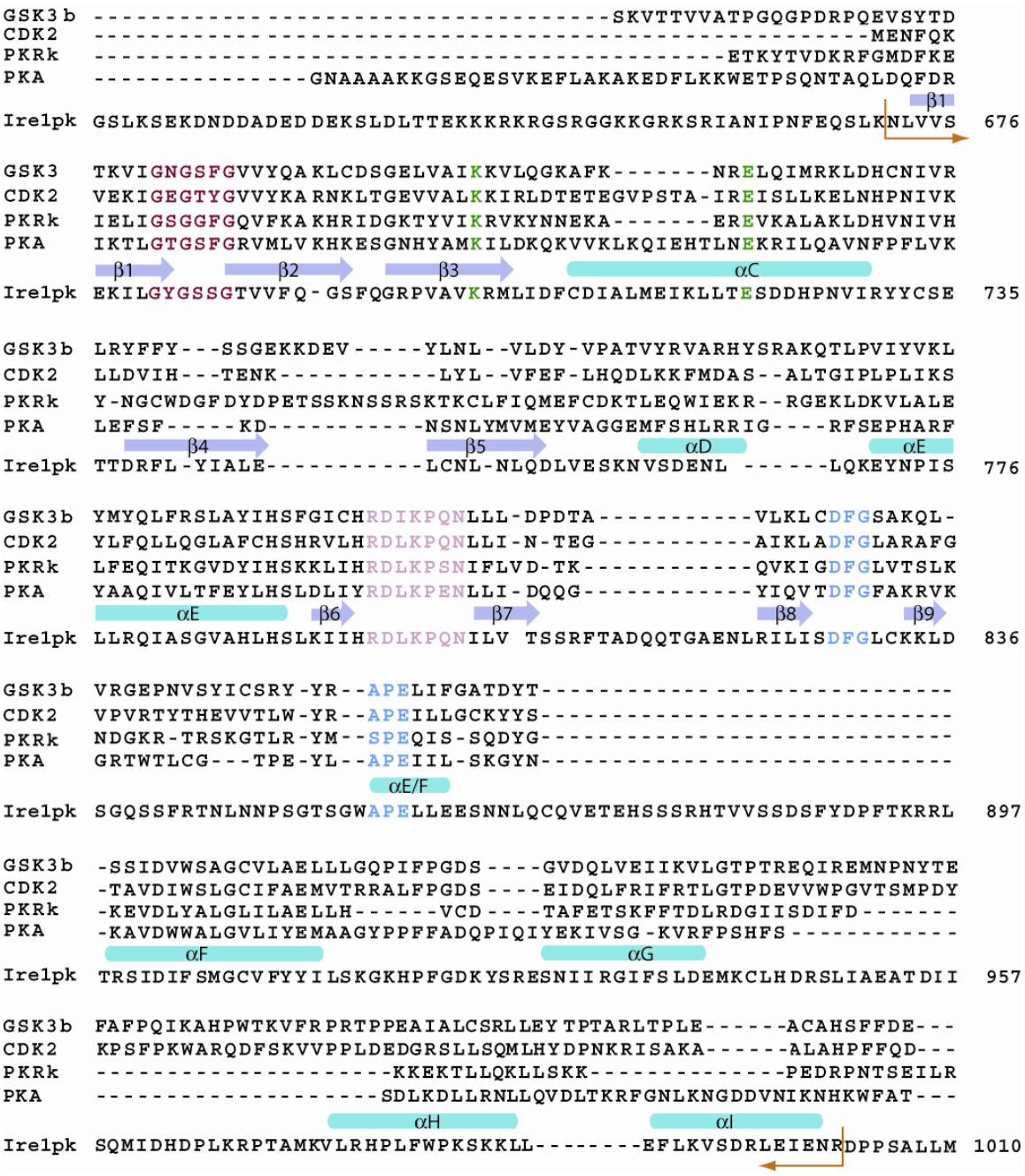


Figure 5.4. Sequence and structure based alignment of the kinase domain of Ire1p with the kinase cores of GSK3β, CDK2, PKR and PKA. Beta sheets are denoted by violet arrows and alpha helices by cyan cylinders. The glycine-rich loop is colored red, the invariant lys-glu ion pair green, the catalytic loop pink and the N- and C- terminal anchors of the activation segment blue.

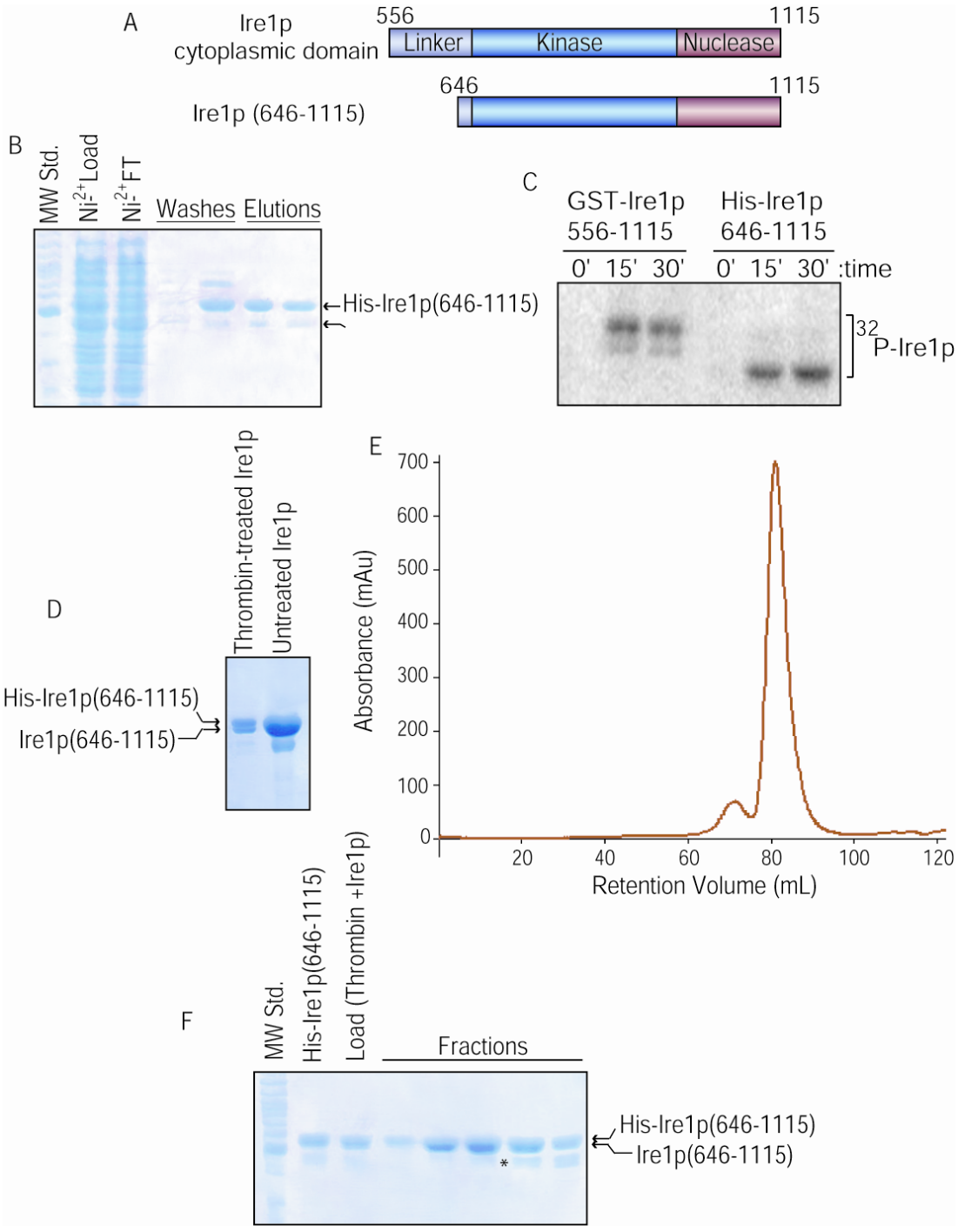


Figure 5.5. Purification of His-Ire1p(646-1115). A. Schematic representation of the construct generated. B. SDS-PAGE analysis of purification of His-Ire1p (646-1115) by Ni2+ affinity chromatography. C. Activity assay of His-Ire1p (646-1115) in comparison to full length cytosolic Ire1p. D. SDS-PAGE analysis of thrombin treatment of His-Ire1p (646-1115). E. Size-exclusion profile and F. corresponding SDS-PAGE analysis of thrombin-treated Ire1p (646-1115). * denotes the lower molecular weight species.

confirming that a non-kinase core segment was deleted at the N-terminus (Figure 5.5C). However, efforts to remove the poly-Histidine tag of this protein by thrombin were unsuccessful, as thrombin treatment yielded a mixture of equal amounts of His- tagged and untagged protein (Figure 5.5D). Careful analysis of the purified His-Ire1p(646-1115) protein showed a lower molecular weight species which appeared to have co-migrated with Ire1p in size exclusion (Figures 5.5E and 5.5F). This species was also affected by thrombin treatment in a manner similar to His-Ire1p(646-1115), indicating that it was an Ire1p-related fragment with an intact N-terminus but truncated C-terminus. We estimated that this C-terminal truncation fragment differs in size with His-Ire1p(646-1115) by only 0.5kDa and therefore designed the construct Ire1p(646-1077).

His-Ire1p(646-1077) purified as a doublet even after three chromatographic steps and was unsuitable for crystallization trials (Figure 5.6A). However, we decided to perform limited proteolysis experiments with this relatively unstable protein to determine a stable core fragment of Ire1p which could be isolated for crystallization. Trypsin digestion of His-Ire1p(646-1077) yielded a stable fragment, whose molecular weight was determined by mass-spectrometric analysis (Figures 5.6B and 5.6C). This 37.6kDa fragment corresponds to the kinase core domain of Ire1p (Figure 5.4, sequence of Ire1p within the orange arrows). The kinase core of Ire1p, $Ire1p_k$, was successfully purified on a large scale by tryptic digestion of His-Ire1p(646-1077) and was active in an in vitro kinase assay (Figures 5.6 D and E). Crystallization trials were conducted with Ire1p(646-1115) and $Ire1p_k$ using commercial sparse matrix screens and grid screens of common precipitants such as PEG 4000, Ammonium Sulphate and MPD.

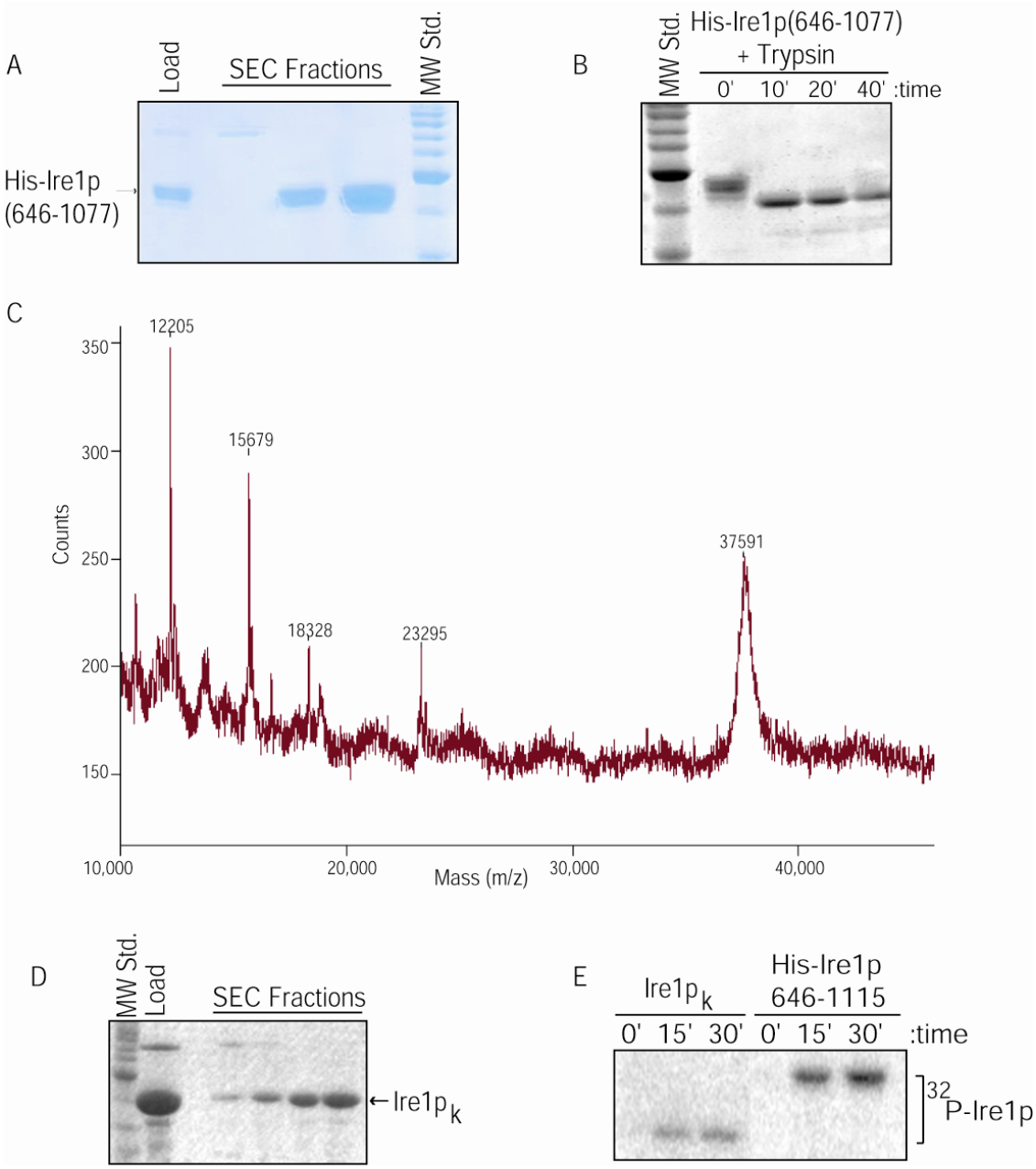


Figure 5.6. Identification and purification of a stable kinase core of Ire1p, Ire1p$_k$. A. SDS-PAGE analysis of size-exclusion fractions of the parent protein, His-Ire1p (646-1077). B. SDS-PAGE analysis and C. Mass-spectrometric analysis of trypsin digestion of His-Ire1p(646-1077). D. SDS-PAGE analysis of size-exclusion fractions of Ire1p$_k$. E. In vitro kinase assay of Ire1p$_k$ in comparison to Ire1p (646-1115).

Since we did not obtain any hits from our initial crystallization trials, we decided to generate several other constructs of Ire1p for our crystallization studies. We reasoned that although the kinase and nuclease core domains are sufficient for the activity and overall stability of the protein, the flanking regions might be essential for crystal packing and thus determine the ability of these constructs to crystallize. The fact that the N-terminal poly-His tag of Ire1p(646-1115) could not be removed by thrombin cleavage, even upon using excess enzyme, seemed to indicate that this region is not free as expected, but may be buried or involved in contact with some other domain. Since we had no other experimental evidence to this effect, an empirical approach was adopted to design these constructs. Most of the constructs encompass both the kinase and the nuclease domains of Ire1p and vary in the lengths of their N-terminal extension (Figure 5.7A). We also generated a construct corresponding to the kinase domain of Ire1p with additional flanking residues at the N- and C- termini (Ire1p656-1001). All these constructs of Ire1p purified very well and are active in an *in vitro* kinase assay (Figure 5.7B). Surprisingly, the N-terminal poly-His tag could not be removed even in the longest, viable construct of Ire1p (Ire1p 570-1115). Crystallization trials were conducted in parallel with all the purified fragments of Ire1p, as for His-Ire1p(646-1115).

2. Crystallization of Ire1p

Micro crystals of His-Ire1p(606-1115)/ATP:$MgCl_2$ were obtained from 1.6M Sodium Citrate, pH 6.5 at 20°C by the micro batch crystallization method (Figure 5.8A). These micro crystals were reproduced by the hanging drop vapor diffusion method in a

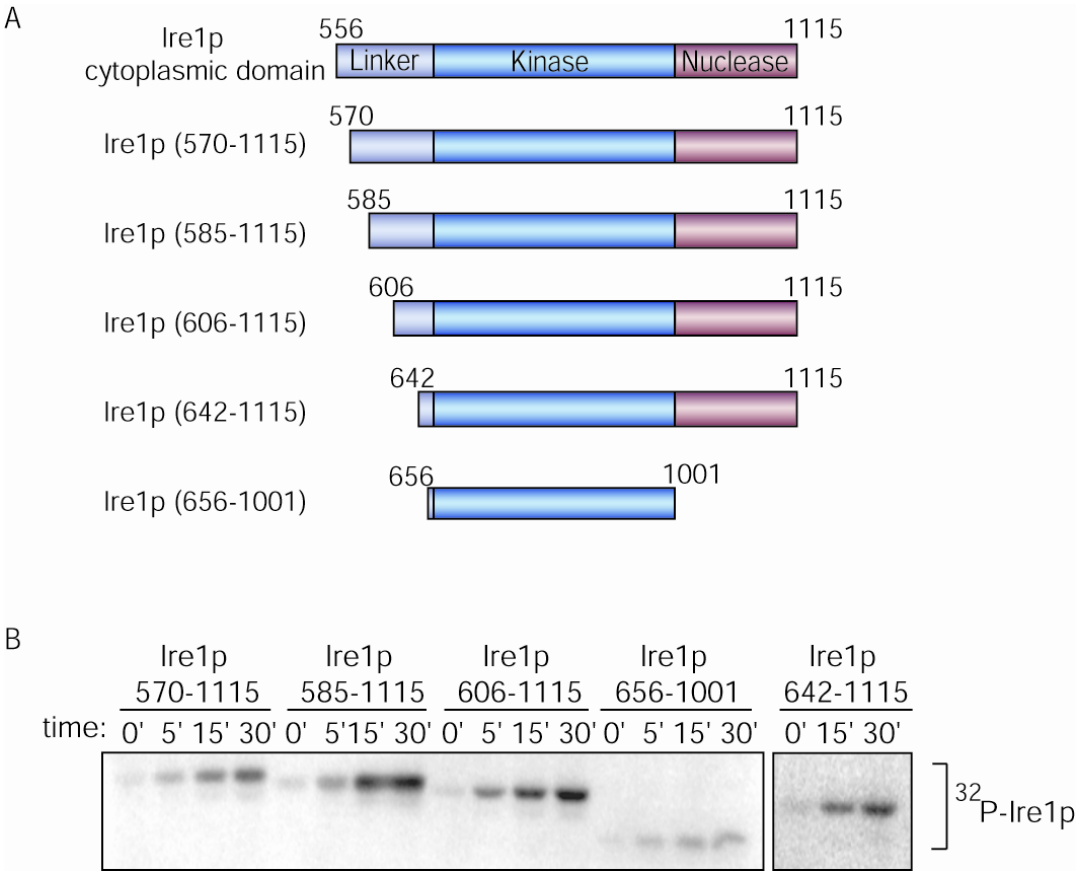


Figure 5.7. Design of Ire1p constructs for crystallization. A. Schematic representation and B. In vitro kinase assay of the constructs of Ire1p purified for crystallization trials

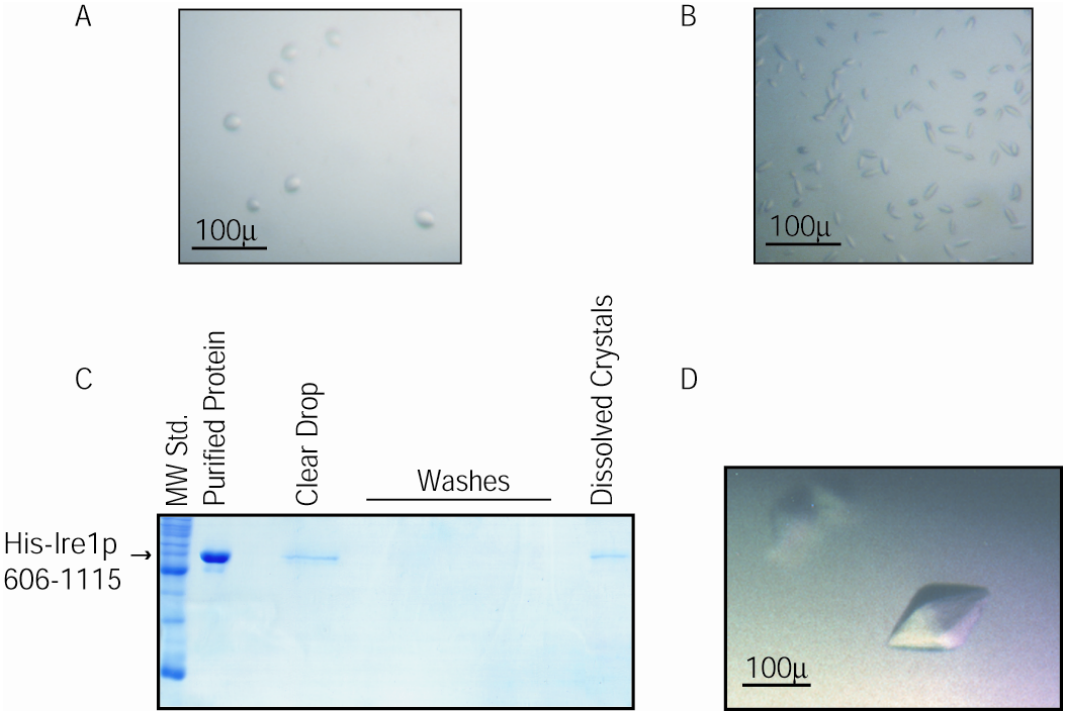


Figure 5.8. Crystallization of His-Ire1p(606-1115):ATP/$MgCl_2$. A. Micro crystals of Ire1p grown in 1.6M Sodium Citrate, pH 6.5 by the micro-batch crystallization method and B. the hanging drop vapor diffusion method. C. SDS-PAGE analysis of dissolved crystals of His-Ire1p(606-1115). D. Single crystal of His-Ire1p(606-1115):ATP/$MgCl_2$ grown from 1M Sodium Citrate (pH 6.5) and 2mM DTT.

similar condition and analyzed by SDS-PAGE to confirm the presence of Ire1p(606-1115) (Figures 5.8 B and C). The crystallization conditions were further optimized to 1.0-1.1M Sodium Citrate, pH 6.5 and 2mM DTT to consistently yield single crystals (Figure 5.8D). The concentration and pH of Sodium Citrate were critical to crystallization of Ire1p, though DTT concentrations could vary from 1 – 10mM without significantly affecting the crystallization process. Crystals of Ire1p(606-1115) were obtained in identical conditions in the presence of ADP:$MgCl_2$ and AMPPNP:$MgCl_2$ and also in absence of any nucleotide. Crystals were also obtained with His-Ire1p(585-1115) and His-Ire1p(570-1115) in presence and absence of nucleotides (ATP, ADP and AMPPNP) (Figure 5.9A). The morphology of the crystals in each case was very similar to those obtained from His-Ire1p(606-1115)/ATP:$MgCl_2$. The crystals from all the different protein-nucleotide complexes exhibited the same behavior in that they grew normally to a maximum size of 150μ × 50μ × 50μ and then started deteriorating with an increase in size. It appeared as though each small crystal served as a nucleation point for several other crystal plates. Crystals of different sizes and at various stages of growth were harvested from the drops, cryo-protected, flash frozen in liquid N_2 and tested at the UCSD X-ray facility for diffraction. None of the crystals showed diffraction even at a low resolution upon prolonged exposure to X-rays. Several cryo-protectants were tested and crystals were also mounted at room temperature to ensure that lack of diffraction was not due to improper freezing and handling of the crystals. Since these crystals were not suitable for diffraction, we screened a wide variety of additives to improve the crystals of Ire1p. Additives such as urea marginally improved the size of the crystals of Ire1p(606-1115) but not their quality (Figure 5.9B). We also attempted to crystallize Ire1p in

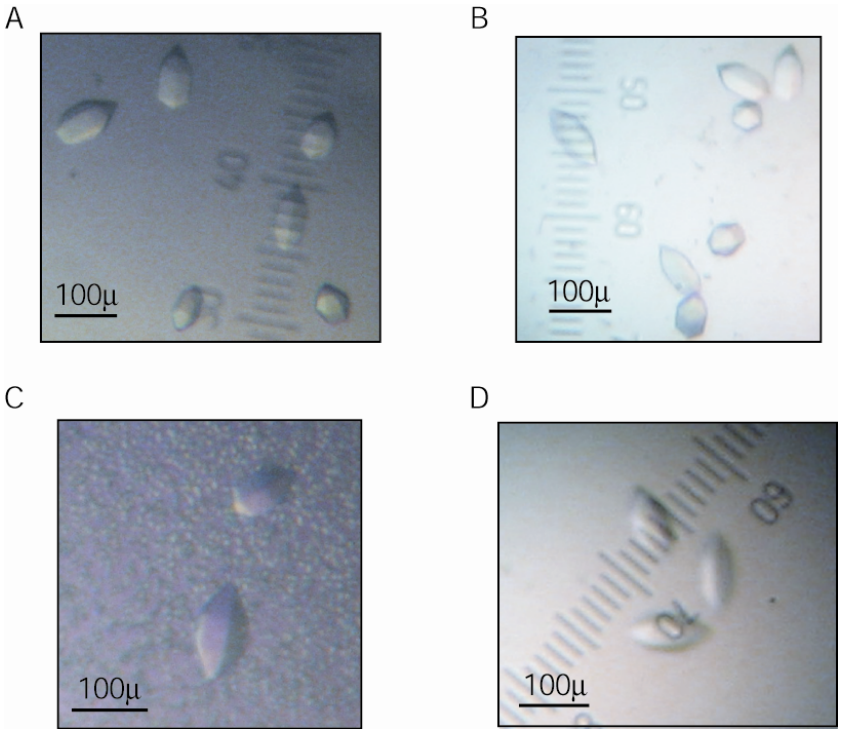


Figure 5.9. Crystals of Ire1p obtained from different constructs and conditions. A. Crystals of His-Ire1p(585-1115):ATP/$MgCl_2$ grown from 1M Sodium citrate (pH 6.5) and 2mM DTT. B. Crystals of His-Ire1p(606-1115) grown from conditions in A + 10mM urea as an additive. C. Crystals of His-Ire1p(585-1115):ATP/$MgCl_2$ grown from 1.5M Sodium Malonate (pH 6.5) and 2mM DTT. D. Crystals of His-Ire1p(585-1115) grown from conditions in C + 3% methanol as an additive.

alternate, but closely related conditions. 1.5M Sodium Malonate, pH 6.5 and 2mM DTT proved to be an effective crystallization condition for Ire1p (Figure 5.9C). The crystals obtained from Sodium Malonate were of a similar morphology and behavior as those obtained from Sodium Citrate and could be improved in size very slightly with alcohol-based additives (Figure 5.9D). No hits were obtained from any other crystallization conditions in the commercial screens tested.

3. Modification of protein for crystallization

Since we did not obtain diffracting crystals from His-Ire1p(606-1115) and His-Ire1p(585-1115) from any condition, we reasoned that proper packing in the crystals must be inhibited either by a stretch of unstructured residues in the protein or due to heterogeneity in the protein preparation. As mentioned earlier, we failed to remove the N-terminal poly-His tag of Ire1p by proteolytic cleavage. We suspected that the poly-His tag may interfere with crystallization in some way, and thus decided to purify an untagged Ire1p(606-1115). We expressed Ire1p(606-1115) as a fusion protein with a self-cleavable intein tag which had a chitin binding domain. The resultant protein, Ire1p (606-1115)-CBD was purified by affinity chromatography on a chitin column, followed by self cleavage of the intein tag in presence of excess DTT and size-exclusion chromatography (Figures 5.10 A and B). Ire1p(606-1115) crystallized in conditions similar to that of its His-tagged counterpart, although the crystals were smaller and lacked well-defined edges (Figures 5.10 C and D). These crystals did not diffract either and could not be improved any further.

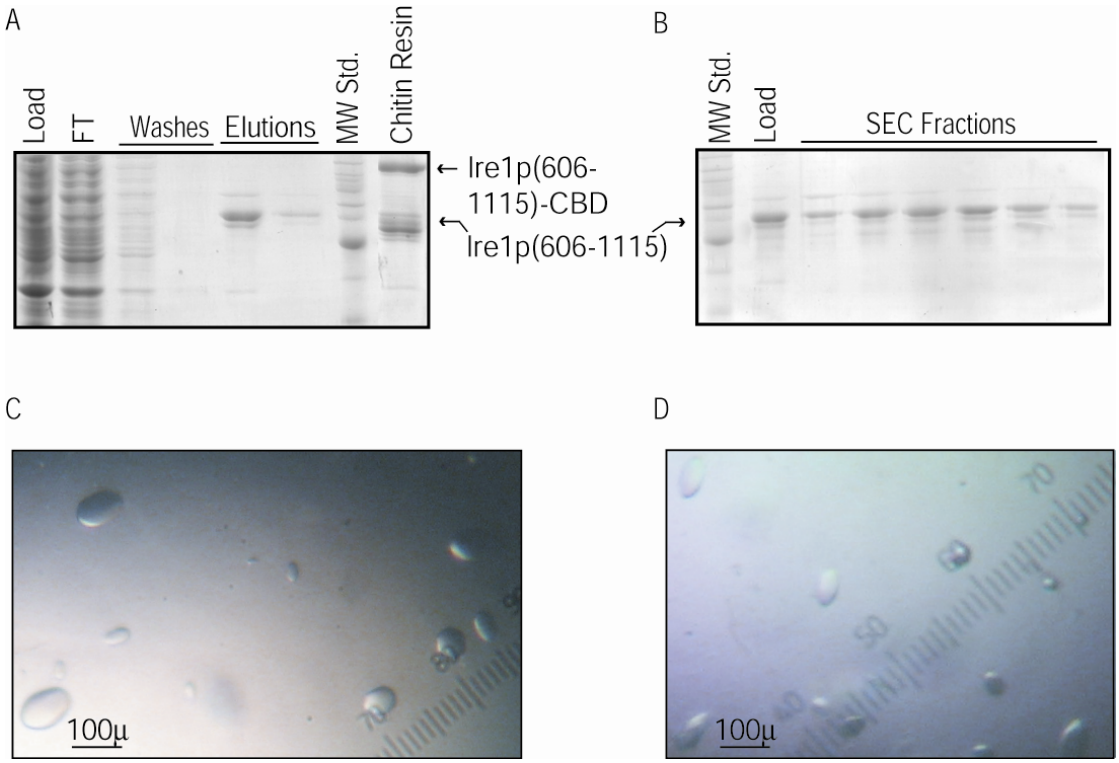


Figure 5.10. Purification and crystallization of Ire1p(606-1115). A. SDS-PAGE analysis of chitin affinity purification and intein-mediated cleavage of Ire1p. B. SDS-PAGE analysis of size-exclusion fractions of Ire1p(606-1115). C. Crystals of Irep(606-1115) grown from 1.1M Sodium Citrate (pH 6.5) and 2mM DTT. D. Crystals of Ire1p(606-1115) grown from 1.5M Sodium Malonate (pH 6.5) and 2mM DTT.

Heterogeneity in a protein preparation can often be due to lack of uniformity in post-translational modification of the target protein. Ire1p is a kinase known to auto-phosphorylate and is thought to contain several other phosphorylation sites in its linker region, in addition to those in its activation segment. It is possible that these sites are only partially phosphorylated when the protein is expressed in *E. coli*, thus contributing to its heterogeneity. To address this problem, we made two sets of mutations in His-Ire1p(606-1115) – one in which the catalytic aspartate residue D797 was mutated to asparagine (Ire1p606-1115 D797N) thus abolishing the kinase activity of Ire1p, and the other where the putative phosphorylation sites in the activation segment S840, S841 and T843 were mutated to aspartates (Ire1p606-1115 SST-D) rendering it constitutively active. The mutants were designed to convert Ire1p to a fully unphosphorylated or phosphorylated form. Both these mutants were purified for crystallization trials (Figures 5.11 A and B). While Ire1p(606-1115 SST-D) was a highly soluble protein whose poly-His tag could be removed by thrombin cleavage, Ire1p(606-1115 D797N) was much less soluble and could not be concentrated readily. Nevertheless, His-Ire1p(606-1115 SST-D), Ire1p(606-1115 SST-D) and His-Ire1p(606-1115 D797N) crystallized in similar conditions (0.9- 1.2M Sodium Citrate + 2mM DTT). Crystals of Ire1p(606-1115 D797N) were very small, possibly due to the low concentration of protein in the drop, and were not tested for diffraction (Figure 5.11C). The crystals of His- and untagged Ire1p(606-1115 SST-D) grew to a maximum size of 180μ × 40μ × 40μ and were flash frozen for data collection (Figure 5.11D). These crystals behaved like His-Ire1p(606-1115) crystals and did not diffract at the UCSD X-ray facility. Both His-Ire1p(606-1115 SST-D) and His-Ire1p(606-1114 D797N) behave as expected in an in vitro kinase assay (Figure

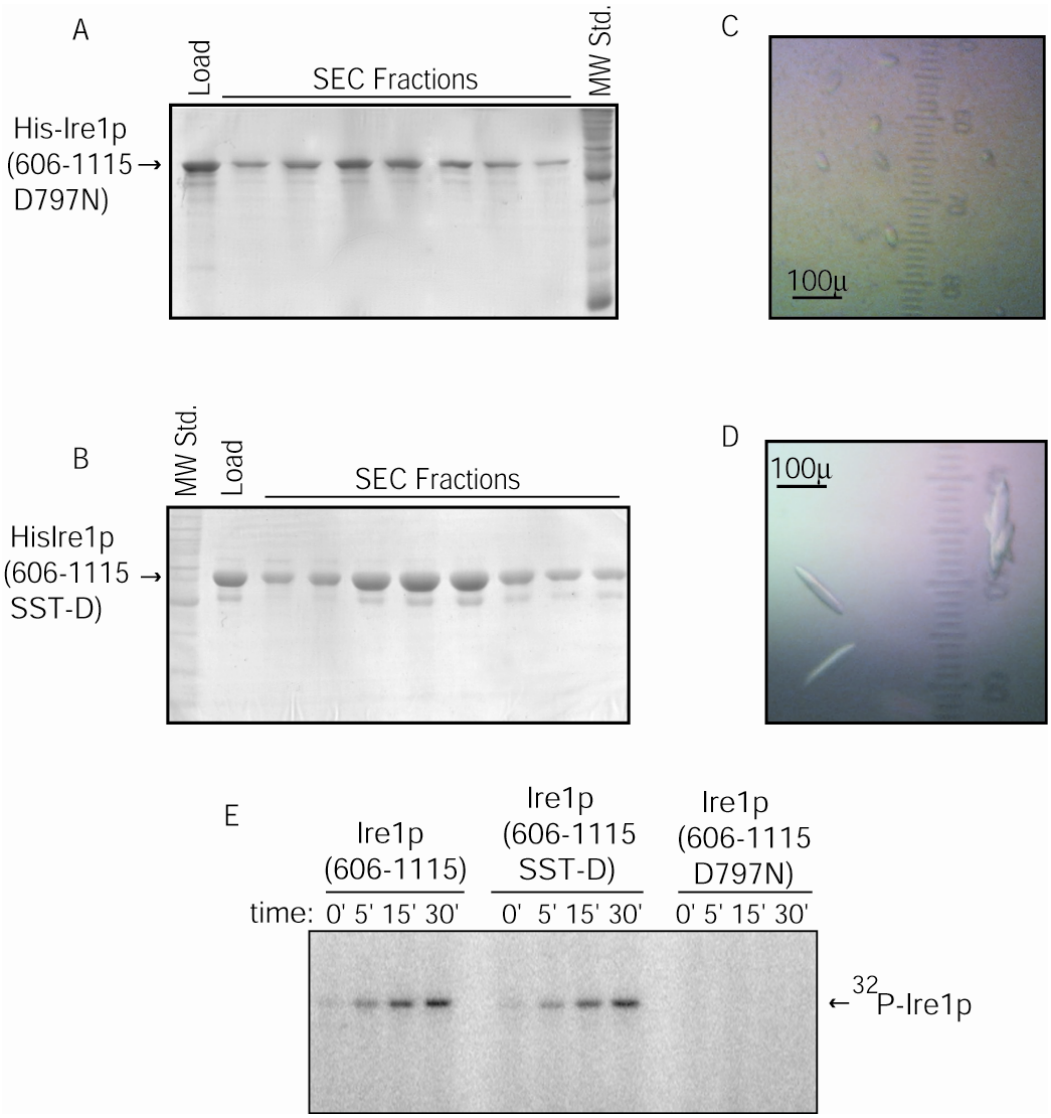
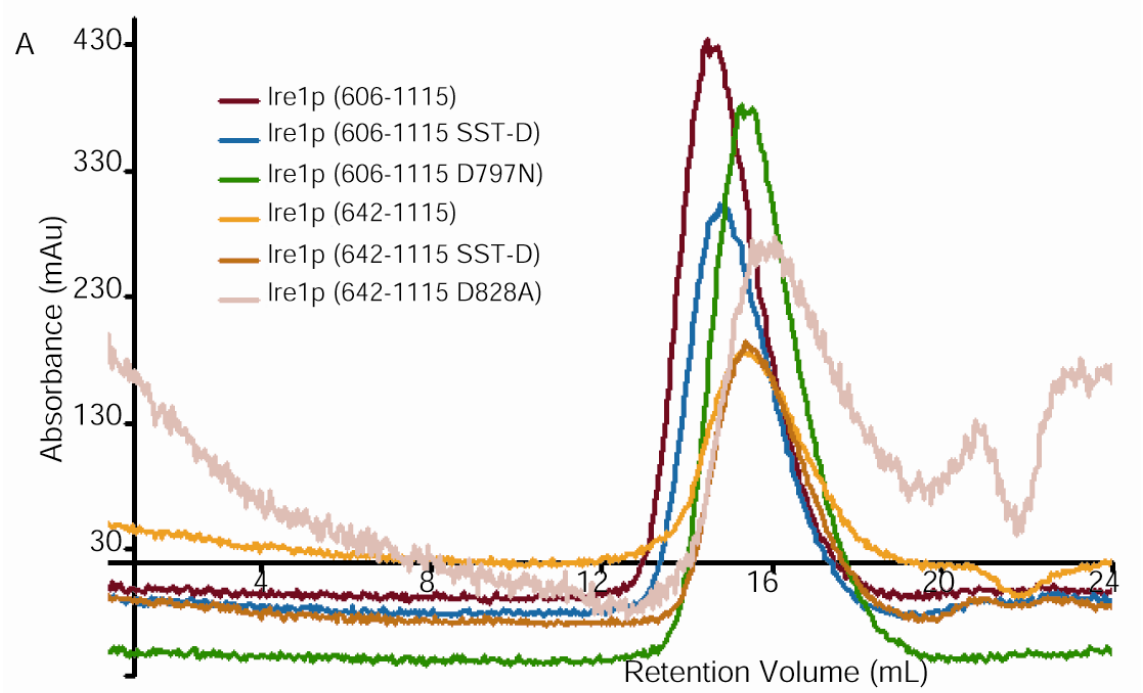


Figure 5.11. Purification and crystallization of His-Ire1p(606-1115) mutants. A. SDS-PAGE analysis of size-exclusion fractions of His-Ire1p(606-1115 D797N). B. SDS-PAGE analysis of size-exclusion fractions of His-Ire1p(606-1115 SST-D). C. Crystals of Ire1p(606-1115 D797N) grown in 1.2M Sodium Citrate (pH 6.5) and 2mM DTT. D. Crystals of His-Ire1p(606-1115 SST-D) grown in 0.9M Sodium Citrate (pH 6.5) and 2mM DTT. E. In vitro kinase assay of the mutant His-Ire1p in comparison to the corresponding wild-type.

5.11E). His-Ire1p(606-1115 SST-D) is as active as His-Ire1p(606-1115) while His-Ire1p(606-1115 D797N) shows no kinase activity in vitro.

4. Oligomeric State of Ire1p

Full-length Ire1p is a monomeric protein in its inactive state and is known to dimerize upon UPR induction. Following dimerization of the ER luminal domain of Ire1p, the cytoplasmic domain undergoes dimerization and trans-autophosphorylation. In order to gain further insight into the biochemical behavior of the constructs of Ire1p generated for crystallization, we wanted to determine the oligomeric state of each of the proteins purified. Equal amounts of all Ire1p proteins were injected on to an Analytical Superdex 200 column along with molecular weight standards. The molecular weight of each protein was calculated from its retention volume (Figure 5.12A). Interestingly, we found that the cytoplasmic domain of Ire1p has an inherent tendency to dimerize as well. We show that Ire1p(606-1115) and Ire1p(606-1115 SST-D) are dimers in solution, while Ire1p(606-1115 D797N) has an apparent molecular weight of 1.5X its monomeric molecular weight and thus is in rapid equilibrium between a monomer and a dimer (Figure 5.12B). We also noted differences in oligomeric states among mutants (catalytically inactive and constitutively active) and the corresponding wild-type version of a shorter construct of Ire1p, Ire1p(642-1115). Ire1p(642-1115) and Ire1p(642-1115 SST-D) are in rapid equilibrium between a monomer and dimer, but Ire1p(642-1115 D828A) is a monomer in solution (Figure 5.12B).



B

Constructs	Retention Volume (mL)	Molecular Weight (kDa)		Oligomeric State
		Expt.	Theor (monomer)	
Ire1p (606-1115)	14.60	126.7	60.7	Dimer
Ire1p (606-1115 SST-D)	14.71	121.6	60.7	Dimer
Ire1p (606-1115 D797N)	15.57	84.8	60.7	Monomer/Dimer
Ire1p (642-1115)	15.62	84.1	56.8	Monomer/Dimer
Ire1p (642-1115 SST-D)	15.65	84.1	56.8	Monomer/Dimer
Ire1p (642-1115 D828A)	16.22	65.0	56.8	Monomer

Figure 5.12. Oligomeric states of the different constructs of Ire1p. A. Overlay of size-exclusion profiles of His-Ire1p(606-1115) wild type and mutants (D797N and SST-D) and His-Ire1p(642-1115) wild type and mutants (D828A and SST-D). B. Experimental molecular weights and oligomeric states of each of the proteins as determined from their retention volumes in size-exclusion and molecular weight standards.

D. Discussion

In this chapter we report the expression, purification and crystallization of specific constructs of Ire1p. Although we were able to successfully express and purify several N- and C-terminal truncations of Ire1p, only a small subset of these constructs could be crystallized. These crystals did not show any diffraction at the UCSD X-ray facility and could not be improved in size and quality by varying the usual crystallization parameters.

Contrary to what was expected, the shortest constructs of Ire1p which consisted of the kinase–nuclease core domains or the kinase core alone failed to crystallize in any condition, while the longer constructs of Ire1p with the 60-70 residues long linker readily crystallized once a crystallization condition was identified. This implies that the N-terminal linker is either involved in crystal packing or interacts with some other part of the protein and serves to stabilize it. The linker has several serine and threonine residues which are potential substrates for Ire1p kinase activity. It is not known how many of these residues are actually phosphorylated in Ire1p or if they have any functional significance in Ire1p activity. It is unlikely that the linker plays a role in protein stability since the constructs of Ire1p lacking the linker are biochemically well-behaved and not prone to aggregation or degradation. However, the length of the linker of Ire1p required for crystallization can vary by at least 35 amino acids, since Ire1p(570-1115) and Ire1p(606-1115) crystallize in the same conditions with similar morphology. This makes the role of the linker somewhat elusive in crystal packing.

Studies on several protein kinases whose structures have been solved both in the active and inactive states show that a kinase has distinct conformations in these two states [78]. Often an inactive form of the kinase is more amenable to crystallization, while in

some case a fully active form of the kinase can be readily crystallized [175]. With this in mind together with the possible heterogeneity within a protein due to differential phosphorylation, we generated two mutant forms of Ire1p(606-1115): Ire1p(606-1115 D797N) which was a catalytically inactive form of the Ire1p kinase and Ire1p(606-1115 SST-D) which would render Ire1p constitutively active as a kinase. We hoped that either of these two mutant proteins would be in a conformation favorable for crystallization of the protein. Surprisingly both the mutant proteins crystallized in conditions very similar to the wild type, indicating that the activation loop, whose conformation varies significantly from an active to an inactive state of the kinase, does not contribute very much to the overall crystal packing of Ire1p.

In light of the current observations, it is difficult to predict which region of Ire1p dominates its crystal packing and why the crystals grown thus far fail to show any diffraction. The most likely explanation is that there is a lack of order within the crystal, probably due to a high degree of structural flexibility within some region of the protein. Thus it is important to determine this flexible region first, before proceeding with crystallization trials on any other construct of Ire1p. Our sequence alignment analyses show that Ire1p has a 30 residue long insert between helices αE/F and αF. This insert does not exist in any other member of the Ire1 family (*S. pombe* Ire1 has a 15 residue long insert) or in any of the protein kinases shown in the multiple sequence alignment. No biochemical information has been forthcoming about the role of this insert region in kinase/nuclease activity of Ire1p. We reason that this insert might be a flexible region hindering the proper packing within Ire1p crystals.

The oligomeric state of a protein may also contribute to its crystallization by enhancing protein-protein interactions which mediate/enhance crystal packing. Our studies on the oligomeric states of the different Ire1p constructs indicate that the cytosolic domain of Ire1p has an intrinsic tendency to dimerize. Dimerization of Ire1p apparently depends both on its N-terminal linker and its phosphorylation state. While Ire1p(606-1115) and Ire1p(606-1115 SST-D) are dimers in solution, an unphosphorylated form of the same construct (Ire1p 606-1115 D797N) is in rapid equilibrium between a monomer and a dimer. Also Ire1p(642-1115) and Ire1p(642-1115 SST-D) are in monomer-dimer equilibrium, while Ire1p(642-1115 D828A) which is the corresponding unphosphorylated form is a monomer. Previous biochemical studies have shown that following UPR induction and dimerization of the luminal domain of Ire1p, the cytosolic domain of Ire1p undergoes dimerization and autophosphorylation. It is possible that the linker of Ire1p mediates this dimerization and that phosphorylation stabilizes this dimer in some way. A genetic screen by Welihinda et al. shows that an yeast phosphatase Ptc2p interacts with Ire1p in a phosphorylation dependent manner and leads to dephosphorylation and inactivation of Ire1p and downregulation of UPR [176]. Dephosphorylation of Ire1p by Ptc2p may lead to destabilization of the dimer and restore Ire1p to its monomeric inactive state. If the dimeric form of Ire1p is more amenable to crystallization, then it is not surprising that the longer constructs which retain a significant portion of the N-terminal linker can be crystallized instead of the shorter constructs. This theory however does not explain why there is no difference between crystallization of the catalytically inactive (D797N) and the constitutively active (SST-D) mutants of Ire1p.

VI: Discussion

A. Identification of an unusual docking element in SR proteins

Our studies on ASF/SF2 and SRPK1 have led to the identification of a docking motif on ASF/SF2. We show that the docking motif is partly responsible for the high affinity interaction of ASF/SF2 with SRPK1 and the mechanism and extent of phosphorylation of ASF/SF2. The docking motif is located between the second RRM and the RS domain and is predominantly basic in nature. Subsequent structural studies on a complex of SRPK1 and ASF/SF2 indicate that part of the RS domain of ASF/SF2 can also bind the docking groove of SRPK1 through the same electrostatic interactions as seen in the structure of the SRPK1:peptide:ADP complex. Biochemical studies with the yeast proteins Sky1p and Npl3p have identified multiple docking motifs in Npl3p, all of which are capable of binding the docking groove of Sky1p [110]. These studies indicate that SR proteins may contain more than one kinase docking motif. The presence of multiple docking motifs suggests a 'sliding' mechanism of docking and phosphorylation, where transient docking at a particular position places a specific segment of the RS domain in the active site of the kinase, leading to its phosphorylation. Multiple docking motifs, strategically located with respect to the RS domain, would then ensure rapid and accurate phosphorylation of the SR proteins. We refer to these collective docking motifs in SR proteins as a 'mobile docking element'.

We speculate that the mobile docking element also determines whether or not a particular SR protein shuttles between the nucleus and the cytoplasm. Non-shuttling SR proteins such as SC35 and SRp40 have extensive RS domains, with longer contiguous tracts of RS peptides. In contrast shuttling SR proteins, such as ASF/SF2, SRp20 and 9G8, have shorter RS domains with fewer uninterrupted stretches of RS dipeptides.

Since a tract of multiple RS dipeptides can function as a docking motif, the span of the mobile docking element in nuclear-retained SR proteins is more expansive than that of shuttling SR proteins. This suggests that nuclear-retained SR proteins are phosphorylated by SRPK1 to a greater extent as compared to shuttling SR proteins. Although a nuclear phosphatase specific for SR proteins is yet to be identified in mammals, it is clear that partial dephosphorylation of SR proteins is essential for their nuclear export since SR proteins interact with the export complex p15/Tap only in their hypophosphorylated states [44]. It is possible that due to extensive phosphorylation of the RS domain of non-shuttling SR proteins, the dephosphorylation is not effective enough to generate a hypophosphorylated species compatible with nuclear export. In support of this idea, studies by Lin et al. show that hyperphosphorylation mimic mutants of ASF/SF2 do not shuttle between the nucleus and cytoplasm and the nuclear-retention signal of SC35 functions by conferring phosphatase resistance on RS domains [133]. Together these observations suggest that in the nucleus, shuttling and non-shuttling SR proteins are sorted based on their phosphorylated states; hypophosphorylated SR proteins are exported whereas extensively phosphorylated SR proteins are retained in the nucleus. Thus, the mobile docking element, which governs the extent of RS domain phosphorylation, may indirectly influence the nucleo-cytoplasmic shuttling ability of an SR protein. The nucleo-cytoplasmic shuttling ability of an SR protein is also related to its other cellular functions which are independent of splicing. Hypophosphorylated ASF/SF2, which is exported out of the nucleus, can interact with cytoplasmic mRNA and mediate its translation [46]. It has been proposed that hyperphosphorylation of ASF/SF2

reduces its binding to cytoplasmic mRNA and thus affects its translation activity. SC35, as expected, does not bind cytoplasmic mRNA or participate in mRNA translation.

B. Understanding the interaction between SR proteins and RNA sequence elements

Our studies have established that ASF/SF2 binds ESE sequences in pre-mRNA with low affinity but high specificity. We have also identified an unusual region of ASF/SF2 that is important for its ESE recognition function. While the individual RRM domains of ASF/SF2 do not possess any RNA binding affinity, the linker between the two RRMs, and specific residues therein, contributes significantly to the affinity and specificity of ESE recognition. The NMR solution structure of the single RRM containing SR protein, SRp20, bound to a 4-nucleotide ESE aptamer delineated specific residues within the RRM domain that are involved in RNA binding [128]. Sequence alignment of SRp20 and 9G8 (a single RRM SR protein) with the individual RRM domains of ASF/SF2 shows that all the RNA binding residues of SRp20 are conserved in the RRM1 domain of ASF/SF2. In contrast, several of the aromatic RNA-binding residues, particularly those within the RNP1 and RNP2 motifs, are absent in the RRM2 domain of ASF/SF2 (Figure 6.1). The RRM2 domain was previously identified to be a non-canonical or 'pseudo' RRM of ASF/SF2. Surprisingly, our in vitro assays show that when tethered independently to the inter-RRM linker, only the RRM2 domain of ASF/SF2 shows some affinity for RNA. Although an enhancement of RNA binding is observed upon addition of RRM1 to the linker-RRM2 protein, RRM1 alone has no affinity for ESE-RNA. The functional role of RRM1 of ASF/SF2 is somewhat elusive, since it does not appear to be directly involved in RNA or SRPK1 binding.

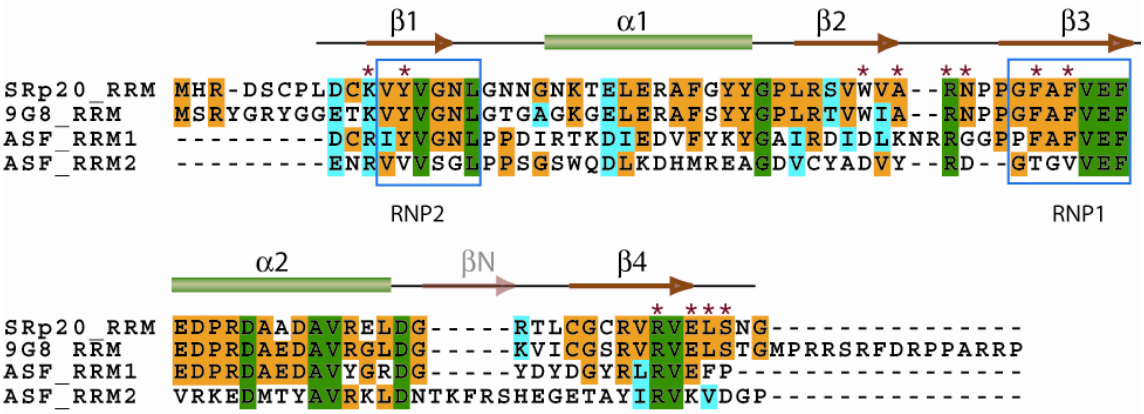


Figure 6.1. Sequence alignment of individual RRM domains of SR proteins. The RRM domains of SRp20 and 9G8 have been aligned with each RRM domain of ASF/SF2. Secondary structural elements were assigned based on the NMR solution structures of SRp20, 9G8, RRM1 of ASF/SF2 and the X-ray crystal structure of RRM2 in the SRPK1:ASF/SF2 complex. Fully conserved residues are colored green, identical residues yellow and similar residues cyan. The RNA binding residues of SRp20 are indicated by '*'.

It appears that weak binding to specific RNA sequence elements in vitro is a common phenomenon for auxiliary splicing factors such as SR proteins and hnRNPs. All studies (including ours) that attempt to determine the affinity of these proteins for RNA have been carried out on isolated systems, where the only components are the specific protein of interest and its cognate RNA. These systems, although useful for determining the absolute affinity, fail to account for the contribution of other cellular interacting partners in the process of RNA recognition by these proteins. It is now well known that several steps in gene expression, such as transcription, splicing and mRNA export are functionally coupled to each other. Hence, it is likely that the nascent mRNA transcript is bound by factors which can interact with the splicing machinery and directly 'load' the mRNA transcript onto them. Such a 'loading' mechanism would serve to increase the apparent affinity of RNA sequence elements for splicing factors. To this end, a study by Li et al. suggests that the phosphorylated-C-terminal domain of RNA polymerase II (CTD-P) enhances recruitment of ASF/SF2 to nascent transcripts [177]. It is possible that SR proteins interact with CTD-P through their RS domains, which facilitates their recruitment to RNA. It would be worthwhile to determine if factors involved in transcription can directly interact with auxiliary splicing factors like SR proteins and hnRNPs.

C. Regulation of splicing factors by phosphorylation: a common theme in diverse splicing mechanisms

The work described here broadly addresses the role of phosphorylation in regulation of a particular class of splicing factors, SR proteins. Our work has defined the

role of differential phosphorylation in spatial regulation of SR proteins. The modulation of splicing activity of SR proteins by phosphorylation has also been clearly established in other studies.

Interestingly, the regulation of splicing factors by phosphorylation extends beyond those involved in spliceosome-mediated mRNA splicing. Phosphorylation is thought to play a critical role in functional regulation of the UPR splicing factor, Ire1p. Although induction of UPR was known to cause autophosphorylation of the kinase domain of Ire1p, the molecular link between phosphorylation and splicing activity was not known. An earlier study by Papa et al. suggested that the key regulatory event in splicing by Ire1p is a conformational change in the kinase domain, triggered by its nucleotide binding [153]. More recent studies indicate that the conformation of the activation loop of the kinase domain is the molecular switch for Ire1p splicing (A.C and M.N., manuscript in preparation). In absence of ER stress, the activation loop of Ire1p is thought to be in an 'off' state. Autophosphorylation of residues within the activation loop bring about a conformational change in this loop (generating an 'on'-state), which is communicated to the nuclease domain, leading to its activation. Once ER protein folding homeostasis is restored, the activation loop is dephosphorylated, leading to its reversal to the 'off'-state and subsequent inactivation of the nuclease domain. In support of this model, activation loop mutants of Ire1p, which mimic the phosphorylated state or stabilize the 'on'-state, show sustained splicing activity. These studies, once again, establish the role of phosphorylation in regulation of mRNA splicing.

References

1. Sidrauski, C., J.S. Cox, and P. Walter, *tRNA ligase is required for regulated mRNA splicing in the unfolded protein response.* Cell, 1996. **87**(3): p. 405-13.

2. Sidrauski, C. and P. Walter, *The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response.* Cell, 1997. **90**(6): p. 1031-9.

3. Staley, J.P. and C. Guthrie, *Mechanical devices of the spliceosome: motors, clocks, springs, and things.* Cell, 1998. **92**(3): p. 315-26.

4. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

5. Davis, C.A., et al., *Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast.* Nucleic Acids Res, 2000. **28**(8): p. 1700-6.

6. Padgett, R.A., et al., *Splicing of messenger RNA precursors.* Annu Rev Biochem, 1986. **55**: p. 1119-50.

7. Johnson, J.M., et al., *Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.* Science, 2003. **302**(5653): p. 2141-4.

8. Krawczak, M., J. Reiss, and D.N. Cooper, *The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences.* Hum Genet, 1992. **90**(1-2): p. 41-54.

9. Jurica, M.S. and M.J. Moore, *Pre-mRNA splicing: awash in a sea of proteins.* Mol Cell, 2003. **12**(1): p. 5-14.

10. Zahler, A.M., et al., *SR proteins: a conserved family of pre-mRNA splicing factors.* Genes Dev, 1992. **6**(5): p. 837-47.

11. Graveley, B.R., *Sorting out the complexity of SR protein functions.* RNA (New York), 2000. **6**(9): p. 1197-1211.

12. Wang, J., S.H. Xiao, and J.L. Manley, *Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing.* Genes Dev, 1998. **12**(14): p. 2222-33.

13. Ma, X. and F. He, *Advances in the study of SR protein family.* Genomics Proteomics Bioinformatics, 2003. **1**(1): p. 2-8.

14. Henry, M., et al., *Potential RNA binding proteins in Saccharomyces cerevisiae identified as suppressors of temperature-sensitive mutations in NPL3.* Genetics, 1996. **142**(1): p. 103-15.

15. Windgassen, M. and H. Krebber, *Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast.* EMBO Rep, 2003. **4**(3): p. 278-83.

16. Krainer, A.R., G.C. Conway, and D. Kozak, *The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites.* Cell, 1990. **62**(1): p. 35-42.

17. Wang, J. and J.L. Manley, *Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways.* Rna, 1995. **1**(3): p. 335-46.

18. Valcarcel, J. and M.R. Green, *The SR protein family: pleiotropic functions in pre-mRNA splicing.* Trends Biochem Sci, 1996. **21**(8): p. 296-301.

19. Mayeda, A., et al., *Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements.* Mol Cell Biol, 1999. **19**(3): p. 1853-63.

20. Blencowe, B.J., *Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases.* Trends Biochem Sci, 2000. **25**(3): p. 106-10.

21. Lam, B.J. and K.J. Hertel, *A general role for splicing enhancers in exon definition.* Rna, 2002. **8**(10): p. 1233-41.

22. Jamison, S.F., et al., *U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements.* Nucleic Acids Res, 1995. **23**(16): p. 3260-7.

23. Tange, T.O. and J. Kjems, *SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain.* J Mol Biol, 2001. **312**(4): p. 649-62.

24. Tarn, W.Y. and J.A. Steitz, *Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps.* Proc Natl Acad Sci U S A, 1995. **92**(7): p. 2504-8.

25. Roscigno, R.F. and M.A. Garcia-Blanco, *SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome.* Rna, 1995. **1**(7): p. 692-706.

26. Zuo, P. and J.L. Manley, *The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3363-7.

27. Graveley, B.R. and T. Maniatis, *Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing.* Mol Cell, 1998. **1**(5): p. 765-71.

28. Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins.* Mol Cell, 2001. **8**(6): p. 1351-61.

29. Smith, C.W. and J. Valcarcel, *Alternative pre-mRNA splicing: the logic of combinatorial control.* Trends Biochem Sci, 2000. **25**(8): p. 381-8.

30. Eperon, I.C., et al., *Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1.* Mol Cell Biol, 2000. **20**(22): p. 8303-18.

31. Zahler, A.M., et al., *Distinct functions of SR proteins in alternative pre-mRNA splicing.* Science (Washington D C), 1993. **260**(5105): p. 219-222.

32. Wang, H.Y., et al., *SC35 plays a role in T cell development and alternative splicing of CD45.* Mol Cell, 2001. **7**(2): p. 331-42.

33. Xu, X., et al., *ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle.* Cell, 2005. **120**(1): p. 59-72.

34. Krainer, A.R., G.C. Conway, and D. Kozak, *Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells.* Genes Dev, 1990. **4**(7): p. 1158-71.

35. Roth, M.B., A.M. Zahler, and J.A. Stolk, *A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription.* J Cell Biol, 1991. **115**(3): p. 587-96.

36. Stojdl, D.F. and J.C. Bell, *SR protein kinases: The splice of life.* Biochemistry and Cell Biology, 1999. **77**(4): p. 293-298.

37. Xiao, S.H. and J.L. Manley, *Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing.* Genes Dev, 1997. **11**(3): p. 334-44.

38. Cao, W. and M.A. Garcia-Blanco, *A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding.* J Biol Chem, 1998. **273**(32): p. 20629-35.

39. Cao, W., S.F. Jamison, and M.A. Garcia-Blanco, *Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro.* Rna, 1997. **3**(12): p. 1456-67.

40. Xiao, S.H. and J.L. Manley, *Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2.* Embo J, 1998. **17**(21): p. 6359-67.

41. Kanopka, A., et al., *Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins.* Nature, 1998. **393**(6681): p. 185-7.

42. Mermoud, J.E., P. Cohen, and A.I. Lamond, *Ser-Thr-specific Protein Phosphatases Are Required for Both Catalytic Steps of Pre-Mrna Splicing.* Nucleic Acids Research, 1992. **20**(20): p. 5263-5269.

43. Mermoud, J.E., P.T.W. Cohen, and A.I. Lamond, *Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism.* EMBO (European Molecular Biology Organization) Journal, 1994. **13**(23): p. 5679-5688.

44. Huang, Y., T.A. Yario, and J.A. Steitz, *A molecular link between SR protein dephosphorylation and mRNA export.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9666-70.

45. Lai, M.C. and W.Y. Tarn, *Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins.* J Biol Chem, 2004. **279**(30): p. 31745-9.

46. Sanford, J.R., et al., *Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor.* Proc Natl Acad Sci U S A, 2005. **102**(42): p. 15042-7.

47. Koizumi, J., et al., *The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs).* J Biol Chem, 1999. **274**(16): p. 11125-31.

48. Yeakley, J.M., et al., *Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors.* J Cell Biol, 1999. **145**(3): p. 447-55.

49. Caceres, J.F., G.R. Screaton, and A.R. Krainer, *A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm.* Genes Dev, 1998. **12**(1): p. 55-66.

50. Allemand, E., et al., *Distinctive features of Drosophila alternative splicing factor RS domain: implication for specific phosphorylation, shuttling, and splicing activation.* Mol Cell Biol, 2001. **21**(4): p. 1345-59.

51. Cazalla, D., et al., *Nuclear export and retention signals in the RS domain of SR proteins.* Mol Cell Biol, 2002. **22**(19): p. 6871-82.

52. Lai, M.C., R.I. Lin, and W.Y. Tarn, *Transportin-SR2 mediates nuclear import of phosphorylated SR proteins.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10154-9.

53. Yun, C.Y., et al., *Phosphorylation-dependent and -independent nuclear import of RS domain-containing splicing factors and regulators.* J Biol Chem, 2003. **278**(20): p. 18050-5.

54. Misteli, T. and D.L. Spector, *Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors.* Mol Biol Cell, 1996. **7**(10): p. 1559-72.

55. Gui, J.-F., W.S. Lane, and X.-D. Fu, *A serine kinase regulates intracellular localization of splicing factors in the cell cycle.* Nature (London), 1994. **369**(6482): p. 678-683.

56. Colwill, K., et al., *The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution.* Embo J, 1996. **15**(2): p. 265-75.

57. Gilbert, W. and C. Guthrie, *The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA.* Mol Cell, 2004. **13**(2): p. 201-12.

58. Gui, J.F., et al., *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**(23): p. 10824-8.

59. Kuroyanagi, N., et al., *Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles.* Biochem Biophys Res Commun, 1998. **242**(2): p. 357-64.

60. Wang, H.Y., et al., *SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells.* J Cell Biol, 1998. **140**(4): p. 737-50.

61. Nikolakaki, E., et al., *Cloning and characterization of an alternatively spliced form of SR protein kinase 1 that interacts specifically with scaffold attachment factor-B.* J Biol Chem, 2001. **276**(43): p. 40175-82.

62. Siebel, C.W., et al., *Conservation in budding yeast of a kinase specific for SR splicing factors.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5440-5.

63. Tang, Z., M. Yanagida, and R.J. Lin, *Fission yeast mitotic regulator Dsk1 is an SR protein-specific kinase.* J Biol Chem, 1998. **273**(10): p. 5963-9.

64. Ding, J.H., et al., *Regulated cellular partitioning of SR protein-specific kinases in mammalian cells.* Mol Biol Cell, 2006. **17**(2): p. 876-85.

65. Ngo, J.C., et al., *SR protein kinase 1 is resilient to inactivation.* Structure, 2007. **15**(1): p. 123-33.

66. Nolen, B., et al., *The structure of Sky1p reveals a novel mechanism for constitutive activity.* Nat Struct Biol, 2001. **8**(2): p. 176-83.

67. Duncan, P.I., et al., *Alternative splicing of STY, a nuclear dual specificity kinase.* J Biol Chem, 1995. **270**(37): p. 21524-31.

68. Misteli, T., J.F. Caceres, and D.L. Spector, *The dynamics of a pre-mRNA splicing factor in living cells.* Nature, 1997. **387**(6632): p. 523-7.

69. Misteli, T., et al., *Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo.* J Cell Biol, 1998. **143**(2): p. 297-307.

70. Sacco-Bubulya, P. and D.L. Spector, *Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing.* J Cell Biol, 2002. **156**(3): p. 425-36.

71. Duncan, P.I., et al., *In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase.* Mol Cell Biol, 1997. **17**(10): p. 5996-6001.

72. Duncan, P.I., et al., *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**(2): p. 300-8.

73. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

74. Taylor, S.S., et al., *Structural framework for the protein kinase family.* Annu Rev Cell Biol, 1992. **8**: p. 429-62.

75. Knighton, D.R., et al., *Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.* Science, 1991. **253**(5018): p. 407-14.

76. Hanks, S.K. and A.M. Quinn, *Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members.* Methods Enzymol, 1991. **200**: p. 38-62.

77. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-82.

78. Nolen, B., S. Taylor, and G. Ghosh, *Regulation of protein kinases; controlling activity through activation segment conformation.* Mol Cell, 2004. **15**(5): p. 661-75.

79. Zheng, J., et al., *Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor.* Biochemistry, 1993. **32**(9): p. 2154-61.

80. Hubbard, S.R., *Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog.* Embo J, 1997. **16**(18): p. 5572-81.

81. Tarricone, C., et al., *Structure and regulation of the CDK5-p25(nck5a) complex.* Mol Cell, 2001. **8**(3): p. 657-69.

82. Schulze-Gahmen, U. and S.H. Kim, *Structural basis for CDK6 activation by a virus-encoded cyclin.* Nat Struct Biol, 2002. **9**(3): p. 177-81.

83. Jeffrey, P.D., et al., *Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex.* Nature, 1995. **376**(6538): p. 313-20.

84. Russo, A.A., P.D. Jeffrey, and N.P. Pavletich, *Structural basis of cyclin-dependent kinase activation by phosphorylation.* Nat Struct Biol, 1996. **3**(8): p. 696-700.

85. Shi, Z., K.A. Resing, and N.G. Ahn, *Networks for the allosteric control of protein kinases.* Curr Opin Struct Biol, 2006. **16**(6): p. 686-92.

86. Remenyi, A., M.C. Good, and W.A. Lim, *Docking interactions in protein kinase and phosphatase networks.* Curr Opin Struct Biol, 2006. **16**(6): p. 676-85.

87. Biondi, R.M. and A.R. Nebreda, *Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions.* Biochem J, 2003. **372**(Pt 1): p. 1-13.

88. Pellicena, P. and J. Kuriyan, *Protein-protein interactions in the allosteric regulation of protein kinases.* Curr Opin Struct Biol, 2006. **16**(6): p. 702-9.

89. Sharrocks, A.D., S.H. Yang, and A. Galanis, *Docking domains and substrate-specificity determination for MAP kinases.* Trends Biochem Sci, 2000. **25**(9): p. 448-53.

90. Chang, C.I., et al., *Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b.* Mol Cell, 2002. **9**(6): p. 1241-9.

91. Lee, T., et al., *Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry.* Mol Cell, 2004. **14**(1): p. 43-55.

92. Zhou, T., et al., *Docking interactions induce exposure of activation loop in the MAP kinase ERK2.* Structure, 2006. **14**(6): p. 1011-9.

93. Barsyte-Lovejoy, D., A. Galanis, and A.D. Sharrocks, *Specificity determinants in MAPK signaling to transcription factors.* J Biol Chem, 2002. **277**(12): p. 9896-903.

94. Remenyi, A., et al., *The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network.* Mol Cell, 2005. **20**(6): p. 951-62.

95. Jacobs, D., et al., *Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase.* Genes Dev, 1999. **13**(2): p. 163-75.

96. Galanis, A., S.H. Yang, and A.D. Sharrocks, *Selective targeting of MAPKs to the ETS domain transcription factor SAP-1.* J Biol Chem, 2001. **276**(2): p. 965-73.

97. Dar, A.C., T.E. Dever, and F. Sicheri, *Higher-order substrate recognition of eIF2alpha by the RNA-dependent protein kinase PKR.* Cell, 2005. **122**(6): p. 887-900.

98. Bhattacharyya, R.P., et al., *The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway.* Science, 2006. **311**(5762): p. 822-6.

99. Velazquez-Dones, A., et al., *Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty.* J Biol Chem, 2005. **280**(50): p. 41761-8.

100. Aubol, B.E., et al., *Processive phosphorylation of alternative splicing factor/splicing factor 2.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12601-6.

101. Hartmann, A.M., et al., *Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors.* Mol Cell Neurosci, 2001. **18**(1): p. 80-90.

102. Prasad, J., et al., *The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing.* Mol Cell Biol, 1999. **19**(10): p. 6991-7000.

103. Lamond, A.I. and D.L. Spector, *Nuclear speckles: a model for nuclear organelles.* Nat Rev Mol Cell Biol, 2003. **4**(8): p. 605-12.

104. Maris, C., C. Dominguez, and F.H.T. Allain, *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* FEBS Journal, 2005. **272**(9): p. 2118-2131.

105. Birney, E., S. Kumar, and A.R. Krainer, *Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors.* Nucleic Acids Research, 1993. **21**(25): p. 5803-5816.

106. White, A., et al., *Molecular basis of MAPK-activated protein kinase 2:p38 assembly.* Proc Natl Acad Sci U S A, 2007. **104**(15): p. 6353-8.

107. Bourne, Y., et al., *Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1.* Cell, 1996. **84**(6): p. 863-74.

108. Song, H., et al., *Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2.* Mol Cell, 2001. **7**(3): p. 615-26.

109. Dajani, R., et al., *Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex.* Embo J, 2003. **22**(3): p. 494-501.

110. Lukasiewicz, R., et al., *The RGG domain of Npl3p recruits Sky1p through docking interactions.* J Mol Biol, 2007. **367**(1): p. 249-61.

111. Tazi, J., et al., *Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing.* Nature (London), 1993. **363**(6426): p. 283-286.

112. Shi, Y., B. Reddy, and J.L. Manley, *PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins.* Mol Cell, 2006. **23**(6): p. 819-29.

113. Fu, X.-D., *The superfamily of arginine/serine-rich splicing factors.* RNA (New York), 1995. **1**(7): p. 663-680.

114. Sanford, J.R., et al., *A novel role for shuttling SR proteins in mRNA translation.* Genes Dev, 2004. **18**(7): p. 755-68.

115. Rossi, F., et al., *Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I.* Nature (London), 1996. **381**(6577): p. 80-82.

116. Blaustein, M., et al., *Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT.* Nat Struct Mol Biol, 2005. **12**(12): p. 1037-44.

117. Zuo, P. and J.L. Manley, *Functional domains of the human splicing factor ASF/SF2.* Embo J, 1993. **12**(12): p. 4727-37.

118. Liu, H.X., M. Zhang, and A.R. Krainer, *Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins.* Genes Dev, 1998. **12**(13): p. 1998-2012.

119. Kohtz, J.D., et al., *Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors.* Nature, 1994. **368**(6467): p. 119-24.

120. Wang, J., et al., *Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes.* Nucleic Acids Res, 2005. **33**(16): p. 5053-62.

121. Ghigna, C., et al., *Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene.* Mol Cell, 2005. **20**(6): p. 881-90.

122. Shen, H., J.L. Kan, and M.R. Green, *Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly.* Mol Cell, 2004. **13**(3): p. 367-76.

123. Shen, H. and M.R. Green, *A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly.* Mol Cell, 2004. **16**(3): p. 363-73.

124. Philipps, D., et al., *Arginine/serine repeats are sufficient to constitute a splicing activation domain.* Nucleic Acids Res, 2003. **31**(22): p. 6502-8.

125. Shen, H. and M.R. Green, *RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans.* Genes Dev, 2006. **20**(13): p. 1755-65.

126. Caceres, J.F. and A.R. Krainer, *Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains.* Embo J, 1993. **12**(12): p. 4715-26.

127. Shamoo, Y., N. Abdul-Manan, and K.R. Williams, *Multiple RNA binding domains (RBDs) just don't add up.* Nucleic Acids Res, 1995. **23**(5): p. 725-8.

128. Hargous, Y., et al., *Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8.* Embo J, 2006. **25**(21): p. 5126-37.

129. Singh, R. and J. Valcarcel, *Building specificity with nonspecific RNA-binding proteins.* Nat Struct Mol Biol, 2005. **12**(8): p. 645-53.

130. Deo, R.C., et al., *Recognition of polyadenylate RNA by the poly(A)-binding protein.* Cell, 1999. **98**(6): p. 835-45.

131. Handa, N., et al., *Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein.* Nature, 1999. **398**(6728): p. 579-85.

132. Sickmier, E.A., et al., *Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65.* Mol Cell, 2006. **23**(1): p. 49-59.

133. Lin, S., et al., *Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation.* Mol Cell, 2005. **20**(3): p. 413-25.

134. Ellgaard, L. and A. Helenius, *QUALITY CONTROL IN THE ENDOPLASMIC RETICULUM.* Nature Reviews Molecular Cell Biology, 2003. **4**(3): p. 181.

135. Hammond, C., I. Braakman, and A. Helenius, *Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control.* Proc Natl Acad Sci U S A, 1994. **91**(3): p. 913-7.

136. Hammond, C. and A. Helenius, *Quality control in the secretory pathway: retention of a misfolded viral membrane glycoprotein involves cycling between the ER, intermediate compartment, and Golgi apparatus.* J Cell Biol, 1994. **126**(1): p. 41-52.

137. Gething, M.-J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33.

138. Kaufman, R.J., *Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls.* Genes Dev, 1999. **13**(10): p. 1211-33.

139. Harding, H.P., et al., *Transcriptional and translational control in the Mammalian unfolded protein response.* Annu Rev Cell Dev Biol, 2002. **18**: p. 575-99.

140. Patil, C. and P. Walter, *Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals.* Curr Opin Cell Biol, 2001. **13**(3): p. 349-55.

141. Mori, K., et al., *A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins.* Embo J, 1992. **11**(7): p. 2583-93.

142. Kohno, K., et al., *The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum.* Mol Cell Biol, 1993. **13**(2): p. 877-90.

143. Travers, K.J., et al., *Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation.* Cell, 2000. **101**(3): p. 249-58.

144. Nikawa, J. and S. Yamashita, *IRE1 encodes a putative protein kinase containing a membrane-spanning domain and is required for inositol phototrophy in Saccharomyces cerevisiae.* Mol Microbiol, 1992. **6**(11): p. 1441-6.

145. Cox, J.S., C.E. Shamu, and P. Walter, *Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase.* Cell, 1993. **73**(6): p. 1197-206.

146. Mori, K., et al., *A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus.* Cell, 1993. **74**(4): p. 743-56.

147. Okamura, K., et al., *Dissociation of Kar2p/BiP from an ER sensory molecule, Ire1p, triggers the unfolded protein response in yeast.* Biochem Biophys Res Commun, 2000. **279**(2): p. 445-50.

148. Kimata, Y., et al., *Genetic evidence for a role of BiP/Kar2 that regulates Ire1 in response to accumulation of unfolded proteins.* Mol Biol Cell, 2003. **14**(6): p. 2559-69.

149. Kimata, Y., et al., *A role for BiP as an adjustor for the endoplasmic reticulum stress-sensing protein Ire1.* J Cell Biol, 2004. **167**(3): p. 445-56.

150. Credle, J.J., et al., *On the mechanism of sensing unfolded protein in the endoplasmic reticulum.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18773-84.

151. Shamu, C.E. and P. Walter, *Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus.* Embo J, 1996. **15**(12): p. 3028-39.

152. Welihinda, A.A. and R.J. Kaufman, *The unfolded protein response pathway in Saccharomyces cerevisiae. Oligomerization and trans-phosphorylation of Ire1p (Ern1p) are required for kinase activation.* J Biol Chem, 1996. **271**(30): p. 18181-7.

153. Papa, F.R., et al., *Bypassing a kinase activity with an ATP-competitive drug.* Science, 2003. **302**(5650): p. 1533-7.

154. Hassel, B.A., et al., *A dominant negative mutant of 2-5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon.* Embo J, 1993. **12**(8): p. 3297-304.

155. Dong, B. and R.H. Silverman, *2-5A-dependent RNase molecules dimerize during activation by 2-5A.* J Biol Chem, 1995. **270**(8): p. 4133-7.

156. Naik, S., J.M. Paranjape, and R.H. Silverman, *RNase L dimerization in a mammalian two-hybrid system in response to 2',5'-oligoadenylates.* Nucleic Acids Res, 1998. **26**(6): p. 1522-7.

157. Zhou, A., B.A. Hassel, and R.H. Silverman, *Expression cloning of 2-5A-dependent RNAase: a uniquely regulated mediator of interferon action.* Cell, 1993. **72**(5): p. 753-65.

158. Cox, J.S. and P. Walter, *A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response.* Cell, 1996. **87**(3): p. 391-404.

159. Niwa, M., et al., *Genome-scale approaches for discovering novel nonconventional splicing substrates of the Ire1 nuclease.* Genome Biol, 2005. **6**(1): p. R3.

160. Nojima, H., et al., *Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe.* Nucleic Acids Res, 1994. **22**(24): p. 5279-88.

161. Ruegsegger, U., J.H. Leber, and P. Walter, *Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response.* Cell, 2001. **107**(1): p. 103-14.

162. Gonzalez, T.N., et al., *Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway.* Embo J, 1999. **18**(11): p. 3119-32.

163. Zhang, K. and R.J. Kaufman, *Signaling the unfolded protein response from the endoplasmic reticulum.* J Biol Chem, 2004. **279**(25): p. 25935-8.

164. Tirasophon, W., A.A. Welihinda, and R.J. Kaufman, *A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells.* Genes Dev, 1998. **12**(12): p. 1812-24.

165. Wang, X.Z., et al., *Cloning of mammalian Ire1 reveals diversity in the ER stress responses.* Embo J, 1998. **17**(19): p. 5708-17.

166. Yoshida, H., et al., *XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor.* Cell, 2001. **107**(7): p. 881-91.

167. Calfon, M., et al., *IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA.* Nature, 2002. **415**(6867): p. 92-6.

168. Tirasophon, W., et al., *The endoribonuclease activity of mammalian IRE1 autoregulates its mRNA and is required for the unfolded protein response.* Genes Dev, 2000. **14**(21): p. 2725-36.

169. Iwawaki, T., et al., *Translational control by the ER transmembrane kinase/ribonuclease IRE1 under ER stress.* Nat Cell Biol, 2001. **3**(2): p. 158-64.

170. Hollien, J. and J.S. Weissman, *Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response.* Science, 2006. **313**(5783): p. 104-7.

171. Harding, H.P., Y. Zhang, and D. Ron, *Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase.* Nature, 1999. **397**(6716): p. 271-4.

172. Haze, K., et al., *Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress.* Mol Biol Cell, 1999. **10**(11): p. 3787-99.

173. Shi, Y., et al., *Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control.* Mol Cell Biol, 1998. **18**(12): p. 7499-509.

174. Ye, J., et al., *ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs.* Mol Cell, 2000. **6**(6): p. 1355-64.

175. Nagar, B., et al., *Structural basis for the autoinhibition of c-Abl tyrosine kinase.* Cell, 2003. **112**(6): p. 859-71.

176. Welihinda, A.A., et al., *Protein serine/threonine phosphatase Ptc2p negatively regulates the unfolded-protein response by dephosphorylating Ire1p kinase.* Mol Cell Biol, 1998. **18**(4): p. 1967-77.

177. Li, X. and J.L. Manley, *Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability.* Cell, 2005. **122**(3): p. 365-78.