

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Insights into the Role of a Disordered N-Terminus in the Functional Regulation of CLK1 Kinase

Permalink

<https://escholarship.org/uc/item/2406c9hs>

Author

George, Athira

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Insights into the Role of a Disordered N-Terminus in the Functional Regulation of CLK1 Kinase

A dissertation submitted in partial satisfaction of the requirements for the
degree of Doctor of Philosophy

in

Chemistry

by

Athira George

Committee in charge:

Professor Joseph Adams, Chair
Professor Patricia Jennings, Co-Chair
Professor Seth Cohen
Professor Gourisankar Ghosh
Professor Partho Ghosh
Professor Roger Sunahara

2021

Copyright

Athira George, 2021

All rights reserved.

The dissertation of Athira George is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents, for all the love and support.

EPIGRAPH

‘Life can only be understood backwards, but it must be lived forwards’

-Soren Kierkegaard

TABLE OF CONTENTS

Dissertation Approval Page.....	iii
Dedication.....	iv
Epigraph	v
Table of Contents	vi
List of Abbreviations.....	ix
List of Figures.....	xii
Acknowledgements	xiv
Vita	xvii
Abstract of the Dissertation.....	xviii
Chapter 1 Introduction.....	1
1.1. Alternative Splicing and Proteomic Diversity.....	2
1.2. Errors in Splicing and Human Diseases	3
1.3. Spliceosome: Assembly and Catalysis	5
1.4. SR Proteins are Essential Splicing Factors.....	7
1.5. Phosphorylation Regulates SR Protein Function.....	12
1.6. CLKs and SRPKs phosphorylate SR Proteins.....	13
1.6.1. SRPKs.....	13
1.6.2. CLKs.....	14
1.6.3. CLK-SRPK complex	18
1.7. N-terminus as a Modulator of CLK1 Function	18
1.7.1. N-terminus is important for nuclear localization.....	19
1.7.2. N-terminus induces hyper-phosphorylation of SRSF1	19
1.7.3. N-terminus mediated oligomerization is a substrate selection mechanism	22
1.8. Goals of the Dissertation.....	23
References	24
Chapter 2 Materials and Methods	29
2.1. Materials.....	30
2.2. Methods.....	30
2.2.1. Expression and purification of recombinant proteins	30
2.2.2. Mutagenesis	32
2.2.3. Generation of baculoviruses and protein expression in Hi5 cells...	32
2.2.4. Transfection and confocal imaging.....	34
2.2.5. Immunofluorescence.....	35
2.2.6. Biochemical Fractionation	35
2.2.7. Immunoprecipitations	35
2.2.8. Pulldown assays	36
2.2.9. Size Exclusion Chromatography and Dynamic Light Scattering ...	36

2.2.10. H/D Exchange and Mass Spectrometry	37
2.2.11. Single turnover kinase assays	39
2.2.12. Steady state kinetic assays	39
Chapter 3 Nuclear Localization of CLK1	41
3.1. Introduction.....	42
3.2. Results.....	46
3.2.1. CLK1 does not have a classical NLS.....	46
3.2.2. Identifying regions important for nuclear localization	48
3.2.3. Nuclear entry is mediated by the nuclear transport system for SR Proteins	51
3.2.4. CLK1 binding to TRN-SR2 is mediated by phospho-SRSF1	52
3.2.5. CLK1 uses the phospho-NLS on SRSF1 for nuclear entry	54
3.2.6. CLK1 nuclear import correlates with SRSF1 binding affinity	56
3.2.7. Nuclear import of CLK1(Δ N).....	58
3.3. Discussion	60
3.4. Conclusions.....	64
References.....	66
Chapter 4 Oligomerization of CLK1	70
4.1. Introduction	71
4.2. Results	75
4.1.1. Oligomerization is retained in single-block CLK1 N-terminal deletions	75
4.1.2. N-terminal residues directly flanking the kinase domain do not induce CLK1 oligomerization.....	77
4.1.3. Self-association of the N-terminus is length dependent but sequence insensitive.....	77
4.1.4. Aromatic residues drive N-N interactions.....	81
4.1.5. N-terminus interacts with CLK1 kinase domain.....	83
4.1.6. N-K interactions drive oligomerization in CLK1(Y-L)	85
4.1.7. Electrostatic interactions drive N-K interactions	87
4.1.8. The kinase domain is protected from solvent exchange by the N- terminus.....	89
4.3. Discussion	96
4.4. Conclusions	102
References	103
Chapter 5 Phosphorylation kinetics of CLK1	107
5.1. Introduction	108
5.2. Results	113

5.2.1. Deletion of any single block does not impair SRSF1 hyper-phosphorylation or binding	113
5.2.2. Deletion of any two blocks does not affect hyper-phosphorylation but impairs binding	117
5.2.3. SRPK1 activates CLK1 deletions.....	120
5.2.4. Low ionic strength enhances hyper-phosphorylation.....	124
5.2.5. Weakened oligomerization in CLK1 impairs hyper-phosphorylation.....	127
5.3. Discussion.....	130
5.4. Conclusions.....	134
References.....	135
 Chapter 6 Conclusions	 137
6.1. Conclusions.....	138
6.2. Future directions	139
References.....	141

LIST OF ABBREVIATIONS

RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
SMA	Spinal Muscular Atrophy
SMN	Survival Motor Neuron
ASOs	Anti-Sense Oligos
BCL	B cell Lymphoma
snRNPs	small nuclear Ribonucleoproteins
U2AF	U2 Auxiliary Factor
RRM	RNA Recognition Motif
ESE	Exonic Splicing Enhancer
ISE	Intronic Splicing Enhancer
hnRNP	heterogenous nuclear Ribonucleoprotein
NMR	Nuclear Magnetic Resonance
CLK	Cdc-2 like Kinases
NMD	Non-sense Mediated Decay
TRN-SR2	Transportin-SR2
PP1	Protein Phosphatase 1
SRPK	SR Protein Kinase
MAPK	Mitogen Activated Protein Kinase
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis

NLS	Nuclear Localization Sequence
DLS	Dynamic Light Scattering
SEC	Size Exclusion Chromatography
MBP	Myelin Basic Protein
GSH	Glutathione
GST	Glutathione S-Transferase
GTP	Guanosine Triphosphate
IBB	Importin β Binding Domain
ERK2	Extracellular Signal Regulated Kinase-2
SID	Spacer Insert Domain
RFP	Red Fluorescent Protein
GFP	Green Fluorescent Protein
Hsps	Heat Shock Proteins
CaMKII	Ca ²⁺ / Calmodulin (CaM)-dependent protein kinase II
RIP2	Receptor Interacting Protein Kinase 2
CARD	Caspase Activation and Recruitment Domain
NOD1	Nucleotide-binding Oligomerization Domain-containing protein 1
IDR	Intrinsically Disordered Region
LLPS	Liquid-Liquid Phase Separation
FUS	Fused in Sarcoma
Rh	Hydrodynamic Radius
ND	Non-Deuterated
FD	Fully Deuterated

ESI	Electron Spray Ionization
HPLC	High Performance Liquid Chromatography
MALDI	Matrix Assisted Laser Desorption Ionization
K_m	Michaelis constant
K_i	Inhibition constant
PDB	Protein Data Bank
TEM	Transmission Electron Microscopy
HGPS	Hutchinson-Gilford Progeria Syndrome
DMD	Duchenne Muscular Dystrophy

LIST OF FIGURES

Figure 1.1: Proteomic diversity arises from different Alternative Splicing events.....	4
Figure 1.2: Splicing happens through two trans-esterification reactions.....	6
Figure 1.3: Assembly of Spliceosome	8
Figure 1.4: SR proteins have RRM and RS domains	10
Figure 1.5: CLKs and SRPKs phosphorylate SR proteins.....	16
Figure 1.6: Subcellular localization of SRSF1 is regulated by RS domain phosphorylation levels	17
Figure 1.7: Disordered N-terminus of CLK1 modulates its subcellular localization	20
Figure 1.8: N-terminus regulates oligomerization and substrate phosphorylation	21
Figure 3.1: The classical nuclear import mechanism.....	43
Figure 3.2: CLK1 does not have a short, classical NLS	47
Figure 3.3: Mapping broad regions in the N-terminus necessary for nuclear localization.....	49
Figure 3.4: CLK1 nuclear import is mediated by TRN-SR2	53
Figure 3.5: CLK1-TRN-SR2 interaction is mediated by phospho-SRSF1	55
Figure 3.6: Disruption of the SRSF1 NLS drives CLK1 into the cytoplasm	57
Figure 3.7: Deletion of residues in the CLK1 N-terminus weakens SRSF1 binding	59
Figure 3.8: Mutation of NLS3 impairs the nuclear import of CLK1(Δ N), but not CLK1.....	61
Figure 3.9: Summary of CLK1 nuclear import mechanism	63
Figure 4.1: Interactions driving protein association	73
Figure 4.2: Single block deletion constructs of CLK1 can form oligomers	76
Figure 4.3: Oligomerization of CLK1 two-block deletion constructs	78

Figure 4.4: Oligomerization of GST-N terminus constructs	80
Figure 4.5: Aromatic interactions mediate N-N interaction	82
Figure 4.6: N-terminus interacts with kinase domain.....	84
Figure 4.7: Oligomerization of CLK1(Y-L).....	87
Figure 4.8: N-K interactions are driven by electrostatic interactions	88
Figure 4.9: H/D Exchange experiment	90
Figure 4.10: Shift in the mass envelope of peptide 441-450 in CLK1 and CLK1(Δ N) .	92
Figure 4.11: Deuterium incorporation plots.....	94
Figure 4.12: Kinase domain is protected in the presence of N-terminus: Deuterium uptake heat maps.....	95
Figure 4.13: Alignment of N-terminal sequences of all the four CLK1 isoforms	98
Figure 4.14: N-N and N-K interactions drive CLK1 oligomerization.....	101
Figure 5.1: CLK1 N-terminus induces hyper-phosphorylation	114
Figure 5.2: Single block deletion mutants retain substrate binding and hyper-phosphorylation.....	116
Figure 5.3: CLK1 two-block deletions retain hyper-phosphorylation but substrate binding is impaired	119
Figure 5.4: SRPK1 binding induces the formation of larger CLK1 oligomers	121
Figure 5.5: SRPK1 activates CLK1 single block deletions	122
Figure 5.6: SRPK1 activates CLK1 two-block deletions	123
Figure 5.7: Hyper-phosphorylation is enhanced under lower salt conditions	126
Figure 5.8: Hyper-phosphorylation is impaired in CLK1(Y-L)	128
Figure 5.9: SRSF1 hyper-phosphorylation is not mediated by π - π stacking interactions between the enzyme and the substrate	129
Figure 5.10: Summary of the correlation between quaternary structure and phosphorylation kinetics of CLK1	132

ACKNOWLEDGEMENTS

My journey through graduate school has had its fair share of peaks and valleys, but getting to the finish line wouldn't be a reality without some wonderful people. First and foremost, I should thank my advisor Prof. Joseph Adams for being a wonderful mentor and a patient guide. He has always appreciated my little successes and has been supportive through my failures. I also owe my co-advisor Prof. Patricia Jennings for bringing me to Adams lab and for all her guidance and mentorship throughout graduate school. Both Joe and Pat, in addition to being brilliant scientists, are also compassionate mentors and I am lucky to have them as my scientific parents.

I would like to thank my committee members Prof. Gourisankar Ghosh, Prof. Partho Ghosh, Prof. Seth Cohen, and Prof. Roger Sunahara for being a part of my doctoral committee and for enriching my work with their thoughtful suggestions.

I would like to thank Brandon Aubol for his pleasant company, mentorship and all the moral support throughout the last four years. Brandon trained me on all experimental techniques when I first started out in the lab. His expertise as a bench scientist, to get long experiments to work in the least amount of time possible is something that amazes me even today. I also owe a great deal to Kendra Hailey, who was an inevitable part of the Jennings lab. My fateful meeting with Kendra is what eventually brought me to Adams lab. Kendra's selfless generosity and willingness to help others is very unique. I am grateful for all the time and effort that she spent training me on H/D exchange experiments and also for the warmth and care that she had for me.

I thank Laurent Fattet from Yang lab for all the help with confocal microscopy. One of the privileges of working in Adams lab has been our close proximity to Sunahara Lab. I have

received extensive help from Daniel Meyer, Pranesh Venkatraman, and Nicolas Villanueva with baculovirus-insect cell expression systems. Sheng Li and Sarah Urata from the DXMS facility spent a lot of time training me on mass spectrometry and helping me run the samples on the LC-MS instrument. I have put them through hard times when my sample precipitated and clogged their pepsin column so I know, writing a thank you note on this page is the least that I could do for them.

When I look back into my graduate school life, there is also one more person that I will always remember with profound gratitude-Prof. Judy Kim. Judy held my back and offered help in every possible way at a time when I hit the rock bottom of my grad school life. Beyond a shadow of doubt, I can say that if she didn't support me through those hard times, I wouldn't have written this dissertation today.

The role that my undergraduate institution played in anchoring me to science and research is immeasurable. My educators from college have been very influential in shaping my career. Prof. George Thomas, my master's thesis advisor deserves a special mention here as his persistent encouragement was one of the pivotal reasons that led me to pursue a career in research.

There have been moments in my life, when this sudden spur of realization hits me that I am truly blessed for the friends that I have. Life outside lab in La Jolla was fun, thanks to my friends Ahanjit, Mariam, Sophia, Shruthi, Shouvik, and Sridip. I could always count on them for any kind of support. I consider myself very privileged for the kind of friendship that I still have with some of my friends from college- my friendship with them have lasted for more than ten years and there isn't a better group of friends that I will ever have a deeper connection with.

Words fail to express my gratitude to my family. My parents and my brother have been

my constant source of emotional support over all these years, despite being on the other side of the planet. The depth of their unconditional love and affection knows no bounds and I wouldn't be where I am today without them.

Chapter 3, in part, is a reprint of the material as it appears in George A, Aubol BE, Fattet L, Adams JA. “Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate”, *J Biol Chem.* **2019**, *294*, 9631–41. The dissertation author was a primary investigator and author of this paper.

Chapter 4 contains unpublished material with co-authors. George, Athira; Li, Sheng; Adams, Joseph A. The dissertation author was a primary investigator and author of this paper.

Chapters 4 and 5, is currently being prepared for submission for publication of the material: George A, Aubol BE, Adams JA. The dissertation author was a primary investigator and author of this paper.

VITA

Education

Ph.D. in Chemistry **2015-2021**
University of California San Diego

BS-MS Dual Degree in Chemistry **2010-2015**
Indian Institute of Science Education and Research Thiruvananthapuram

Publications

1. **George A**, Aubol BE, Fattet L, Adams JA. Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate. *J Biol Chem.* 2019;294(24):9631–41.
2. Thirunavukkuarasu S, **George A**, Thomas A, Thomas A, Vijayan V, Thomas KG. InP Quantum Dots: Probing the Active Domain of Tau Peptide Using Energy Transfer. *J. Phys. Chem. C* 2018; 122(25):14168–76.
3. Churchfield LA, **George A**, Tezcan FA. Repurposing proteins for new bioinorganic functions. *Essays Biochem.* 2017; 245–58.
4. Gupta H, Badarudeen B, **George A**, Thomas GE, Gireesh KK, Manna TK. Human SAS-6 C-Terminus Nucleates and Promotes Microtubule Assembly in Vitro by Binding to Microtubules. *Biochemistry.* 2015;54(41):6413–22.
5. Franke M, Leubner S, Dubavik A, **George A**, Savchenko T, Pini C, et al. Immobilization of pH-sensitive CdTe Quantum Dots in a Poly (acrylate) Hydrogel for Microfluidic Applications. *Nanoscale Res Lett.* 2017; 12, 314.

ABSTRACT OF THE DISSERTATION

Insights into the Role of a Disordered N-Terminus in the Functional Regulation of CLK1 Kinase

by

Athira George

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Joseph Adams, Chair
Professor Patricia Jennings, Co-Chair

SR proteins are a family of splicing factors that plays important roles in regulating alternative splicing. The function of SR proteins is heavily regulated by the extent of phosphorylation of its C-terminal RS domain. Cdc-2 like kinases (CLKs) are known to hyperphosphorylate the RS domains and control the splicing functions of SR proteins. CLKs have a folded kinase domain and an N-terminus that is predicted to be disordered. The N-terminus plays vital roles in regulating the function of CLK1 in various aspects.

The studies presented here elucidate how the N-terminus modulates three different aspects of CLK1 function: sub-cellular localization, self-association to form oligomers, and substrate phosphorylation. Although prior studies had shown that the N-terminus was important for modulating the nuclear import of CLK1, the mechanism for this change in subcellular localization

had largely been uninvestigated. Here, we show that CLK1 lacks a short, classical nuclear localization sequence (NLS), indicating that the nuclear import is not mediated by the classical importin α/β system. Instead, we show that CLK1 enters the nucleus by forming a complex with its physiological substrate SRSF1, an SR protein prototype, in the cytoplasm and transportin SR-2 (TRN-SR2) imports the kinase-substrate complex into the nucleus. Previous studies from our laboratory had shown that the N-terminus induces oligomerization of CLK1, which helps the kinase select its physiological substrates over non-physiological ones. The nature of the interactions underlying this oligomerization was investigated and our results show that CLK1 oligomerization is driven not only by self-association of the N-terminus (N-N interactions) but also by interactions between the N-terminus and the kinase domain (N-K interactions). While interactions between the N-termini are mediated solely by aromatic residues, interactions between the N-terminus and kinase domain are electrostatic in nature. Lastly, we also investigated the role of the N-terminus in regulating SRSF1 hyper-phosphorylation. Our results show a strong correlation between CLK1 quaternary structure and substrate phosphorylation activity. While substrate binding affinity is solely regulated by the length of the N-terminus, the velocity of hyper-phosphorylation is tightly regulated by the quaternary structure of CLK1 oligomers. Our studies demonstrate that the N-terminus of CLK1 is highly versatile. It is important not only for recognizing a broad range of RS domains for essential SR protein hyper-phosphorylation but also for CLK1 nuclear localization through substrate “piggybacking.”

Chapter 1

Introduction

1.1. Alternative Splicing and Proteomic Diversity

The flow of genetic information from genes to proteins, as proposed by the ‘central dogma’ in biology, was thought to go via a ‘one gene-one RNA-one protein’ mechanism (1). One of the challenges to our understanding of this dogma was posed by the surprising discovery that the protein coding units on eukaryotic DNA were not present as a continuous stretch (2). Instead, the protein coding sequences were interrupted by intervening, silent DNA that are not expressed in the translated product. The need for having this ‘genes in pieces’ architecture was a long-standing conundrum in molecular biology as it was also observed that the intervening sequences, later called introns, were much longer than the expressed sequences called exons. The presence of introns also necessitates a process and a complex machinery to excise the introns and stitch together the exons to make mature RNA. Such complex modifications involving cutting and ligation of RNA, a process termed splicing, were first reported in adenoviruses (3,4). Later, the first evidence for splicing in eukaryotic cells came with the discovery of different isoforms of the hormone calcitonin (5). It has been almost four decades since this discovery and we now understand that splicing is ubiquitous in the eukaryotic world and is the primary mechanism that allows for the generation of a diverse set of proteins to be made from a limited set of genes (6–8).

Two types of mRNA splicing can happen depending on how splice sites are selected (8) (Fig. 1.1A). In constitutive splicing, all introns are excised and exons are ligated in the order in which they appear on the pre-mRNA transcript. However, in alternative splicing, splicing sites are variably selected resulting in multiple mRNA transcripts from a single pre-mRNA. This variable selection can happen through different kinds of splicing events which include exon skipping, intron retention, alternative selection of 5’ or 3’ splice sites, inclusion of 5’ or 3’ untranslated regions (UTRs), or mutual exclusion of exons (9,10). The mRNA transcripts generated from different

alternative splicing events translate into different protein isoforms which can have different structure, stability, expression, localization and sometimes even result in drastically different phenotypes (11) (Figure 1.1A). Alternative splicing is the primary process credited to be responsible for proteomic diversity, as more than 90% of protein coding genes in human genome are alternatively spliced (8).

1.2. Errors in Splicing and Human Diseases

The caveat of having such an intricate system to enhance proteomic diversity is that any small defects in splicing can result in lethal mutations in the mRNA transcript, leading to diseases such as cystic fibrosis, frontotemporal dementia, Parkinson's, various kinds of laminopathies and cancer (12–14). Some of the splicing related disorders arise from mutations in the pre-mRNA transcript, while others arise due to the mutations in the proteins that bind the pre-mRNA forming the splicing machinery (15). Understanding alternative splicing and its regulation mechanisms are of vital importance as they can be useful in developing therapeutic approaches targeting these splicing disorders (16). Stemming from the realization that about 15% of genetic diseases arise from splicing defects, two kinds of approaches have been made to target spliceopathies: (a) Using antisense oligos (ASOs) that bind to specific RNA sequences and redirect splicing machinery and (b) small molecules that alter the binding of RNA binding proteins. A notable example for a disease that has been successfully treated through designed ASOs is spinal muscular atrophy (SMA). SMA is a disease that is characterized by the progressive loss of motor neurons leading to paralysis and mortality. The root of SMA has been tracked down to the formation of a truncated, unstable form of the survival motor neuron (SMN) protein. The unstable variant of SMN is made as a result of a mutation in exon 7, leading to its skipping (Figure 1.1B). ASOs designed to target the pre-mRNA of the SMN gene have been shown to enhance Exon 7 inclusion, generate the stable

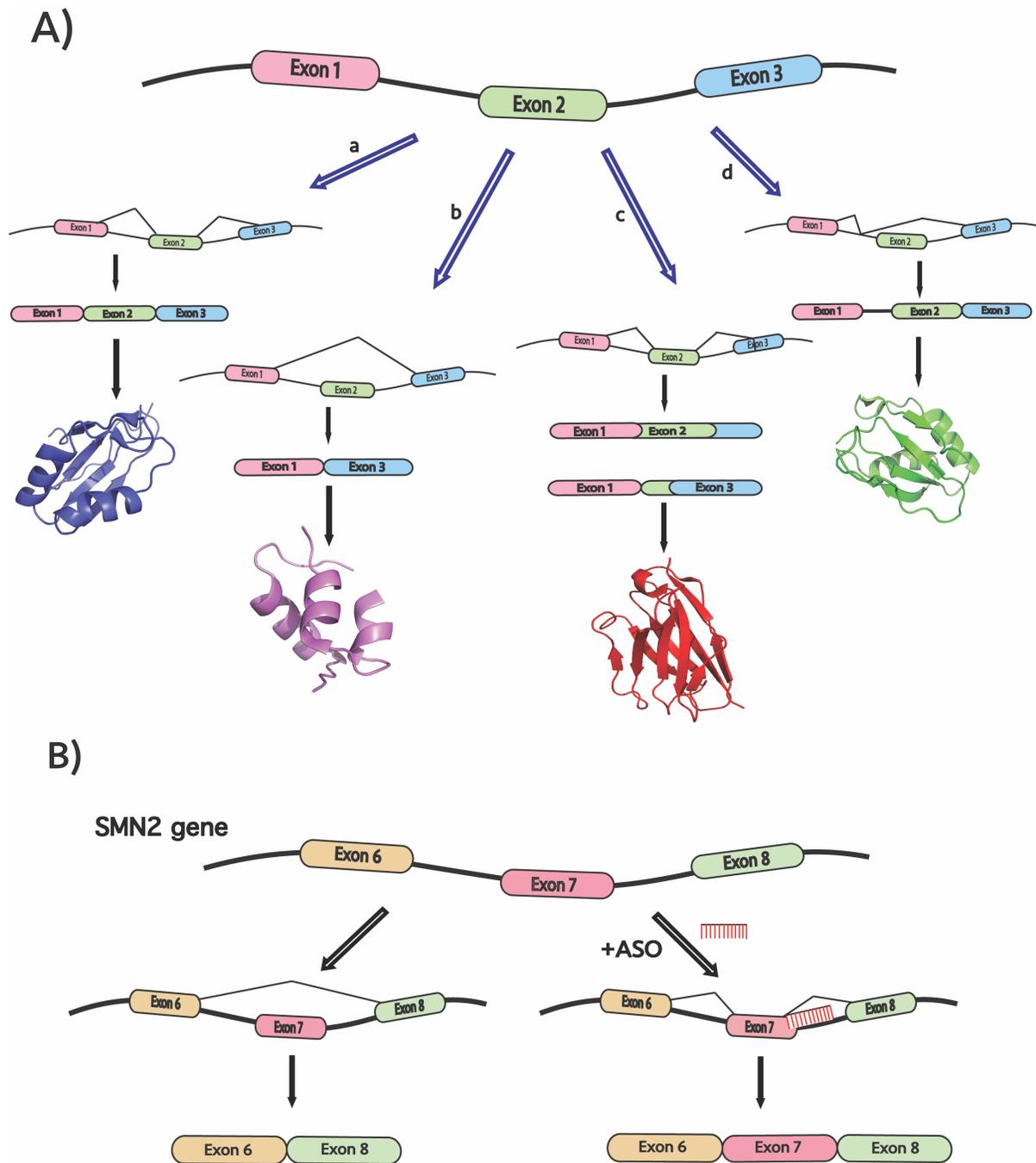


Figure 1.1: **A)** Proteomic diversity arises from different Alternative Splicing events: (a) Exon inclusion (b) Exon skipping (c) Alternative 5' and 3' splice sites and (d) Intron retention **B)** ASOs promote inclusion of Exon 7 in SMN2 gene.

full-length SMN protein, and improve motor neuron function (17,18). The second therapeutic approach, using small molecule drugs, primarily act on proteins that form the splicing machinery. Amiloride, a small molecule drug used to treat high blood pressure, was initially identified in a screen targeted to select for splicing modulators of BCL-X (B-cell lymphoma) gene (15). Amiloride treated cells showed altered expression and phosphorylation levels of several splicing factors and regulatory kinases. Because splicing is one of the critical steps in gene expression, drugs inhibiting splicing have also been reported to have anti-tumorigenic activity (15).

1.3. Spliceosome: Assembly and Catalysis

Splicing happens at the spliceosome, a macromolecular complex composed of more than 100 proteins and five snRNPs (19). The spliceosome is the largest ribonucleoprotein (RNP) machine in the nucleus and catalyzes the excision of introns and ligation of exons through the meticulously coordinated association, rearrangement, and dissociation of its components (9). Understanding the spliceosomal machinery and its function has sparked considerable interest in the scientific community and attempts have been made to gather insights into the structural details of this complex. Owing to the tremendous advances recently made in electron microscopy, many of the attempts made towards structural characterization of the spliceosomal complex trapped in its various intermediate states have yielded outstanding results. Consequently, our understanding of the spliceosomal machinery has taken significant leaps in the past decade (19–23).

The process of splicing happens through two trans-esterification reactions. In the first step, the 2' OH of a conserved adenosine located on the intron being spliced makes a nucleophilic attack on the phosphate at the 5' end of the exon, resulting in the release of the 5'exon and the concomitant formation of a looped structure of the intron known as the intron lariat. In the second trans-esterification, the 3' OH from the detached exon at the 5' splice site attacks the phosphate at

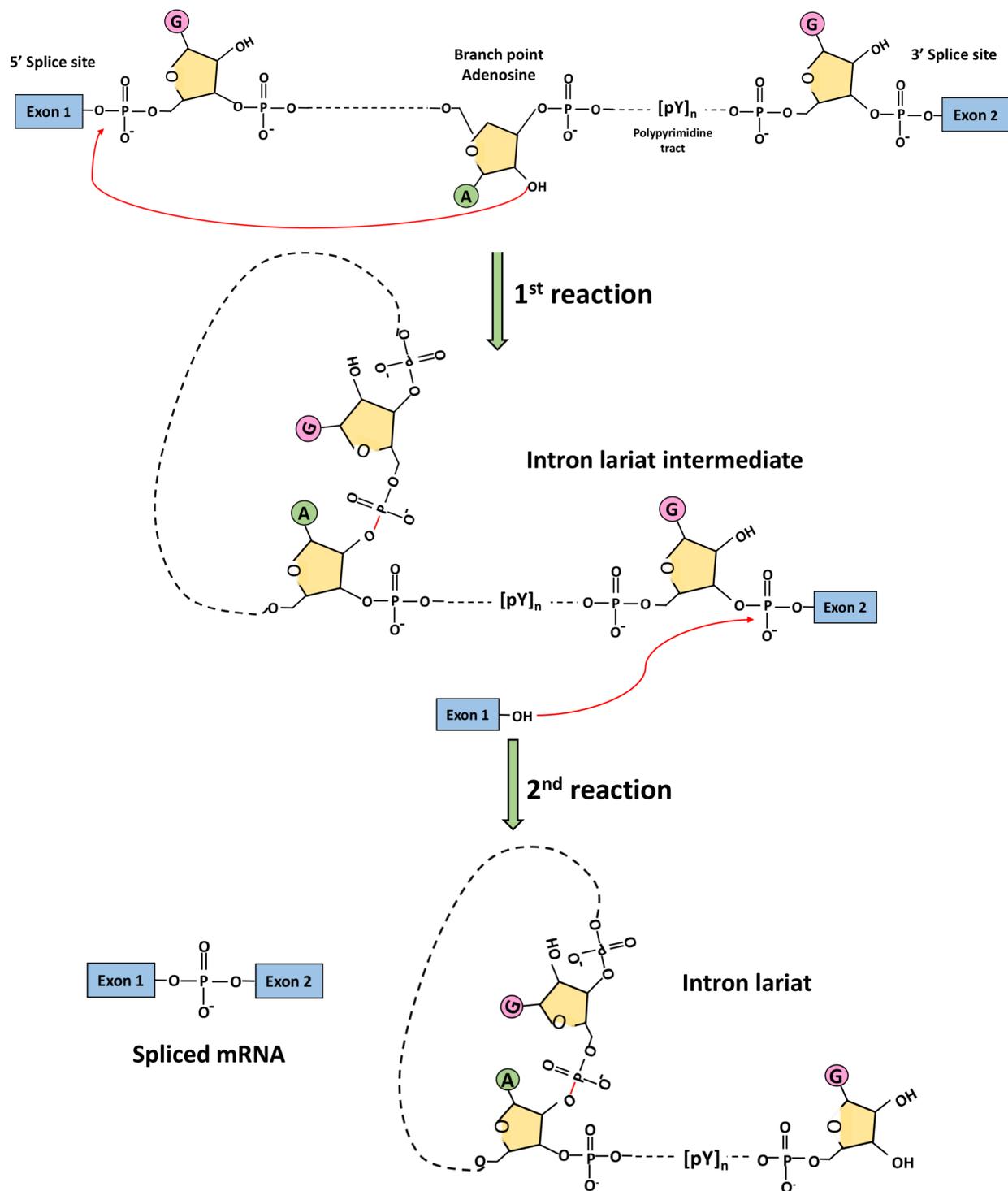


Figure 1.2: Splicing happens through two trans-esterification reactions: The first reaction leads to the formation of an intron lariat intermediate and a free exon. In the second reaction, the two exons are connected through a phosphodiester bond and the intron is released.

the 3' splice site resulting in the formation of a phosphodiester bond between the two exons and the release of the intron lariat (9) (Figure 1.2).

The initiation and arguably the most important step in splicing begins with the correct identification of authentic 5' and 3' splice sites over cryptic sites. The exon-intron boundaries in metazoan pre-mRNAs are characterized by the presence of several consensus sequences that help define exon-intron boundaries and the binding of spliceosomal proteins guided by base-pairing interactions. Apart from the 5' and 3' splice sites, two conserved sequences located a few nucleotides upstream of the 3' splice site, known as the branch point sequence (BPS) and a polypyrimidine tract (rich in uridines and cytosines) also act as reactive sites. The initiation of spliceosome assembly happens by the binding of U1 snRNP at the 5' splice site guided by base-pairing interactions (Fig. 1.3A). The two subunits of the heterodimer U2 Auxiliary factor (U2AF), U2AF65 and U2AF35 also bind to the BPS and the 3' splice site, respectively, leading to the formation of the pre-spliceosomal E complex. Next, U2 snRNP binds to the BPS to form the A complex. In subsequent steps, the U4/U5.U6 tri-snRNP is recruited to form the B complex in an ATP-dependent manner. The B complex undergoes several rearrangements to form the activated complex B* that catalyzes the first transesterification reaction. The resulting complex with the looped-out intron lariat bound to the 3' splice site is the C complex which subsequently carries out the second transesterification reaction (Fig.1.3B). The intron lariat then diffuses from the complex and new complexes are assembled at the new sites for further splicing (9) (Figure 1.3A).

1.4. SR Proteins are Essential Splicing Factors

In addition to snRNPs, one family of proteins that is known to play essential roles in defining intron-exon boundaries is the SR protein splicing factors (24,25). They were first identified as a group of factors that could restore splicing activity in splicing-deficient HeLa cell

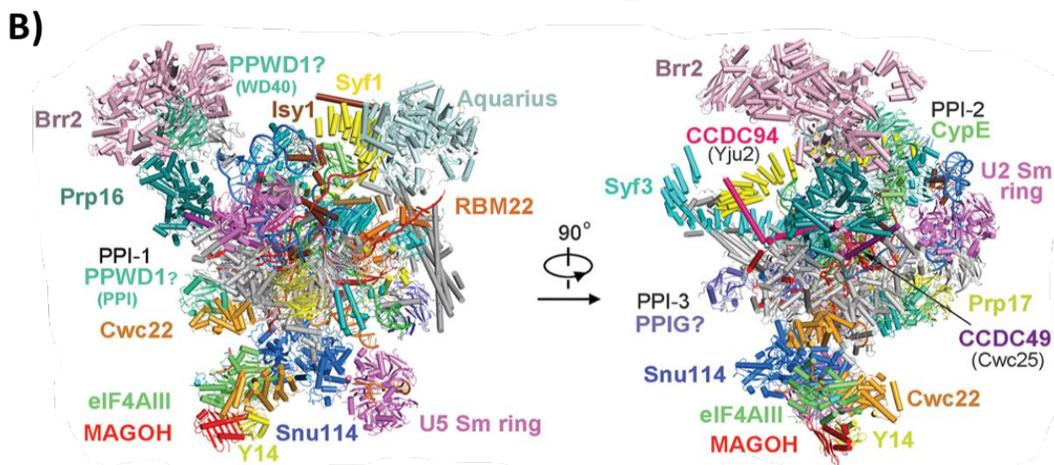
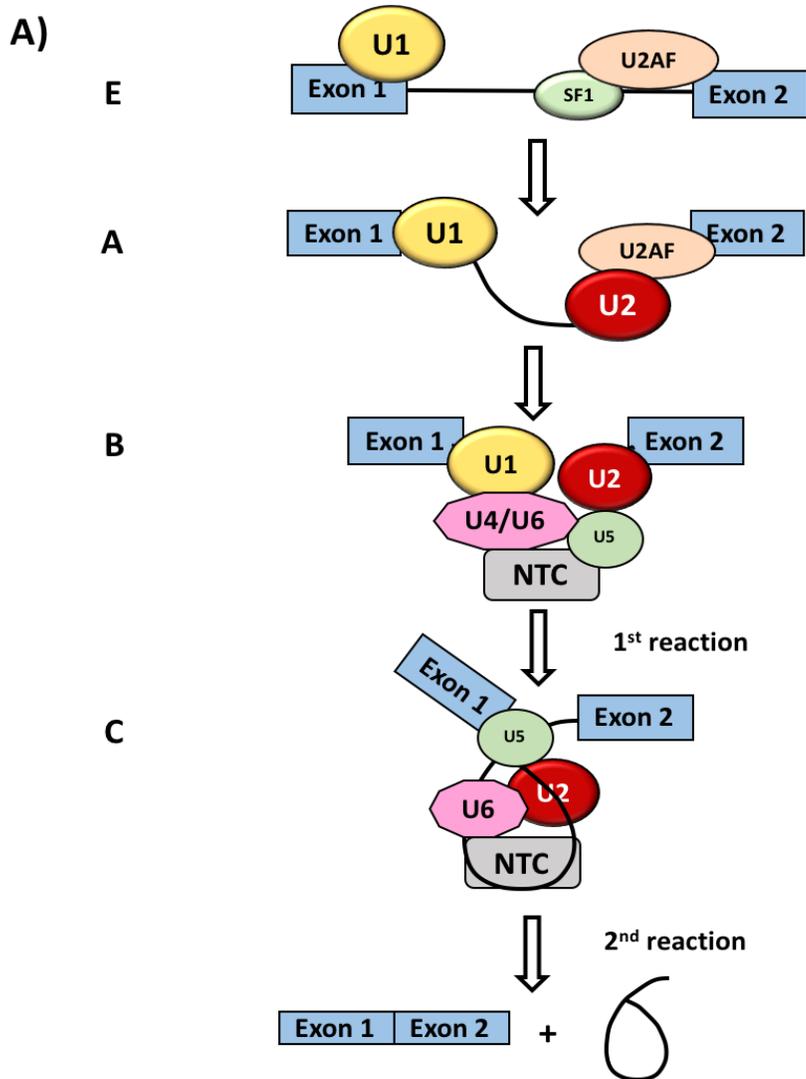


Figure 1.3: Assembly of spliceosome **A)** Assembly and major structural rearrangements in the spliceosome as it transitions from E→A→B→C complex to facilitate the two transesterification reactions. **B)** Structure of the human spliceosome C complex. (Adapted from 23).

S100 extracts (26). The SR protein family consists of 12 proteins, all characterized by the presence of one or two N-terminal RNA recognition motifs (RRMs), a glycine-rich linker and a C-terminal RS domain (Figure 1.4A,B). The RRM of SR proteins bind to the exon-intron boundaries by recognizing exonic splicing enhancers (ESE) or intronic splicing enhancers (ISE), which are sequences located near exon-intron boundaries that promote exon inclusion. There are two models proposed to explain how SR proteins regulate alternative splicing. One is a recruitment model where SR proteins recruit U1 snRNP to the 5' site and U2AF at the 3' site helping to stabilize the early spliceosomal complex. SR proteins bound to the 5' and 3' splice sites interact through their RS domains and bring the splice sites together by looping out the intervening intronic or exonic sequences (24). The second model is an inhibitory model where SR proteins antagonize the activity of splicing repressors like hnRNPs (27). hnRNPs bind to splicing silencer sequences known as Exonic and Intronic Splicing silencers and promote exon skipping. The ratio of SR protein to hnRNPs present at the splice sites are decisive factors in determining the inclusion or exclusion of exons (27). The relative expression levels of SR proteins and hnRNPs also allow for the tissue-specific expression of different isoforms of a protein (24).

Structures of the RRM of some SR proteins have been solved using crystallography and NMR spectroscopy (28). SRSF1 is the archetypal SR protein and has been studied extensively. The structure of RRM1 in SRSF1 follows a canonical RNA binding fold comprising four beta sheets and two alpha helices. Some SR proteins, including SRSF1, have a second RRM commonly referred to as pseudo RRM that are thought to enhance the RNA binding specificity. The SRSF1 pseudo RRM has a non-canonical RNA recognition mechanism that binds RNA through residues on an α helix rather than β strands (Figure 1.4C) (29).

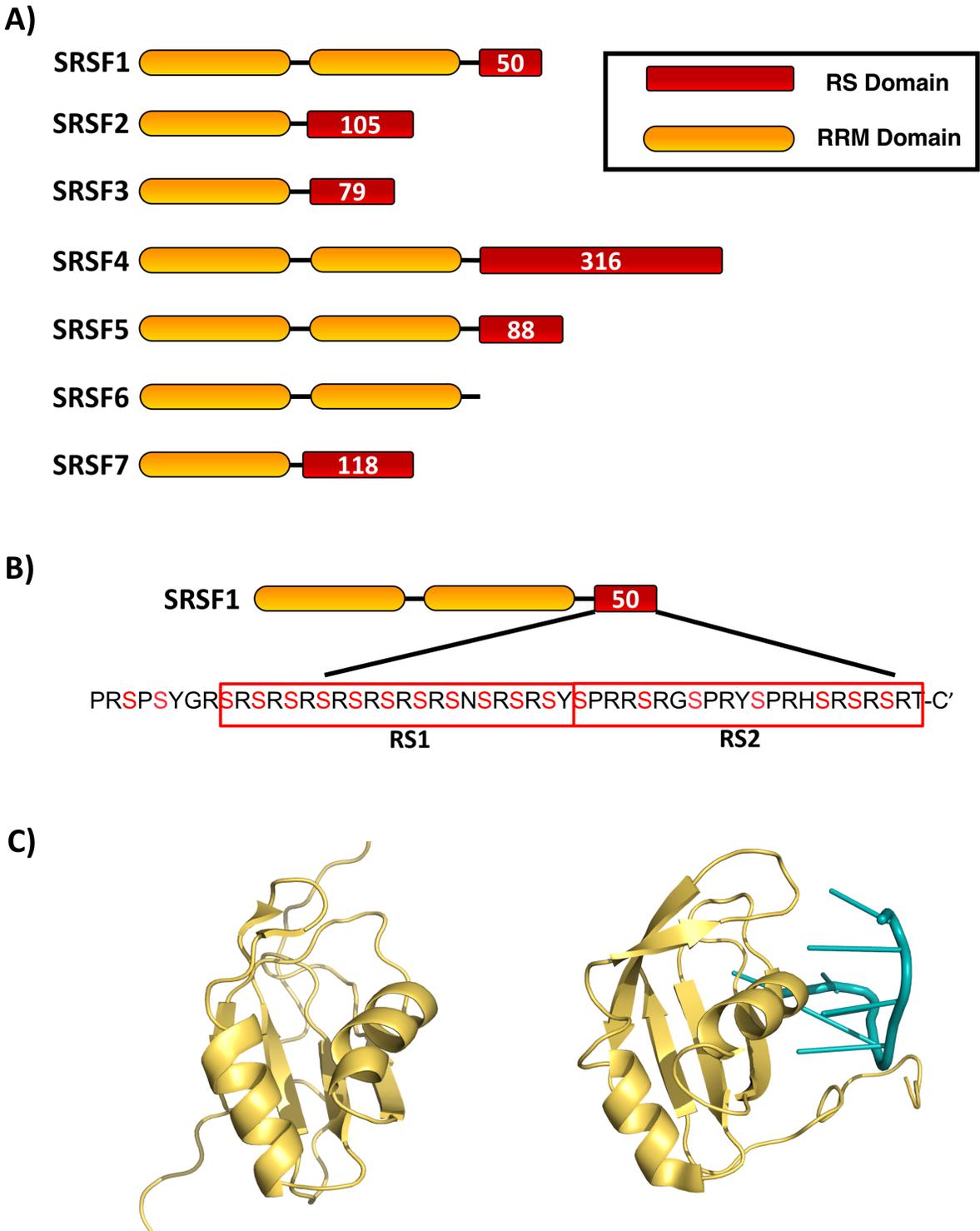


Figure 1.4: SR proteins have RRMs and RS domains **A)** Domain organization of some SR proteins. **B)** The RS domain in SRSF1 showing sequences on RS1 and RS2 **C)** Crystal structures of SRSF1 RRM1 (left) and RRM2 bound to RNA (right). (PDB IDs: 1X4A and 2M8D).

The RS domain, named after the arginine-serine dipeptide repeats that constitute at least 40% of the domain, is a distinguishing feature of the SR protein family. Attempts made towards structural characterization of full-length SR proteins have been impeded by the aggregation prone and insoluble nature of the RS domains. Sequence analyses predict that RS domains are intrinsically disordered (30). Molecular dynamics simulations performed to determine the most stable conformation of a peptide with eight RS dipeptide repeats suggested a stable helical conformation in the unphosphorylated form. In the fully phosphorylated form, however, an ‘arginine claw’ structure where a central phosphate group surrounded by arginine residues was predicted to be the most stable conformation (31). Interestingly, although these simulations predicted some helicity, no appreciable helicity was observed for RS domains in circular dichroism experiments in either phosphorylated or unphosphorylated forms (32). Recently, the disordered nature of the RS domain was also confirmed through NMR experiments (28). The RS domain is also known to self-associate extensively and bind other proteins with RS-like domains including U1snRNP and the Cdc2- like kinases (CLKs) (24).

In addition to playing critical roles in regulating alternative splicing, SR proteins are also known to be important for a variety of other functions, making them vital for survival. Even though SR protein function often overlaps, knockouts of SRSF1, 2 and 3 are known to result in embryonic lethal phenotypes, suggesting a possible non-redundancy in function (33). SR proteins may also serve as proto-oncogenes. In one example, overexpression of SRSF1 at even low levels has been shown to lead to enhanced cell proliferation and suppressed apoptotic pathways ultimately leading to the development of tumors (34). SR proteins also play indispensable roles in virtually all aspects of mRNA maturation, including mRNA export, regulation of alternative polyadenylation of mRNAs, and nonsense-mediated decay (NMD) of mRNPs (33). The roles of SR proteins in

enhancing genome stability are also well studied. SR proteins stabilize DNA by preventing the formation of R-loops, which are hybrids of DNA and RNA. In the absence of SRSF1, cells exhibit a hyper-mutagenic phenotype (34). Even though most SR proteins are predominantly localized to the nucleus, some members of the SR protein family can shuttle from the nucleus to the cytoplasm by associating with nuclear export factors like TAP (35). These shuttling SR proteins are also known to transport spliced mRNA along with them to the cytoplasm making the processed mRNA available for translation. Shuttling SR proteins are also known to regulate translation in the cytoplasm (35).

1.5. Phosphorylation Regulates SR Protein Function

Several reports show that SR proteins undergo extensive post translational modifications in cells. For SRSF1, extensive phosphorylation of the RS domain and arginine methylation on the linker connecting the two RRMs have been reported (36–38). These post-translational modifications act as regulatory mechanisms for a wide array of functions. Paradoxically, both hypo- and hyper-phosphorylation of SR proteins have been shown to inhibit splicing (33). On one hand, SR proteins need to be hyper-phosphorylated to get recruited from their storage sites to splicing sites but they also need to be de-phosphorylated later for efficient catalysis of the transesterification reactions (39). Such findings suggest that the phosphoryl content of SR protein must be carefully regulated for splicing function. Overall, reversible phosphorylation and dephosphorylation of the serine residues in the RS domains have been shown to modulate the SR protein subcellular localization, RNA binding affinities and association with early spliceosomal components (33).

Although SR proteins are translated in the cytoplasm, they must enter the nucleus for splicing function. The nuclear import of SR proteins is controlled by the nuclear import mediator

transportin SR-2 (TRN-SR2) (40,41). TRN-SR2 binds phosphorylated residues on its cargo proteins and transports the cargo into the nucleus in an ATP-dependent manner (42). Dephosphorylation of RS1 by protein phosphatase 1 (PP1) is essential for the nuclear export of SR proteins, which also acts as an export mechanism for some mRNAs. Inside the nucleus, the SR proteins are found in membrane-less storage organelles known as nuclear speckles. Hyperphosphorylation of the RS domain induces the dissociation of SRSF1 from nuclear speckles to splicing sites (37). In addition to phosphorylation, arginine methylation on the inter-RRM linkers is also known to affect the subcellular localization of shuttling SR proteins (38). The three methylated arginines R90, R93 and R117 on SRSF1 are known to be the binding sites of export factor TAP (28). The subcellular localization of SR proteins is, thus, carefully regulated by the interplay of kinases and phosphatases.

1.6. CLKs and SRPKs Phosphorylate SR Proteins.

There are two major families of protein kinases that are known to phosphorylate the RS domain of SR proteins: The SR protein kinase (SRPK1-3) and cdc2-like kinase (CLK 1-4) (36,37). These two kinases differ in their subcellular localization, substrate specificities and phosphorylation kinetics (43,44).

1.6.1. SRPKs: In resting cells, SRPKs are predominantly found in the cytoplasm, and translocate into the nucleus under certain conditions, such as osmotic stress or epidermal growth factor (EGF) signaling. Studies have revealed that chaperones like Hsp 70 and Hsp 90 act as cytoplasmic anchors for SRPK1, which are shed upon EGF signaling leading to the nuclear translocation of the kinase (45). SRPKs have a 250-residue spacer insert domain (SID) that bifurcates the N-lobe and the C-lobe of its canonical kinase domain (Fig. 1.5 A). The deletion of

the SID only made minimal effects on the activity of the kinase, but resulted in complete nuclear translocation, implying that the chaperones acting as cytoplasmic anchors bind to the SID (46).

The crystal structure of SRPK1 bound to a truncated form of its substrate SRSF1 has been solved using crystallography (32,47) (Fig. 1.5C). The structure shows the RS dipeptides on SRSF1 bound to a negatively charged docking groove on SRPK1 that is partly formed from the residues forming the MAPK insert domain. Previous studies have shown that SRPK1 processively phosphorylates the N-terminal region of SRSF1- termed RS1 (48). The docking groove facilitates the processive phosphorylation of RS1 by sequentially feeding the dipeptide repeats into the active site. After 5-8 residues on the RS1 gets phosphorylated processively, the β 4 strand on the RRM2 of SRSF1 unfolds from the mechanical stress of the accumulating negative charge and binds the docking groove resulting in the dissociation of the phosphorylated product (49). It has also been shown through engineered protease foot printing experiments that SRPK1 phosphorylation of RS1 is strictly directional (50). SRPK1 binds to an initiation box on the C-terminal end of RS1 (residues 221-225 on SRSF1) and phosphorylation follows in a C to N terminal direction until the product release (50). SRPK1 can also phosphorylate RS dipeptides in the RS2 domain. However, the RS2 phosphorylation is distributive and is much slower than RS1 phosphorylation. As a result, single turnover experiments with SRPK1 show biphasic kinetics- a fast burst phase representing RS1 phosphorylation and a 100-fold slower phase representing RS2 phosphorylation (51). The activation loop in SRPK is relatively short and adopts a conformation that allows binding of the substrate even in the unphosphorylated state making the kinase constitutively active (47).

1.6.2. CLKs: There are four major isoforms of CLKs (CLK1-4) in humans that phosphorylate SR proteins. CLKs belong to the LAMMER kinase family because of the highly conserved 'EHLAMMERILG' sequence present in all four isoforms (52,53). CLKs are dual

specificity kinases that can phosphorylate S/T residues on its substrates and autophosphorylate Y residues (54). Structurally, CLKs differ from SRPKs by lacking an insert domain that bifurcates the N and C lobes. Instead, a unique feature of the CLK family is the presence of a long N-terminus that is predicted to be disordered based on sequence analyses (Figures 1.5A & 1.7A-C). The N-terminus is also classified as an RS-like domain owing to the presence of RS dipeptides and a lack of sequence diversity (55). The crystal structures of the kinase domains of CLK1 and CLK3 lacking their N-termini have been solved using X-ray crystallography (56) (Fig. 1.5B). One of the striking observations made from the CLK1/3 kinase domain structures is the absence of a docking groove as seen in SRPK1. There are two insertions in the CLK family kinase domain that mask potential docking motifs on the kinase domain. The first insertion is a β hairpin loop that masks accessibility to the structural analogue of the ‘D motif’ or the docking motif found in MAP kinases. The β hairpin loop is also located at the site where the insert domain is seen in SRPKs. The second insertion is a helix α H that is positioned over the LAMMER motif making it solvent inaccessible and ruling out the possibility of this motif interacting with CLK substrates (56). The absence of a folded docking groove makes the question of how CLKs bind and phosphorylate their substrates very intriguing.

Results from single turnover kinetics and lysine footprinting experiments suggest that CLK1 can phosphorylate SRSF1 at about 18-20 sites without any regiospecificity or directionality (57). The phosphorylation velocities measured also indicate that CLK1 is a much slower kinase in comparison to SRPK1. However, CLK1 is unique in its ability to phosphorylate three Ser-Pro (SP) dipeptides in RS2 in addition to phosphorylating the RS1 dipeptide repeats in SRSF1 (58). Hyperphosphorylation at SP dipeptides manifests as a gel shift when SRSF1 is phosphorylated with ^{32}P -labeled ATP and separated using low percentage SDS-PAGE. SP dipeptide phosphorylation plays

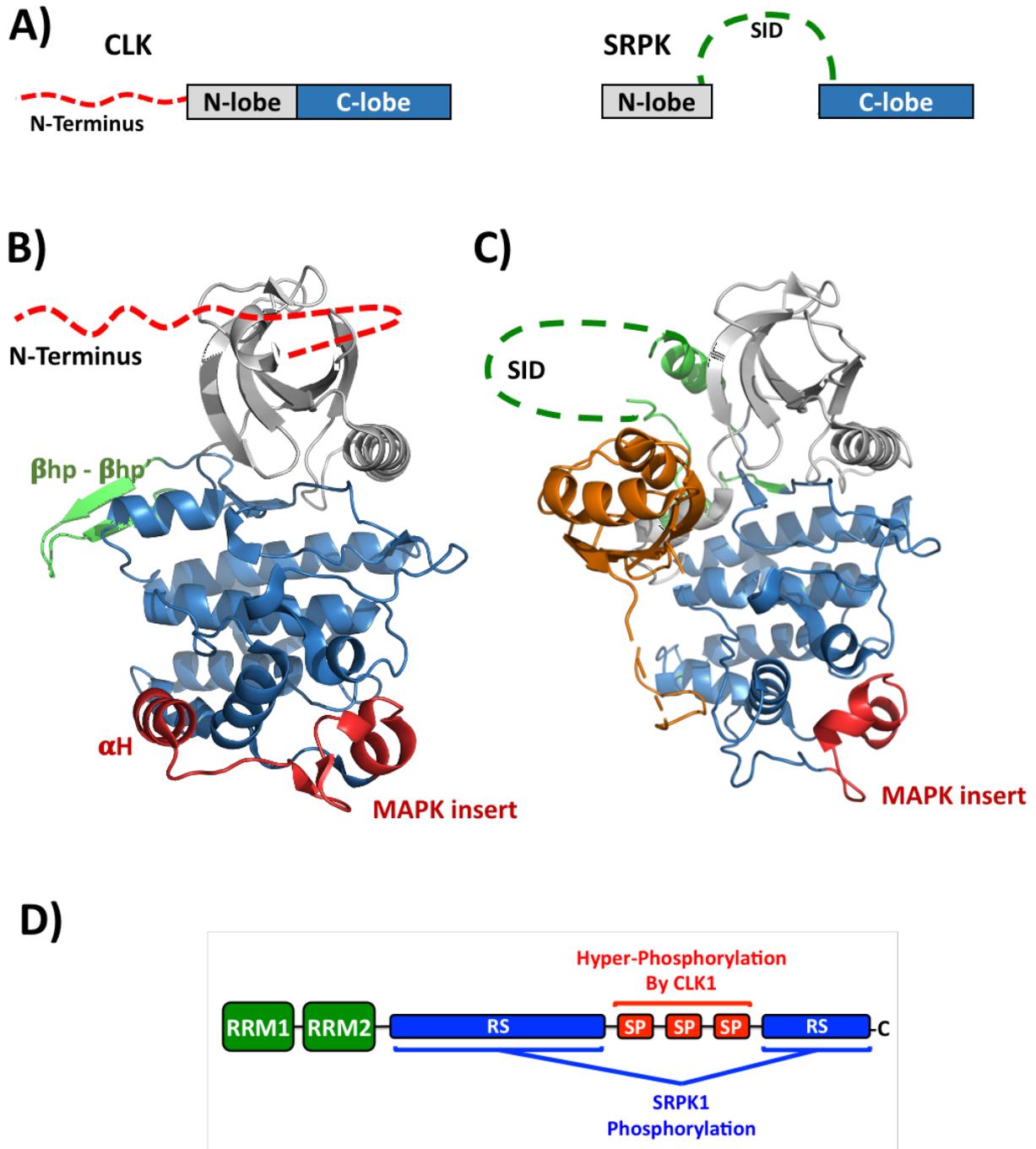


Figure 1.5: CLKs and SRPKs phosphorylate SR proteins **A)** Domain organization of CLKs and SRPKs. **B)** Crystal structure of CLK1 kinase domain (PDB ID: 1Z57) **C)** Crystal structure of SRPK1 kinase domain bound to RRM2 and RS1 of SRSF1 (PDB ID: 3BEG). In both B and C, the N and C lobes are shown in gray and blue respectively. Insertions unique to CLK1, the β hairpin loop (green) and MAPK insert (red) block access to potential docking grooves. The RRM2 and the RS1 (orange) make contacts with SRPK **D)** Phosphorylation sites of CLK and SRPK on SRSF1.

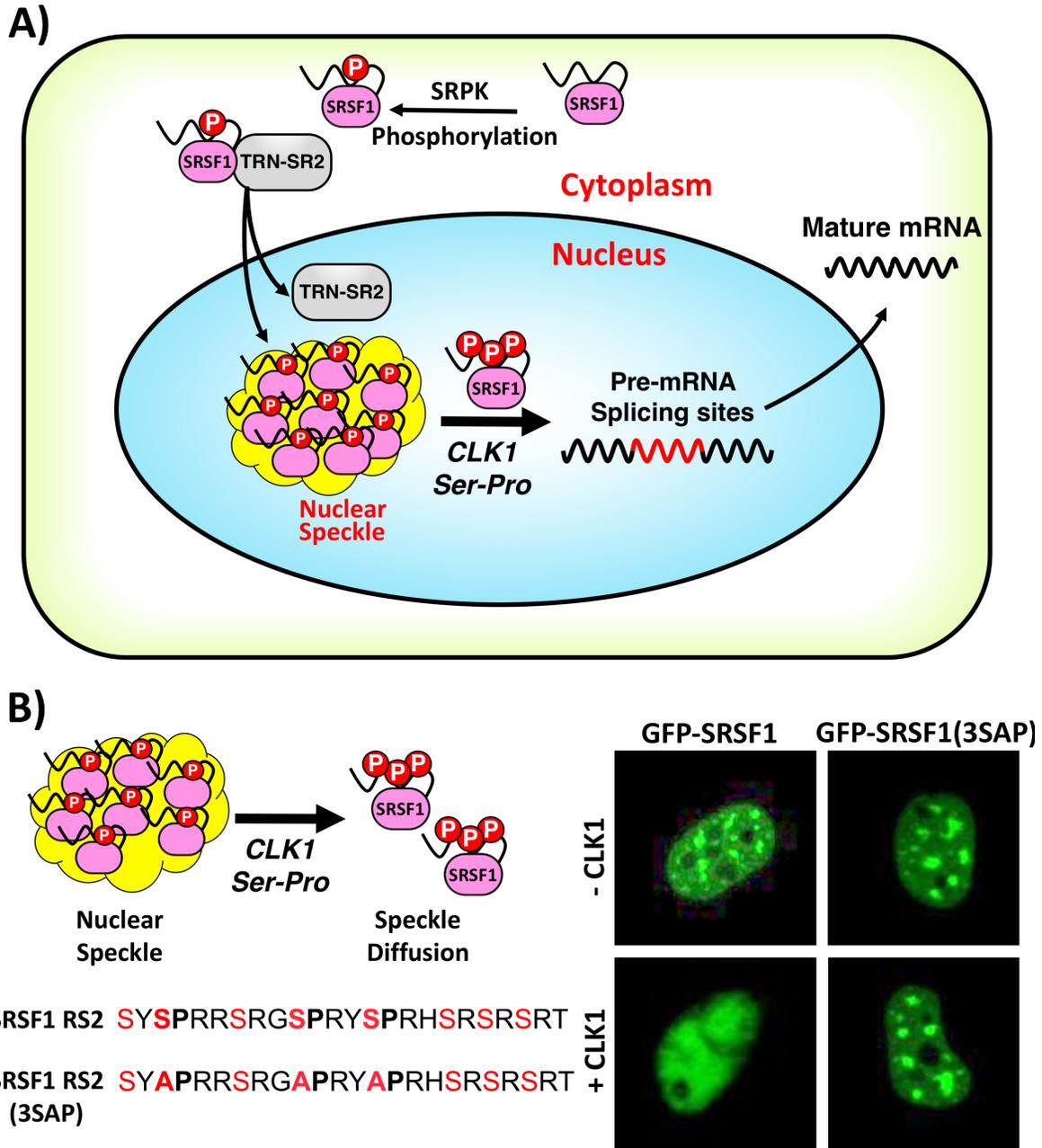


Figure 1.6: Subcellular localization of SRSF1 is regulated by RS domain phosphorylation levels. **A)** Scheme summarizing SRSF1 localization: SRPK phosphorylation drives the nuclear import of SRSF1 mediated by TRN-SR2. Hyper-phosphorylation by CLK1 leads to the mobilization of SRSF1 from speckles to splicing sites. **B)** Ser-Pro phosphorylation leads to speckle diffusion. Overexpression of CLK1 facilitates diffusion of GFP-SRSF1 speckles, but not of GFP-SRSF1 (3SAP) (Adapted from 58).

pivotal roles in regulating the subnuclear distribution of SRSF1. Phosphorylation of SP dipeptides by CLK1 results in the recruitment of SRSF1 from nuclear speckles to splicing sites. In HeLa cells transfected with GFP-SRSF1, the over expression of RFP-CLK1 leads to the dissociation of GFP-SRSF1 from speckles. However, if all three serines in the SP dipeptides in RS2 are mutated to alanines, overexpression of RFP-CLK1 does not lead to speckle diffusion (58) (Figure 1.6D). These investigations indicate that CLK1-dependent SP phosphorylation likely induces unique conformational changes that free SRSF1 from storage speckles for splicing action.

1.6.3. CLK-SRPK complex: While phosphorylated SR proteins are known to promote splice-site recognition, one of the intriguing observations made was that the overexpression of CLK1 led to weakened splice-site selection (59). One possible explanation for this aberrant observation stems from the fact that CLK1 binds to both phosphorylated and unphosphorylated forms of SRSF1 with high affinity (60). Post phosphorylation, SRSF1 stays bound to CLK1, thus making it unavailable to bind splice sites. Surprisingly, our laboratory showed that this problem is circumvented by CLK1 forming a complex with SRPK1 through its N-terminus. SRPK1 serves as a release factor by peeling off the CLK1 N-terminus from SRSF1, enabling the dissociation of phosphorylated SRSF1 (61). Overexpression of CLK1 can lead to enhanced nuclear levels of SRPK1, making CLK1 a nuclear anchor for SRPK1. The dual kinase system also shows enhanced activity towards SP dipeptide phosphorylation. The two kinases, thus, form a symbiotic relationship in the nucleus facilitating enhanced SRSF1 phosphorylation. Very recently, the CLK1 N-terminus docking motif on the SRPK1 kinase domain was also identified. CLK1 binds an acidic surface on SRPK1 kinase domain formed from highly conserved residues on the α helix G (62).

1.7. N-terminus as a Modulator of CLK1 Function

Since CLKs lack a folded docking groove, the possibility of the disordered N -terminus

serving as an SR protein recruitment domain was explored. Using a disordered domain as a docking motif is uncommon among kinases, as most kinases use the ‘two pronged’ substrate recognition mechanism have their docking motifs in the well-folded domains (63,64). In the absence of the N-terminus, many of the functions of CLK1 are significantly compromised, even though the N-terminus by itself does not possess any of the structural features essential for phosphorylation.

1.7.1. N-terminus is important for nuclear localization: Since CLK1 is a strictly nuclear kinase, it was assumed that its N-terminus contained a classical nuclear localization sequence (NLS). A previous study performed on mouse CLK1 (CLK/STY) indicates that while the full length CLK/STY localized exclusively to the nucleus, a deletion construct lacking the first 60 residues expressed in both the nucleus and cytoplasm (65) (Figure 1.7D). Within the nucleus, the N-terminus was also found to be important for the localization of CLK1 in speckles. Our laboratory showed that a kinase-inactive CLK1 accumulated in the nuclear speckles, but a mutant form lacking the N-terminus, CLK1(Δ N), but containing an NLS was found to be diffused throughout the nucleus (Figure 1.7E). Thus, both the nuclear and speckle localization of CLK1 is carefully controlled by the N-terminus (66).

1.7.2. N-terminus induces hyper-phosphorylation of SRSF1: CLK1 and CLK1(Δ N) differ drastically in their phosphorylation kinetics. While CLK1 adds about 18 phosphates on the RS domain, CLK1(Δ N) does not phosphorylate beyond 6 sites (Figure 1.8A). This effect may have its basis in altered binding affinities since CLK1 binds SRSF1 with much higher affinity (more than 20-fold) than CLK1(Δ N), based on K_m values. Pull-down assays also show that CLK1 interacts stably with SRSF1 whereas CLK1(Δ N) does not (60). The measured K_m values for ATP are comparable for CLK1 and CLK1(Δ N), suggesting that the N-terminus does not affect the ATP

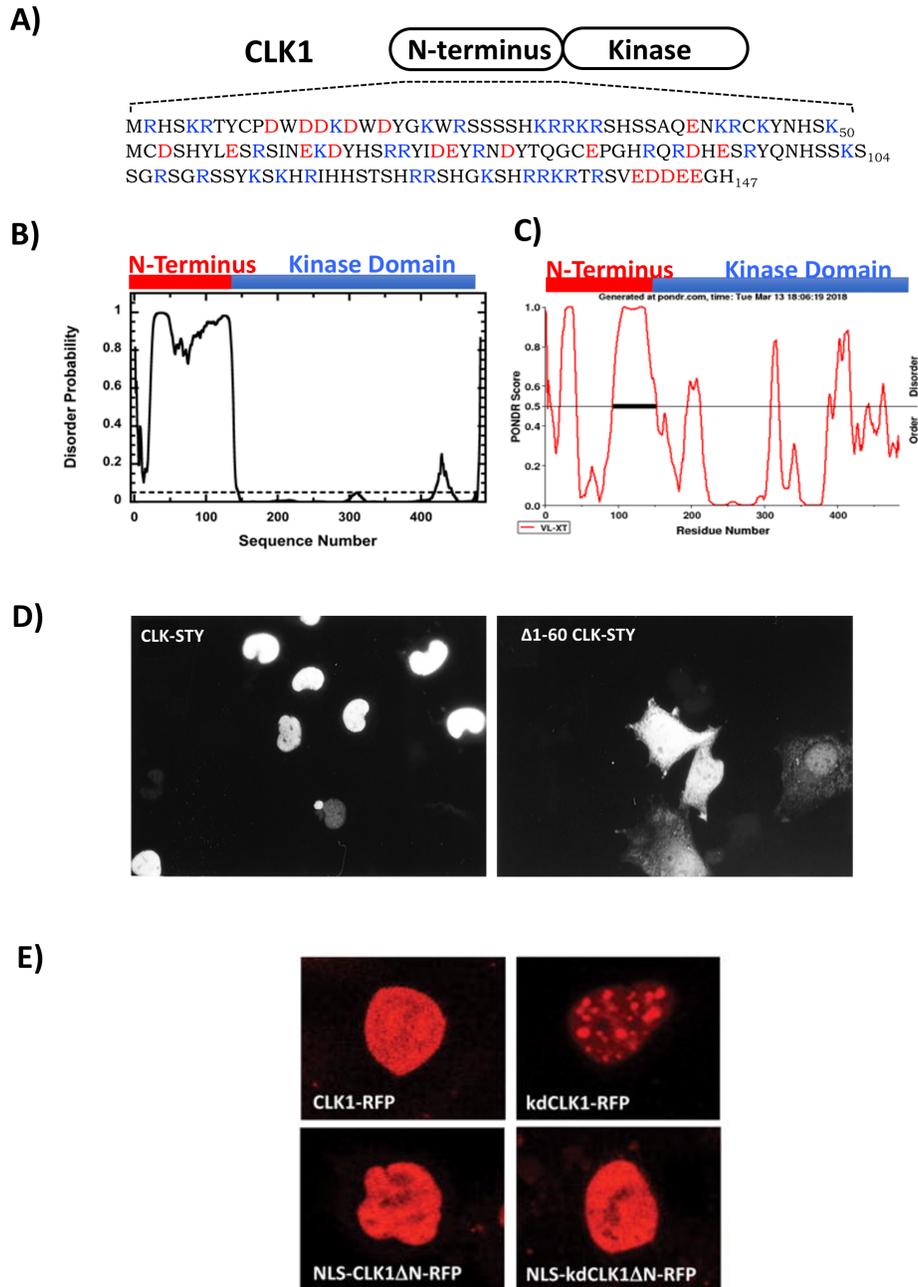
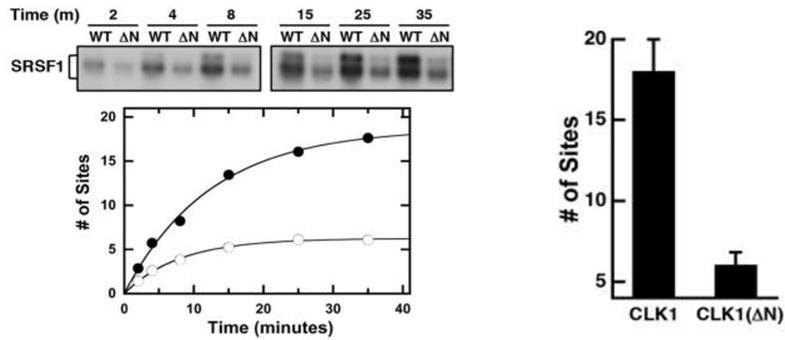
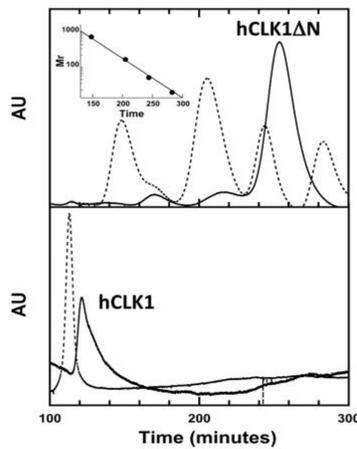


Figure 1.7 Disordered N-terminus of CLK1 modulates its subcellular localization **A)** CLK1 N-terminus **B,C)** DISOPRED (B) and PONDIF (C) predictions showing that the N-terminus is disordered. **D)** N-terminus is necessary for the nuclear import of CLK-STY: deletion of first 60 residues result in cytoplasmic localization (Adapted from 65). **E)** N-terminus is necessary for speckle localization: kinase dead CLK1 accumulates in speckles, but kinase dead Δ N CLK1 is diffused. (Adapted from 66).

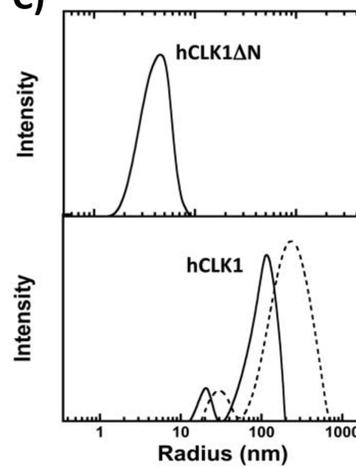
A)



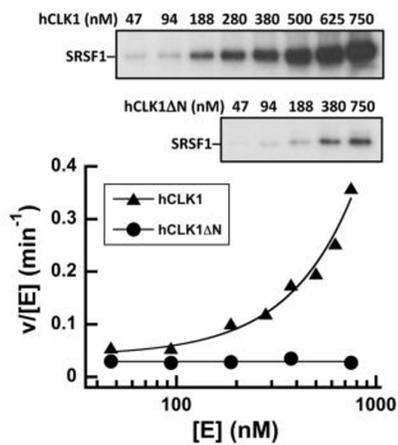
B)



C)



D)



E)

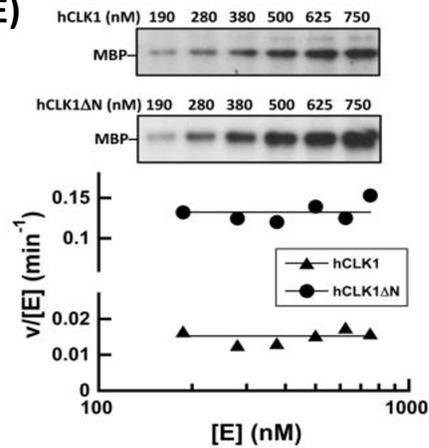


Figure 1.8 N-terminus regulates oligomerization and substrate phosphorylation **A)** CLK1 N-terminus induces hyper-phosphorylation of SRSF1 (Adapted from 65). **B,C)** Size Exclusion (B) and DLS plots (C) showing that N-terminus induces oligomerization **D,E)** Specific activity as a function of enzyme concentration with SRSF1 (D) and MBP (E) as substrates. (Adapted from 66).

binding. The SP dipeptide phosphorylation characteristics of CLK1 are completely lost upon deletion of the N-terminus. Taken together, these results indicate that the N-terminus in CLK1 regulates SR protein binding affinity and hyper-phosphorylation of SRSF1, a critical step for splicing activation (60).

1.7.3. N-terminus mediated oligomerization is a substrate selection mechanism: A characteristic feature of RS domains or RS-like domains is their ability to self-associate and form oligomers. Prior studies from our laboratory showed that the presence of the N-terminus also makes CLK1 form oligomers as is observed through Dynamic Light Scattering (DLS) and Size Exclusion Chromatography (SEC). While CLK1 showed a hydrodynamic radius of about 100 nm and eluted in the void volume of an S300 column, CLK1(Δ N) showed a hydrodynamic radius of 4 nm and eluted as a monomer on S300 column (66) (Figure 1.8 B, C). The physiological relevance of oligomerization was revealed by comparing the specific activity of the enzyme ($v/[E]$) as a function of increasing concentrations of the enzyme. For a typical enzyme, the specific activity should remain constant with increasing enzyme concentration. However, for CLK1, specific activity increased sigmoidally with increasing enzyme concentration when SRSF1 was used as the substrate. On the other hand, CLK1(Δ N) showed a constant specific activity towards SRSF1 at all concentrations tested (Figure 1.8D). Specific activity vs enzyme concentration plots for both CLK1 and CLK1(Δ N) with myelin basic protein (MBP) as a substrate showed a constant specific activity. Interestingly, CLK1(Δ N) was more active than CLK1 towards MBP, which is a non-physiological substrate (Figure 1.8E). Taken together, these studies suggest that the N-terminus-mediated oligomerization is a substrate selection mechanism and facilitates the recognition of physiological over nonphysiological substrates (66).

1.8. Goals of the Dissertation

Although the disordered N-terminus is essential for regulating various functions of CLK1 including SR protein recognition and nuclear import, the underlying mechanisms for these functions are poorly understood. However, studying the role of the CLK1 N-terminus at a biochemical level has long been a challenge owing to its tendency toward aggregation and poor expression. Our lab has overcome these expression problems so that large amounts of pure, full-length CLK1 can be obtained and studied at detailed kinetic and structural levels. CLK1 is important for mobilizing SR proteins from storage speckles to splicing sites and, thus, is important to understand at a fundamental level. Regulating CLK1 function can help regulate splicing as CLK1 phosphorylation is essential for the recruitment of SRSF1 to splicing sites. Having known that the N-terminus-mediated oligomerization of CLK1 is a substrate selection mechanism, elucidating the structural and functional details of oligomerization gathers significance. Insights into the structural details of CLK oligomerization can be useful as oligomerization enhances CLK1 activity. The goals of this dissertation can be broadly summarized as (1) Investigate how the disordered N-terminus mediates the nuclear import of CLK1, (2) Elucidate the structural details of CLK1 oligomerization, and (3) Understand how the N-terminus enhances SRSF1 phosphorylation.

References

1. Crick F. Central Dogma. *Nature*. 1970;227:561–3.
2. Gilbert W. Why genes in pieces? *Nature*. 1978;271(5):501.
3. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*. 1977;74(8):3171–5.
4. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*. 1977;12(1):1–8.
5. Rosenfeld MG, Lin CR, Amara SG, Stolarsky L, Roos BA, Ong ES, Evans, RM. Calcitonin mRNA polymorphism: Peptide switching associated with alternative RNA splicing events. *Proc Natl Acad Sci U S A*. 1982;79(6 I):1717–21.
6. Berka AJ. Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci U S A*. 2016;113(4):801–5.
7. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol*. 2014;6(6).
8. Wang Y, Liu J, Huang B, Xu Y-M, Li J, Huang L-F, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. Mechanism of alternative splicing and its regulation. *Biomed Reports*. 2015;3(2):152–8.
9. Hertel KJ. Spliceosomal pre-mRNA splicing. Vol. 1126, *Methods in Molecular Biology*. 2014. 45–54 p.
10. Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol*. 2007;8(4):349–57.
11. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. Function of alternative splicing. *Gene*. 2013;514(1):1–30.
12. Faustino NA, Cooper T a, Andre N. Pre-mRNA splicing and human disease. *Genes Dev*. 2003;17(4):419–37.
13. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. *Biochim Biophys Acta - Mol Basis Dis*. 2009;1792(1):14–26.
14. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19–32.
15. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med*. 2012;18(8):472–82.

16. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 2005;579(9):1900–3.
17. Passini MA, Bu J, Richards AM, Kinnecom C, Sardi SP, Stanek LM, Hua Y, Rigo F, Matson J, Hung G, Kaye, EM, Shihabuddin LS, Krainer AR, Bennett CF, Cheng SH. Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Sci Transl Med.* 2011;3(72).
18. Hua Y, Sahashi K, Rigo F, Hung G, Horev G, Bennett CF, Krainer AR. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature.* 2011;478(7367):123–6.
19. Papasaikas P, Valcárcel J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci.* 2016;41(1):33–45.
20. Yan C, Wan R, Shi Y. Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb Perspect Biol.* 2019;11(1).
21. Kotlajich M V., Crabb TL, Hertel KJ. Spliceosome Assembly Pathways for Different Types of Alternative Splicing Converge during Commitment to Splice Site Pairing in the A Complex. *Mol Cell Biol.* 2009;29(4):1072–82.
22. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol.* 2014;15(2):108–21.
23. Zhan X, Yan C, Zhang X, Lei J, Shi Y. Structure of a human catalytic step I spliceosome. *Science (80-).* 2018;359(6375):537–45.
24. Long JC, Caceres JF. The SR protein family of splicing factors: Master regulators of gene expression. *Biochem J.* 2009;417(1):15–27.
25. Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. SR Proteins in Vertical Integration of Gene Expression from Transcription to RNA Processing to Translation. *Mol Cell.* 2009;35(1):1–10.
26. Krainer AR, Conway GC, Kozak D. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* 1990;4(7):1158–71.
27. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell.* 2001;8(6):1351–61.
28. Tintaru AM, Hautbergue GM, Hounslow AM, Hung ML, Lian LY, Craven JC, Wilson, SA. Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.* 2007;8(8):756–62.

29. Cléry A, Sinha R, Anczuków O, Corrionero A, Moursy A, Daubner GM, Valcarcel J, Krainer AR, Allain FHT. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc Natl Acad Sci U S A*. 2013;110(30).
30. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res*. 2006;34(1):305–12.
31. Hamelberg D, Shen T, Andrew McCammon J, Stroud RM. A proposed signaling motif for nuclear import in mRNA processing via the formation of arginine claw. *Proc Natl Acad Sci U S A*. 2007;104(38):14947–51.
32. Ngo JCK, Giang K, Chakrabarti S, Ma CT, Huynh N, Hagopian JC, Dorrestein PC, Fu XD, Adams JA, Ghosh, G. A Sliding Docking Interaction Is Essential for Sequential and Processive Phosphorylation of an SR Protein by SRPK1. *Mol Cell*. 2008;29(5):563–76.
33. Oeffinger M. *The Biology of mRNA : Structure and Function*.
34. Das S, Krainer AR. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res*. 2014;12(9):1195–204.
35. Twyffels L, Gueydan C, Kruys V. Shuttling SR proteins: More than splicing factors. *FEBS J*. 2011;278(18):3246–55.
36. Gui JF, Lane WS, Fu XD. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*. 1994;369(6482):678–82.
37. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J*. 1996;15(2):265–75.
38. Sinha R, Allemand E, Zhang Z, Karni R, Myers MP, Krainer AR. Arginine Methylation Controls the Subcellular Localization and Functions of the Oncoprotein Splicing Factor SF2/ASF. *Mol Cell Biol*. 2010;30(11):2762–74.
39. Naro C, Sette C. Phosphorylation-mediated regulation of alternative splicing in cancer. *Int J Cell Biol*. 2013;2013.
40. Lai MC, Lin RI, Huang SY, Tsai CW, Tarn WY. A human importin- β family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem*. 2000;275(11):7950–7.
41. Lai M-C, Lin R-I, Tarn W-Y. Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci*. 2001;98(18):10154–9.

42. Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Oztop I, Lee K, Pye VE, Cosnefroy O, Snijders AP, KewalRamani VN, Fassatu A, Engelman A, Cherepanov, P. Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci.* 2014;111(7):2728–33.
43. Colwill K, Feng LL, Yeakley JM, Gish GD, Ca JF, Pawson T, Fu XD. SRPK1 and Clk / Sty Protein Kinases Show Distinct Substrate Specificities for Serine / Arginine-rich Splicing Factors *. *1996;271(40):24569–75.*
44. Ghosh G, Adams JA. Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J.* 2011;278(4):587–97.
45. Ding JH, Zhing XY, Hagopian JC, Cruz MM, Ghosh G, Feramisco J, Adams JA, Fu XD. Regulated Cellular Partitioning of SR Protein-specific Kinases in Mammalian Cells. *Mol Biol Cell.* 2008;17(February):876–85.
46. Zhong XY, Ding JH, Adams JA, Ghosh G, Fu XD. Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes Dev.* 2009;23(4):482–95.
47. Ngo JCK, Gullingsrud J, Giang K, Yeh MJ, Fu XD, Adams JA, McCammon JA, Ghosh G. SR Protein Kinase 1 Is Resilient to Inactivation. *Structure.* 2007;15(1):123–33.
48. Aubol BE, Chakrabarti S, Ngo J, Shaffer J, Nolen B, Fu XD, Ghosh G, Adams JA. Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc Natl Acad Sci U S A.* 2003;100(22):12601–6.
49. Huynh N, Ma CT, Giang N, Hagopian J, Ngo J, Adams J, Ghosh G. Allosteric interactions direct binding and phosphorylation of ASF/SF2 by SRPK1. *Biochemistry.* 2009;48(48):11432–40.
50. Ma CT, Velazquez-Dones A, Hagopian JC, Ghosh G, Fu XD, Adams JA. Ordered Multi-site Phosphorylation of the Splicing Factor ASF/SF2 By SRPK1. *J Mol Biol.* 2008;376(1):55–68.
51. Ma CT, Hagopian JC, Ghosh G, Fu XD, Adams JA. Regiospecific Phosphorylation Control of the SR Protein ASF/SF2 by SRPK1. *J Mol Biol.* 2009;390(4):618–34.
52. Yun B, Farkas R, Lee R, Rabinow L. The Doa locus encodes a member of a new protein kinase family and is essential for eye and embryonic development in *Drosophila melanogaster*. *Genes Dev.* 1994;8(10):1160–73.
53. Ben-David Y, Letwin K, Tannock L, Bernstein A, Pawson T. A mammalian protein kinase with potential for serine/threonine and tyrosine phosphorylation is related to cell cycle regulators. *EMBO J.* 1991;10(2):317–25.

54. Prasad J, Manley JL. Regulation and Substrate Specificity of the SR Protein Kinase Clk/Sty. *Mol Cell Biol.* 2003;23(12):4139–49.
55. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J.* 2009;417(1):15–27.
56. Bullock AN, Das S, Debreczeni JÉ, Rellos P, Fedorov O, Niesen FH, Guo K, Papagrigoriou, Amos AL, Cho S, Turk, BE, Ghosh G, Knapp S. Kinase Domain Insertions Define Distinct Roles of CLK Kinases in SR Protein Phosphorylation. *Structure.* 2009;17(3):352–62.
57. Ma CT, Ghosh G, Fu XD, Adams JA. Mechanism of Dephosphorylation of the SR Protein ASF/SF2 by Protein Phosphatase 1. *J Mol Biol.* 2010;403(3):386–404.
58. Keshwani MM, Aubol BE, Fattet L, Ma C-T, Qiu J, Jennings PA, Fu XD, Adams, JA. Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochem J.* 2015;466(2):311–22.
59. Prasad J, Colwill K, Pawson T, Manley JL. The Protein Kinase Clk/Sty Directly Modulates SR Protein Activity: Both Hyper- and Hypophosphorylation Inhibit Splicing. *Mol Cell Biol.* 1999;19(10):6991–7000.
60. Aubol BE, Plocinik RM, Keshwani MM, Mcglone ML, Hagopian JC, Ghosh G, Fu XD, Adams JA. N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. *Biochem J.* 2014;462(1):143–52.
61. Aubol BE, Wu G, Keshwani MM, Hertel KJ, Maliheh M, Fattet L, Fu X, Adams JA. Release of SR Proteins from CLK1 by SRPK1 : A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol Cell.* 2016;218–28.
62. Aubol BE, Fattet L, Adams JA. A conserved sequence motif bridges two protein kinases for enhanced phosphorylation and nuclear function of a splicing factor. *FEBS J.* 2020;1:1–16.
63. Adams JA. Kinetic and Catalytic Mechanisms of Protein Kinases. 2001;(858).
64. Goldsmith EJ, Akella R, Min X, Zhou T, Humphreys JM. Substrate and docking interactions in serine/threonine protein kinases. *Chem Rev.* 2007;107(11):5065–81.
65. Duncan PI, Howell BW, Marius RM, Drmanic S, Douville EMJ, Bell JC. Alternative splicing of STY, a nuclear dual specificity kinase. *J Biol Chem.* 1995;270(37):21524–31.
66. Keshwani MM, Hailey KL, Aubol BE, Fattet L, McGlone ML, Jennings PA, Adams, JA. Nuclear protein kinase CLK1 uses a non-Traditional docking mechanism to select physiological substrates. *Biochem J.* 2015;472(3):329–38.

Chapter 2

Materials and Methods

2.1. Materials

Tris HCl, glycerol, NaCl, dithiothreitol (DTT), isopropyl β -D-1-thiogalactopyranoside (IPTG), LB broth, phenylmethylsulphonyl fluoride (PMSF), glycerol, N-morpholino propane sulphonic acid (MOPS), 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES), phoenix imaging films, bovine serum albumin (BSA), nitrocellulose membranes, glutathione (GSH), imidazole, triton-X, 3-[(3Cholamidopropyl) dimethylammonio]-1-propanesulfonate hydrate (CHAPS), β -mercaptoethanol (β -ME), dulbecco's modified eagle medium (DMEM), dulbecco's phosphate buffered saline (DPBS), foetal bovine serum (FBS), Ni-nitrilo triacetic acid (NTA) resin, GSH-immobilized resin, formic Acid, NP-40 detergent, sodium dodecyl sulfate (SDS), tris buffered saline-tween20 (TBST), luria bertani (LB) media, ampicillin, kanamycin, gentamycin, tetracycline, and bluogal were obtained from Thermo fischer scientific. Fugene transfection reagent was obtained from Promega, cellfectin II and bacmid isolation kits from invitrogen, anti-FLAG and anti-his tag antibodies were from Cell Signaling Technology, anti-RFP, and anti-TRN-SR2 antibodies from Abcam, anti-GST from Biolegend and anti-GAPDH antibody from R&D Systems. ^{32}P -ATP was purchased from NEN Products. D_2O , deuterated HEPES, DCl, NaOD, and deuterated DTT were obtained from Cambridge Isotope Laboratories. Protease Inhibitor tablets were obtained from Roche. SiRNA for TRN-SR2 was obtained from Bioneer.

2.2. Methods

2.2.1. Expression and purification of recombinant proteins

All GST-N-terminus constructs were cloned in pGEX vectors and transformed into BL21 DE3 *E.coli* cells for expression. The cells were grown at 37°C in 2 L in LB Broth media supplemented with ampicillin (100 $\mu\text{g}/\text{mL}$) until the optical density (O.D) at 600 nm reached 0.6. The temperature was brought down to 25°C and then induced with 0.5 mM IPTG. The cells were

harvested 4 hours after IPTG induction by centrifuging at 4000 rpm for 10 minutes. The cells were lysed in lysis buffer (50 mM Tris HCl at pH 8.0, 500 mM NaCl, 1 mM DTT and 0.025% TritonX-100) by sonicating in the presence of 1 mM PMSF and protease inhibitor cocktail. The lysate was centrifuged at 13,000 rpm for 30 minutes and the supernatant was incubated with GSH-agarose Resin prewashed with lysis buffer at 4°C for one hour. The lysate supernatant was collected and the resin was washed with three column volumes of the lysis buffer and the protein was eluted in elution buffer (50 mM Tris HCl at pH 8.0, 500 mM NaCl, 1 mM DTT and 20 mM GSH). The elution was repeated until most of the protein was washed off from the resin and the purity of the eluted fractions were analyzed by running the fractions on 12% SDS-PAGE. Excess GSH was removed from the proteins by dialyzing the eluted fractions overnight at 4°C against GSH-free dialysis buffer (50 mM Tris HCl at pH 8.0, 500 mM NaCl, 1 mM DTT and 10% glycerol). The protein concentrations were estimated using a Bradford reagent measured against BSA as a standard. The dialyzed fractions were frozen and stored at -80°C until further use.

His-CLK1(Δ N) gene was cloned in pET-28a vector from Genscript and transformed into BL-21 DE3 *E. coli* strain. Cells were grown in LB media supplemented with 50 μ g/mL kanamycin. Cells were grown at 25°C for 4 hours post-induction with 1 mM IPTG. Cells were harvested and lysed in lysis buffer (50 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT and 0.04% CHAPS) and the soluble fraction was isolated by centrifugation. The soluble fraction was incubated with Ni-NTA Resin for one hour at 4°C, loaded onto a column and then washed with two column volumes of wash buffer 1 (50 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT, 0.04% CHAPS, and 5mM Imidazole) and one column volume of wash buffer 2 (50 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT, 0.04% CHAPS, and 30 mM Imidazole). The protein was then eluted in 10 mL of elution buffer (50 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT, 0.04% CHAPS, and 300 mM

Imidazole). The excess imidazole was then removed by dialyzing overnight against imidazole-free dialysis buffer (50 mM Tris HCl pH 7.5, 500 mM NaCl, 1 mM DTT, 10% glycerol). The dialyzed protein concentrations were estimated using Bradford, purity analyzed using SDS-PAGE, and flash frozen in small aliquots and stored at -80°C until further use.

2.2.2. Mutagenesis

All mutations were made using Quikchange Lightning site-directed mutagenesis kits. For a single point mutation, overlapping primers of equal lengths (~ 20-25 nucleotides long) were used. On the other hand, for multisite mutagenesis, non-overlapping primers with a long forward primer (~ 60 nucleotides long) and a shorter reverse primer (~ 20 nucleotides) were used. All components were added in amounts recommended by the manufacturer (Agilent Technologies) and annealing temperatures were adjusted according to the melting temperatures of the primers used. The PCR amplification product was DpnI digested for 5 minutes to remove the parent template DNA and then transformed into XL-10 ultragold cells provided with the kit. 2 μ L of β -ME was added to competent cells to enhance the efficiency of transformation. Mutant plasmids were isolated from colonies formed on LB-agar plates supplemented with antibiotics. Plasmids were sequenced with T7 forward or reverse primers to confirm the presence of the desired mutation.

2.2.3. Generation of baculoviruses and protein expression in Hi5 cells

All recombinant CLK1 constructs used, namely CLK1, CLK1(Δ 1), CLK1(Δ 2), CLK1(Δ 3), CLK1(Δ 12), CLK1(Δ 13), CLK1(Δ 23), and CLK1(Y-L) were expressed and purified from Hi5 cells using baculoviruses. DNA encoding the protein of interest cloned into pFastBac vectors were bought from Genscript. The plasmids were transformed into DH-10 Bac *E.coli* cells and plated on LB-Agar plates supplemented with kanamycin (50 μ g/mL), gentamycin (7 μ g/mL), tetracycline

(10 µg/mL), IPTG (4 µg/mL) and blueo-gal (100 µg/mL). The plates were incubated at 37°C for 48 hours for the colonies to form and express distinct blue/white coloration. The white colonies that have the gene of interest recombined in the bacmid were re-streaked on a new plate along with a blue colony and incubated for an additional 48 hours. Cells from the white colonies were cultured overnight in LB media supplemented with kanamycin, gentamycin, and tetracycline and harvested by centrifugation. Bacmid isolation was done using the Bacmid isolation kit from Thermofisher following the manufacturer's protocols. The bacmid concentrations were estimated using a nanodrop prior to transfection.

Sf9 cells were grown to a density of 1.5×10^6 - 2.5×10^6 cells/mL in serum free Sf900 II-SFM in a sterile incubator at 25°C. In a 6-well plate, 8×10^5 Sf9 cells were plated in 2 mL of Sf900 media and left at 25°C for 1 hour. In two separate tubes, 1 µg of bacmid DNA was diluted in 100 µL Sf900 media and 5 µL of cellfectin-II transfection reagent was diluted in 100 µL media. The bacmid solution was then added to the cellfectin-II solution and incubated for 45 minutes for the transfection complexes to form. The bacmid-cellfectin complexes are added to the plated Sf9 cells in the 6-well plates and then left overnight in the incubator at 25°C. The media in the plate is then replaced with 2 mL of fresh Sf900 SFM media with 1% FBS. The cells were checked for signs of infection (cessation of growth, granular appearance, lowered confluency, enlarged nuclei, detachment from the bottom of the plate) after 72 hours of transfection and then the supernatant was collected (P1 viral stock) after 5 days of transfection. The transfected cells were collected, lysed in lysis buffer in the presence of protease inhibitor cocktail and the lysates were probed with α-His tag antibody in western blots to check for protein expression.

Once protein expression was confirmed, the P1 viruses were amplified by infecting 30 mL culture of Sf9 cells with 50 µL of the P1 stock. The amplified P2 viral stock was collected after 72

hours of the infection. The P2 viral stock was used to infect Hi5 cells grown to a density of 2×10^6 cells/mL on a larger culture (200 mL) to express and purify recombinant protein. Cells were harvested after 48 hours of infection and resuspended in lysis buffer (50mM Tris HCl pH 7.5, 500mM NaCl, 1mM DTT, 0.04% CHAPS) along with DNAase, RNAase, PMSF and protease inhibitor cocktail. The lysate after sonication was then incubated with Ni-NTA resin for one hour at 4°C. The washes, elution and dialysis were done as described previously with His-CLK1(Δ N) purification in section 2.2.1.

2.2.4. Transfection and Confocal imaging

HeLa or HEK 293 cells were plated to 50% confluency 24 hours before transfection in complete growth medium (DMEM with 10% FBS) in MatTek polylysine plates. Prior to transfection, the cells were washed with DPBS and replaced with fresh media. Plasmid DNA to be transfected was prepared at a concentration of 0.020 μ g/ μ L in optiMEM media. Fugene:DNA ratio was always kept at 3:1, and the volumes added were as recommended by the manufacturer for each plate type. The fugene: DNA complexes were mixed by pipetting and incubated for 5 minutes in the tissue culture hood. The fugene: DNA complexes were added to cells and the plates were left at 37°C in the incubator for 24 hours until confocal imaging. Si-RNA knockdown transfections were performed by co-transfecting CLK-RFP and siRNA (50 pmol for a 12 well plate) using lipofectamine 2000. All transfected cells were imaged using Olympus FV1000 microscope. All live cell imaging was done in 35 mm plates and all fixed cell imaging were done in cover slips. Direct fluorescence of RFP was imaged using HeLa cells fixed on cover slips that was stained with DAPI without permeabilization.

2.2.5. Immunofluorescence

For immunofluorescence experiments, HeLa cells were plated on coverslips placed in 12 well plates. Transfected cells were washed with PBS, fixed using 1% paraformaldehyde. Cells were washed three times in PBS to remove the excess formaldehyde, and then permeabilized by incubating with TritonX-100 for 20 minutes at 4°C. To minimize non-specific binding of primary antibody, the cells were then blocked using 20% goat serum for one hour at 25°C. The cells were then incubated with anti-myc antibody overnight at 4°C, incubated with fluorescent-labeled anti-mouse secondary antibody for one hour in the dark, and then stained with DAPI before imaging.

2.2.6. Biochemical Fractionation

Cell fractionation to isolate the nuclear and cytoplasmic fractions were performed using cell fractionation kits from cell signaling technology. Transfected HeLa cells were trypsinized, collected, and resuspended in PBS. A small fraction of cells (20%) were kept aside to check for protein expression. The remaining cells were fractionated to separate the cytoplasmic and nuclear fractions using the cytoplasm isolation buffer and the cytoskeleton/nucleus isolation buffer respectively. The collected fractions were analyzed using Western blots and probed using α -RFP antibody. α -GAPDH and α -histone controls were performed to ensure that the fractionation was efficient. For Image-J analysis of protein bands on the Western blots, a profile plot was generated for each band by drawing a rectangle frame around the bands. The relative band intensities were calculated by integrating the area under the curve for each band.

2.2.7. Immunoprecipitations

Transfected cell lysates of HEK or HeLa cells (200 μ L) were incubated with 10 μ L immobilized antibody conjugated beads with gentle rocking overnight at 4°C. The beads were washed with PBS (500 μ L) five times. Proteins still bound to the beads were eluted by heating at

95°C with SDS-gel loading dye for 5 minutes. The immunoprecipitated proteins were then analyzed using Western blots and probed using α -FLAG, α -Myc or α -CLK1 antibodies. Resin controls or IgG controls were done to rule out non-specific binding.

2.2.8. Pulldown assays

GSH-tagged agarose resins were pre-washed in pulldown buffer (50mM Tris HCl pH 7.5, 75mM NaCl, 0.1% Nonidet P40). GST-tagged proteins were incubated with His-tagged proteins for one hour at 4°C with gentle rocking. GST controls were done by adding equal concentration of purified, recombinant GST and resin controls only had pulldown buffer. Pre-washed GSH-Agarose resin (30 μ L) was added and incubated for another one hour at 4°C with gentle rocking. The GSH resin was then washed three times with 200 μ L of pulldown buffer and then eluted with SDS gel loading dye at 95°C for 5 minutes. The eluted mixtures were run on 12% SDS gel, stained with Instant Blue overnight and de-stained in water. The fraction bound was calculated by ImageJ.

2.2.9. Size Exclusion Chromatography and Dynamic Light Scattering

The oligomer-monomer distribution of all GST-N constructs and purified CLK1 constructs were analyzed using Size Exclusion Chromatography (SEC) and Dynamic Light Scattering (DLS). SEC was performed on a Biologic duoflow fine performance liquid chromatography (FPLC) system maintained at 4°C. The proteins were run on an S200 gel filtration column in running buffer (50mM Tris HCl pH 7.5, 500 mM NaCl, and 1 mM DTT) with a flow rate of 0.7 mL/min) for 30 minutes. The 280 nm UV signal was monitored to track protein elution. The eluted proteins were collected in 200 μ L fractions in 96-well plate and then analyzed using coomassie stained SDS PAGE gels to detect proteins. SEC standard was run to calculate the molecular weights of proteins at given elution fractions.

DLS spectra were collected on a DynaPro instrument and the data collected were analyzed using Dynamics software. The protein concentrations for all monomeric constructs were 50 μM . Spectra for all oligomeric constructs were collected at the highest attainable concentration for that construct (10 μM for CLK1, 3 μM for GST-N and 25 μM for GST-N ($\Delta 3$). All samples were in 50 mM Tris HCl pH 7.5, 500 mM NaCl, 1mM DTT and 10% glycerol. Prior to data collection, all samples were centrifuged at 12,000 rpm for 10 minutes to remove any large aggregates and dust particles. All samples were recorded at 25°C and 20 acquisitions were collected for each sample. The collected spectra were then fit to an autocorrelation curve and the average hydrodynamic radius was obtained by fitting to a regularization histogram on DynaPro software.

2.2.10. H/D Exchange and Mass Spectrometry

CLK1 and CLK1(ΔN) were purified as mentioned in section 2.2.1. Both proteins were dialyzed into lower salt HEPES buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 10% glycerol and 1 mM DTT) overnight at 4°C. The proteins were spun down at 12,000 rpm for 15 minutes to remove any aggregates. The protein concentrations were estimated using Bradford assay. Both protein concentrations were 8 μM before initiating the reaction.

The D₂O buffer was prepared (50 mM deuterated HEPES, 1 mM deuterated DTT, 250 mM NaCl in D₂O). The pD of the buffer was adjusted to 7.5 using DCl or NaOD. The quench condition was optimized by comparing the fragmentation patterns obtained with different quench conditions of varying guanidinium hydrochloride concentrations. The quench buffer chosen for exchange reaction was 0.24 M guanidinium hydrochloride, 0.24% formic acid and 40% glycerol. The H/D exchange reaction was initiated by adding 110 μL of protein to 330 μL of D₂O buffer. Aliquots were withdrawn (40 μL each) from the reaction at regular time intervals of 10 s, 30 s, 60 s, 5 min, 15 min, 40 min, 60 min, and 120 min and rapidly mixed with 10 μL of quench buffer in glass vials.

The vial caps were sealed using a crimper and immediately frozen on dry ice. Non-deuterated (ND) control was prepared by diluting 10 μL of protein with 30 μL of H_2O buffer and quenched with 10 μL of quench buffer. The fully-deuterated control (FD) was prepared by exchange of pepsin digested peptides in D_2O buffer for 24 hours. To generate the FD control, samples of both CLK1 and CLK1(ΔN) were first prepared according to ND conditions and collected after on-column pepsin digestion. The digested peptides were concentrated on a speed vac and allowed to exchange in D_2O buffer for 24 hours by mixing 10 μL of digested peptides with 0.8% deuterated formic acid. The FD samples were then quenched in 10 μL of quench buffer, frozen on dry ice and then analyzed using liquid chromatography and mass spectrometry (LC-MS). All samples were prepared in duplicates and stored at -80°C until mass spectrometric analysis.

The samples were loaded onto auto sampler, digested on a pepsin 16 column, and peptides were separated on a C-18 reverse phase HPLC column. The FD samples were analyzed by bypassing the pepsin digestion. The peptides were separated by running a linear gradient of acetonitrile running from 0 to 40% acetonitrile in 0.8% formic acid. The separated peptides were characterized using an electron-spray ionization (ESI) mass spectrometer. The identity of the peptides were determined using the SEQUEST program. The predicted and experimentally observed peptide mass envelopes were matched using the DXMS program and verified manually based on their monoisotopic mass, charge on the peptide, and expected elution time on the chromatogram. A collision check was performed using manual inspection to ensure that no two peptides were assigned to the same peptide envelope. Peptides common to both CLK1 and CLK1(ΔN) were matched using a fragmatcher and the rates of incorporation of deuterium were compared between the two proteins. The increase in centroid mass of the peptide envelopes were monitored and the number of deuteriums incorporated were calculated and plotted in deuterium uptake plots.

The deuterium incorporation heat maps were plotted using % deuterium incorporation for representative peptides as a function of time. The % of deuterium incorporation was calculated according to the equation

$$\% D \text{ uptake} = \frac{\text{Centroid Mass of peptide} - \text{Centroid Mass of ND peptide}}{\text{Centroid Mass of FD peptide} - \text{Centroid Mass of ND peptide}} * 100\%$$

2.2.11. Single turnover kinase assays

Activity assay buffer was prepared as 100 mM MOPS pH 7.2, 5 mg/mL BSA, 10 mM MgCl₂, 1 μM enzyme, 0.2 μM SRSF1, 100 μM ATP. ³²P labelled ATP was added into the master mix to yield a final specific activity of 4000-8000 cpm/mol. The reaction master-mixes and the enzymes were pre-incubated at 37°C for 5 minutes before initiating the reaction. The reactions were initiated by adding 1 μM enzyme into the master mix. Aliquots of 10 μL each were withdrawn from the reaction mixture and then quenched using 10 μL of the SDS loading buffer. The phosphorylated substrate was separated from unreacted ³²P-ATP by SDS-PAGE on a 10% gel. The gels were dried overnight in the hood and upper and lower bands were cut from the dried gel and counted in the ³²P channel using a liquid scintillant. The product formed was plotted as a function of time and the reaction velocities and maximum amplitudes were obtained by fitting to a single exponential function.

2.2.12. Steady state kinetic assays

All steady state kinetic assays were performed at 25°C in activity assay buffer (100 mM MOPS pH 7.2, 5 mg/mL BSA, 10 mM MgCl₂, 25 μM ATP, 100 nM enzyme). The SRSF1 concentration ranges were varied from 100 nM to 2000 nM. The ³²P-ATP concentration was adjusted to keep the specific activity at 4000-8000 cpm/mol. The initial velocities were obtained

by quenching the reaction after 5 minutes of the reaction initiation. The phosphorylated SRSF1 was separated from unreacted ^{32}P -ATP by running the reaction mixtures on a 12% SDS gel. A plot of initial velocities vs SRSF1 concentration was made and the data were fitted to the quadratic function below to obtain K_m and V_{max} values:

$$v = \frac{V_{max}}{2[E]} * (([E] + [S] + K_m) - \sqrt{([E] + [S] + K_m)^2 - 4[E][S]})$$

where $[E]$ =enzyme concentration and $[S]$ =SRSF1 concentration.

Chapter 3

Nuclear Localization of CLK1

Abstract

CLKs are exclusively nuclear kinases but the underlying mechanism controlling such strict subcellular localization is not well understood. Early investigations suggested that the CLKs may contain a classical nuclear localization signal on their N-termini but this proposal has not been rigorously tested. In this chapter, we investigate the nuclear import of CLK1 and find that the kinase lacks a short, classical nuclear localization signal as once proposed. Instead, CLK1 is imported into the nucleus by a ‘piggyback mechanism’ where the kinase binds to its substrate SRSF1 in the cytoplasm, and the karyopherin TRN-SR2 brings the kinase-substrate complex into the nucleus using the nuclear localization signal on the substrate.

3.1 Introduction

The compartmentalization of biomolecules into the nucleus and cytoplasm is not only one of the major defining features of eukaryotic cells but also is a vital means for partitioning and regulating various cellular functions (1). For instance, the nucleus is the storage compartment for genetic information and a center for DNA transcription and mRNA splicing whereas the cytoplasm contains the ribosomal machinery for protein translation and numerous metabolic functions. The nuclear and cytoplasmic compartments are separated by a double-walled nuclear membrane which is interrupted by nuclear pores (2,3). Passive diffusion of cargo across the nuclear pores happens only for smaller sized proteins, with a predicted upper limit of about 40 kDa (4). Bigger protein cargoes that have nuclear functions need to be actively transported across the nuclear membrane at the expense of energy (4). Hence, the presence of distinct compartments also necessitates designated molecular machinery responsible for the proper transport of biomolecules across these compartments (4).

Passage across the nuclear membrane is mediated by a family of proteins known as β -karyopherins. In human cells, around 20 different karyopherins have been identified and can be classified into two groups: the importins that transport cargo into the nucleus and the exportins that transport cargo out of the nucleus (1). There are many different nuclear import mechanisms identified so far. The most well characterized is the classical nuclear import mechanism that is mediated by the heterodimer importin- α/β . Cargo proteins for the importin- α/β family usually carry a highly basic lysine/arginine-rich nuclear localization signal (NLS). Importin- α acts as an adaptor protein that recognizes the NLS on cargo proteins using its armadillo (ARM) domain, a repeated structural unit composed of two α helices separated by a hairpin turn (5). The N-terminal domain of importin- α has an importin- β binding domain (IBB) that binds importin β forming the

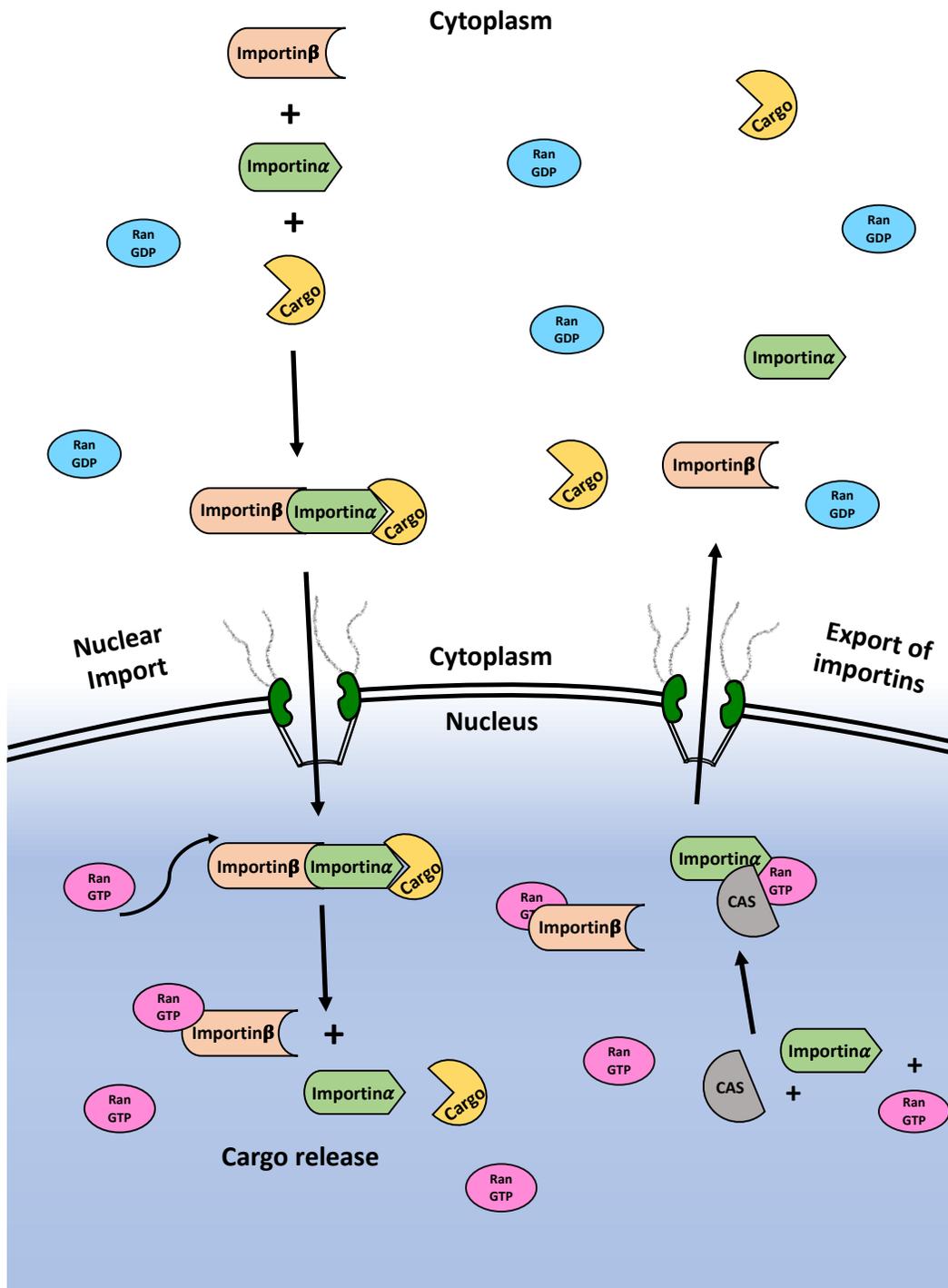


Figure 3.1: The classical nuclear import mechanism.

functional, trimeric complex composed of the cargo protein and importin- α/β heterodimer (1) (Fig.3.1). This ternary complex is subsequently translocated into the nucleus through the interaction of importin- β with nucleoporins that line the nuclear pores. One of the critical factors determining the directionality of transport is the Ran-GTP gradient. Ran-GTP concentration is, at least, 10-fold higher in the nucleus than in the cytoplasm. Once the importin-cargo complex reaches the nucleus, the binding of Ran-GTP triggers cargo release, and the importin- α/β heterodimer is then shuttled back to the cytoplasm by CAS (cellular apoptosis susceptibility gene)-Ran-GTP complex (Fig.3.1) (5–7).

In addition to the above classical mechanism, a few other nuclear import mechanisms are known. Several examples of direct recruitment of cargo proteins by importin- β without any adaptor proteins are known. For example, importin-7, another member of the karyopherin family, binds Ser-Pro-Ser or Thr-Pro-Thr motifs when both the serine and threonine residues are phosphorylated in the cargo proteins (8). Even before the classical mechanism was fully understood, hnRNPA1 and several other RNA binding proteins were found to have a non-classical NLS, with a consensus PY motif at the C-terminal end (9). Transportin-1 (TNPO1) was later identified as the karyopherin that recognizes PY-type NLSs on cargoes (10). Transportin SR-2 (TRN-SR2) also known as transportin-3 (TNPO3), or importin-12 is yet another karyopherin that binds to cargo proteins possessing stretches of phosphorylated residues (11). The prototypical SR protein SRSF1 is known to be imported into the nucleus by TRN-SR2 in a phosphorylation-dependent manner (12–14). TRN-SR2 has also been identified as the karyopherin used by retroviruses for nuclear entry in the host cells to integrate viral genome into the host genome. Hence, TRN-SR2 has recently gathered scientific attention with potential applications in viral therapeutics (15,16).

Interestingly, one protein can have multiple nuclear import mechanisms (17). ERK2 is a protein kinase that is important in various signaling pathways with its subcellular localization being one of the key parameters that regulate its function. ERK2 has been reported to have multiple nuclear import mechanisms. A few studies have shown that the phosphorylation of an SPS motif in the MAP kinase insert domain leads to importin-7 mediated nuclear import of ERK2 (8,18). However, another study showed that ERK2 dimerization upon phosphorylation of two residues in its activation loop leads to nuclear import (19,20). Surprisingly, isoforms of the same protein can have very different nuclear import mechanisms. For example, SRPK1 is anchored in the cytoplasm by chaperones like Hsp70 that bind to the spacer insert domain (SID) that bifurcates the traditional kinase domain (21). Under stress signaling, the chaperones are shed from the SID, leading to its nuclear import. However, the SRPK2 isoform has an importin-7 binding site that interacts with S494 and S497 in the SID when phosphorylated, a post-translational modification that drives nuclear import (22).

Although SRPKs are mostly cytoplasmic kinases that enter the nucleus in a regulated manner, the CLK family is largely found in the nucleus. Over-expression studies of four isoforms of mouse CLK (CLK/STY) and three isoforms of human CLKs (CLK1-4) have shown that these protein kinases are exclusively nuclear based on confocal microscopy (23–25). Furthermore, endogenous CLK1 is also found expressed in the nucleus of cells suggesting that these other studies are not artefacts arising due to protein over-expression (26). Such localization studies along with the observations that CLKs avidly phosphorylate SR proteins support a role for these enzymes in the regulation of nuclear splicing. How CLKs enter the nucleus was originally explored in studies by Duncan et.al more than two decades ago (27). Using immunofluorescence experiments, they showed that overexpressed mouse CLK1 (CLK/STY) was exclusively nuclear whereas a

deletion mutant lacking the first 60 residues of the 150-residue N-terminus (Δ 1-60 CLK/STY) was localized to both the nucleus and cytoplasm (Fig.1.7D) (27). This was thought to be due to the presence of a putative NLS on its N-terminus, thus relegating CLKs to the cargos recognized by the classic importin system. In this study, we show that this long-held view is incorrect and that nuclear import of human CLK1 does not utilize a classic NLS but rather utilizes a novel ‘piggyback’ mechanism, by binding to its substrate SRSF1 (28).

3.2 Results

3.2.1 CLK1 does not have a classical NLS.

Previous studies suggested that the CLK1 N-terminus is vital for nuclear import. To confirm that the nuclear import of CLK1 is regulated by its N-terminus, we first looked at the subcellular localization of an RFP-tagged CLK1 and CLK1(Δ N) in HeLa cells using confocal microscopy. We placed the RFP tags on all constructs at the C-terminus of the gene of interest to avoid any potential interference with the disordered N-terminal domain of CLK1. We found that while CLK1-RFP localized exclusively to the nucleus, CLK1(Δ N)-RFP expressed in both the nucleus and the cytoplasm in confocal imaging experiments (Fig. 3.2A,B). We confirmed these results using biochemical fractionation and quantitated the amount of CLK1 to be 95% nuclear, whereas CLK1(Δ N) was only 58% nuclear (Fig. 3.2C). Next, we tried to identify the residues that might be functioning as classical NLSs on the N-terminus using cNLS mapper (29). The search initially resulted in two monopartite sequences, titled NLS1 and NLS2 (Fig. 3.2A). To test the roles of these NLSs, we mutated NLS1 and NLS2 to glycines and looked at their nuclear localization using confocal imaging and fractionation experiments. We found that both mutants CLK1(Δ NLS1)-RFP and CLK1(Δ NLS2)-RFP were exclusively nuclear (Fig 3.2 B,C). Next, we wished to determine if the two NLSs worked cooperatively and mutated both NLSs together in

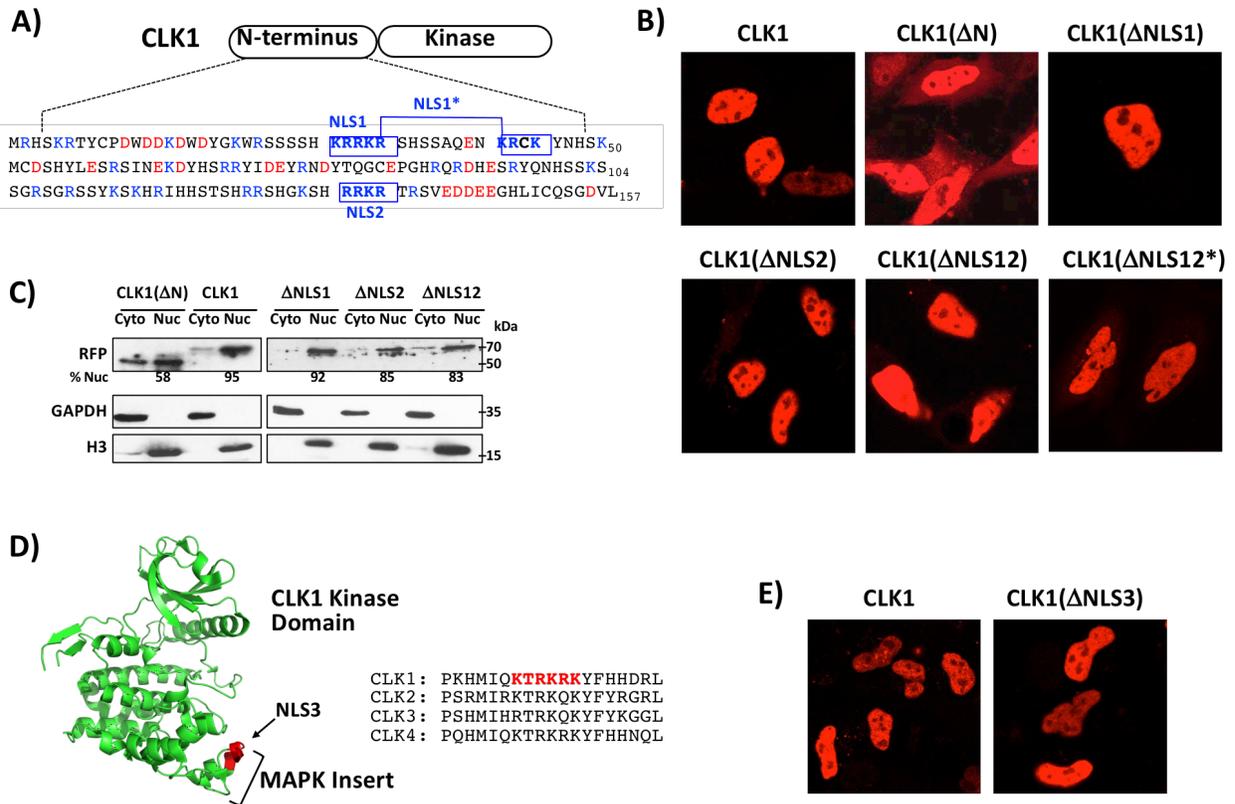


Figure 3.2: CLK1 does not have a short, classical NLS. **(A)** The CLK1 N-terminus with putative NLSs highlighted in blue. **(B)** Confocal images of HeLa cells transfected with CLK1-RFP, CLK1(ΔN)-RFP and different NLS mutants where K/R residues in the putative NLSs are mutated to glycines. **(C)** Fractionation immunoblots of HeLa cells transfected with CLK1-RFP, CLK1(ΔN)-RFP, and various NLS mutants. The % Nuclear protein quantitated using Image J are indicated below each band. GAPDH and histone controls are included. **(D)** NLS3 on the crystal structure of CLK1(ΔN) (PDB ID:1Z57). CLK1 sequence is aligned with other isoforms of CLK. **(E)** Confocal imaging of live HeLa cells transfected with CLK1-RFP and the mutant CLK1(ΔNLS3)-RFP where R/K residues in NLS3 are mutated to alanine residues.

the construct CLK1(Δ NLS12)-RFP. Results from both confocal imaging and fractionation experiments indicated that CLK1(Δ NLS12)-RFP was largely nuclear (Fig. 3.2 B,C). A further search using cNLS mapper identified a putative bipartite NLS which we named NLS1*. We then mutated NLS1* along with NLS2 in the construct CLK1(Δ NLS12*)-RFP and studied its effect on subcellular localization. Even with all potential classical NLSs mutated out, CLK1(Δ NLS12*)-RFP showed minimal impairment of nuclear entry (Fig.3.2B). Overall, these findings indicate that the CLK1 N-terminus does not likely contain a classical NLS. We next expanded our search for potential classical NLSs in the kinase domain finding a small stretch of basic residues on the MAP kinase like insert domain of CLK1 titled NLS3. We mutated NLS3 to alanines on CLK1-RFP generating the construct CLK1(Δ NLS3)-RFP and transfected it into HeLa cells to test for its effect on nuclear localization. We found that the CLK1(Δ NLS3)-RFP also localized largely in the nucleus (Fig. 3.2D,E). Taken together, these results suggest that CLK1 does not possess a short, classical NLS and, thus, likely uses a non-canonical mechanism for nuclear import.

3.2.2 Identifying regions important for nuclear localization

Having shown that none of the specific mutations in short, basic NLS's impaired the nuclear localization of CLK1, we decided to target broader regions of the N-terminus for mutagenesis. For simplicity, we divided the roughly 150-residue N-terminus into three blocks of approximately 50 residues each (Fig 3.3A). We deleted one block of residues at a time in constructs CLK1(Δ 1)-RFP, CLK1(Δ 2)-RFP and CLK1(Δ 3)-RFP and studied their subcellular localization in fixed HeLa cells. Surprisingly, all three single block deletions localized in the nucleus (Fig. 3.3B). These results suggest that nuclear import of CLK1 is not mediated by any specific sequence within the 50-residue blocks in the N-terminus. We further generated deletion mutants of two blocks at a time in two new constructs CLK1(Δ 13)-RFP and CLK1(Δ 23)-RFP and found that these constructs

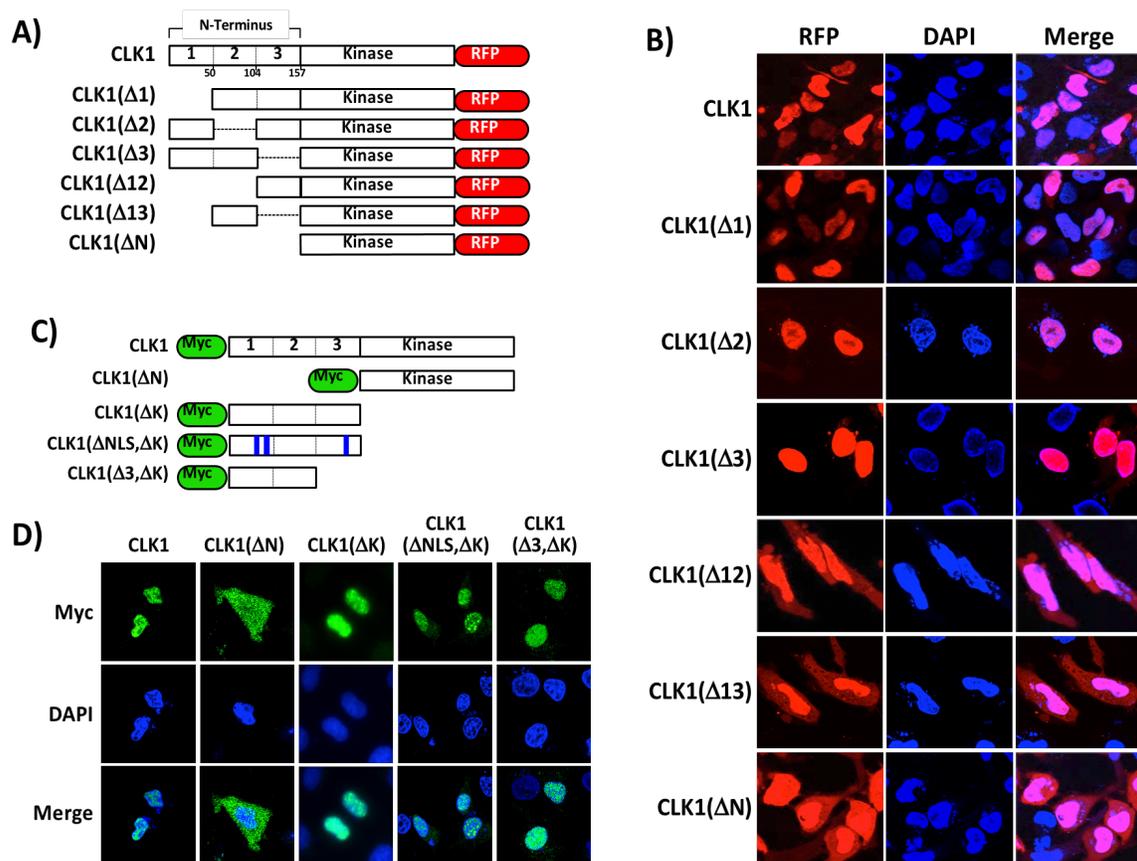


Figure 3.3: Mapping broad regions in the N-terminus necessary for nuclear localization. **(A)** Deletion constructs made in the CLK1-RFP vector. The N-terminus was divided into three blocks of roughly fifty residues each. **(B)** Confocal imaging of fixed HeLa cells expressing CLK-RFP and various deletion mutants from panel (A). Cells were fixed with paraformaldehyde and stained with DAPI for visualization of the nucleus. **(C)** Myc-tagged CLK1 N-terminus constructs used for studying the localization of the N-terminus in the absence of the kinase domain. **(D)** Immunofluorescence images of HeLa cells overexpressing Myc-CLK1 constructs probed with α -myc antibody. Myc-CLK1 and Myc-(Δ N) CLK1 are imaged as controls.

expressed both in the cytoplasm and the nucleus, similar CLK1(Δ N)-RFP (Fig. 3.3B). These results suggest that any two-block combination of N-terminal residues is sufficient for nuclear import.

To support the idea that the CLK1 N-terminus is solely essential for nuclear localization of CLK1, we next determined if the N-terminus of CLK1 is sufficient for nuclear entry. We generated several myc-tagged N-terminal constructs (Fig. 3.3C). Since the myc tag lacks intrinsic fluorescence signal unlike the RFP tags, we probed its localization using immunofluorescence in fixed cells using anti-myc antibodies. As a control, we tested the localization of myc-CLK1 and myc-CLK1(Δ N) and confirmed that their localization was in agreement with that observed for CLK1-RFP and CLK1(Δ N)-RFP (Fig. 3.3D). Next, we determined the subcellular localization of a form of CLK1 lacking the kinase domain (CLK1(Δ K)) to address whether the kinase domain plays a role in nuclear localization. We found that CLK1(Δ K) was completely nuclear, suggesting that the N-terminus alone can gain complete access to the nucleus in the absence of the kinase domain (Fig. 3.3D). We wished to explore whether the nuclear import of CLK1(Δ K) may be due to the classical NLS's on the N-terminus that are otherwise inaccessible to the Importin α/β system in the presence of the kinase domain. To address this, we tested the localization of a new construct myc-CLK1(Δ NLS, Δ K) that lacks the three classic NLS's identified by cNLS mapper and found that this construct was still nuclear based on immunofluorescence (Fig. 3.3D). We further investigated what happens with the sequential deletion of a single block from CLK1(Δ K) (CLK1(Δ 3, Δ K)) and found that the deletion of block 3 still did not affect the nuclear localization of the N-terminus by itself (Fig. 3.3D). Taken together, these results indicate that the nuclear import of CLK1 is not mediated by a short basic sequence often observed in traditional nuclear

import systems involving Importin α/β . Instead, the localization sequence is rather broad and appears to be diffused throughout the N-terminus.

3.2.3 CLK1 nuclear entry is mediated by the nuclear transport system for SR proteins

Several studies published about two decades ago established that SR protein nuclear import is mediated by Transportin-SR2 (TRN-SR2) (12,13). TRN-SR2 binds with high affinity to phosphorylated as opposed to unphosphorylated SR proteins, thereby establishing that the nuclear import of SR proteins is a phosphorylation-dependent phenomenon. Prior studies have established that, in the case of the prototype SR protein SRSF1, the RS1 segment in the C-terminal RS domain gets extensively phosphorylated in the cytoplasm by SRPK1 (30). These phosphorylation events provide the key recognition element for TRN-SR2 binding to SRSF1 (31). An X-ray structure for TRN-SR2 (TNPO3) was solved with a truncated form of SRSF1 (ASF/SF2) that shows critical interactions between a cluster of arginines in the transportin and the phosphates on Ser-207 and Ser-209 in the RS1 segment (14). Since the CLK1 N-terminus is also an RS-like domain that contains several known phosphorylation sites (S61, T138, S140), we considered the possibility that CLK1 could be imported into the nucleus by TRN-SR2, using a recognition mechanism similar to that for SR proteins. Furthermore, a previous study identified CLK2 and CLK3 as binding partners of TRN-SR2 using co-immunoprecipitation and proteomic mass spectrometry underscoring the possibility that CLKs may gain access to the nucleus in the same manner as SR proteins (14).

To test the above idea, we first investigated if CLK1 interacts with TRN-SR2 using immunoprecipitation assays. We overexpressed and immunoprecipitated FLAG-tagged CLK1 and CLK1(Δ N) and looked for endogenous TRN-SR2 using western blots. We found that CLK1-FLAG interacted with endogenous TRN-SR2 but not CLK1(Δ N), implying that the kinase binds

TRN-SR2 via its N-terminus, a result expected if the transportin associates directly with the phosphorylated N-terminus (Fig 3.4.E). We next wished to study if the subcellular localization of CLK1 was altered by TRN-SR2 down-regulation. To investigate this idea, we tested the effect of siRNA knockdown of TRN-SR2 on CLK1 localization using both confocal microscopy and biochemical fractionation. We first confirmed that siRNA transfection lowered endogenous TRN-SR2 levels by comparing the lysates of siRNA treated and untreated HeLa cells (Fig. 3.4D). Confocal microscopy experiments indicate that while CLK1-RFP localized to the nucleus in untreated cells, as expected, CLK1-RFP localized to both the cytoplasm and nucleus upon siRNA knockdown (Fig. 3.4A). Similar results were also found using biochemical fractionation experiments. Whereas overexpressed CLK1-FLAG localized to the nucleus, CLK1-FLAG was found in both the cytoplasm and nucleus upon siRNA knockdown (Fig. 3.4B). To rule out the possibility of cytoplasmic localization resulting from overexpression, we also confirmed that the siRNA knockdown of TRN-SR2 also resulted in the cytoplasmic localization of endogenous CLK1 in fractionation experiments (Fig. 3.4C). Overall, these results suggest that the nuclear import of CLK1 is dependent on TRN-SR2 and the CLK1 N-terminus.

3.2.4 CLK1 binding to TRN-SR2 is mediated by phospho-SRSF1

It was previously shown in our lab that CLK1 forms a very tight complex with SRSF1 independent of the RS domain phosphorylation state (32). Since phospho-SRSF1 is imported into the nucleus through TRN-SR2, we explored the possibility that CLK1 could also gain entry to the nucleus owing to its tight association with the SR protein in a so-called “piggyback” mechanism. To address this exciting possibility, we wondered whether both the SR protein and CLK1 nuclear localization are simultaneously regulated by TRN-SR2.

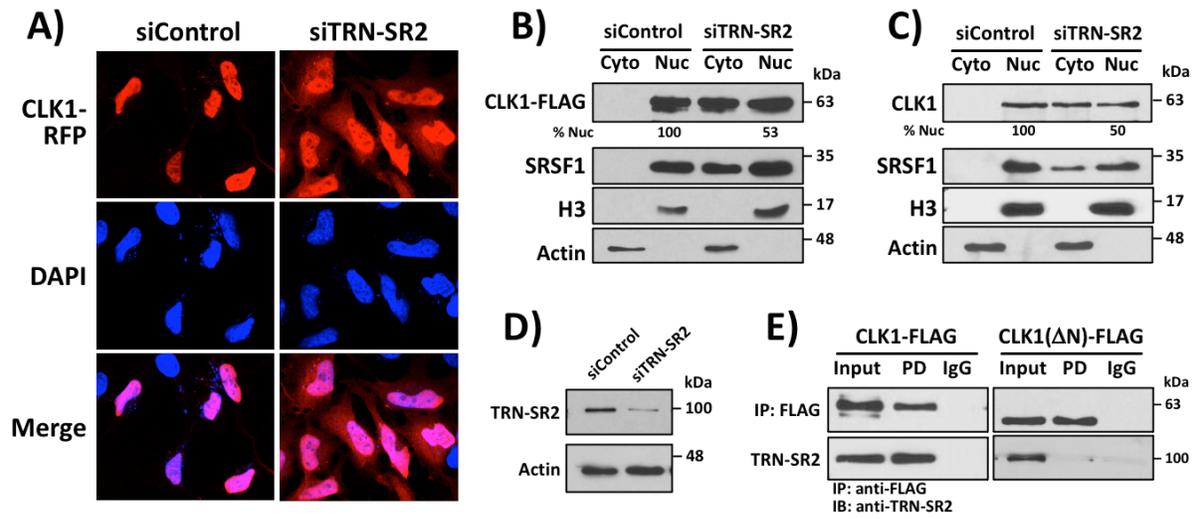


Figure 3.4: CLK1 nuclear import is mediated by TRN-SR2. **(A)** Confocal images of HeLa cells transfected with CLK1-RFP in the presence and absence of TRN-SR2 siRNA. **(B)** Biochemical fractionation immunoblots of HeLa cells transfected with CLK1-FLAG upon TRN-SR2 siRNA knockdown. The siRNA treatment results in cytoplasmic expression of CLK1-FLAG. % Nuclear protein as calculated by ImageJ integration is included. **(C)** Fractionation of HEK cells upon TRN-SR2 siRNA knockdown showing cytoplasmic expression of endogenous CLK1 in the absence of TRN-SR2. **(D)** Immunoblots comparing the TRN-SR2 expression in the presence and absence of siRNA. **(E)** Co-immunoprecipitation assays done with HeLa cell lysates overexpressing CLK1-FLAG and CLK1(Δ N)-FLAG. CLK1-FLAG interacts with TRN-SR2 but CLK1(Δ N)-FLAG does not.

We first showed that siRNA knockdown of TRN-SR2 resulted in the cytoplasmic localization of both CLK1-RFP and GFP-SRSF1 using confocal microscopy (Fig. 3.5A). Next, we wished to determine if the interaction of CLK1 with TRN-SR2 is a direct or an indirect binding event mediated by SRSF1. To get started, we first verified that the binding of SRSF1 to TRN-SR2 is phosphorylation dependent. We immunoprecipitated endogenous TRN-SR2 from cytoplasmic fractions of HeLa cells rather than whole cells to remove SRSF1 and CLK1 for our binding assays

(Fig.3.5B). We then performed pulldown assays with recombinant GST-tagged SRSF1 (GST-SRSF1) and immunoprecipitated TRN-SR2 in the presence and absence of catalytic amounts of SRPK1. We found that GST-SRSF1 interacted with TRN-SR2 only in the presence of SRPK1 and ATP suggesting that the binding of SRSF1 with TRN-SR2 happens only when the RS domain is phosphorylated (Fig. 3.5C).

Next, we wished to understand if the binding of CLK1 to TRN-SR2 was a direct or indirect interaction mediated by SRSF1. We performed pulldown assays with TRN-SR2 immunoprecipitated from cytoplasmic fractions and observed that the TRN-SR2 interacted with recombinant His-CLK1 only in the presence of SRPK1-phosphorylated GST-SRSF1 (Fig. 3.5D). To confirm this, we also looked at the binding of immunoprecipitated His-CLK1 with endogenous TRN-SR2 from cytoplasmic fractions and found that the binding does not happen without SRPK1 phosphorylation of GST-SRSF1 (Fig. 3.5E). We also showed using pulldown assays with GST-SRSF1 and His-CLK1 that CLK1 binds SRSF1 in both its phosphorylated and unphosphorylated states (Fig. 3.5F). Overall, these results suggest that the interaction of CLK1 with TRN-SR2 is mediated by phosphorylated SRSF1. Hence, SR proteins act as adaptor proteins by carrying CLK1 on its back as it gets imported into the nucleus by TRN-SR2, thereby leading to a ‘piggyback’ mechanism of nuclear import for CLK1.

3.2.5 CLK1 uses the phospho-NLS on SRSF1 for nuclear entry

Having shown that the nuclear import of CLK1 is dependent on phosphorylated SRSF1, we were next interested in knowing if disruption of the nuclear import of the SR protein would also impair the nuclear import of CLK1. Previous studies showed that SRPK1-dependent phosphorylation of about 8 serines on the N-terminal half of the RS domain (RS1) is important for binding to TRN-SR2 and subsequent nuclear import (14). We mutated those 8 serines to alanines

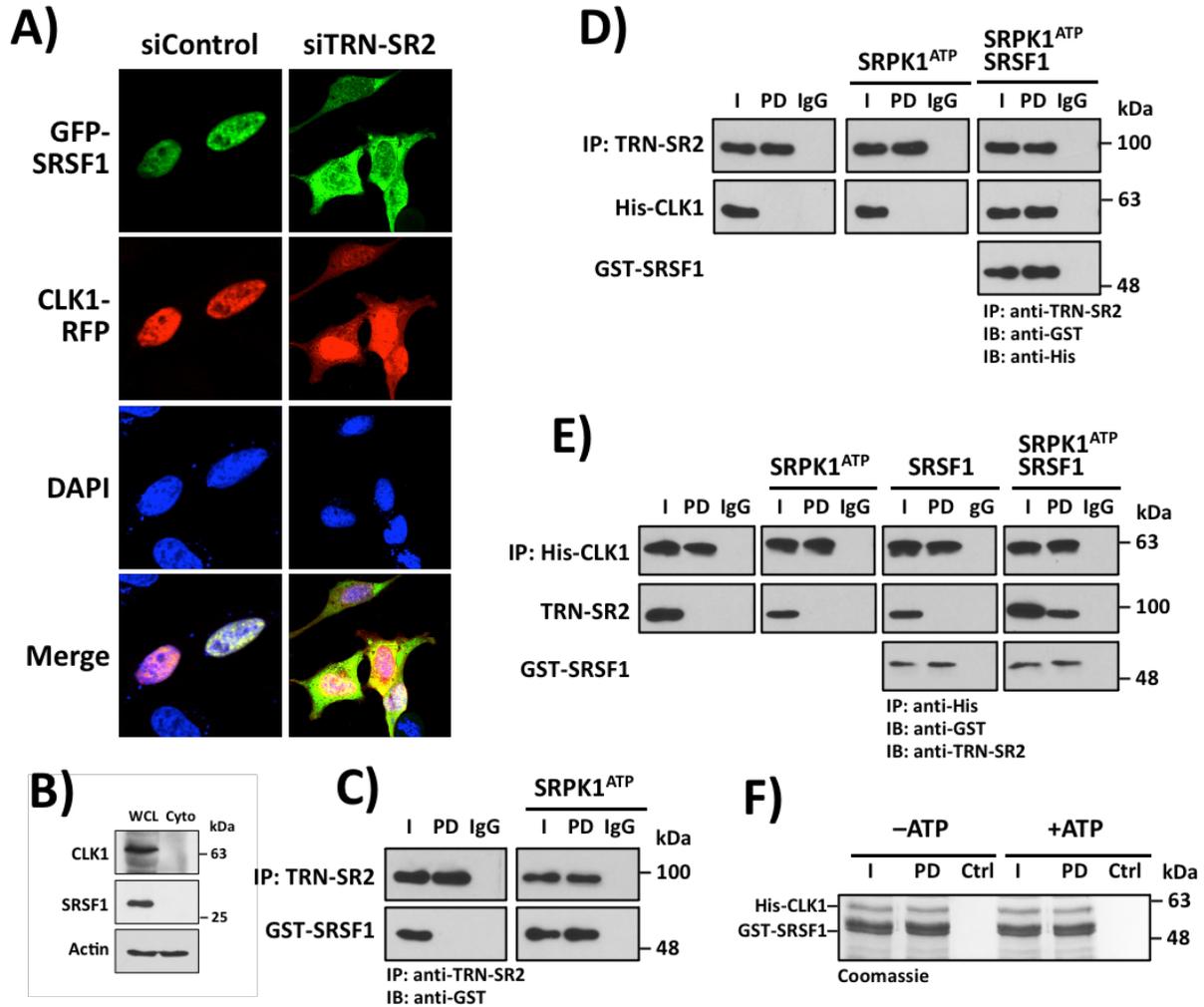


Figure 3.5: CLK1-TRN-SR2 interaction is mediated by phospho-SRSF1. **(A)** Confocal imaging of fixed HeLa cells showing colocalization of CLK1-RFP and GFP-SRSF1 upon siRNA knockdown of TRN-SR2. **(B)** Cytoplasmic fraction of HeLa cells lack detectable levels of SRSF1. **(C)** Pull-down assay of GST-SRSF1 and phospho-GST-SRSF1 with TRN-SR2 immunoprecipitated from cytoplasmic HeLa cells. SRPK1-phosphorylated GST-SRSF1 pulls down TRN-SR2 but unphosphorylated GST-SRSF1 does not. **(D)** Co-immunoprecipitation assays showing that TRN-SR2 immunoprecipitated from cytoplasmic fractions interact with His-CLK1 only in the presence of SRPK1-phosphorylated GST-SRSF1. TRN-SR2 was immunoprecipitated using α -TRN-SR2 antibody conjugated resin and IgG refers to a resin lacking the antibody. **(E)** Immunoprecipitated recombinant His-CLK1 interacts with TRN-SR2 from HeLa cell cytoplasmic fractions only in the presence of SRPK1-phosphorylated GST-SRSF1. Recombinant His-CLK1 was added to cytoplasmic fraction and immunoprecipitated with α -His antibody. IgG refers to agarose resin control without the antibody. **(F)** Pull-down assays with phosphorylated and unphosphorylated GST-SRSF1 and His-CLK1 showing that His-CLK1 binds to GST-SRSF1 in both phosphorylated and unphosphorylated forms.

in the construct GFP-SRSF1^{mut} and monitored its subcellular localization using confocal microscopy (Fig. 3.6A). As expected, we found that GFP-SRSF1^{mut} primarily localized to the cytoplasm while GFP-SRSF1 was exclusively nuclear. More importantly, when we overexpressed CLK1-RFP along with GFP-SRSF1^{mut}, we found that CLK1-RFP also colocalized to the cytoplasm (Fig. 3.6B). We corroborated these results from the confocal microscopy with biochemical fractionation experiments and showed that GFP-SRSF1^{mut} overexpression also results in the cytoplasmic accumulation of endogenous CLK1 (Fig. 3.6D). In order to ensure that the mutations in GFP-SRSF1^{mut} did not affect CLK1 binding, we immunoprecipitated CLK1-FLAG and looked at its interaction with GFP-SRSF1 and GFP-SRSF1^{mut}. We found that CLK1-FLAG interacted with GFP-SRSF1 and GFP-SRSF1^{mut}, indicating that the Ser-to-Ala mutations did not impair CLK1 binding (Fig. 3.6C). Overall, these results suggest that CLK1 uses the phospho-NLS on SRSF1 to gain nuclear access and that disruption of this NLS impairs nuclear import of CLK1.

3.2.6 CLK1 nuclear import correlates with SRSF1 binding affinity

Having shown in prior studies that the CLK1 N-terminus is required for SRSF1 binding and phosphorylation (32), we next wished to determine if weakened SR protein affinity would result in diminished CLK1 nuclear import efficiency. We purified recombinant GST-tagged CLK1 N-terminus (GST-N) and several deletion mutants and looked at their binding to His-SRSF1 in pulldown assays (Fig. 3.7A). We observed that immunoprecipitated His-SRSF1 robustly pulled down GST-N whereas it bound less potently to GST-N(Δ 3) and did not interact with GST-N(Δ 23) (Fig. 3.7A). Owing to co-migration with His-SRSF1, we had to perform western blots to visualize the pulldown efficiency of GST-N(Δ 23). These results also complement the previous findings

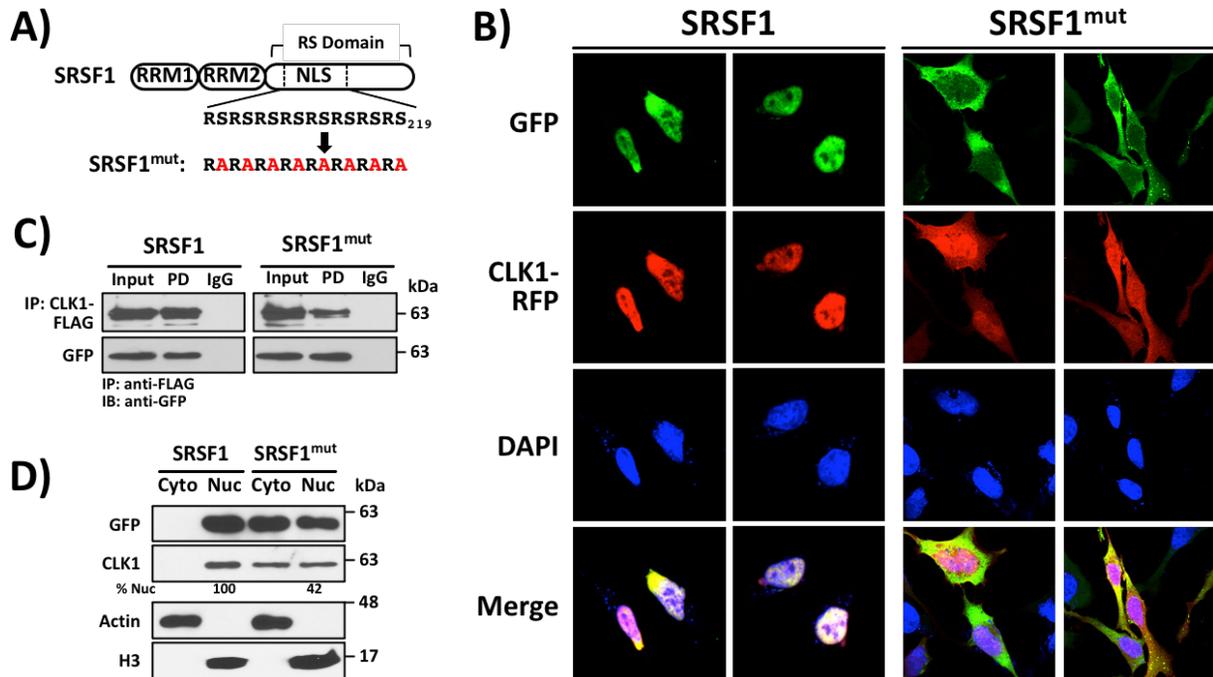


Figure 3.6: Disruption of the SRSF1 NLS drives CLK1 into the cytoplasm. **(A)** Ser-to-Ala mutations in RS1 of SRSF1 leads to the cytoplasmic localization of SRSF1. **(B)** Confocal imaging of HeLa cells co-transfected with GFP-SRSF1 and CLK1-RFP show both proteins localized to the nucleus whereas GFP-SRSF1^{mut} and RFP-CLK1 both localized to the cytoplasm. **(C)** Co-immunoprecipitation assays show that Ser-to-Ala mutations do not impair the binding of CLK1 to SRSF1. CLK1-FLAG was immunoprecipitated from HeLa cell lysates using α -FLAG antibody. Both GFP-SRSF1 and GFP-SRSF1^{mut} binds to CLK1-RFP. **(D)** Biochemical fractionation shows cytoplasmic localization of endogenous CLK1 when GFP-SRSF1^{mut} is overexpressed. GFP-SRSF1 was immunoprecipitated using α -GFP and bound endogenous CLK1 was probed using α -CLK1 antibody. The % nuclear protein quantitated using Image J integrations are shown at the bottom.

that phosphorylation is not required for the interaction of SRSF1 and CLK1 N-terminus as both His-SRSF1 and the GST-N constructs are bacterially expressed and are unphosphorylated. We further corroborated these results by performing pulldown assays with GST-SRSF1 and several His-CLK1 deletion constructs. We purified full-length His-CLK1 and His-CLK1(Δ 12) from insect cells using baculovirus expression systems and His-CLK1 (Δ N) from bacteria. We found that

while immunoprecipitated GST-SRSF1 pulled down full-length His-CLK1, it bound with lower affinity to His-CLK1(Δ 12) and did not bind to His-CLK1(Δ N) (Fig. 3.7B). Thus, the pull-down results indicated that with progressive deletion of the N-terminus, the binding affinity of GST-SRSF1 weakened. To ensure that the progressive reduction in the binding affinity was not an artifact arising from recombinant, purified proteins, we overexpressed several FLAG-CLK1 deletion constructs in HeLa cells mirroring those made in the CLK1-RFP constructs, and looked at their interaction with endogenous SRSF1 in co-immunoprecipitation assays. We observed that whereas endogenous, immunoprecipitated SRSF1 interacted robustly with FLAG-CLK1, it bound less efficiently to FLAG-CLK1(Δ 1) and FLAG-CLK1(Δ 12) and did not bind to FLAG-CLK1(Δ N) (Fig. 3.7C). The decrease in binding affinity also mirrors the nuclear localization of the deletion constructs since a deletion of 100 or more residues resulted in the cytoplasmic expression of CLK1-RFP. Taken together, these results suggest that the high-affinity interaction of the CLK1 N-terminus with SRSF1 drives the nuclear entry of CLK1.

3.2.7 The nuclear import of CLK1(Δ N).

Although deletion of the N-terminus significantly impairs nuclear import of CLK1, approximately 60% of overexpressed CLK1(Δ N)-RFP still localized to the nucleus (Fig.3.2). This finding was somewhat surprising since the molecular mass of CLK1(Δ N)-RFP is 62 kDa and is not expected to readily pass through the nuclear pore to this extent by a passive transport mechanism (2). To investigate this further, we considered the possibility that NLS3 in CLK1(Δ N), while not important for nuclear localization of the full-length CLK1 (Fig. 3.1E), may play a role upon deletion of the N-terminus as such deletion could expose this sequence for a classical transport mechanism with Importin α/β . To test this possibility, we generated a new mutant kinase by mutating the NLS3 residues to alanines in CLK1(Δ N, Δ NLS3)-RFP and comparing its

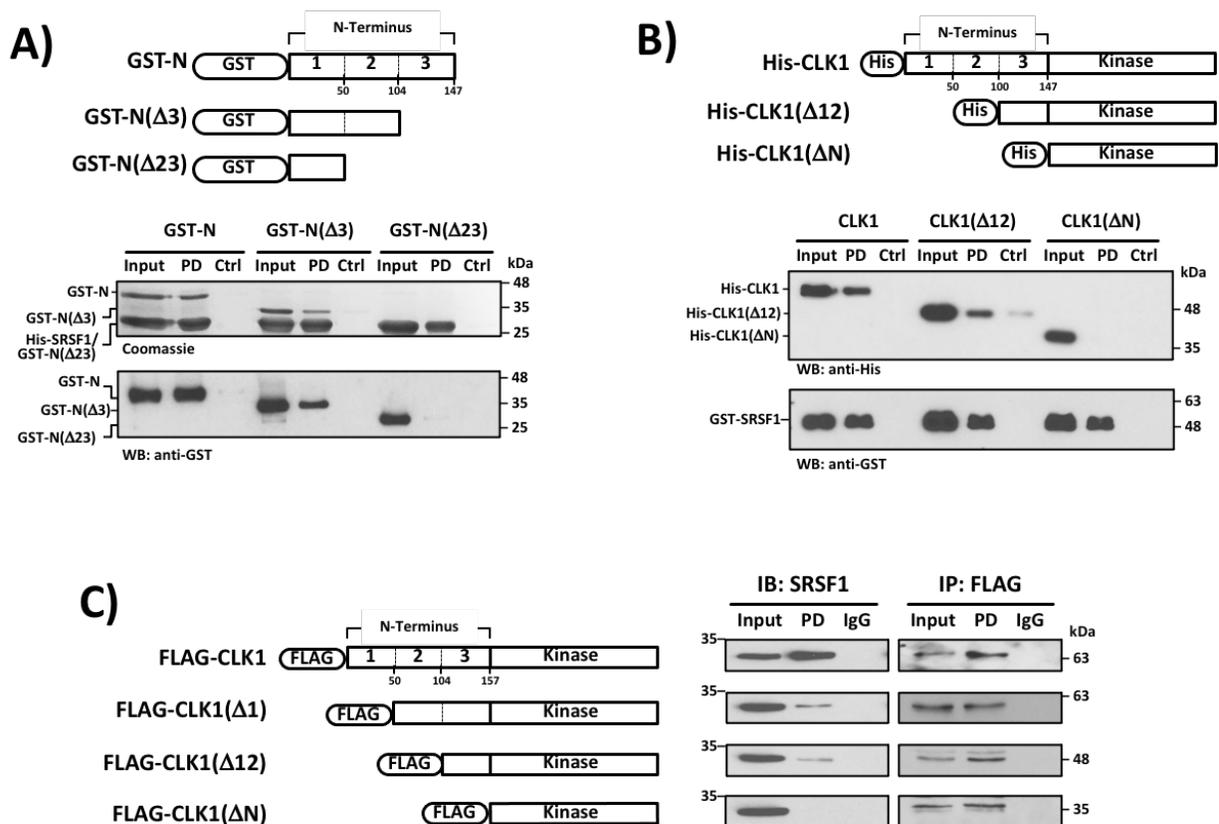


Figure 3.7: Deletion of residues in the CLK1 N-terminus weakens SRSF1 binding. **(A)** Pull-downs show weakened interaction of GST N deletion constructs with His-SRSF1. Deletion mutants of GST-N-terminus constructs are shown above the pull-downs. **(B)** Pull-downs of GST-SRSF1 with His-CLK1 deletion constructs. The His-CLK1 deletion constructs are shown at the top. **(C)** Co-immunoprecipitations showing that the fraction of endogenous SRSF1 in HeLa cells bound to overexpressed FLAG-CLK1 deletion constructs is lowered with sequential deletion of N-terminal residues.

subcellular localization to other forms of RFP-tagged CLK1 (Fig. 3.8A). Confocal microscopy on HeLa cells transfected with CLK1(ΔN)-RFP showed expression in both the nucleus and the cytoplasm, compared to full-length CLK1-RFP and CLK1(ΔNLS3)-RFP that are fully nuclear (Fig. 3.8B). While CLK1(ΔN, ΔNLS3)-RFP was expressed in both the cytoplasm and nucleus, it

was more expressed in the cytoplasm compared to CLK1(Δ N)-RFP (Fig. 3.8B). Quantification of RFP signal from the nucleus and cytoplasm showed that, on an average, roughly 50% of CLK1(Δ N)-RFP was nuclear but less than 30% of CLK1(Δ N, Δ NLS3)-RFP was nuclear (Fig.3.8B). These results suggest that NLS3 can alter the subcellular distribution of CLK1(Δ N), but not CLK1, implying that NLS3 has minimal impact on the nuclear localization of CLK1. Taken together, these findings suggest that the N-terminus is the major driver of nuclear import for CLK1 through the “piggyback” mechanism but deletions may induce other effects on localization that need to be considered.

3.3 Discussion

Unlike most protein kinases that use docking grooves in either their kinase domains or a neighboring structured domain for substrate recognition, CLKs depart from this classical arrangement by using an N-terminus with RS-like character that lacks structure to bind the RS domains of SR proteins (33). Prior studies have shown that the latter interactions, while still not understood at a structural level, are highly stable and, in the case of SRSF1, generate a very low K_m of about 100 nM, a value close to that for SRPK1 which contains a well-structured docking groove for RS domain binding (33,34). Thus, CLKs use disorder-disorder interactions to efficiently target SR proteins, phosphorylating Ser-Pro dipeptides for splicing function. In this new study, we identify a new function for the high-affinity binding of CLK1 to SRSF1. Here, we show that the stable interaction of CLK1 and its SR protein substrate is not only important for splicing factor activation but also necessary for the efficient nuclear import of CLK1. We show that through its high-affinity interactions with SRSF1, CLK1 gains entry to the nucleus using the phospho-NLS on the RS domain of the SR protein in a “piggyback” mechanism (Fig. 3.9).

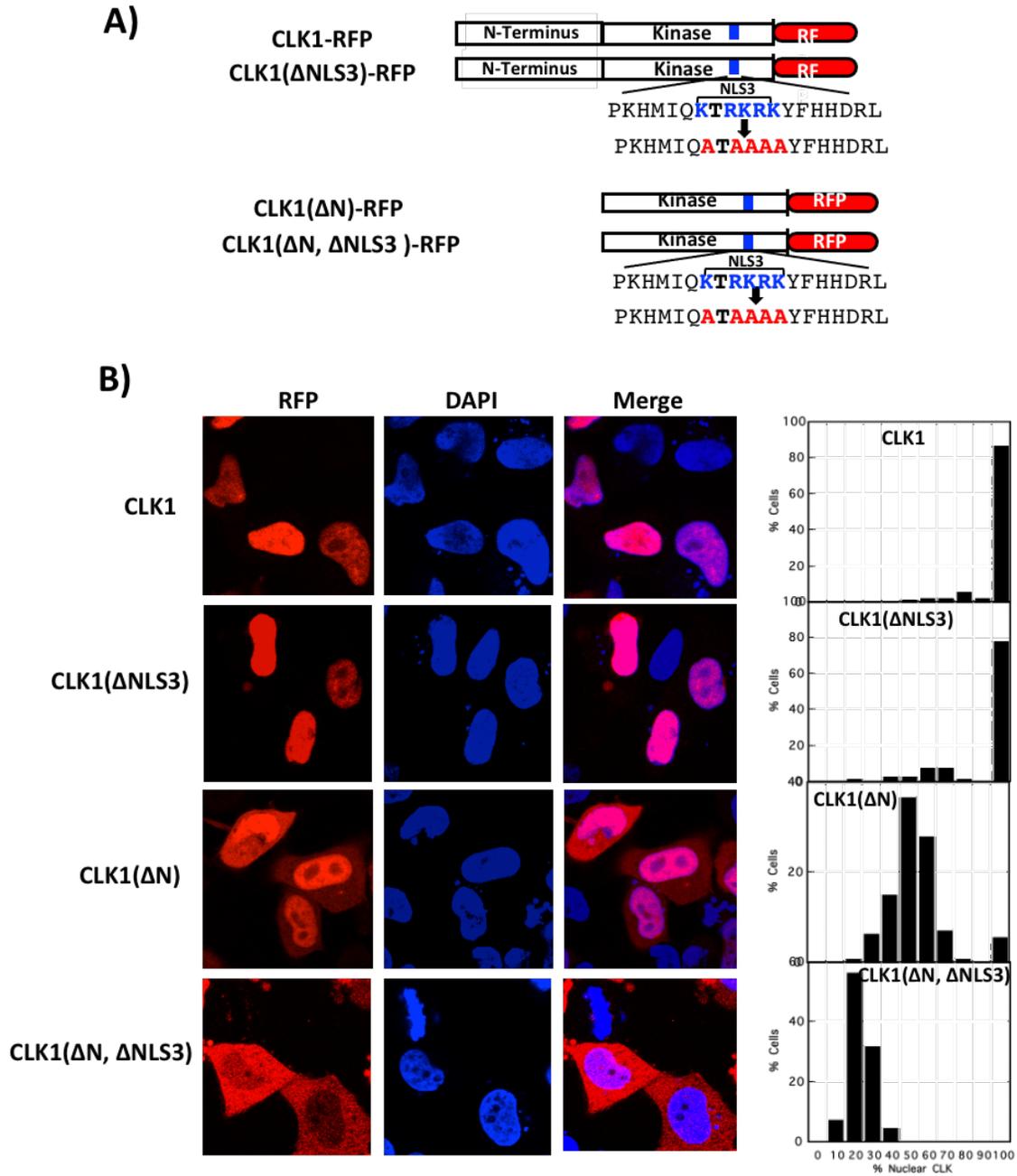


Figure 3.8. Mutation of NLS3 impairs the nuclear import of CLK1(Δ N), but not CLK1. **(A)** NLS3 mutant constructs **(B)** Confocal imaging of fixed HeLa cells expressing various CLK1-RFP mutants.

The majority of the nuclear proteins are imported into the nucleus via a well characterized, classical nuclear import pathway. In this pathway, Importin α recognizes a short NLS on the cargo, which is mostly composed of a small patch of basic residues (1,5,6). Interestingly, although the CLK1 N-terminus has several classical NLSs, our results show that these sequences are not vital for nuclear import. We speculate that there could be two reasons why these basic NLS's are not functional in a classic Importin α/β mechanism. First, since CLK1 N-terminus forms a high-affinity complex with the RS domain in its substrate SRSF1, the classical NLSs may not be accessible for Importin α binding. Second, the strong ability for CLK1 to self-associate and form large oligomers through its N-terminus might also mask the classical NLSs blocking Importin α association. Although we have found that these sequences play no role in nuclear transport, it is possible that under certain cellular conditions where SRSF1 binding or oligomerization is hindered, the classical NLSs on the N-terminus might serve as active binding sites for Importin α leading to nuclear import through the classical mechanism.

While the N-terminus has been shown to be a critical modifier of CLK1 subcellular localization, our results also demonstrate that the CLK1(Δ N)-RFP construct is not excluded from the nucleus, a result we originally anticipated based on its mass (62 kDa). Since there can be multiple nuclear import mechanisms for any protein, we do not exclude the possibility that alternative mechanisms may exist for CLK1 upon N-terminal deletion. Even though our immunoprecipitation and pulldown assays do not indicate a strong interaction between CLK1(Δ N) and SRSF1, there could be transient interactions as we have previously noted that the CLK1(Δ N) does phosphorylate SRSF1 albeit with minimal efficiency (32). Another possibility is the unmasking of a latent signal in the absence of the N-terminus, as it has also been well established from prior studies that the removal of the N-terminus drastically changes the quaternary structure

of CLK1 from an oligomer to a monomer (33). Based on our observation that removal of a putative NLS (NLS3) in CLK1 lacking its N-terminus leads to increased cytoplasmic localization, we speculate the kinase domain may contain a weak, basic NLS that supports some nuclear import.

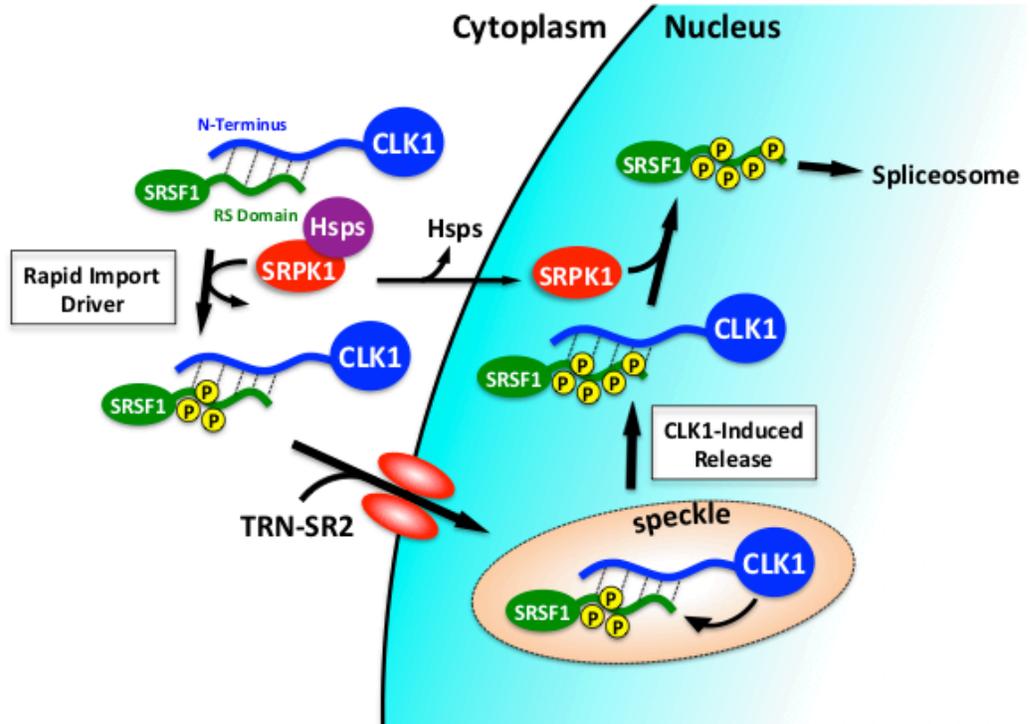


Figure 3.9. Summary of CLK1 nuclear import mechanism. In the cytoplasm, SRSF1 is phosphorylated in the RS domain by SRPK1. CLK1 forms a high affinity complex with phosphorylated SRSF1. The CLK1-SRSF1 complex is imported into the nucleus by TRN-SR2, where they accumulate in membrane-less storage organelles called nuclear speckles. Hyper-phosphorylation of SRSF1 in the nucleus by CLK1 results in the mobilization of the complex, but CLK1 binds tightly to phosphorylated SRSF1 hindering the product release. SRPK1, being predominantly cytoplasmic, is imported into the nucleus upon shedding its anchoring Heat Shock Proteins (Hsps). Nuclear SRPK1 binds to the CLK1-SRSF1 complex, releasing SRSF1 to the spliceosome.

Another implication of having such a piggyback mechanism for nuclear import is potential unidentified functions for CLK1 in the cytoplasm. SR proteins, being nucleo-cytoplasmic shuttling proteins, have long been known to play critical roles in exporting processed mRNAs by transporting mRNAs into the cytoplasm making them available for translation (35). In certain mouse tissues, SRSF1 has been shown to localize to cytoplasmic granules along with CLK1 and SRPK1 (36). Since we showed that CLK1 can colocalize with SRSF1 in the cytoplasm in the absence of an NLS on SRSF1, it is also possible that the high affinity of CLK1 to SRSF1 might also drive CLK1 to the cytoplasm if triggered by environmental factors. This raises the possibility of CLKs performing additional functions in the cytoplasm.

3.4. Conclusions

In this study, we deciphered the nuclear import mechanism of human CLK1. The nuclear localization of CLK1 was previously thought to be due to the classical NLSs present on the N-terminus. Here, we showed that the nuclear import is mediated by the karyopherin TRN-SR2 although CLK1 does not directly interact with the transportin. Instead, TRN-SR2 binds SRSF1, the physiological substrate of CLK1, which then associates with TRN-SR2, forming a ternary complex (Fig. 3.8). Although CLK1 binds SRSF1 with high affinity whether its RS domain is phosphorylated or not, only SRPK1-phosphorylated SRSF1 binds TRN-SR2 and can be transported into the nucleus. Such “piggybacking” has been reported to be the nuclear import mechanism for a few multi-subunit protein complexes such as the ribosomal machinery, the RNA polymerase II and transcription factor-II D (37–39). These proteins were identified as binding partners of importin- α using proteomic analyses but did not possess a classical NLS (38,40). While “piggybacking” as a nuclear import mechanism has been demonstrated before, what makes CLK1 nuclear import unique is that the binding of CLK1 to its substrate is completely mediated by

disordered domains. Nuclear import, in general, is a highly ordered transition. However, CLK1 can harness disorder-disorder interactions for this highly ordered transition from the cytoplasm to the nucleus.

Chapter 3, in part, is a reprint of the material as it appears in George A, Aubol BE, Fattet L, Adams JA. “Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate”, *J Biol Chem.* **2019**, *294*, 9631–41. The dissertation author was a primary investigator and author of this paper.

References

1. Marfori M, Mynott A, Ellis JJ, Mehdi AM, Saunders NFW, Curmi PM, Forwood JK, Boden M, Kobe B. Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta - Mol Cell Res.* 2011;1813(9):1562–77.
2. Kim YH, Han ME, Oh SO. The molecular mechanism for nuclear transport and its application. *Anat Cell Biol.* 2017;50(2):77–85.
3. Fahrenkrog B, Aebi U. The nuclear pore complex: Nucleocytoplasmic transport and beyond. *Nat Rev Mol Cell Biol.* 2003;4(10):757–66.
4. Freitas N, Cunha C. Mechanisms and Signals for the Nuclear Import of Proteins. *Curr Genomics.* 2009;10(8):550–7.
5. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA. Importin α : A multipurpose nuclear-transport receptor. *Trends Cell Biol.* 2004;14(9):505–14.
6. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: Definition, function, and interaction with importin α . *J Biol Chem.* 2007;282(8):5101–5.
7. Kutay U, Ralf Bischoff F, Kostka S, Kraft R, Görlich D. Export of importin α from the nucleus is mediated by a specific nuclear transport factor. *Cell.* 1997;90(6):1061–71.
8. Chuderland D, Konson A, Seger R. Identification and Characterization of a General Nuclear Translocation Signal in Signaling Proteins. *Mol Cell.* 2008;31(6):850–61.
9. Siomi H, Dreyfuss G. A nuclear localization domain in the hnRNP A1 protein. *J Cell Biol.* 1995;129(3):551–60.
10. Twyffels L, Gueydan C, Kruys V. Transportin-1 and Transportin-2: Protein nuclear import and beyond. *FEBS Lett.* 2014;588(10):1857–68.
11. Nardozi JD, Lott K, Cingolani G. Phosphorylation meets nuclear import: A review. *Cell Commun Signal.* 2010;8:1–17.
12. Lai MC, Lin RI, Tarn WY. Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A.* 2001;98(18):10154–9.
13. Kataoka N, Bachorik JL, Dreyfuss G. Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol.* 1999;145(6):1145–52.
14. Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Oztop I, Lee K, Pye VE, Cosnefroy

- O, Snijders AP, KewalRamani VN, Fassatu A, Engelman A, Cherepanov, P. Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci.* 2014;111(7):2728–33.
15. Logue EC, Taylor KT, Goff PH, Landau NR. The Cargo-Binding Domain of Transportin 3 Is Required for Lentivirus Nuclear Import. *J Virol.* 2011;85(24):12950–61.
 16. Demeulemeester J, Blokken J, Houwer S, Dirix L, Klaassen H, Marchand A, Chaltin P, Christ F, Debyser Z. Inhibitors of the integrase-transportin-SR2 interaction block HIV nuclear import. *Retrovirology.* 2018;15(1):1–13.
 17. Mühlhäusser P, Müller EC, Otto A, Kutay U. Multiple pathways contribute to nuclear import of core histones. *EMBO Rep.* 2001;2(8):690–6.
 18. Ranganathan A, Yazicioglu MN, Cobb MH. The nuclear localization of ERK2 occurs by mechanisms both independent of and dependent on energy. *J Biol Chem.* 2006;281(23):15645–52.
 19. Yazicioglu MN, Goad DL, Ranganathan A, Whitehurst AW, Goldsmith EJ, Cobb MH. Mutations in ERK2 binding sites affect nuclear entry. *J Biol Chem.* 2007;282(39):28759–67.
 20. Khokhlatchev A V., Canagarajah B, Wilsbacher J, Robinson M, Atkinson M, Goldsmith E, Cobb, MH. Phosphorylation of the MAP kinase ERK2 promotes its homodimerization and nuclear translocation. *Cell.* 1998;93(4):605–15.
 21. Zhong XY, Ding JH, Adams JA, Ghosh G, Fu XD. Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes Dev.* 2009;23(4):482–95.
 22. Lee G, Zheng Y, Cho S, Jang C, England C, Dempsey JM, Yu Y, Liu X, He L, Cavaliere PM, Chavez A, Zhnag E, Isik M, Couvillon A, Dephoure NE, Blackwell TK, Yu JJ, Rabinowitz JD, Cantley LC, Blenis J. Post-transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. *Cell.* 2017;171(7):1545-1558.e18.
 23. Keshwani MM, Aubol BE, Fattet L, Ma CT, Qiu J, Jennings PA, Fu XD, Adams JA. Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochem J.* 2015;466:311–22.
 24. Duncan PI, Stojdl DF, Marius RM, Scheit KH, Bell JC. The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. *Exp Cell Res.* 1998;241(2):300–8.
 25. Nayler O, Stamm S, Ullrich A. Characterization and comparison of four serine- and arginine-rich (SR) protein kinases. *Biochem J.* 1997;326(3):693–700.

26. Aubol BE, Wu G, Keshwani MM, Movassat M, Fattet L, Hertel KJ, Maliheh M, Fattet L, Fu X, Adams JA. Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol Cell*. 2016;63(2):218–28.
27. Duncan PI, Howell BW, Marius RM, Drmanic S, Douville EMJ, Bell JC. Alternative splicing of STY, a nuclear dual specificity kinase. *J Biol Chem*. 1995;270(37):21524–31.
28. George A, Aubol BE, Fattet L, Adams JA. Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate. *J Biol Chem*. 2019;294(24):9631–41.
29. Lin J rong, Hu J. SeqNLS: nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. *PLoS One*. 2013;8(10).
30. Aubol BE, Chakrabarti S, Ngo J, Shaffer J, Nolen B, Fu XD, Ghosh G, Adams JA. Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc Natl Acad Sci U S A*. 2003;100(22):12601–6.
31. Lai MC, Lin RI, Huang SY, Tsai CW, Tarn WY. A human importin- β family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem*. 2000;275(11):7950–7.
32. Aubol BE, Plocinik RM, Keshwani MM, Mcglone ML, Hagopian JC, Ghosh G, Fu XD, Adams, JA. N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. *Biochem J*. 2014;462(1):143–52.
33. Keshwani MM, Hailey KL, Aubol BE, Fattet L, McGlone ML, Jennings PA, Adams JA. Nuclear protein kinase CLK1 uses a non-Traditional docking mechanism to select physiological substrates. *Biochem J*. 2015;472(3):329–38.
34. Ngo JCK, Giang K, Chakrabarti S, Ma CT, Huynh N, Hagopian JC, Dorrestein PC, Fu XD, Adams JA, Ghosh, G. A Sliding Docking Interaction Is Essential for Sequential and Processive Phosphorylation of an SR Protein by SRPK1. *Mol Cell*. 2008;29(5):563–76.
35. Twyffels L, Gueydan C, Kruys V. Shuttling SR proteins: More than splicing factors. *FEBS J*. 2011;278(18):3246–55.
36. Saitoh N, Sakamoto C, Hagiwara M, Agredano-Moreno LT, Jiménez-García LF, Nakao M. The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Mol Biol Cell*. 2012;23(6):1115–28.
37. Go D. Mediate Nuclear Import of Ribosomal Proteins in Mammalian Cells. 1998;17(15):4491–502.
38. Bange G, Murat G, Sinning I, Hurt E, Kressler D. New twist to nuclear import: When two

- travel together. *Commun Integr Biol.* 2013;6(4).
39. Mosammaparast N, Jackson KR, Guo Y, Brame CJ, Shabanowitz J, Hunt DF, Pemberton LF. Nuclear import of histone H2A and H2B is mediated by a network of karyopherins. *J Cell Biol.* 2001;153(2):251–62.
 40. Tessier TM, Macneil KM, Mymryk JS. Piggybacking on classical import and other non-classical mechanisms of nuclear import appear highly prevalent within the human proteome. *Biology (Basel).* 2020;9(8):1–20.

Chapter 4

Oligomerization of CLK1

Abstract

CLK1 is known to self-associate through its N-terminus forming large oligomers but investigating the structural details of oligomer formation has been challenging owing to the intrinsic disorder of the N-terminus and its aggregation-prone nature. Studies presented in this chapter shed new light on the key interactions that drive oligomerization in this protein kinase. Our results show that the CLK1 oligomer forms as a result of two interaction modes: self-association of the N-termini (N-N) and N-terminus-kinase domain interactions (N-K) among neighboring monomers. The N-N interactions are driven exclusively by aromatic residues in the N-terminus (tyrosines) whereas the N-K interactions are largely electrostatic in nature. Either mode appears to be sufficient to drive lower levels of self-association but both modes combined fully stabilize robust oligomer formation in CLK1.

4.1 Introduction

The self-association of proteins (homo-oligomerization) into different forms of assemblies has been known to regulate several biochemical functions including transcription, gene expression, ion transport across membranes, enzyme activation, and cell adhesion processes (1,2). Homo-oligomerization has been proposed to be one of nature's mechanisms to generate large complexes without having to expand genomic diversity (3). Transitions between different oligomeric states are often controlled for the optimal functioning of cells. Indeed, even slight misregulation of assembly mechanisms leading to higher or lower order oligomeric states can lead to diseases (4). The formation of insoluble amyloid fibrils due to the aggregation of the tau protein leading to the neurodegenerative Alzheimer's disease is one of the most well-cited examples of toxic, misregulated protein association (5).

Research over the last two decades has highlighted many distinct forms of protein homo-oligomerization. The assembly of hemoglobin tetramers from its α and β subunits was one of the earliest identified examples of quaternary assemblies enhancing protein function (4). Some protein kinases use self-association as an activation mechanism. For example, CaMKII (Ca²⁺/ Calmodulin (CaM)-dependent protein kinase II) is a serine/threonine kinase that is known to form oligomers with approximately 12 subunits (6–8). This oligomerization brings together monomeric units in a way that facilitates inter-subunit autophosphorylation in a Ca²⁺-dependent manner. This autophosphorylation activates the kinase and helps it maintain its activity even in a Ca²⁺-deficient environment (8). The receptor interacting protein kinase 2 (RIP2) plays critical roles in detecting bacterial infections and also uses a similar oligomerization mechanism for catalytic activation (9,10). The caspase activation and recruitment domain (CARD) in RIP2 oligomerizes forming long, filamentous structures upon the binding of other regulatory proteins like nucleotide-binding

oligomerization domain-containing proteins 1 and 2 (NOD1 and NOD2). This oligomerization promotes autophosphorylation on the activation loop activating RIP2 for pro-inflammatory signaling pathways (9).

While most initial studies on protein assemblies were centered around interactions between well-folded structured domains, recent studies have shifted their focus towards oligomerization mediated by intrinsically disordered regions (IDRs) (11–13). IDRs have high propensity to drive protein-protein interactions owing to their multivalency, flexibility and sequence-insensitivity (14). In recent years, protein oligomerization mostly mediated by IDRs has been shown to generate liquid-liquid phase separation (LLPS), a biophysical process that drives the formation of membrane-less organelles such as P-granules, stress granules, and nucleoli (15–17). These organelles provide a dynamic platform where the components can diffuse in and out of these biomolecular condensates in response to cellular cues (18).

Understanding the interactions that drive the formation of quaternary structures associated with protein oligomers is of critical importance for understanding both protein assembly and function. A quick glance at the protein assembly mechanisms reported in the literature suggests that there can be great diversity in the nature of the interactions leading to protein association (2). Such interactions can be covalent (e.g., disulfide bonds), or non-covalent in nature. In the latter category, a diverse set of contacts may be found. For example, hydrophobic, aromatic, electrostatic, polar and hydrogen bonding interactions are often found to stabilize oligomeric structures (Fig.4.1) (19). Many of these interactions may be critical for normal cellular function and can also play a role in human diseases. For example, hydrophobic interactions between the R3 domains in the tau protein drive pathological fibril formation (5). Hydrophobic interactions drive the self-assembly of several transcription factors leading to the formation of structural motifs

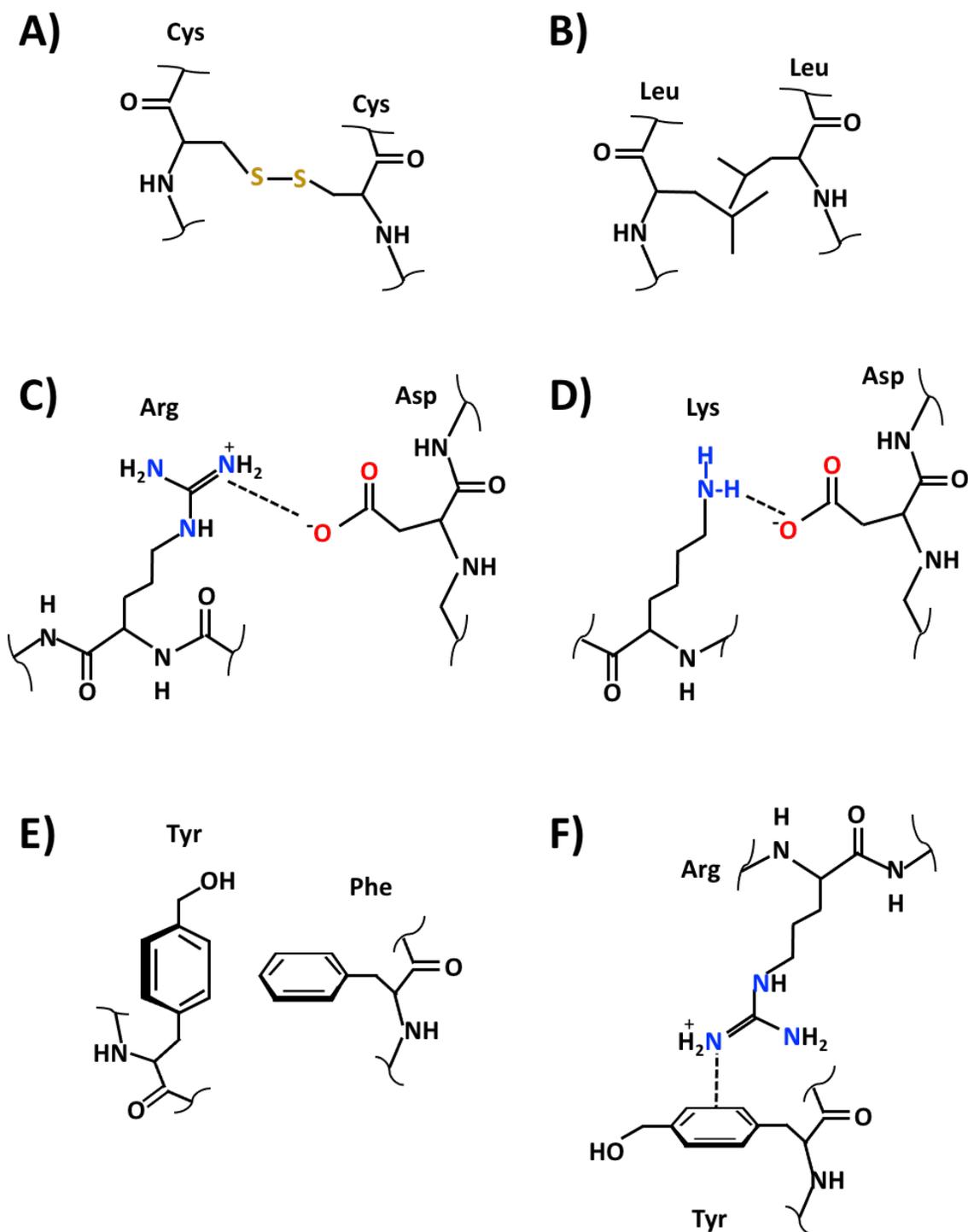


Figure 4.1. Interactions driving protein association. (A) Covalent binding through disulfide bonds (B) Hydrophobic packing (C) Electrostatic interactions between oppositely charged side chains (D) Hydrogen bonding (E) π - π stacking between two aromatic residues (F) Cation- π interaction between a positively charged side chain and an aromatic residue.

known as ‘leucine zippers’ at the interface (20). Analogous to the concept of leucine zippers, Max Perutz put forth the idea of ‘polar zippers’ where electrostatic interactions between appropriately spaced charged residues lead to protein self-association, as is seen in the assembly of eight subunits in *Ascaris* hemoglobin (21–23). Electrostatic interactions between the low complexity domains in U1-70K and SR proteins drive their phase separation leading to the formation of nuclear speckles (24,25). Aromatic interactions can be of two types: (a) π -stacking interactions where two aromatic residues interact in a T-shape or slip stacked orientation or (b) cation- π interactions where a positive charged moiety (like in arginine residue) interacts with a π -electron cloud from a tryptophan/tyrosine/phenylalanine residue on a different subunit (Fig. 4.1E,F). Both kinds of aromatic interactions are ubiquitous in protein-protein interfaces (26,27). For example, glutathione S-transferase (GST) has been shown to dimerize through π -stacking interactions (28). Cation- π interactions are predicted to be the predominant interaction driving molecular recognition in the Ewing’s sarcoma oncoprotein family of proteins (29). Aromatic interactions also drive liquid-liquid phase separation in the splicing repressing ribonucleoprotein hnRNPA1 and the RNA binding protein fused in sarcoma (FUS) (15,30).

The RS domains in SR proteins are known to be disordered and highly prone to aggregation (31). The N-terminus of CLK1 has been referred to as an RS-like domain due to the high content of RS dipeptides and other disorder-promoting residues (32) (Fig.1.7). Previous studies from our lab showed that CLK1 forms oligomers mediated by its disordered N-terminus (33). This oligomerization also acts as a substrate selection mechanism facilitating the recognition of the physiological substrate SRSF1 over non-physiological substrates (33). However, the structural aspects of CLK1 oligomerization has long been unexplored owing to the challenges arising from the limited solubility and aggregation-prone nature of the N-terminus. The presence of this

disordered domain and the dynamic nature of the oligomer further complicates structural investigations by diffraction techniques such as crystallography or electron microscopy. Here, we investigate the structural details of how the disordered N-terminus leads to CLK1 oligomerization. We showed that oligomerization is not just mediated by self-association of the N-terminus as previously thought (33), but also by interactions between the N-terminus and kinase domain. We also showed that while the N-terminus is highly charged and RS-like, its ability to oligomerize is driven by aromatic interactions.

4.2 Results

4.2.1. Oligomerization is retained in single-block CLK1 N-terminal deletions.

To understand how CLK1 forms oligomers, we first tried to see if there was a specific region in the N-terminus that was vital for oligomerization. For ease of study, we divided the roughly 150-residue N-terminus into three blocks of approximately fifty residues each. We designed deletion mutants of CLK1, called CLK1(Δ 1), CLK1(Δ 2), and CLK1(Δ 3) where one block of approximately fifty residues was deleted at a time (Fig. 4.2A). Since CLK1 cannot be expressed at high levels in bacteria, we made baculoviruses with the genes for the deletion constructs and purified these constructs from baculovirus-infected Hi5 cells (Fig. 4.2B). Next, we looked at the oligomerization status of these single block deletions using DLS and SEC and found that all constructs were oligomeric (Fig. 4.2 C-E). According to previous reports, the hydrodynamic radius of the CLK1 kinase domain by itself, is about 4 nm (33). The hydrodynamic radius of CLK1 was seen to be around 70 nm, although the radius can show some variability owing to the highly dynamic nature of oligomerization. All three single-block deletion constructs of CLK1 showed a radius in the range of 15-30 nm, indicating that they are oligomeric. Consistent with the oligomeric sizes found by DLS, all constructs eluted early on the S200 size exclusion

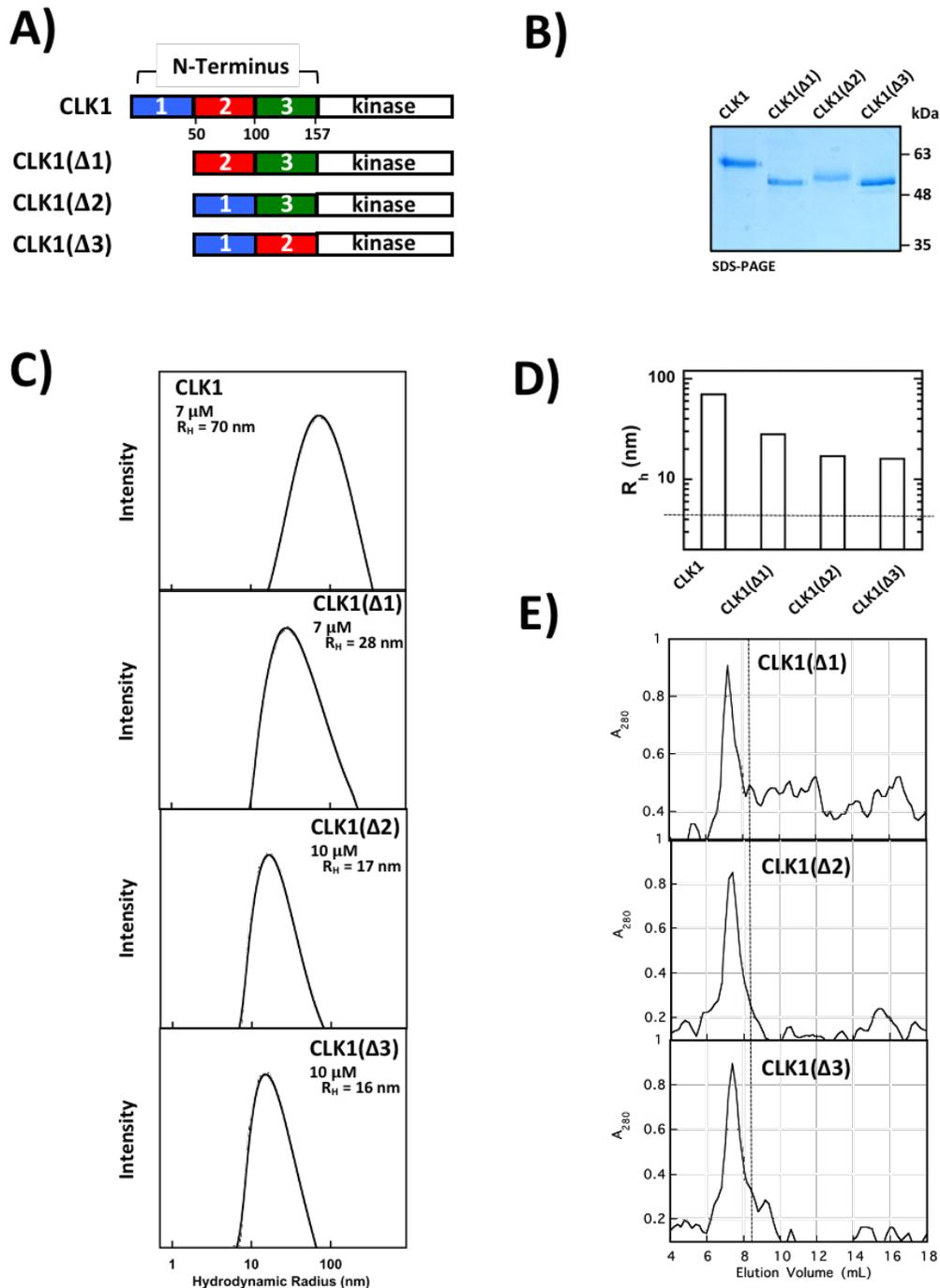


Figure 4.2. Single block deletion constructs of CLK1 can form oligomers. **(A)** CLK1 single block deletion constructs. **(B)** SDS-PAGE showing all constructs **(C)** DLS spectra of all constructs **(D)** Hydrodynamic radii of all single block deletions. The dashed line corresponds to 4 nm, the hydrodynamic radius of CLK1(Δ N). **(E)** SEC chromatograms of all deletion constructs when run on a S200 column. Oligomeric fractions elute before the dashed line on the chromatograms.

column suggesting that the oligomers are, at least, 600kDa in size as the molecular weight of the oligomers are higher than the resolution range of an S200 column. Thus, CLK1 and all single block deletions form oligomers composed of, at least, 10 subunits. These results suggest that there is no specific sequence within the 50-residue blocks of the N-terminus that drives oligomer formation.

4.2.2. N-terminal residues directly flanking the kinase domain do not induce CLK1 oligomerization.

Having shown that the N-terminus remains oligomeric even after 50-residue deletions were made throughout the N-terminus (Fig. 4.2), we next decided to look at larger deletions in CLK1 by designing the constructs CLK1 (Δ 12), CLK1 (Δ 13), and CLK1(Δ 23) (Fig. 4.3A). We were able to express and purify these proteins to homogeneity using baculovirus methods for detailed biophysical studies (Fig. 4.3B). Interestingly, we found that CLK1(Δ 13) and CLK1(Δ 23) which both remove 2/3 of the N-terminus are still oligomeric by DLS and SEC methods (Fig. 4.3C-E). While these results suggest important roles for the remaining blocks in these constructs, it is important to note that these deletions also change the register of these blocks with respect to the kinase domain. Surprisingly, we found that only CLK1(Δ 12) which removes blocks 1 and 2 and keeps block 3 in proper register with the kinase domain is monomeric. These findings suggest that whereas block 1 or 2 induces oligomerization when directly flanking the kinase domain, block 3 connected to the kinase domain cannot induce such oligomerization. These findings suggest that Block 3 is not sufficient to induce oligomerization in CLK1.

4.2.3 Self-association of the N-terminus is length dependent but sequence insensitive.

Since the N-terminus is essential for CLK1 oligomerization, we wished to next determine whether specific regions in this disordered segment are necessary for self-association in the

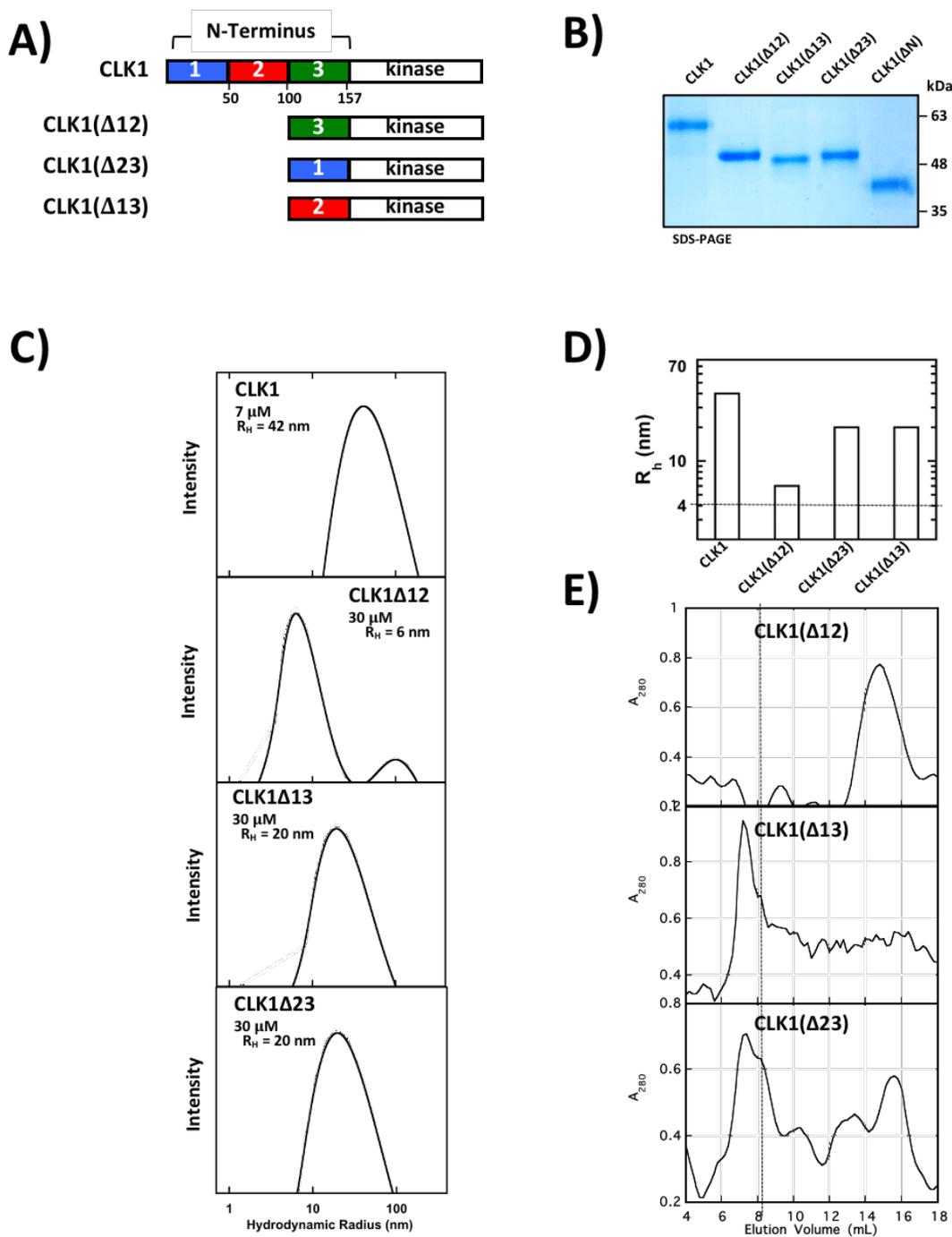


Figure 4.3. Oligomerization of CLK1 two-block deletion constructs **(A)** CLK1 two-block deletion constructs. **(B)** SDS-PAGE **(C)** DLS spectra **(D)** Hydrodynamic radii and **(E)** S200 chromatograms of all two block deletion constructs. The oligomeric fractions elute before the dashed line. CLK1 (Δ 12) is monomeric but CLK1 (Δ 13) and CLK1 (Δ 23) can form oligomers.

absence of the kinase domain. Since the N-terminus has poor solubility when expressed by itself in *E. coli*, we designed a GST-tagged N-terminus along with several additional forms that mimic the block deletions in the full-length CLK1 (Fig. 4.4A). We were able to purify these constructs from *E. coli* using glutathione-agarose affinity chromatography overcoming solubility problems associated with prior His-tagged forms (Fig. 4.4B). We first evaluated the impact of sequential deletion of N-terminal residues on oligomerization using SEC and DLS methods. GST-N, which contains all residues in the N-terminus, was an oligomer that eluted early on the S200 size exclusion column and displayed a hydrodynamic radius of about 70 nm by DLS, consistent with a large oligomeric form of the protein (Fig. 4.4D,E). Deletion of block 3 in GST-N ($\Delta 3$) resulted in an oligomeric protein that had a smaller hydrodynamic radius of 20 nm but still eluted as an oligomer on the S200 column suggesting that the molecular mass of this deletion construct is still greater than 600 kDa. Further deletion of blocks 2 and 3 in GST-N($\Delta 23$) resulted in complete loss of oligomerization as evidenced from DLS and SEC experiments. We found that GST-N($\Delta 23$) displayed a radius and molecular mass similar to that for its carrier protein GST (Fig.4.4D,E). We next tested the oligomerization status of GST-N($\Delta 13$) and GST-N($\Delta 12$) and found that neither construct could form oligomers, similar to GST-N($\Delta 23$) (Fig.4.4 C-E). We also noticed that SEC for oligomeric constructs like GST-N and GST-N($\Delta 3$) showed peaks of lower molecular weights which were not prominent in the DLS spectra. The reason for the dissociation of GST-N constructs into smaller units can be attributed to the dilution of the oligomeric constructs while passing through the column. CLK1 oligomerization is dynamic and has been previously shown to be concentration dependent (33). DLS samples, on the other hand, do not get diluted and hence do not show a significant monomeric peak. In addition, DLS is much more sensitive towards larger species in the sample solution than it is to species of smaller radii. DLS can also detect transient

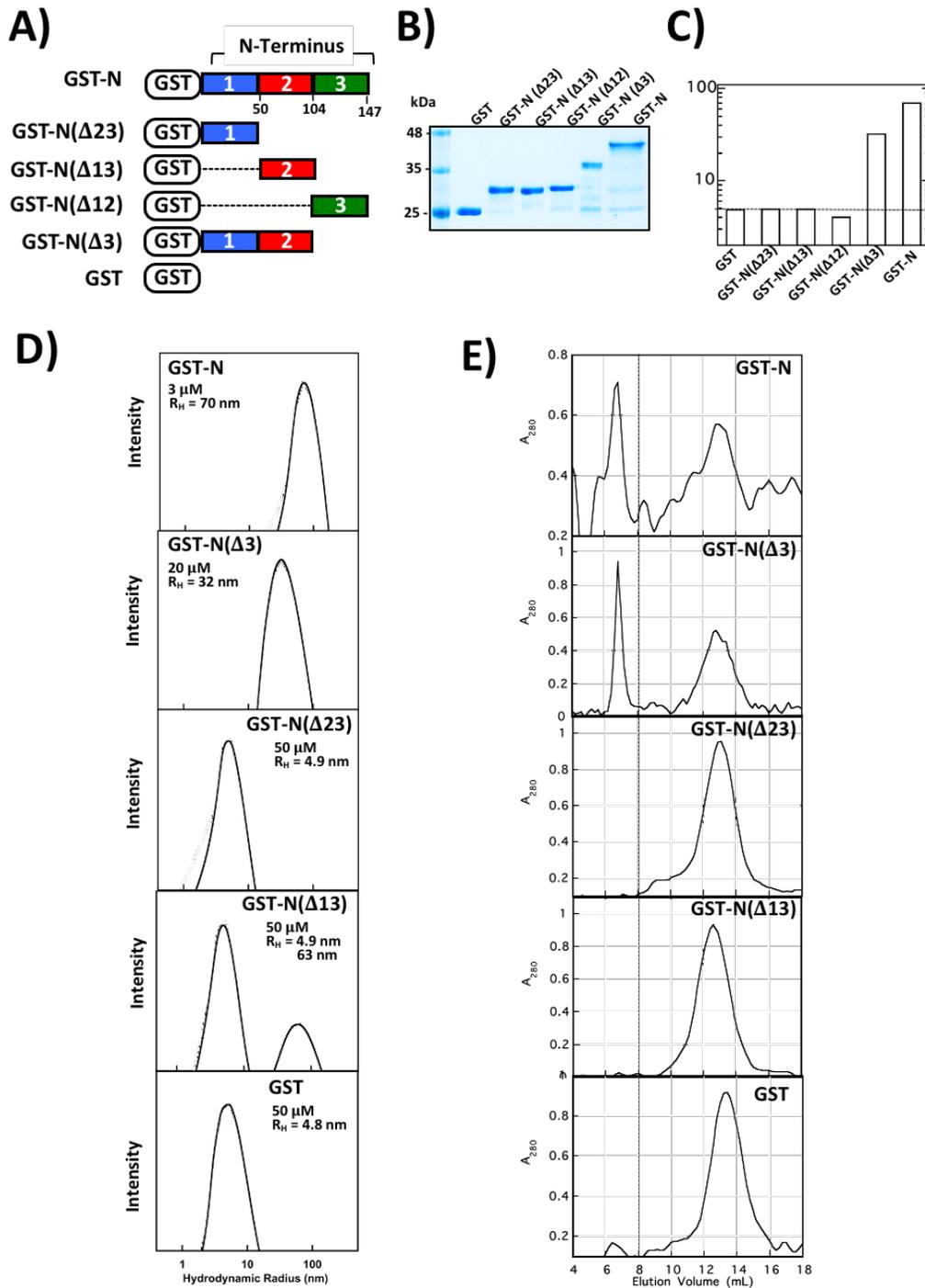


Figure 4.4. Oligomerization of GST-N terminus constructs **(A)** GST-N terminus constructs **(B)** SDS-PAGE showing all purified GST-N constructs. **(C)** Hydrodynamic radii of all GST-N constructs **(D)** DLS and **(E)** SEC of GST-N, GST-N (Δ 3), and GST-N (Δ 23), GST-N(Δ 13). GST is included as a control. Oligomeric constructs elute before the dashed line.

aggregation states present in the sample even when present in minute amounts. Hence, constructs like GST-N(Δ 13) which do not show any oligomerization by SEC, shows peaks of higher molecular masses by DLS (Fig.4.4). Analysis of the mass percentage of the oligomeric peak showed that the oligomeric fraction is three orders of magnitude smaller than the monomeric peak, suggesting that the fraction of oligomers in these constructs is negligible.

Taken together, these results suggest that sequential deletion of the N-terminus weakens oligomerization, implying that the number of residues on the N-terminus can be decisive in determining the extent of self-association. However, oligomerization is not driven by a specific sequence within any block, as all single block constructs do not form oligomers.

4.2.4 Aromatic residues drive N-N interactions.

Having shown the sequence-insensitive nature of N-N interactions, we next wanted to further explore the residues involved in driving self-association. On examining the amino acid sequence of the N-terminus, we noticed that while most of the residues on the N-terminus are disorder-promoting, hydrophilic residues, the sequence is also peppered with structure-promoting aromatic residues, particularly in block 2 (Fig.4.5A). Of the residues that constitute the N-terminus, there are three tryptophans in Block 1 and ten tyrosines distributed throughout the N-terminus. We, therefore, considered the possibility that CLK1 oligomerization may be mediated by aromatic interactions, particularly since there is precedence for such residues in disordered domains that induce protein aggregates. Several RNA binding proteins involved in neurodegenerative disorders have been shown to aggregate via aromatic interactions in their low-complexity domains (30,34). To test if aromatic residues could play a role in CLK1 oligomerization, we generated a GST-tagged N-terminus construct, GST-N (Y-L) where all the

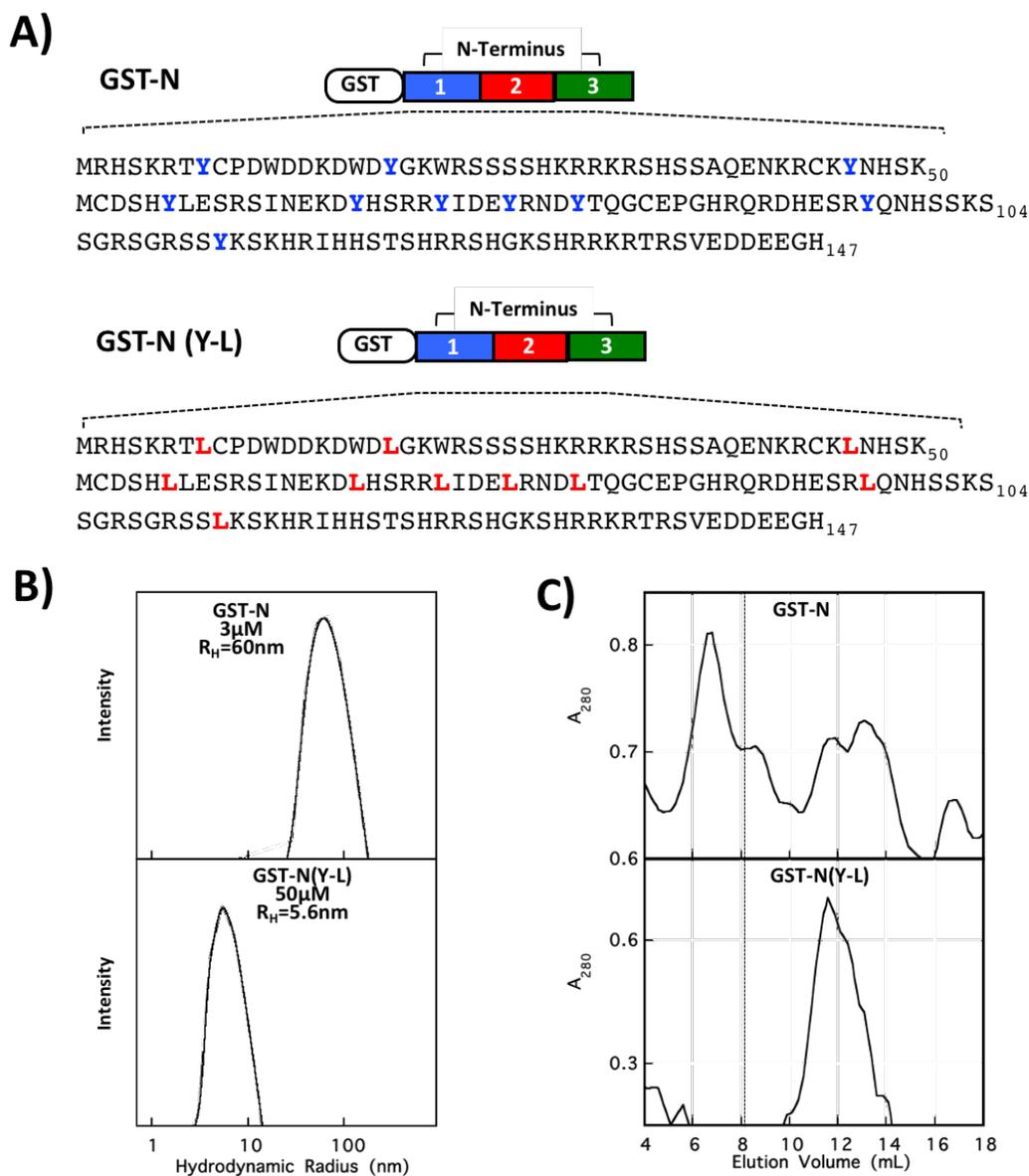


Figure 4.5. Aromatic interactions mediate N-N interaction **(A)** GST-N Terminus and GST-N(Y-L) mutant showing the Tyrosines mutated to Leucines. **(B)** DLS and **(C)** SEC showing of GST-N and GST-N(Y-L) showing that the mutant does not form oligomers. Oligomeric fractions elute before the dashed line.

tyrosines in the N-terminus are mutated to leucines (Fig.4.5A). We first tested if this construct could self-associate and form oligomers and found that this new construct does not self-associate as no oligomers were detected using either DLS or SEC (Fig.4.5B,C). The hydrodynamic radius of the GST-N (Y-L) mutant was only 5.6 nm while the GST-N construct formed oligomers with a hydrodynamic radius of 60 nm. It is interesting to note that GST-N(Y-L), with all three blocks of the N-terminus, showed a hydrodynamic radius of 5.6 nm and is only slightly bigger than the carrier GST that has a radius of 4.8 nm. These results suggest that the N-N interactions that drive oligomerization in the N-terminus may be solely mediated by tyrosines.

4.2.5. N-terminus interacts with the CLK1 kinase domain.

Although we have shown that CLK1 oligomerization is dependent on a disordered N-terminus that can also self-associate (33), it is unclear whether the kinase domain plays a role in forming higher-order CLK1 structures. To test for any interactions between the N-terminus and kinase domain (N-K) in CLK1, pull-down assays were performed using GST-tagged N-terminal constructs and His-tagged CLK1 kinase domain (CLK1(Δ N)). We initially found that GST-N pulls down His-CLK1 suggesting that the N-terminus can interact with the full-length CLK1 (Fig. 4.6A). However, since this interaction could be due to N-N contacts, we next tested if the N-terminus can interact directly with the kinase domain in a pull-down assay using GST-N and His-CLK1(Δ N). We found that GST-N, indeed, pulled down CLK1 kinase domain (Fig.4.6B). We next wished to learn which parts of the N-terminus interact with the kinase domain. Using the sequential deletion constructs GST-N, GST-N(Δ 3) and GST-N(Δ 23) in pull down experiments with His-CLK1 (Δ N), we found that all the three N-terminal constructs interact with the kinase domain (Fig. 4.6C). Although the fraction of bound His-CLK1(Δ N) increased on going from GST-N(Δ 23) to GST-N, we found that GST-N(Δ 3) did not show an enhanced binding compared to GST-N(Δ 23)

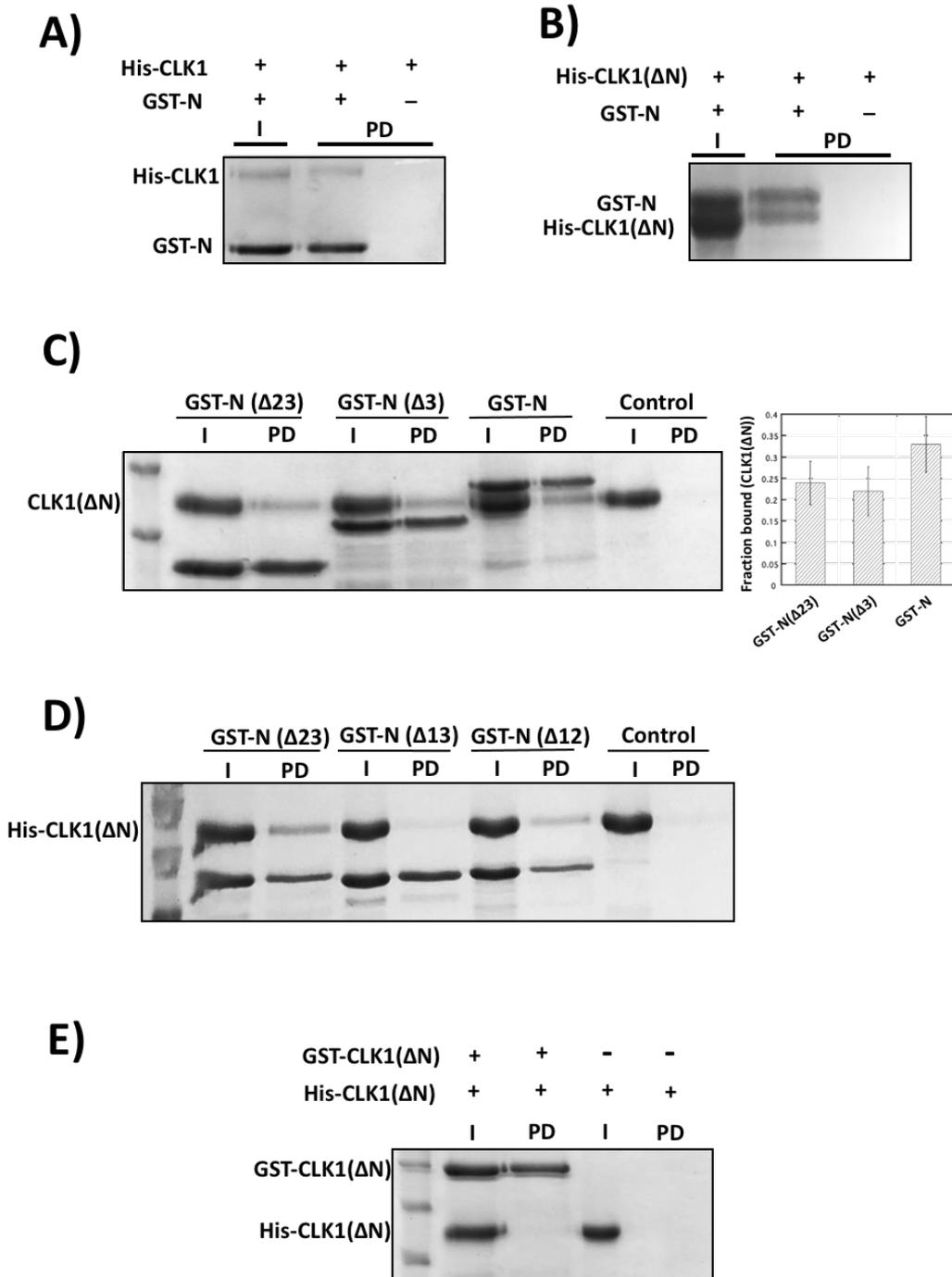


Figure 4.6. N-terminus interacts with kinase domain (A) PD with GST-N and CLK1. (B) PD with GST-N and CLK1(Δ N). (C) PD with GST-N(Δ 23), GST-N(Δ 3), and GST-N with CLK1(Δ N). Fraction of CLK1(Δ N) bound to each GST-N construct. (D) PD with GST-N (Δ 23), GST-N (Δ 13), GST-N (Δ 12) with CLK1 (Δ N). (E) CLK1(Δ N) does not self-associate: PD showing GST-CLK1(Δ N) and His-CLK1(Δ N).

(Fig.4.6C). This suggests that the residues in block 2 likely form minimal-to-no interactions with the kinase domain. To test the binding interactions of individual blocks in the N-terminus, we found that while GST-N(Δ 23) and GST-N(Δ 12) pulled down His-CLK1(Δ N), GST-N(Δ 13) did not (Fig. 4.6D). These findings indicate that block 1 and block 3 interact directly with the kinase domain while block 2 does not, a result in agreement with those from the pull-down experiments in Fig 4.6C. Although we showed previously that His-CLK1(Δ N) does not self-associate (33), we wanted to test if the kinase domain could self-associate in the absence of the N-terminus in our pull-down assays. In a pull-down assay using GST-CLK1(Δ N) and His-CLK1(Δ N), we observed that the two kinase domains did not interact with each other (Fig. 4.6E), in support of our prior findings that no oligomers are detected for His-CLK1(Δ N) via SEC (33). However, we still cannot rule out the possibility of kinase-kinase interactions that can be induced if the N-termini bring kinase domains close together in a large oligomer (K-K interactions induced by N-N interactions). In summary, our pull-down experiments suggest that in addition to direct N-N interactions, N-K interactions, particularly those from blocks 1 and 3, may be possible in the CLK1 oligomer.

4.2.6. N-K interactions drive oligomerization in CLK1 (Y-L)

Since we showed that tyrosines facilitate N-N interactions in the GST-N fusion protein (Fig. 4.5), we next investigated if N-K interactions are also driven by these residues. To test this idea, we performed pull-down assays using GST-N (Y-L) and His-CLK1(Δ N) and found that GST-N (Y-L), while incapable of forming large oligomers, directly pulled down His-CLK1(Δ N) (Fig.4.7A). These findings suggest that while N-N interactions are solely mediated by tyrosines, these aromatic residues played no role in N-K interaction. We next wished to investigate the role of N-N aromatic contacts for oligomerization of full-length CLK1. We expressed and purified a new construct, CLK1(Y-L) from Hi5 cells using our baculovirus expression system where all 10

tyrosine in the N-terminus of CLK1 were mutated to leucines. Based on our previous findings (Figs. 4.5 & 4.6), this construct is expected to promote N-K but not N-N interactions. We first studied its oligomerization status by DLS and found that the mutant forms oligomers although the overall hydrodynamic radius is smaller than that for the wild-type CLK1 (Fig. 4.8B). Based on elution volume on the S200 column, we conclude that the molecular mass of CLK1(Y-L) is still greater than 600 kDa as the mass is higher than the resolution range of an S200 column (Fig. 4.7C). We also looked at the oligomers of this construct using negative stained transmission electron microscopy (TEM). We found that while wild-type CLK1 formed more oligomers with similar sizes, CLK1(Y-L) formed oligomers of varying sizes with high polydispersity (Fig.4.7D). Although these imaging experiments were performed at similar concentration for both CLK1 and CLK1(Y-L) (4.5 μ M), more oligomers were observed on the grid with CLK1 than what was observed for CLK1 (Y-L). This could be indicative of a less stable oligomer when only N-K interactions are present. Since the oligomerization was dynamic, this may be suggestive of a more rapid dissociation of oligomers into monomers if N-N interactions are not present to stabilize the oligomer. These results suggest that N-K interactions alone can drive oligomerization but more efficient oligomer formation is induced when both N-N and N-K interactions are available in CLK1.

4.2.7. Electrostatic interactions drive N-K interactions.

Having shown that N-K interactions alone can drive oligomerization of CLK1, we next tried to characterize the nature of the interaction between the N-terminus and the kinase domain. Since we know that the tyrosines do not play a role in N-K interactions (Fig. 4.7A), we next considered the possibility of N-K interactions being driven by electrostatic interactions. Since only blocks 1 and 3 bind the kinase domain, the interacting residues must be localized to block 1

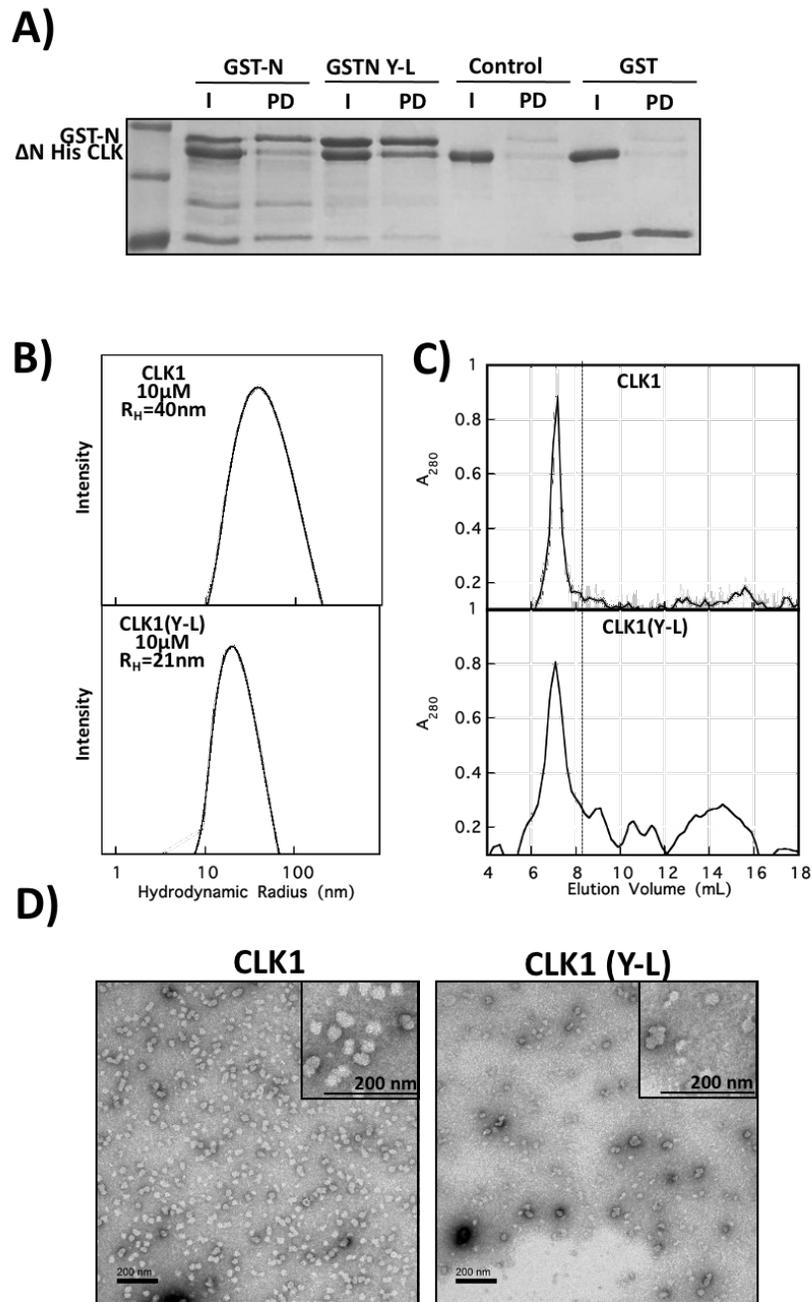


Figure 4.7: Oligomerization of CLK1(Y-L). **(A)** PD of GST-N and GST-N (Y-L) showing that GST-N (Y-L) mutant can interact with CLK1 kinase domain even though the mutant does not have tyrosine residues. **(B)** DLS and **(C)** SEC showing that CLK1(Y-L) forms oligomers. Oligomeric fractions elute before the dashed line. **(D)** Negative stained TEM micrographs of CLK1(Y-L) and CLK1 with a small region magnified shown in the inset. Scale bar=200nm.

A)

GST-N(Δ 23) MRHSKRTYCPDWDDKDWDYGKWRSSSSHKRRKRSHSSAQENKRCKYNHSK₅₀

GST-N(Δ 23, Δ K/R) MRHSKRTYCPDWDDKDWDYGKWRSSSSHGGGGSHSSAQENGGCGYNHSK₅₀

GST-N(Δ 23, Δ D) MRHSKRTYCPGWGGKGWGYGKWRSSSSHKRRKRSHSSAQENKRCKYNHSK₅₀

B)

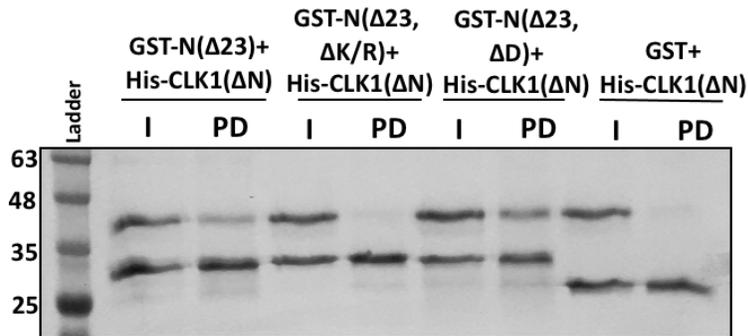


Figure 4.8: N-K interactions are driven by electrostatic interactions. (A) GST-N(Δ 23), GST-N(Δ 23, Δ K/R) and GST-N(Δ 23, Δ D) mutants. (B) PDs of GST-N(Δ 23) mutants with His-CLK1(Δ N).

or 3. However, the interaction between block 3 and the kinase domain is most likely an intramolecular interaction rather than an inter-molecular interaction because CLK1(Δ 12) is a monomer despite having block 3 next to the kinase domain. This brings the focus to block1, which has an acidic patch of residues (D11-D18) and a basic patch (K30-R35). We mutated the aspartate residues to glycines in GST-N(Δ 23, Δ D) and the lysine/arginine residues to glycines in GST-N(Δ 23, Δ K/R) (Fig. 4.8A). To test if the mutants could bind the kinase domain, we performed pull-down assays with His-CLK1(Δ N). Our results showed that while GST-N(Δ 23, Δ D) pulled down His-CLK1(Δ N), GST-N(Δ 23, Δ K/R) mutant did not (Fig. 4.8B). This indicated that the N-K interactions are primarily electrostatic in nature, with the interaction being localized to a small stretch of positively charged residues.

4.2.8. The kinase domain is protected from solvent exchange by the N-terminus

We showed that the N-terminus induces CLK1 oligomer formation not only through N-N interactions, as expected from prior studies (33), but also through N-K interactions. To identify regions in CLK1 that may participate in these inter-protein contacts we used H/D Exchange methods coupled with mass spectrometric detection. In H/D exchange studies, the protein of interest, initially stored in H₂O buffer, is transferred into a D₂O buffer allowing the backbone amide hydrogens to exchange with deuterium atoms as a function of time. Amide hydrogens in solvent accessible or dynamic regions exchange more rapidly with solvent deuterium than the buried or highly structured regions (35) (Fig. 4.9A). To assess deuterium exchange levels, the reaction is stopped at various time points by rapidly lowering the temperature and pH with a low pH quench buffer (Fig. 4.9B). Quenching also unfolds the protein, making it accessible for proteolytic cleavage by pepsin, a protease that optimally functions at the quench pH of 2.5. Since all the downstream processes post-deuterium labeling are done in H₂O buffer, there is a

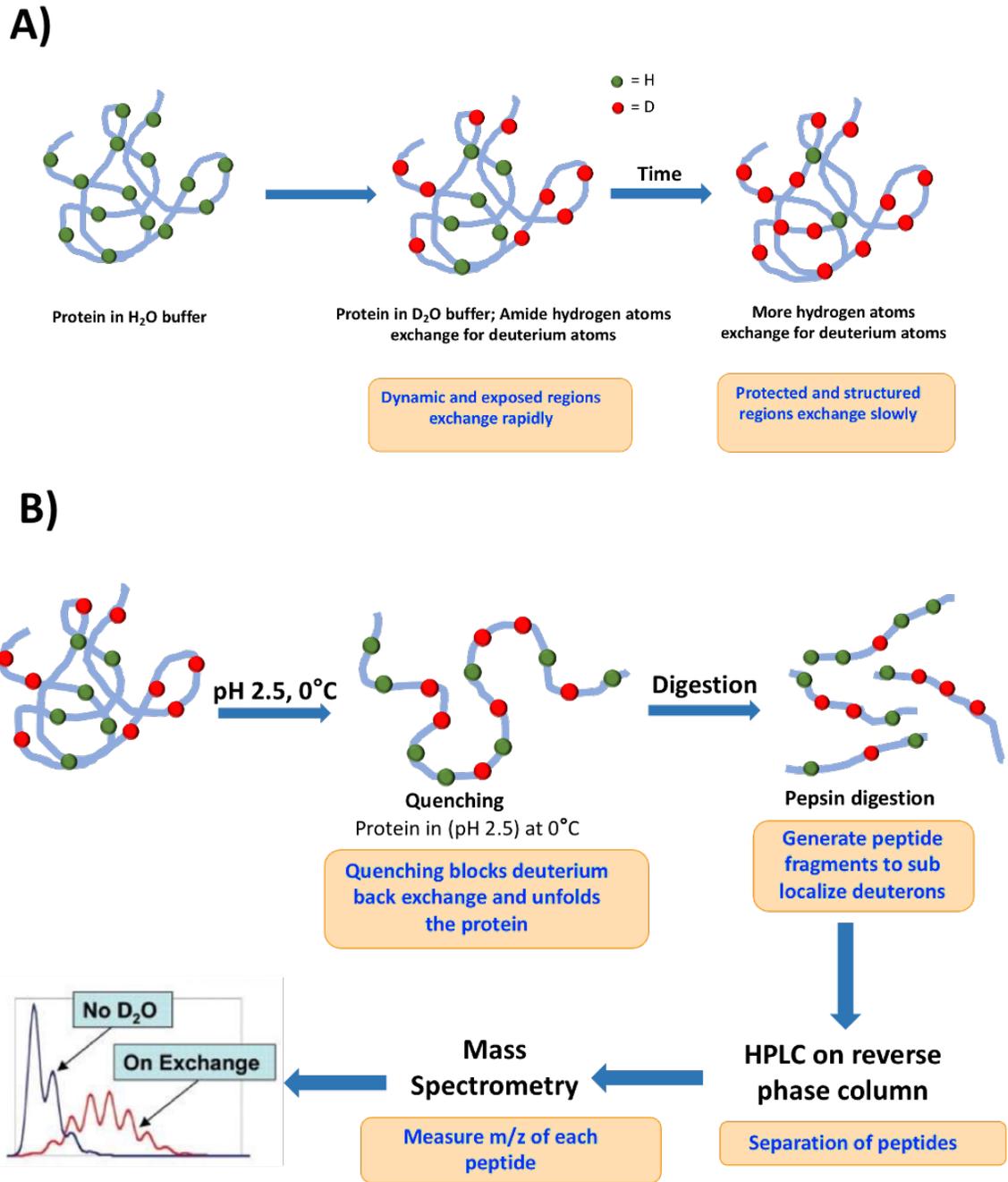


Figure 4.9. (A) H/D Exchange experiment (B) Detecting exchanged Deuterium using LC/MS

possibility for the labeled amide deuteriums to back exchange into hydrogen atoms. To circumvent this problem, throughout the post-exchange phase, the pH is kept low to minimize back exchange. Pepsin digestion generates peptide fragments (probes) that are separated based on their polarity on a reversed-phase HPLC column and identified using an electron-spray ionization (ESI) mass spectrometer (35,36). The increase in the mass of the peptide envelopes can be tracked as a function of time (Fig. 4.9B). The amount of deuterium incorporation is quantitated relative to a fully deuterated (FD) control, which was taken after a 24-hour incubation period of the pepsin-digested peptides in D₂O buffer.

To gain insights into CLK1 oligomeric structure, we performed H/D exchange studies on CLK1 and CLK1(Δ N) which is a monomer and not expected to engage in significant protein-protein interactions. We first compared the fragmentation maps of CLK1 and CLK1(Δ N) generated after pepsin digestion. We were not able to detect any peptides on the N-terminus with high signal intensity. On the kinase domain, around 60 peptides were detected in CLK1, but CLK1(Δ N) generated around 130 peptides. While CLK1(Δ N) generated more peptides than CLK1, we noticed that the peptides on CLK1 were cleaved at the same sites as in CLK1(Δ N), giving us common probes to compare the exchange rates in both the constructs.

We next compared the deuterium incorporation profiles of peptides common to both proteins by tracking the shift in peptide envelopes as a function of time. The non-deuterated peptides showed identical peptide envelopes with the same centroid mass for both CLK1 and CLK1(Δ N) (Fig.4.10). To correct for the back-exchange of amide deuterons to protons during the separation over the C-18 column, we generated experimental FD controls by incubating pre-digested peptides in D₂O buffer for 24 hours. The FD or the ‘back-exchange control’ provides a measure for the loss of deuterons under our experimental conditions for each peptide as back

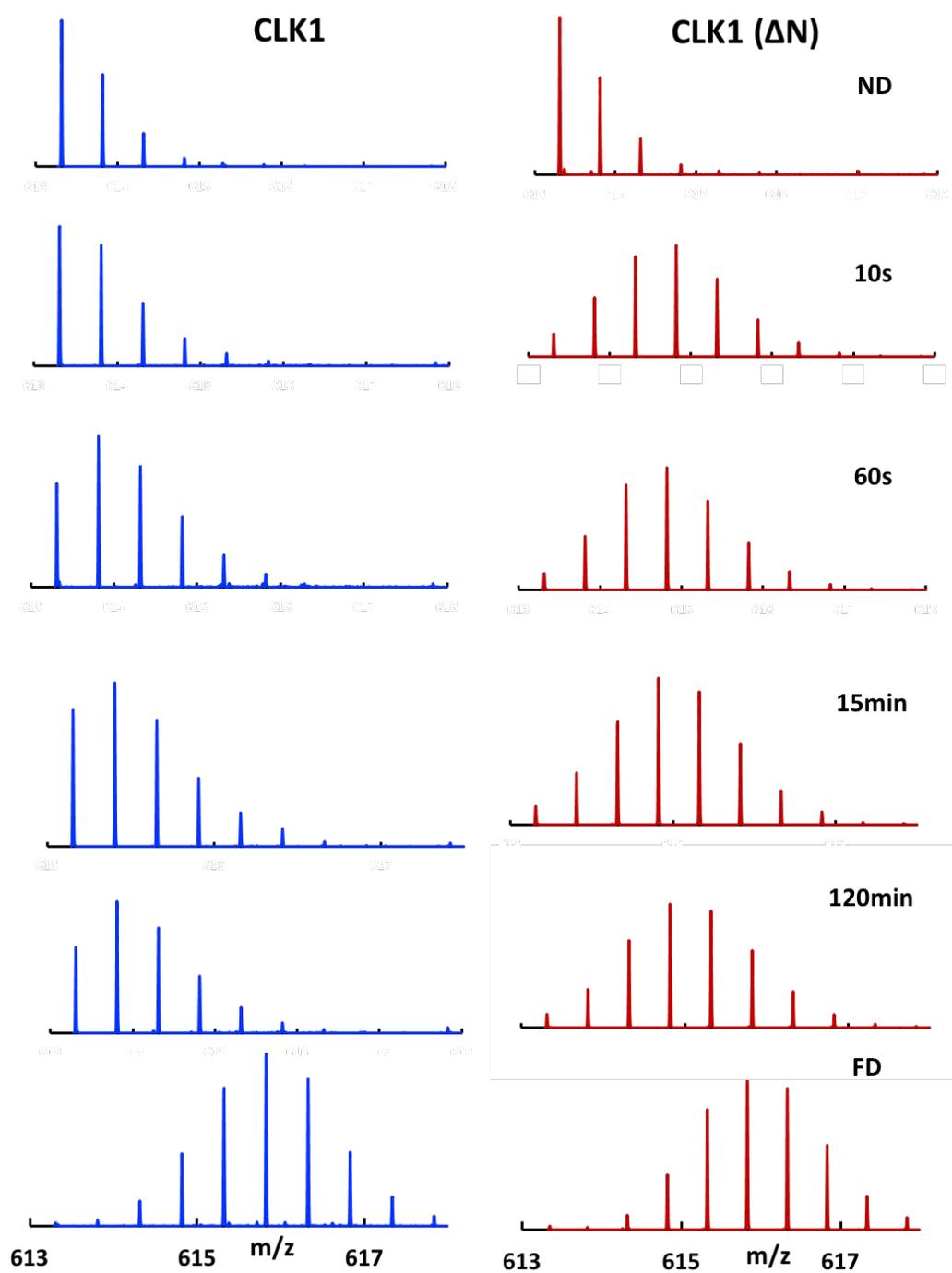


Figure 4.10. Shift in the mass envelope of peptide 441-450 in CLK1 (blue) and CLK1 (ΔN) with increasing time in D_2O buffer. ND: Non deuterated and FD: Fully deuterated.

exchange rates are unique for each peptide due to its sequence-dependent nature. We noticed that FD controls also showed identical centroid masses for both CLK1 and CLK1(Δ N), with the m/z much higher than the non-deuterated (ND) peptides (Fig.4.10). For intermediate time points, the peptide envelopes shifted to higher m/z values with increasing time. In most peptides, the shift in peptide masses was much more pronounced in CLK1(Δ N) than in CLK1 (Fig.4.10).

Next, we quantified and compared the amount of deuterium incorporation in both constructs in deuterium uptake plots. For all peptides, CLK1(Δ N) displayed varying intermediate rates of deuterium incorporation, while those in the kinase domain of CLK1 mostly exchanged very slowly. For every peptide detected, CLK1 showed a much lower deuterium uptake than CLK1(Δ N), even though there were slight differences in the extent of protection (Fig.4.11). The summary of how each region exchanges is shown in the heat maps in Fig 4.12. where the percentage of deuterium incorporation is shown as a function of time. Except for the residues on the C-terminal end, the percentage of deuterium incorporation in CLK1 never went higher than 40% in the timeframe of our kinetic experiment (2 hrs). Peptides on CLK1(Δ N), on the other hand, always showed a deuterium intake higher than 50% at just two hours (Fig.4.12). Certain regions in CLK1(Δ N) showed a rapid deuterium intake within the first 30 seconds of being placed in the D₂O buffer. These peptide fragments could lie in regions where there is high solvent accessibility (eg. the C-terminal end) or high flexibility (eg. the hinge region connecting the N and the C-lobes of the kinase domain).

Overall, our H/D exchange results suggest that when CLK1 forms oligomers, the kinase domain may be largely buried in the oligomer and is, therefore, not accessible to solvent. These results also support our previous results that the CLK1 oligomer also has N-K contacts, as the N-

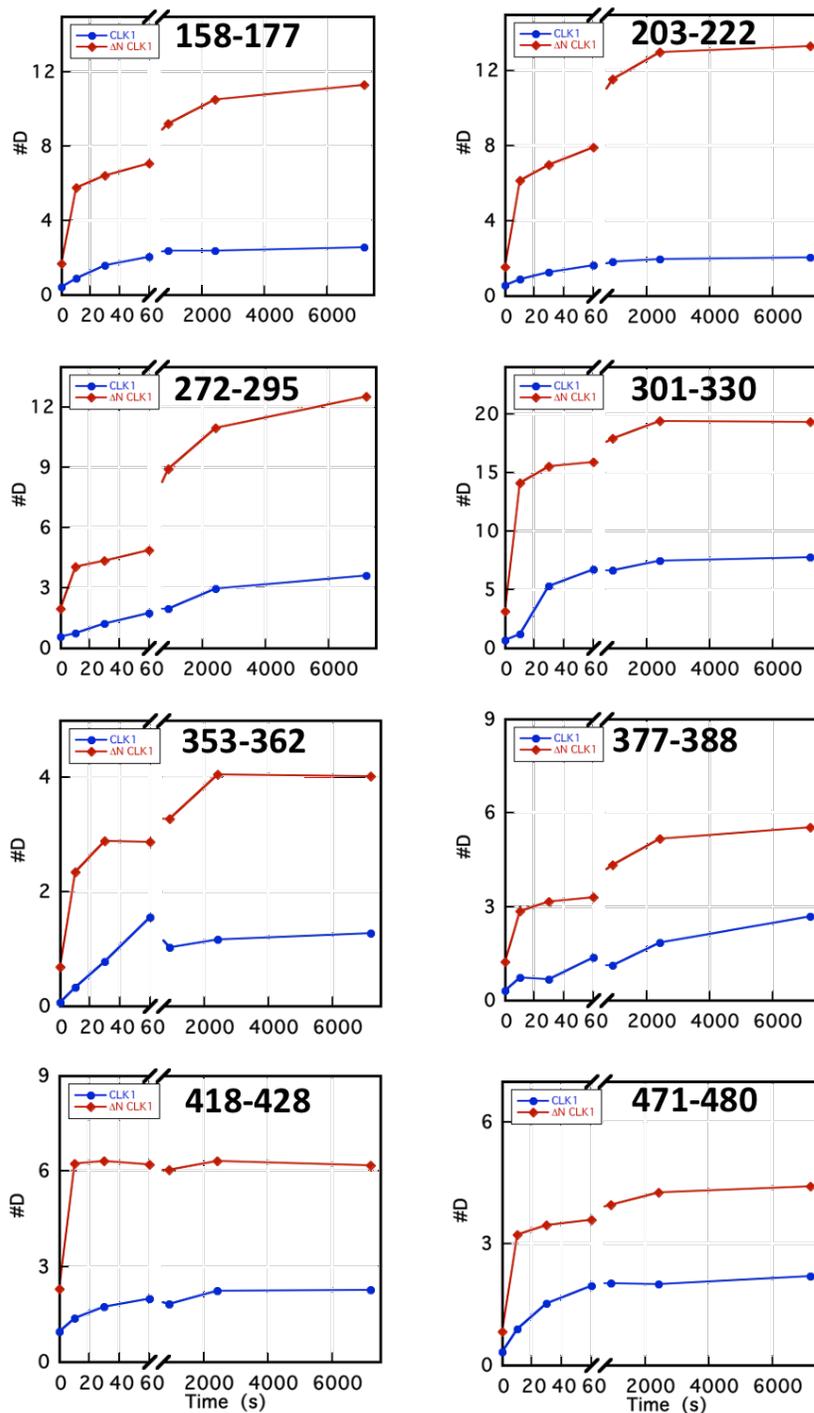
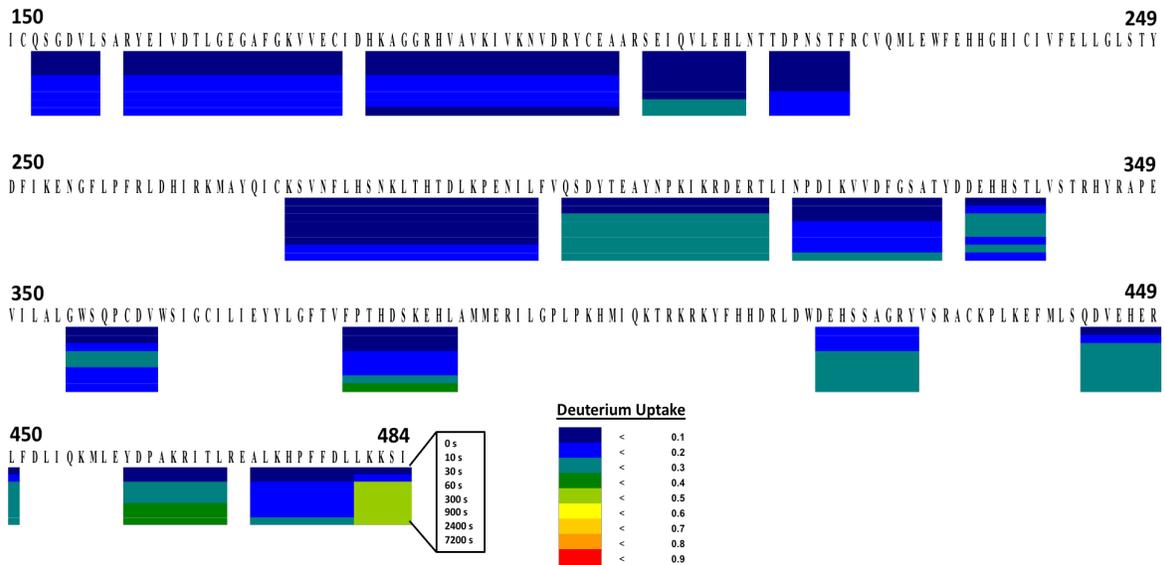


Figure 4.11. Deuterium incorporation plots as a function of time for a few representative peptides. CLK1 uptake is shown in blue and CLK1 (Δ N) uptake is shown in red. All peptides showed a much lower uptake for CLK1.

A)



B)

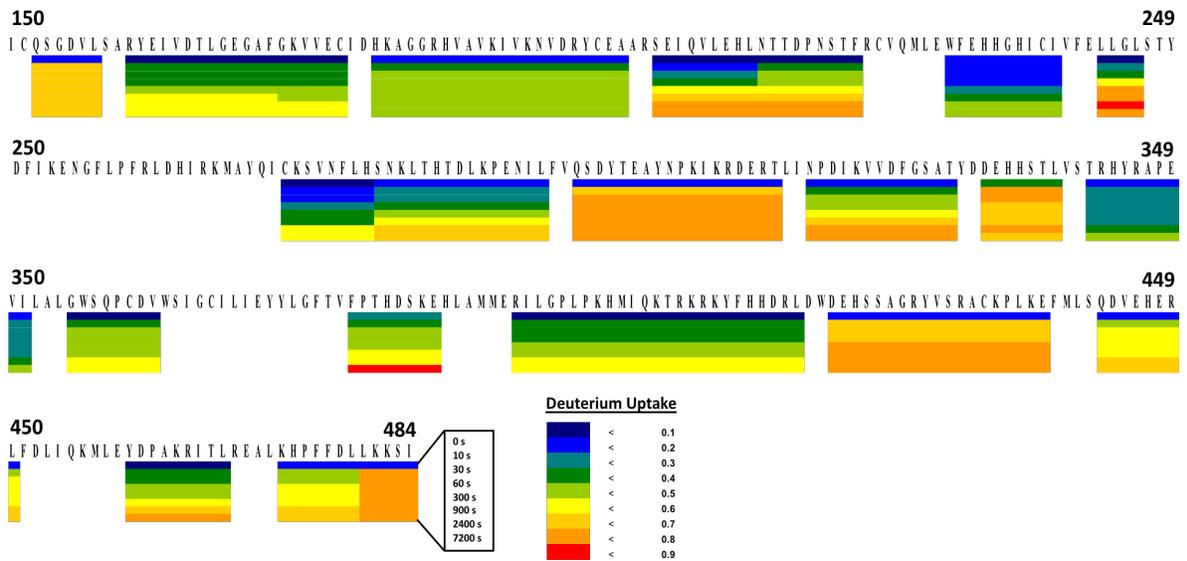


Figure 4.12. Kinase domain is protected in the presence of N-terminus: Deuterium uptake heat maps for (A) CLK1 and (B) CLK1 (Δ N).

terminus binding the kinase domain is reflected in the heavy protection of kinase domain in CLK1.

4.3 Discussion

Protein kinases use a variety of recognition mechanisms for efficient substrate phosphorylation. In general, docking grooves within the kinase core or flanking auxiliary domains secure the protein substrate in position for efficient phosphoryl transfer. However, CLKs are unique among the kinases in that they lack such structured elements and instead use disordered N-terminal extensions for substrate recognition and phosphorylation. Our laboratory showed that both the binding affinity and hyper-phosphorylation rate of the SR protein SRSF1 are heavily compromised in the absence of the N-terminus (37). This enhancement in SR protein phosphorylation induced by a disordered segment is positively coupled to the oligomerization state of CLK1. Furthermore, oligomerization may also facilitate the selection of this physiological substrate over non-physiological substrates. Hence, understanding the structural details of CLK1 oligomerization is of vital importance for understanding the mechanism of SR protein recognition as well as possible means of therapeutic intervention given the link of CLKs to cell proliferation and cancer.

To understand how CLK1 assembles into large oligomers we focused on identifying potential regions in the N-terminus that are indispensable for oligomerization. Interestingly, although the N-terminus is essential for CLK1 oligomerization, there does not appear to be significant sequence conservation among the N-termini within the larger CLK family (Fig. 4.13). In fact, less than 15% of residues on the N-terminus are conserved among all four isoforms. Given this observation, we elected to make broad deletions in the N-terminus to identify potentially relevant sequences. Overall, by studying N-terminal deletions of CLK1 or deletions of a GST fusion construct of the N-terminus, we found that residues in blocks 1 and 2 are particularly

important for the generation of the CLK1 oligomer. It will be interesting to probe further the first 100 residues in the N-terminus to see whether smaller regions in either block or perhaps a sequence that overlaps both blocks are critical for oligomerization of CLK1.

Disorder prediction programs suggest that the N-terminus of CLK1 is an intrinsically disordered domain (IDR). IDRs are often called ‘protein interaction hubs’ owing to their ability to adopt many conformations that can accept a wide range of binding partners (11). The amino acids in IDRs usually lack sequence diversity, with a limited subset of polar and charged residues constituting most of the sequence. The minimal representation of hydrophobic residues in an IDR may prevent the formation of a hydrophobic core and help induce disorder. On comparing the sequences of the N-termini within the CLK family, we find mostly charged (both positive and negative) and polar residues (Fig.4.13). While these sequence analyses might suggest that the N-N interactions are driven by electrostatic forces, our findings on the nature of N-N interactions tell a different story. We found that the N-N interactions are largely driven by aromatic residues. Such protein self-association mediated by aromatic residues in IDRs has been reported in a few other proteins like FUS and hnRNPA1 (15,30). Very recently, a ‘sticker and spacer’ model was proposed to explain the self-association of hnRNPA1 to form liquid condensates (34). According to this model, IDRs have ‘sticker’ residues (in this case aromatic residues) that act as non-covalent intermolecular crosslinkers. Stickers are separated by ‘spacers’ that do not interact with each other. Studies show that a uniform spacing of stickers is necessary to ensure that the strength of the interaction is strong enough to promote oligomerization but weak enough to prevent forming insoluble, solid aggregates (34). Coupling our results that the N-N interactions are driven by tyrosines with the previously known fact that CLK1 is a component of nuclear speckles (33), it is

highly likely that CLK1 oligomerization can also be explained by this sticker and spacer model.

Although previous studies showed that the N-terminus mediates oligomerization, most likely through N-N interactions (33), we found that N-K interactions also play an important role in CLK1 oligomerization. Among the three blocks of the N-terminus probed in our studies, blocks 1 and 3 form dominant interactions with the kinase domain in pull-down assays whereas block 2 shows only minimal contacts. These findings suggest that blocks 1 and 3 could potentially form both intramolecular and intermolecular contacts with the latter supporting CLK1 oligomerization. To address the latter possibility, we took advantage of results from a mutant form of the N-terminus (GST-N(Y-L)) that we found is not capable of forming an oligomeric species. Such findings strongly indicate that N-N contacts are driven by aromatic interactions and that removal of these contacts in the full-length CLK1 should lead to a monomeric kinase. Surprisingly, we found that removal of all tyrosines in the N-terminus did not abolish oligomerization in CLK1. However, tyrosine-to-leucine mutations throughout the N-terminus had an effect on oligomer size. We interpret such findings to mean that the N-terminus induces CLK1 oligomerization through two modes of intermolecular contacts within CLK1. The N-terminus joins CLK1 monomers together through N-N interactions that are driven by aromatic residues and N-K interactions that do not require such residues. An interesting question that arises here is why two different interaction modes are needed to drive oligomerization when either one alone induces oligomer formation. One possible explanation is that the two complementary interactions make the oligomerization more stable. We noticed that CLK1(Y-L) formed fewer and polydispersed oligomers than CLK1 at identical concentrations, a structural switch that might impair catalysis. Since oligomerization is also an activation mechanism, both N-N and N-K interactions might be required for efficient SR protein phosphorylation by CLK1.

Our structural and binding data strongly suggest CLK1 forms large oligomers by exploiting both intrasteric N-N and N-K interactions. To understand how such contacts impact the structure of the catalytic centers in the oligomer we used H/D exchange experiments. Our findings suggest that the N-terminus likely binds the kinase domain since the latter is heavily protected from amide backbone exchange in full-length CLK1 compared to CLK1 lacking its N-terminus (CLK1(Δ N)). The solvent inaccessibility of the kinase domain in the presence of the N-terminus suggests that the oligomer is a compact structure that is brought together by N-N and N-K interactions, an observation also supported by our TEM results. Although we did not detect any K-K interactions in CLK1(Δ N), the N-K interactions in the oligomer might bring together multiple kinase domains in the oligomer. There can be many advantages with the kinase functioning as an oligomeric system rather than a monomeric unit. When the kinase domains are in proximity in such an oligomeric complex, activation loop autophosphorylation might be facilitated, resulting in an enhanced activity. CLK autophosphorylation on activation loop tyrosine residues has been shown before (38,39). Such proximity-induced autophosphorylation is known to be the activation mechanism of many kinases like receptor tyrosine kinases (40) and Ca^{2+} /CamKII (7). Additionally, when the kinase is oligomeric, there are many N-termini on each oligomeric unit that can enhance the efficiency of substrate recruitment.

Based on all the insights that we gathered on the structural details of CLK1 oligomerization in this chapter, we propose a model for the oligomeric structure in Fig. 4.14. The figure summarizes how N-N interactions and blocks 1 and 3-mediated N-K interactions lead to oligomer formation. While N-N interactions are driven by tyrosine residues diffused throughout the N-terminus, N-K interactions are localized to K/R residue patches on block 1.

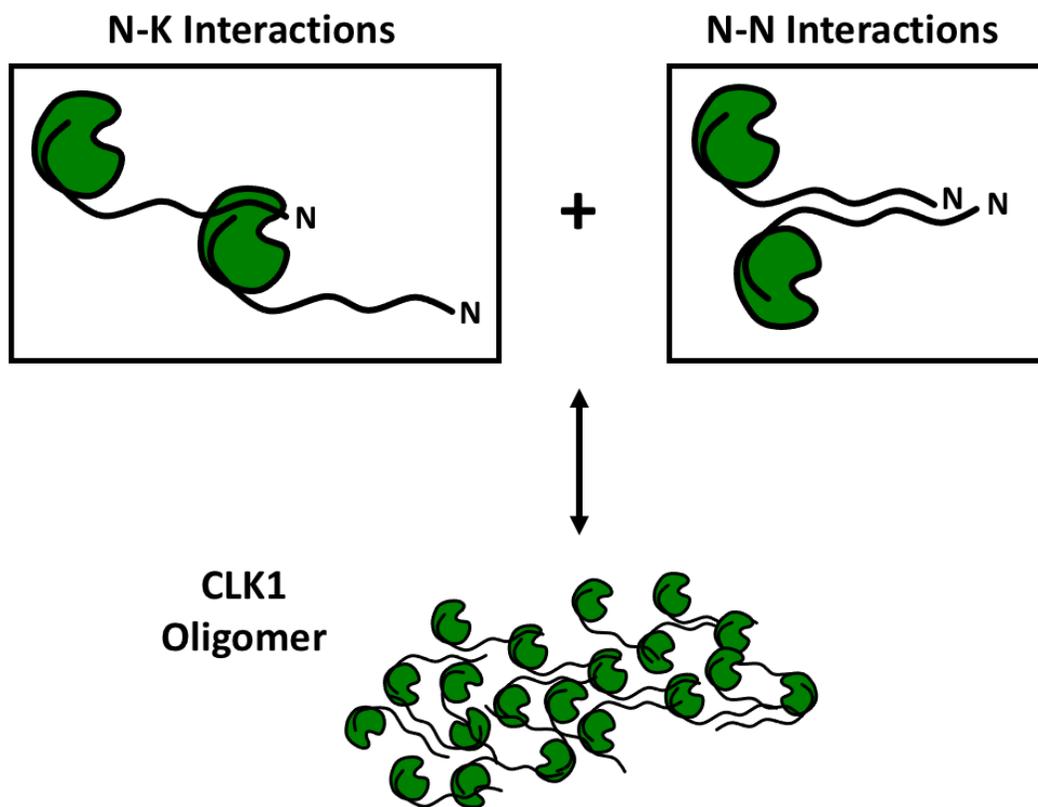


Figure 4.14. N-N and N-K interactions drive CLK1 oligomerization.

4.4. Conclusions

Since the N-terminus was indispensable for oligomerization and because self-association of RS like domains is a well-established fact, the CLK1 oligomerization was thought to be driven solely by N-N interactions. However, in this study, we show that CLK oligomerization is driven by a combination of N-N and N-K interactions. The two interactions are complementary and the absence of one interaction does not nullify oligomer formation in its entirety. We also show that aromatic interactions drive N-N interactions, but not N-K interactions. The two interactions

together lead to the formation of a compact oligomer that makes the kinase domains solvent inaccessible. Since oligomerization is CLK's activation mechanism, how does this oligomerization enhance the substrate phosphorylation becomes an interesting question. The impact of N-terminus mediated oligomerization on SRSF1 phosphorylation will be the focus of our study in the next chapter.

Chapter 4 is currently being prepared for submission for publication of the material: George A, Aubol BE, Adams JA. The dissertation author was a primary investigator and author of this paper.

Chapter 4 contains unpublished material with co-authors. George, Athira; Li, Sheng; Adams, Joseph A. The dissertation author was a primary investigator and author of this paper.

References

1. Hashimoto K, Panchenko AR. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci U S A*. 2010;107(47):20352–7.
2. Gotte G, Libonati M. Protein Oligomerization. *Oligomerization Chem Biol Compd*. 2014;450.
3. Dayhoff JE, Shoemaker BA, Bryant SH, Panchenko AR. Evolution of Protein Binding Modes in Homooligomers. *J Mol Biol*. 2010;395(4):860–70.
4. Liu S. A review on protein oligomerization process. *Int J Precis Eng Manuf*. 2015;16(13):2731–60.
5. Iqbal K, Liu F, Gong C-X, Grundke-Iqbal I. Tau in Alzheimer Disease and Related Tauopathies. *Curr Alzheimer Res*. 2010;7(8):656–64.
6. Lantsman K, Tombes RM. CaMK-II oligomerization potential determined using CFP/YFP FRET. *Biochim Biophys Acta - Mol Cell Res*. 2005;1746(1):45–54.
7. Swulius MT, Waxham MN. Ca²⁺/calmodulin-dependent protein kinases. *Cell Mol Life Sci*. 2008;65(17):2637–57.
8. Bhattacharyya M, Stratton MM, Going CC, McSpadden ED, Huang Y, Susa AC, Elleman A, Cao YM, Pappireddi N, Burkhardt P, Gee CL, Barros T, Schulman H, Williams ER, Kuriyan J. Molecular mechanism of activation-triggered subunit exchange in Ca²⁺/calmodulin-dependent protein kinase II. *Elife*. 2016;5(MARCH2016):1–32.
9. Gong Q, Long Z, Zhong FL, Teo DET, Jin Y, Yin Z, Boo ZZm Zhang Y, Zhang J, Yang R, Bhushan S, Reversade B, Li Z, Wu Bin. Structural basis of RIP2 activation and signaling. *Nat Commun*. 2018;9(1):1–13.
10. Pellegrini E, Signor L, Singh S, Erba EB, Cusack S. Structures of the inactive and active states of RIP2 kinase inform on the mechanism of activation. *PLoS One*. 2017;12(5):1–27.
11. Faust O, Bigman L, Friedler A. A role of disordered domains in regulating protein oligomerization and stability. *Chem Commun*. 2014;50(74):10797–800.
12. Fuertes G, Nevola L, Esteban-Martín S. Perspectives on drug discovery strategies based on IDPs. *Intrinsically Disordered Proteins*. Elsevier Inc.; 2019. 275–327 p.
13. Uversky VN. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol*. 2011;43(8):1090–103.

14. Uversky VN. Intrinsically disordered proteins and their “Mysterious” (meta)physics. *Front Phys*. 2019;7(FEB).
15. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*. 2015;163(1):123–33.
16. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki R, Pappu RV, Brangwynne CP. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*. 2016;165(7):1686–97.
17. Rai AK, Chen JX, Selbach M, Pelkmans L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature*. 2018;559(7713):211–6.
18. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall’Agnese A, Hannett NM, Spille JH, Afeyan LK, Zamudio AV, Shrinivas K, Abraham BJ, Boija A, Decker TM, Rimel JK, Fant CB, Lee TI, Cisse II, Sharp PA, Tatjes DJ, Young RA. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*. 2019;572(7770):543–8.
19. Brinda K V., Vishveshwara S. Oligomeric protein structure networks: Insights into protein-protein interactions. *BMC Bioinformatics*. 2005;6.
20. Baisamy L, Jurisch N, Diviani D. Leucine zipper-mediated homo-oligomerization regulates the Rho-GEF activity of AKAP-Lbc. *J Biol Chem*. 2005;280(15):15405–12.
21. Perutz MF, Staden R, Moens L, De Baere I. Polar zippers. *Curr Biol*. 1993;3(5):249–53.
22. Perutz MF. Polar zippers: their role in human disease. *Pharm Acta Helv*. 1995;69(4):213–24.
23. Perutz MF, Johnson T, Suzuki M, Finch JT. Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A*. 1994;91(12):5355–8.
24. Xue S, Gong R, He F, Li Y, Wang Y, Tan T, Luo SZ. Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci Adv*. 2019;5(11):2–10.
25. Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol Cell*. 2020;77(6):1237-1250.e4.
26. Lanzarotti E, Biekofsky RR, Estrin DA, Marti MA, Turjanski AG. Aromatic-aromatic interactions in proteins: Beyond the dimer. *J Chem Inf Model*. 2011;51(7):1623–33.

27. Elias RD, Ma W, Ghirlando R, Schwieters CD, Reddy VS, Deshmukh L. Proline-rich domain of human ALIX contains multiple TSG101-UEV interaction sites and forms phosphorylation-mediated reversible amyloids. *Proc Natl Acad Sci U S A*. 2020;117(39):24274–84.
28. Vargo MA, Nguyen L, Colman RF. Subunit Interface Residues of Glutathione S-Transferase A1-1 that Are Important in the Monomer-Dimer Equilibrium. *Biochemistry*. 2004;43(12):3327–35.
29. Song J, Ng SC, Tompa P, Lee KAW, Chan HS. Polycation- π Interactions Are a Driving Force for Molecular Recognition by an Intrinsically Disordered Oncoprotein Family. *PLoS Comput Biol*. 2013;9(9).
30. Kang J, Lim L, Lu Y, Song J. A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids. Vol. 17, *PLoS Biology*. 2019. 1–33 p.
31. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res*. 2006;34(1):305–12.
32. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*. 2009;417(1):15–27.
33. Keshwani MM, Hailey KL, Aubol BE, Fattet L, McGlone ML, Jennings PA, Adams JA. Nuclear Protein Kinase CLK1 Uses A Nontraditional Docking Mechanism To Select Physiological Substrates. *Biochem J*. 2015;1:BJ20150903.
34. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Sorrano A, Pappu RV, Mittag T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science (80-)*. 2020;367(6478):694–9.
35. Links DA. *Chem Soc Rev* Hydrogen exchange mass spectrometry for studying protein structure and dynamics. 2011;1224–34.
36. Weis DD, Wales TE, Engen JR, Hotchko M, Ten Eyck LF. Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis. *J Am Soc Mass Spectrom*. 2006;17(11):1498–509.
37. Aubol BE, Plocinik RM, Keshwani MM, Mcglone ML, Hagopian JC, Ghosh G, Fu XD, Adams, JA. N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. *Biochem J*. 2014;462(1):143–52.
38. Ben-David Y, Letwin K, Tannock L, Bernstein A, Pawson T. A mammalian protein kinase with potential for serine/threonine and tyrosine phosphorylation is related to cell cycle regulators. *EMBO J*. 1991;10(2):317–25.

39. Lee K, Du C, Horn M, Rabinow L. Activity and autophosphorylation of LAMMER protein kinases. *J Biol Chem.* 1996;271(44):27299–303.
40. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol Cancer.* 2018;17(1):1–13.

Chapter 5

Phosphorylation kinetics of CLK1

Abstract

The oligomerization of CLK1 is a substrate selection mechanism that allows CLK1 to recognize its physiological substrates over non-physiological substrates. In this chapter, we investigate the impact of CLK1 quaternary structure on its hyper-phosphorylation activity (i.e.-serine-proline phosphorylation). Our studies show that there is a strong correlation between the quaternary structure of CLK1 and SRSF1 hyper-phosphorylation. Although weakening oligomer formation by removing N-K interactions slows the rate of hyper-phosphorylation, severing N-N interactions through tyrosine mutations leads to the complete loss of hyper-phosphorylation activity.

5.1. Introduction

The phosphorylation levels of SR proteins are highly important for splicing function. Early *in vitro* assays showed that both hypo- and hyper-phosphorylation of SRSF1 is inhibitory for splicing of a reporter gene suggesting that SR protein phosphorylation status needs to be delicately balanced (1). SRPK1-dependent phosphorylation in the cytoplasm has been shown to promote migration of SRSF1 to the nucleus where it largely resides in storage speckles. Hyper-phosphorylation of SRSF1 by nuclear CLK1 then releases SRSF1 from these speckles allowing early assembly events in the spliceosome (2). On the other hand, dephosphorylation of SR proteins is necessary for later stages of spliceosome development and for catalysis of the second transesterification reaction. SR protein phosphorylation kinetics of two families of kinases have received considerable attention in the last two decades: SRPKs and CLKs. Structurally, the two families of kinases possess a canonical kinase domain fold with an N-lobe and a C-lobe but differ significantly outside these core elements (3,4). These differences likely play important roles in their biological function with respect to splicing regulation.

Although SRPK1 and CLK1 both target RS domains in SR proteins, their phosphorylation kinetics and substrate specificities are very different (5). Kinetic assays and mass spectrometric detection of phosphorylated peptides of SRSF1 revealed that SRPK1 rapidly phosphorylates about 10 serines while CLK1 phosphorylates up to 20 sites at a much slower rate (6). SRPK1 has a strong preference for the N-terminal residues (RS1) while CLK1 phosphorylates both RS1 and RS2 with equal efficiency (Fig.1.4B and Fig.1.5D) (7). Single-turnover progress curves for SRSF1 phosphorylation are best fit to a biexponential function indicating that SRPK1 phosphorylation happens in two distinct phases: a fast phase with a $t_{1/2}$ of less than 0.3 minutes followed by a much slower phase with a $t_{1/2}$ of about 10 minutes. The fast phase is associated with RS1 phosphorylation

whereas the slow phase is reflective of RS2 phosphorylation. It has also been demonstrated that the two kinases can act synergistically. CLK1 not only can phosphorylate SRSF1 alone but also can phosphorylate RS domains that are pre-phosphorylated by SRPK1 (6). MALDI analyses have also shown that neither CLK1 nor SRPK1 phosphorylates residues on RRM1s suggesting that both kinases strictly target RS domains (8). Although both kinases modify Arg-Ser dipeptides, CLK1 possesses the additional capacity to phosphorylate Ser-Pro dipeptides in SR proteins. Prior work from our lab showed that CLK1 modifies three Ser-Pro dipeptides in the RS2 segment inducing a characteristic hyper-phosphorylation state for SRSF1 viewed as a slow-migrating band on SDS-PAGE (6). This modification has significant effects on cellular function promoting release of SRSF1 from nuclear storage speckles, attachment to U1 snRNP in the early spliceosome, and widespread changes on alternative mRNA splicing patterns (9).

Kinases that phosphorylate more than one site on a substrate can utilize either a processive or a distributive mechanism. In a processive phosphorylation mechanism, the kinase binds the substrate and phosphorylates all sites before releasing the final phospho-product. On the other hand, in a distributive phosphorylation mechanism, the substrate is released from the kinase after each phosphorylation event before re-binding to the active site occurs. Since SR proteins are phosphorylated at numerous sites, specialized kinetic strategies are necessary to evaluate the nature of the phosphorylation mechanism (10). In the start-trap experiment, developed previously in our lab (10), a pre-incubated complex of kinase and substrate is mixed with ATP to initiate phosphorylation (start) and then trapped by the addition of a large amount of kinase dead enzyme (trap). If the kinase is distributive, intermediate phospho-products will be released from the active site and tightly bound by the trap, halting further phosphorylation steps. If the mechanism is processive, no intermediate phospho-products are released and fully phosphorylated substrate is

detected. Start-trap experiments have shown that both SRPK1 phosphorylation of RS1 and CLK1 phosphorylation of the entire RS domain are processive (8). To support a multi-step, processive phosphorylation reaction, the kinase needs to bind with high affinity to the substrate and its phospho-intermediate forms prior to release of the fully phosphorylated product. Indeed, the affinity of SRSF1 to both CLK1 and SRPK1 is unusually high compared to most kinase-substrate pairs with K_m values in the low nanomolar range (6).

Having shown the processive nature of both kinases, the next interesting question is how multi-site phosphorylation occurs. Both SRPKs and CLKs bind RS domains in SR proteins phosphorylating serines that frequently reside in lengthy, consecutive Arg-Ser repeats. To address how phosphates are added, our lab developed an engineered Lys-C footprinting assay to test whether either kinase adds phosphates onto the RS domain of SRSF1 in either a random or directional (N-to-C or C-to-N) manner. In this assay, a lysine residue is placed in the middle of the RS1 domain and the substrate is then digested with Lys-C as a function of phosphorylation (7). The number of phosphates incorporated in both segments is then analyzed showing that SRPK1 phosphorylates RS1 in a C-to-N direction. Using variable cleavage sites throughout the RS domain, our lab showed that SRPK1 initiates phosphorylation at the C-terminal end of RS1, (residues 221-225), whereas CLK1 initiates randomly in the RS domain, thus, displaying no specific directionality (11).

Both CLK1 and SRPK1 bind SRSF1 very tightly with similar K_m values (approx. 100 nM). However, since K_m values are not always reflective of true dissociation constants, competition assays were previously performed to determine K_i values for both kinases. Such studies showed that SRSF1 binds to CLK1 with a 10-fold greater affinity than SRPK1 (6). Measured binding affinities also revealed that while SRPK1 showed a preference for RS1, CLK1 makes extensive

contacts with the entire RS domain, an observation consistent with its broader regiospecificity (6). Additionally, competition assays also found that the deletion of RRM2 results in large decreases in binding affinities to CLK1, suggesting that RRM2 makes contacts with the kinase. A pull-down assay with GST-tagged RRM2 and CLK1 also confirmed that CLK1 interacts with RRM2. For SRPK1, the K_i value for RRM2 was about 25-fold higher than that of CLK1, indicating that interaction of SRPK1 with RRM2, in the absence of the RS domain is very weak. However, an X-ray structure of SRPK1 in complex with RRM2-RS1 (PDB ID: 3BEG) shows that RRM2 makes contacts with the N-lobe of SRPK1, but the RRM2 interaction is most likely facilitated by the docking groove occupancy of RS1 residues.

In addition to differences in phosphorylation regiospecificity within the RS domain, CLK1 and SRPK1 also differ in their rate-limiting steps. SRPK1 has a fast phosphoryl transfer step ($k_p \sim 30 \text{ sec}^{-1}$) and is rate limited by the slower ADP release step ($k_{\text{off}} \sim 0.5 \text{ sec}^{-1}$) (12). Deletion analyses indicated that regions from the disordered regions in the N-terminal extension and the spacer insert domain function cooperatively to facilitate nucleotide exchange (13). SRPK1 phosphorylation is also remarkably ordered. The crystal structure of SRPK1 bound to a substrate peptide revealed an acidic docking groove that is away from the active site (14). Further studies also unraveled a unique ‘feeding’ mechanism mediated by sliding docking interactions (15). In this mechanism, the C-terminal residues of RS1 bind to the docking groove and is sequentially fed into the active site in two-residue increments as phosphorylation continues. As more serines get phosphorylated, the buildup of negative charge leads to the unfolding of $\beta 4$ (residues 190-196) in RRM2 and binding to the docking groove facilitating product release (16). These findings provide a strong mechanistic foundation for the C-to-N directional mechanism observed in LysC footprinting experiments (7). CLK1, on the other hand, has a slow, rate-limiting phosphoryl transfer step most

likely due to the absence of a structured docking groove. The X-ray structures of the CLK kinase domains show unique insertions that block access to any potential docking groove in the C-lobe (4). The absence of a docking groove most likely explains the lack of directionality in CLK1 phosphorylation. In addition to this, CLK1 binds tightly to both phosphorylated and unphosphorylated SRSF1 making product release a daunting task. However, our lab showed previously that release of phosphorylated SRSF1 is greatly facilitated by the binding of SRPK1 and formation of the CLK1-SRPK1 complex. SRPK1 binds tightly to the N terminus of CLK1, thereby severing contacts between the N-terminus and the RS domain and releasing phosphorylated SRSF1 from CLK1 (17).

The 150-residue N-terminus of CLK1, composed of numerous disorder-promoting residues (Fig. 5.1A), is indispensable for SR protein recognition and phosphorylation. In the absence of the N-terminus, SRSF1 binding and phosphorylation are significantly impaired. For example, GST-SRSF1 readily pulls down CLK1, but not CLK1(Δ N) (18). While CLK1 can phosphorylate about 18 sites on SRSF1, CLK1(Δ N) phosphorylates only 6 sites. The N-terminus mediated activation has also been demonstrated for other SR proteins, including SRSF2, SRSF5 and the SR-like protein Tra2 β . In addition to general decreases in phosphoryl content, CLK1(Δ N) is not able to phosphorylate Ser-Pro dipeptides, a feature unique to CLK1. Ser-Pro phosphorylation manifests as a gel shift on the autoradiograms, arising from an altered conformational landscape possibly induced by cis-trans isomerization of prolines. Hyper-phosphorylation of Ser-Pro dipeptides has been shown to be essential for a multitude of downstream functions, starting with speckle dissociation, cooperative binding to pre-mRNA splice sites, and subsequent dephosphorylation by protein phosphatase 1 (PP1) (2,11).

Lacking a docking groove similar to SRPK1, CLK1 instead uses a non-traditional recognition mechanism to select for its physiological substrates, the SR proteins. Prior studies from our lab showed that the N-terminus interacts directly and with high affinity to the RS domain in SRSF1 (19). Interestingly, the N-terminus also induces oligomerization which correlates with higher CLK1 specific activity ($v/[E]$) towards SR proteins compared to other non-physiological substrates like myelin basic protein (MBP) (19). In addition to controlling recognition and Ser-Pro dipeptide specificity, the N-terminus is also essential for recruiting CLK1 into nuclear speckles (19). However, while the role of the N-terminus with regard to these phenomena are well documented, how it controls substrate interactions and the oligomerization state of CLK1 for catalytic function at a residue-specific level is still poorly understood. In this chapter, we will address how the N-terminus activates CLK1 for SRSF1 phosphorylation and how such activation is linked to the oligomerization state of the kinase.

5.2. Results.

5.2.1. Deletion of any single block does not impair SRSF1 hyper-phosphorylation or binding.

In order to investigate how the N-terminus modulates the binding and phosphorylation of SRSF1, we first verified that CLK1 shows a much higher activity than CLK1(Δ N) in single-turnover kinetic assays as previously reported (18). We purified CLK1 from baculovirus infected Hi5 cells and CLK1(Δ N) construct, that has the entire N-terminus deleted, from *E.coli*. We noticed that CLK1, but not CLK1(Δ N), induced a slower-migrating, gel-shifted band (i.e.- hyper-phosphorylation) for SRSF1 confirming that the deletion of the N-terminus completely abolishes Ser-Pro phosphorylation (Fig. 5.1B,C). We also tried to test the importance of the N-terminus for SRSF1 binding by measuring the K_m values of CLK1 and CLK1(Δ N) under steady-state conditions. While we measured a K_m of about 100 nM for CLK1, the K_m for CLK1(Δ N) was

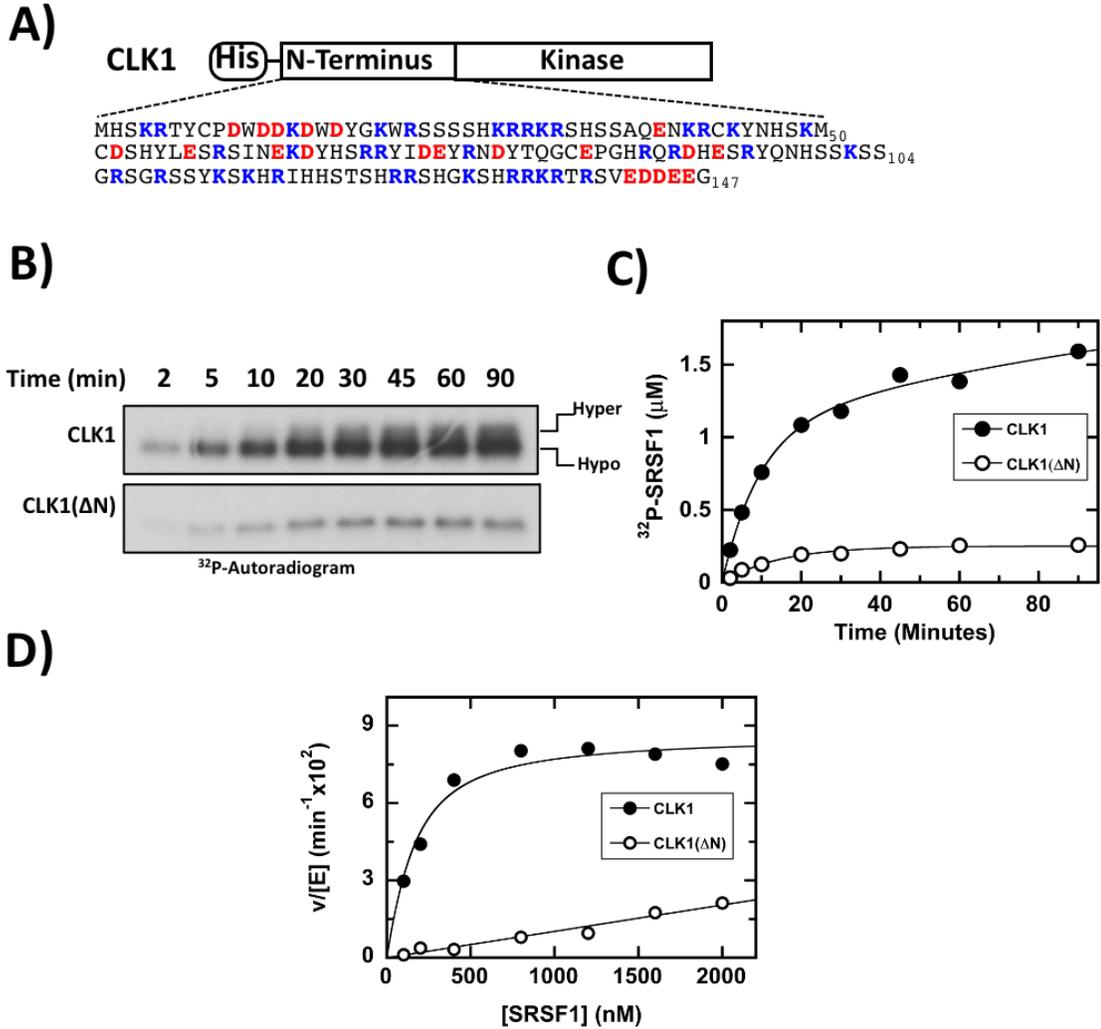


Figure 5.1. CLK1 N-terminus induces hyper-phosphorylation (A) CLK1 N-terminus (B) ³²P-Autoradiograms and single turnover progress curves showing SRSF1 hyper-phosphorylation with CLK1 but not with CLK1 (ΔN). (D) Steady state $v/[E]$ vs [SRSF1] for CLK1 and CLK1 (ΔN).

indeterminate under our experimental conditions as the initial velocity did not saturate in the SRSF1 concentration range tested (Fig. 5.1 D). SRSF1 tends to aggregate and precipitate when added into activity assay mixtures beyond a concentration of 2 μ M, limiting our ability to measure the initial velocity at SRSF1 concentrations above this limit. However, considering this technical shortcoming, it is likely that the N-terminus enhances the observed SRSF1 binding affinity by a minimum of 20-fold. Thus, the deletion of the N-terminus has dramatic impacts on both Ser-Pro hyper-phosphorylation and SRSF1 binding affinity.

Next, we expressed and purified several deletion mutants in insect cells using baculoviruses (Fig. 5.2A). Three single-block deletions CLK1(Δ 1), CLK1(Δ 2), and CLK1(Δ 3) that remove approximately 50 residues each were purified and their activities towards SRSF1 were measured in single-turnover kinetic assays. We observed gel-shifted bands in the autoradiograms of SRSF1 using wild-type and all three single-block deletions, indicating that any two blocks support Ser-Pro hyper-phosphorylation (Fig. 5.2B). Phosphorylated products were quantified and plotted as a function of time (Fig. 5.2C). The curves for total SRSF1 phosphorylation (Total-P) were fitted to double exponential functions whereas those for hyper-phosphorylation (Hyper-P) were fitted to single exponential functions to obtain initial velocities and maximum amplitudes for the latter (Fig. 5.2D). With regard to total phosphorylation, we noticed that the maximum amplitudes for none of the single block deletions were significantly compromised relative to CLK1 although CLK1(Δ 1) seemed to phosphorylate with a smaller initial velocity than that for CLK1 (Fig 5.2C). Since hyper-phosphorylation at Ser-Pro dipeptides is a unique feature of CLK1, we focused attention on the impact of deletions on hyper-phosphorylation. Surprisingly, while CLK1(Δ 1) was slightly slower than CLK1, the deletion of block 2 in CLK1(Δ 2) resulted in an enhanced hyper-phosphorylation

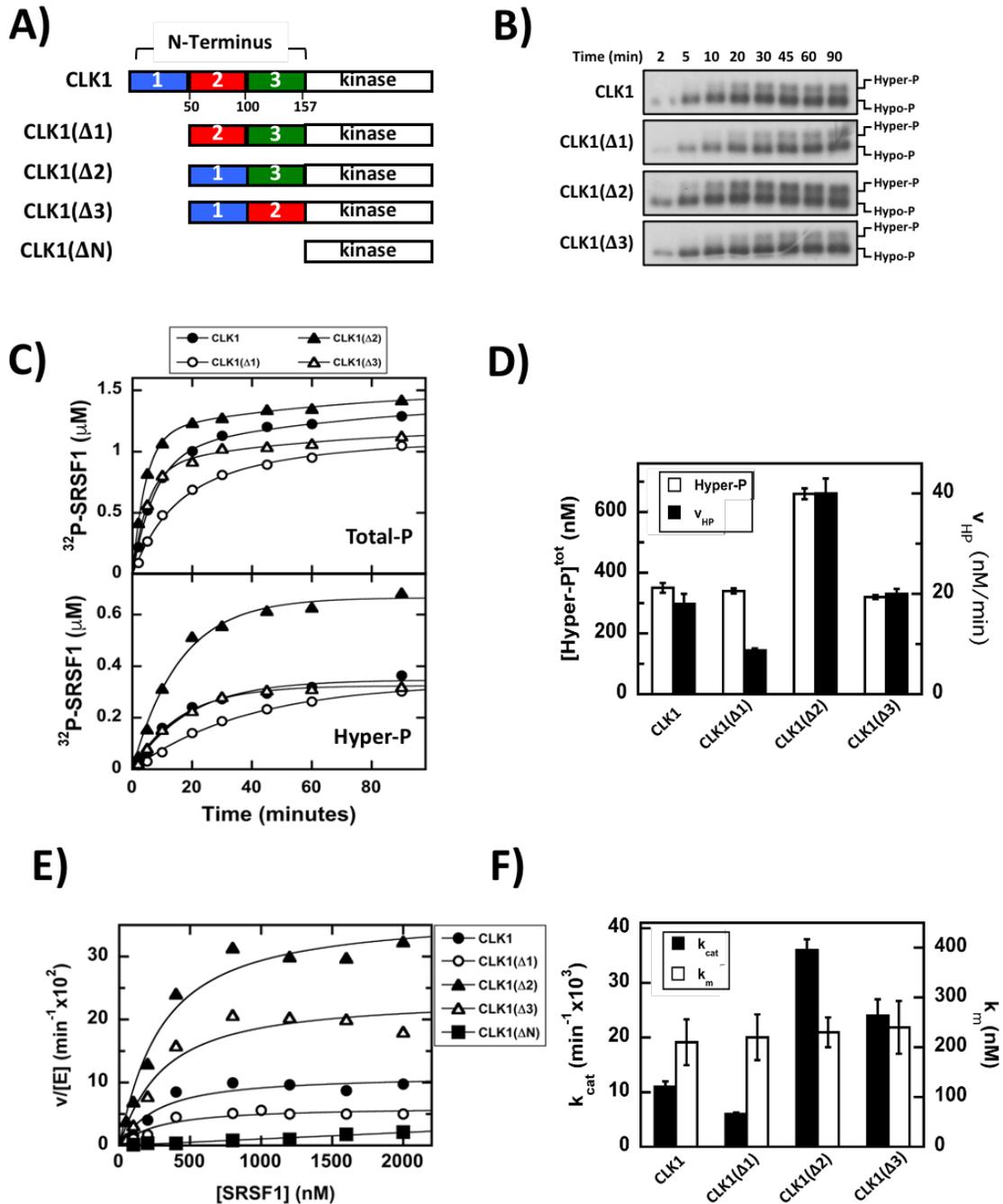


Figure 5.2. Single block deletion mutants retain substrate binding and hyper-phosphorylation (A) Single block deletions (B). 32 P-Autoradiogram of single turnover assays with single block deletions (C) single turnover progress curves of total product (upper panel) and hyper-phosphorylated product (bottom panel) with single block deletions. (D) Velocities and maximum amplitudes from (C) after fitting to single exponential function (E) Steady state $v/[E]$ vs [SRSF1] for CLK1 single block deletions. (F) Measured k_{cat} and K_m values from (E).

both in terms of amplitude and velocity (Fig. 5.2D). This gain-of-function result is particularly remarkable as it suggests that the N-terminus, while a strong activator of SR protein hyper-phosphorylation, may contain some inhibitory sequences.

Next, we measured the SRSF1 K_m values for single-block deletions using steady-state kinetics. The initial velocities were measured by quantifying the amount of phosphorylated product formed 5 minutes after initiating the reaction. The velocities were plotted as a function of increasing SRSF1 concentration (Fig. 5.2E) and the data were fitted to a quadratic equation to obtain k_{cat} and K_m values (Fig. 5.2F). The SRSF1 K_m values of the single-block deletions are comparable to that measured for CLK1, indicating that observed SRSF1 binding affinity is not affected by the deletion of any specific fifty-residue block. This suggests that SRSF1 binding might be sequence-insensitive as any sequence-specific binding will be reflected in a higher K_m . It was observed previously from K_i experiments that the CLK1 N-terminus makes extensive contacts with the RS domain with no specificity for any region within the RS domain (6), also hinting towards sequence insensitive binding. The k_{cat} values, however, showed a few variations between constructs. While CLK1(Δ 1) showed a slightly smaller k_{cat} than CLK1, CLK1(Δ 2) and CLK1(Δ 3) displayed higher k_{cat} values with CLK1(Δ 2) showing the greatest effects. Overall, the data from single-turnover and steady-state kinetic studies suggest that the disordered N-terminus greatly activates the kinase domain of CLK1 toward SRSF1 hyper-phosphorylation but that some sequences may play repressive roles. CLK1(Δ 2) is a more efficient kinase than CLK1, particularly with respect to hyper-phosphorylation. Furthermore, CLK1(Δ 2) is also a more effective catalyst than the wild-type CLK1 based on traditional measurements such as k_{cat}/K_m which monitors phosphorylation rates at limiting substrate concentration.

5.2.2. Deletion of two blocks does not affect hyper-phosphorylation but impairs binding.

Next, we purified three double-block deletion constructs, CLK1(Δ 12), CLK1(Δ 13), and CLK1(Δ 23), and monitored their catalytic activities toward SRSF1 (Fig. 5.3A). Under single-turnover kinetic conditions, we looked for any impairment in hyper-phosphorylation. We found that all three double deletions were capable of Ser-Pro phosphorylation, as could be seen by the presence of an upper band on the autoradiograms (Fig. 5.3B). We next quantified the products formed and then fitted the single-turnover progress curves to single exponential functions to get the velocities (Fig. 5.3C,D). While there was a slight decrease in the initial velocities for CLK1(Δ 12) and CLK1(Δ 13) while monitoring the hyper-phosphorylation phases (Fig. 5.3C, lower panel), we noticed that CLK1(Δ 23) was faster than CLK1. This was a surprising finding as CLK1(Δ 23), which has only fifty residues more than CLK1(Δ N), was faster than CLK1 in hyper-phosphorylation. We also noticed that CLK1(Δ 23) formed almost as much hyper-phosphorylated product as CLK1 did. Interestingly, CLK1(Δ 23) removes block 2 which we found in our single-block deletion experiments to play a repressive role in SRSF1 phosphorylation (Fig. 5.2). Thus, it appears that removal of block 2, even in context of a secondary deletion in block 3, activates the kinase domain relative to wild-type CLK1.

We measured the binding affinities of the two-block deletion mutants with SRSF1 by measuring K_m values in plots of initial velocity vs total substrate concentration. Interestingly, we observed much weaker binding of SRSF1 to CLK1(Δ 12) and CLK1 (Δ 13). The measured K_m for SRSF1 was five-fold higher for CLK1(Δ 13) and CLK1(Δ 12), indicating that if the length of the N-terminus is shortened by 100 residues or more, the SRSF1 binding affinity is weakened. CLK1(Δ 23), however, was an exception to the observed trend of higher K_m values. The binding affinity for CLK1(Δ 23) was comparable to that of CLK1. This is surprising as we have shown

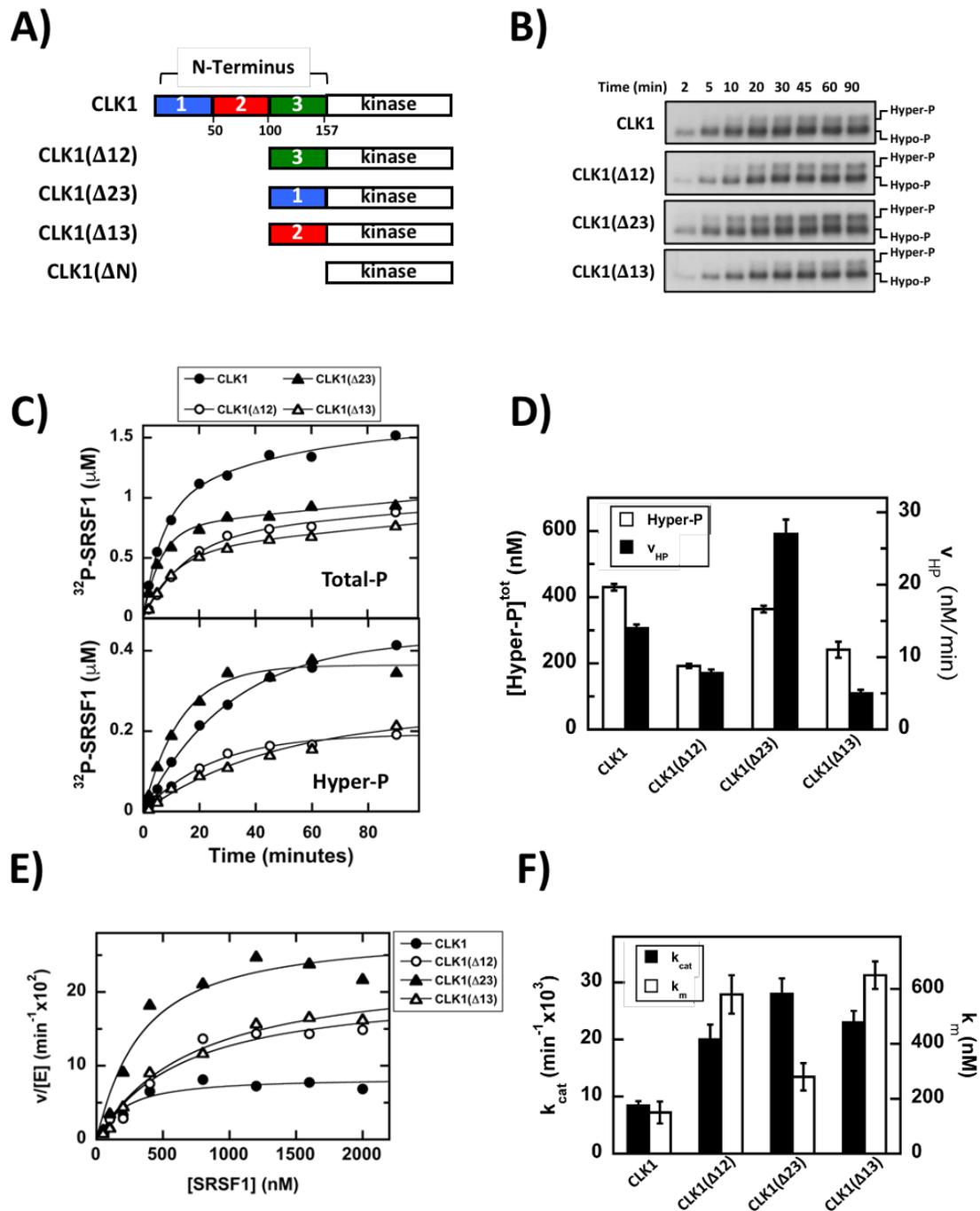


Figure 5.3. CLK1 two-block deletions retain hyper-phosphorylation but substrate binding is impaired. **(A)** Two-block deletions **(B)**. ^{32}P -Autoradiograms of single turnover assays with double block deletions **(C)** Single turnover progress curves of total product (upper panel) and hyper-phosphorylated product (bottom panel) with single block deletions. **(D)** Velocities and maximum amplitudes from **(C)** after fitting to single exponential function **(E)** Steady state $v/[E]$ vs $[\text{SRSF1}]$ for CLK1 two block deletions. **(F)** Measured k_{cat} and K_{m} values from **(E)**.

from pull-down assays that sequential deletion of the N-terminus lowers SRSF1 binding affinity. However, we may have strengthened the interaction with SRSF1 through some non-natural interactions as in CLK1(Δ 23) as block 1 is placed next to the kinase domain in this deletion construct. Another noteworthy finding from this is that all three two-block deletions have a five-fold higher k_{cat} , possibly due to the weak binding. Since CLK1 binds strongly to both phosphorylated and unphosphorylated SRSF1, product release can be inhibitive under steady state conditions. Weakened binding of SRSF1 in two-block deletions can also result in faster release of phosphorylated product manifesting in a higher turnover. Although we observed some differences in SRSF1 binding and Ser-Pro phosphorylation with the sequential deletion of the N-terminus, surprisingly, the deletion mutants retained their hyper-phosphorylation with any fifty-residue stretch of the N-terminus attached to the kinase domain. The need for having a 150-residue long stretch of disordered N-terminus was intriguing when a 50-residue stretch could almost be equally productive. Since, inside cells, CLK1 functions along with SRPK1 as a symbiotic kinase complex, we decided to investigate the impact of N-terminal deletion on SRPK1 activation.

5.2.3. SRPK1 activates CLK1 deletions.

Our laboratory demonstrated that SRPK1 forms a complex with CLK1 that enhances CLK1 mediated hyper-phosphorylation of SR proteins (20,21). Since previous studies have shown a correlation between the quaternary structure of CLK1 and catalytic activity (19), we wished to understand if SRPK1 alters the oligomeric structure of CLK1 in anyway. While CLK1 is known to form oligomers through its disordered N-terminus, SRPK1 is not known to form oligomers. We confirmed these results using negatively stained transmission electron microscopy (TEM) (Fig. 5.4). Next, we looked at a preincubated CLK1:SRPK1 complex (CLK1:SRPK1 \approx 1:3) under the microscope and found that, in the presence of SRPK1, CLK1 formed larger superstructures. While

CLK1 oligomers were about 20 nm in diameter, SRPK1 induced the formation of larger assemblies that appeared to have stitched together several individual units of CLK1 oligomers. Since we know that the CLK1-SRPK1 complex is a much more efficient system than either enzyme individually, it is also possible that the assembly into larger superstructures also contributes to the enhanced efficiency.

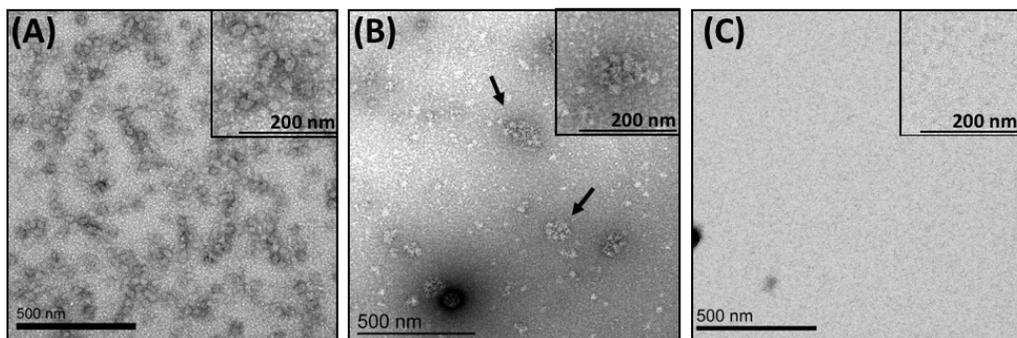


Figure 5.4. SRPK1 binding induces the formation of larger CLK1 oligomers. Negative stained micrographs of (A) CLK1 alone (B) CLK1-SRPK1 complexes and (C) SRPK1 alone. The inset shows the magnified view of a small region in the figure. The arrows show CLK-SRPK complexes. Scale bar: 500 nm.

Since CLK1 binding to SRPK1 is mediated by the N-terminus, we were interested in investigating how the enhancement of hyper-phosphorylation is affected upon deletion of the N-terminus. Single-turnover kinetic assays were performed with CLK1 single-block deletions in the presence and absence of SRPK1 (Fig. 5.5A). We observed that the hyper-phosphorylation phases for all CLK1 deletions are remarkably faster in the presence of SRPK1 (Fig. 5.5B). The autoradiograms showed a predominant upper band within only 10 minutes of reaction time. The

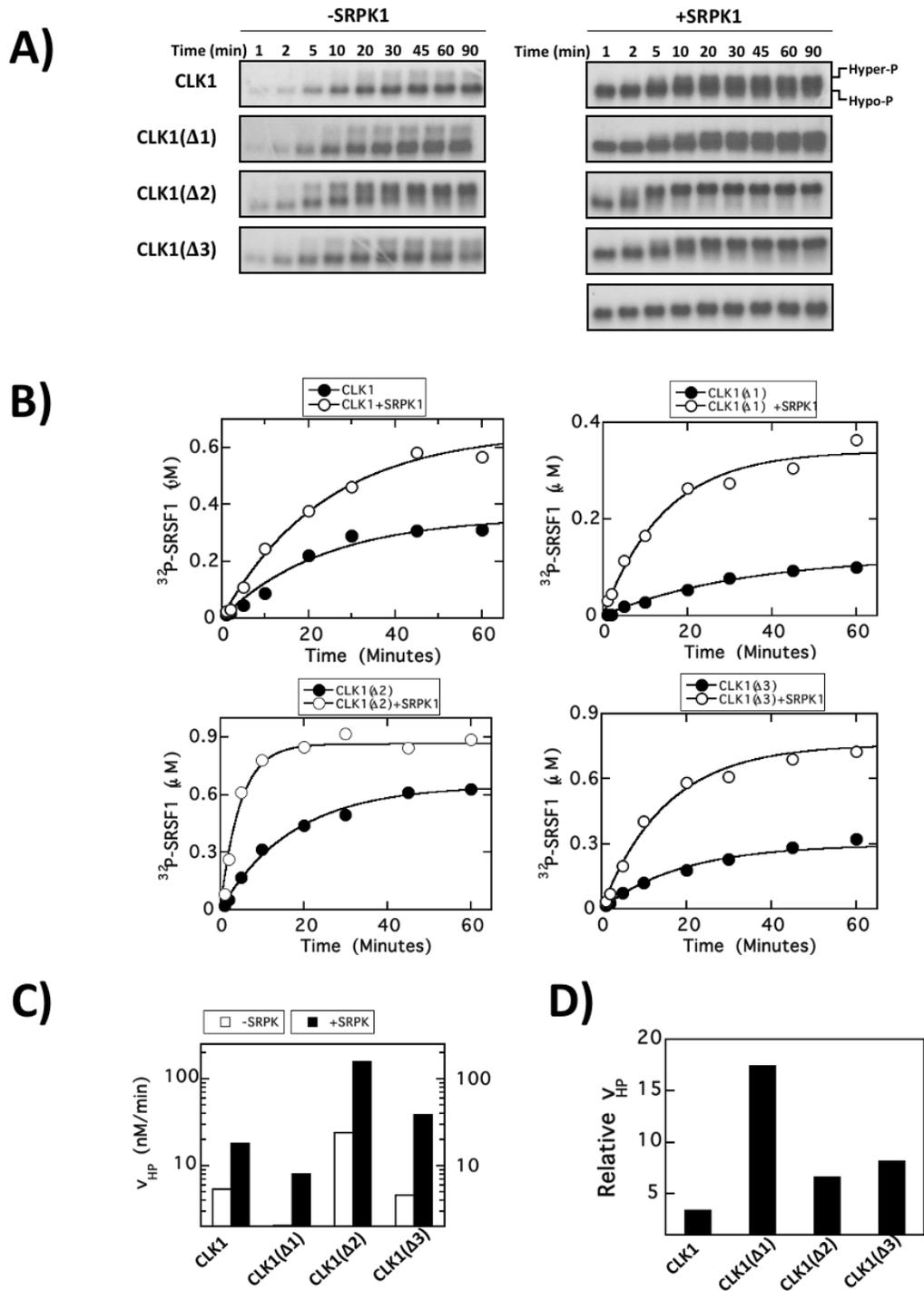


Figure 5.5. SRPK1 activates CLK1 single block deletions (A) 32 P-Autoradiograms of single turnover assays with single block deletions (B). Single turnover progress curves of hyper-phosphorylated product of single block deletions in the presence and absence of SRPK1. (C) Velocities from (B) after fitting to single exponential function (D) Relative enhancement in velocities observed in the presence of SRPK1.

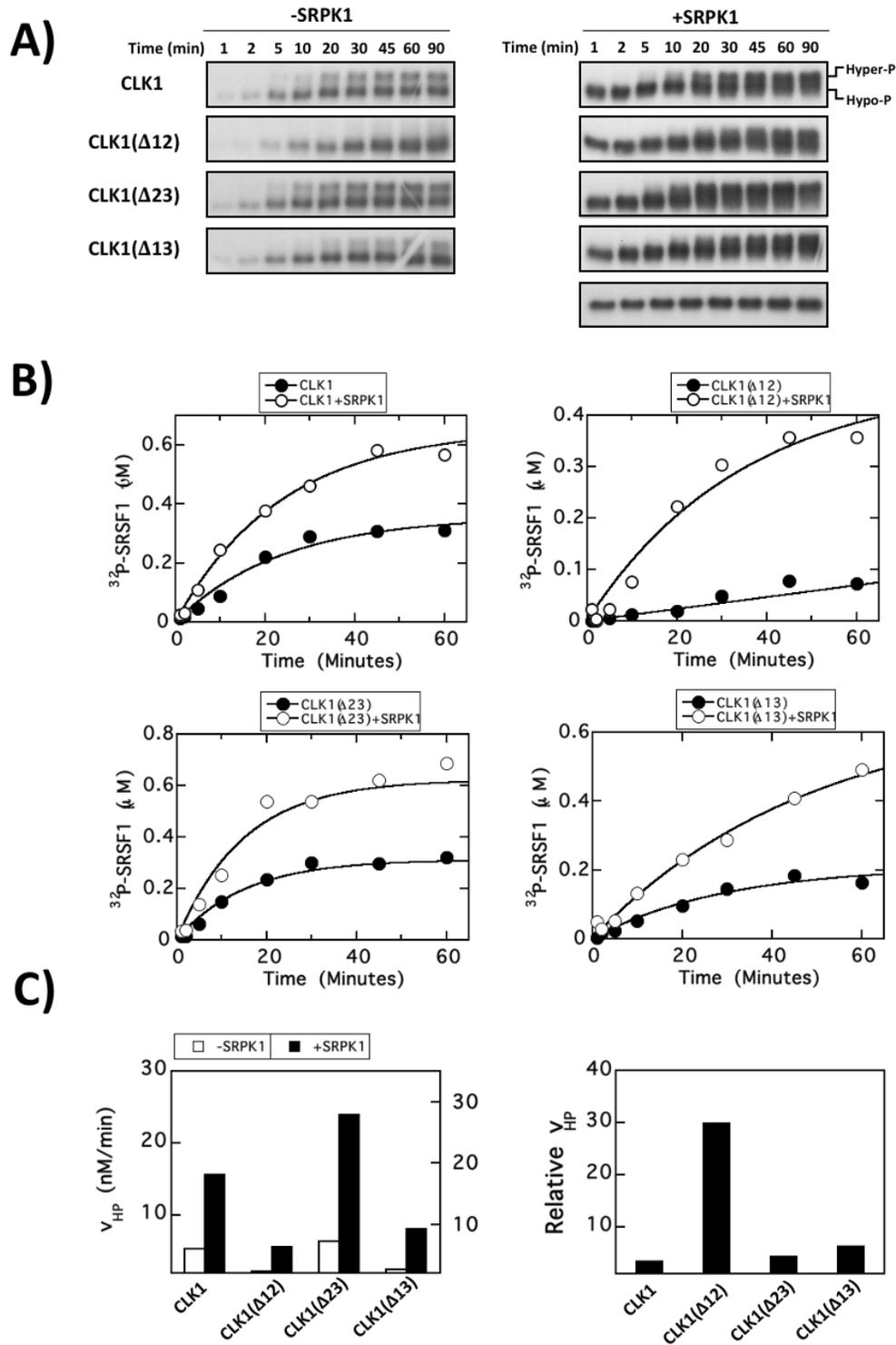


Figure 5.6. SRPK1 activates CLK1 two-block deletions (A) 32 P-Autoradiograms of single turnover assays with two-block deletions (B). Single turnover progress curves of hyperphosphorylated product of two-block deletions in the presence and absence of SRPK1. (C) Velocities from (B) after fitting to single exponential function (D) Relative enhancement in velocities observed in the presence of SRPK1.

relative velocities of the hyper-phosphorylation phases were, at least, 4-fold faster in the presence of SRPK1 for all single-block deletion constructs, with the highest enhancement observed with CLK1(Δ 1) (Fig. 5.5 C,D).

Next, we performed single-turnover kinetic assays with the two-block deletions in the presence and absence of SRPK1 (Fig. 5.6A). All two-block deletions also showed an enhanced hyper-phosphorylation rate in the presence of SRPK1, with the highest enhancement observed for CLK1(Δ 12) (Fig. 5.6B,C). When comparing the effects of both single- and double-block deletions, SRPK1 enhanced the hyper-phosphorylation rates for CLK1(Δ 1) and CLK1(Δ 12) to the greatest extent. Interestingly, these two constructs are also the weakest catalysts with respect to Ser-Pro phosphorylation in the absence of SRPK1. While SRPK1 also enhanced hyper-phosphorylation in constructs like CLK1(Δ 2) and CLK1(Δ 23) that showed strong Ser-Pro phosphorylation in the absence of SRPK1, the relative enhancement was modest. Overall, these results suggest that residues in block 1 are not only important for facilitating the hyper-phosphorylation phase catalyzed by CLK1 alone but also play a key role in SRPK1-dependent enhancement of this hyper-phosphorylation phase.

5.2.4. Low ionic strength enhances hyper-phosphorylation.

Previous studies had shown that CLK1 formed larger oligomers with increasing concentrations and that such quaternary changes correlate with higher SRSF1 phosphorylation efficiency (19). These studies suggest that oligomerization is likely to enhance CLK1 catalytic function by facilitating substrate binding and/or phosphoryl transfer rates. Since both CLK1 and SRSF1 are highly prone to aggregation, both proteins are more stable at high salt (e.g., 500 mM NaCl) for long storage. Complete removal of salt from the buffer results in precipitation of the proteins. We considered studying the impact of lowered salt concentrations on CLK1 activity. All

single-turnover kinetic assays performed so far were done in buffers containing 225 mM NaCl. To test if a lowered salt concentration would affect SRSF1 phosphorylation, we first ran single-turnover kinetic assays on CLK1 and CLK1(Δ N) with 50 mM NaCl in the assay buffer for comparison. We noticed that while lower salt did not have a significant effect on the rate of appearance of the hypo-phosphorylated SRSF1 form, it enormously enhanced both the initial velocity and the amplitude of the hyper-phosphorylation phase (Fig. 5.7 A). Overall, lower buffer salt concentration greatly increased the ability of the hypo-phosphorylated form to shift to the hyper-phosphorylated form. In comparison, lower buffer salt appeared to have no effect on CLK1(Δ N) phosphorylation of SRSF1. Such findings suggest that ionic strength significantly impacts the core catalytic function of CLK1, namely Ser-Pro dipeptide phosphorylation.

Next, we decided to investigate the impact of lower salt on the phosphorylation kinetics of the CLK1 deletions. Since all single-block deletions are oligomeric and did not show an impairment of hyper-phosphorylation in single-turnover kinetic assays, we decided to investigate the effects of lowered salt concentrations on the CLK1 two-block deletions. Among the three two-block deletions, according to results presented in Chapter 4, CLK1(Δ 12) is a monomer, but CLK1(Δ 13) and CLK1(Δ 23) are oligomers with 500 mM NaCl in the buffer. Single-turnover kinetic assays with two-block deletions were performed under high (225 mM NaCl) and low salt concentrations (50 mM NaCl). Similar to CLK1, all two-block deletions showed much faster kinetics at lower salt conditions (Fig. 5.7B). These results indicate that as long as there was a fifty-residue stretch of N-terminus attached to the kinase domain, hyper-phosphorylation was enhanced under lower salt conditions. Although CLK1(Δ 12) was found to be a monomer under high salt conditions, it is possible that CLK1(Δ 12) could form oligomers at lower salt conditions. Our efforts to lower salt concentrations below 150 mM NaCl resulted in precipitation of CLK1(Δ 12),

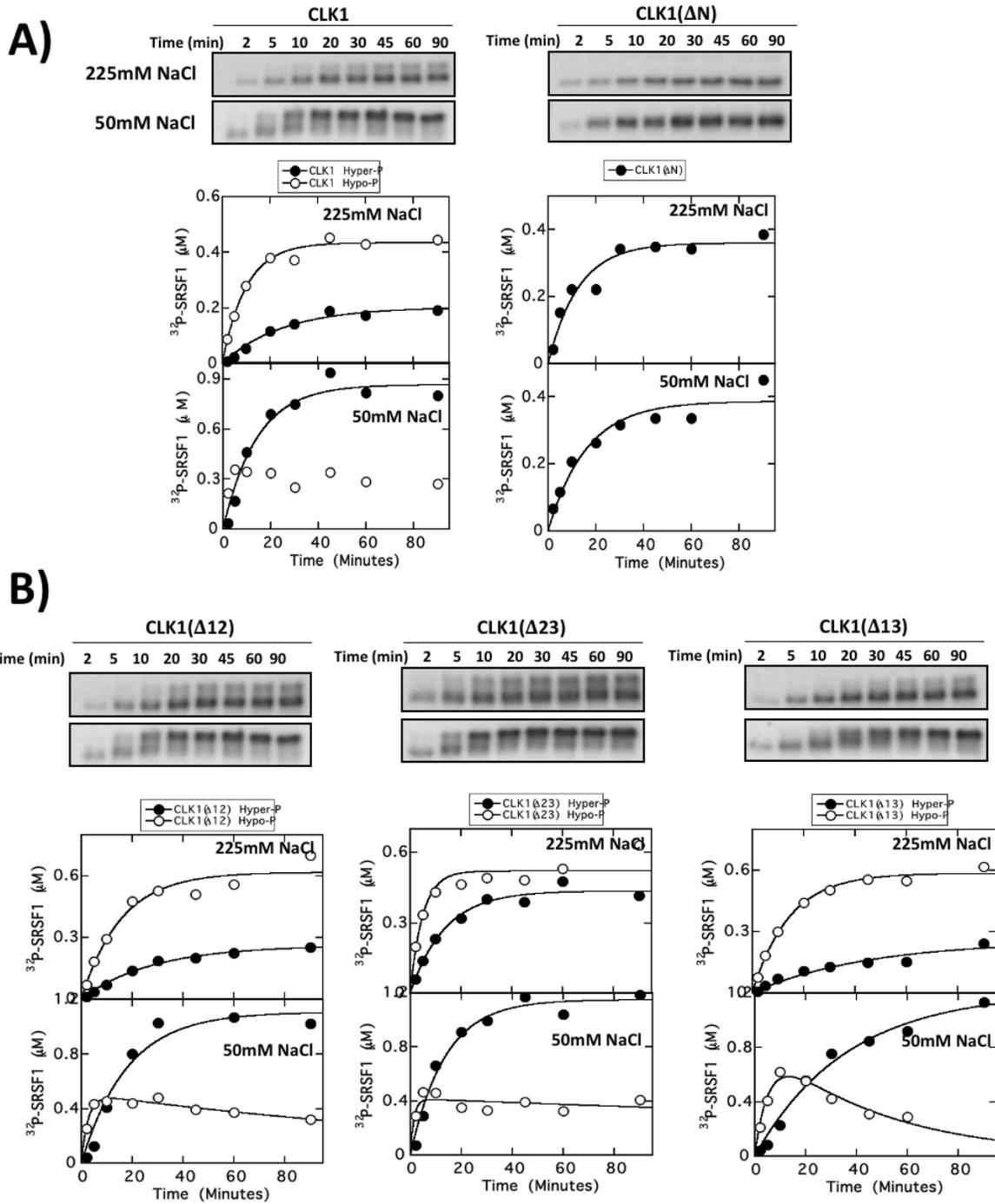


Figure 5.7. Hyper-phosphorylation is enhanced under lower salt concentrations (A) Autoradiograms and single turnover progress curves of CLK1 and CLK1(Δ N) at 225 mM and 50 mM NaCl. (B) Autoradiograms (upper panel) and single turnover progress curves (bottom panel) of CLK1 and CLK1(Δ N) at 225 mM and 50 mM NaCl concentrations.

suggesting that CLK1(Δ 12) is also aggregation prone at lower salt concentrations. Overall, results from salt dependency studies suggest that a lower ionic strength results in enhanced sequential hyper-phosphorylation.

5.2.5. Weakened oligomerization in CLK1(Y-L) impairs hyper-phosphorylation.

Since we showed that enhanced oligomerization correlates with enhanced hyper-phosphorylation, we next decided to verify if the converse is also true: Does weakened oligomerization result in weakened hyper-phosphorylation? From the studies in the previous chapter, we found that CLK1(Y-L) which replaces all tyrosines in the N-terminus with leucines to disrupt N-N contacts still forms oligomers through N-K interactions. Both TEM images and DLS experiments showed that CLK1(Y-L) formed oligomers although such structures were smaller than those for the wild-type CLK1, consistent with the idea that N-K interactions play an important role in oligomerization along with the expected N-N interactions (Fig 4.8). We next tested the impact of removing N-N interactions by studying the phosphorylation kinetics of CLK1(Y-L) in single-turnover and steady-state kinetic conditions. Interestingly, under single-turnover conditions, CLK1(Y-L) completely loses the ability to hyper-phosphorylate and does not phosphorylate any more sites than CLK1(Δ N) (Fig.5.8A). This is a significant impact considering that even two-block deletions of CLK1 with just 50-residue long N-terminus can hyper-phosphorylate, but CLK1(Y-L), with the 150-residue long N-terminus is no better than CLK1(Δ N) in phosphorylation (Fig. 5.8B,C). The SRSF1 K_m measured under steady-state conditions, however, showed that CLK1(Y-L) has comparable binding affinity as wild-type CLK1 (Fig. 5.8D,E). Taken together, these results suggest that in the absence of N-N interactions, Ser-Pro phosphorylation is completely abolished, without affecting SRSF1 binding. We next considered the possibility of Ser-Pro phosphorylation being mediated by π - π stacking interactions between

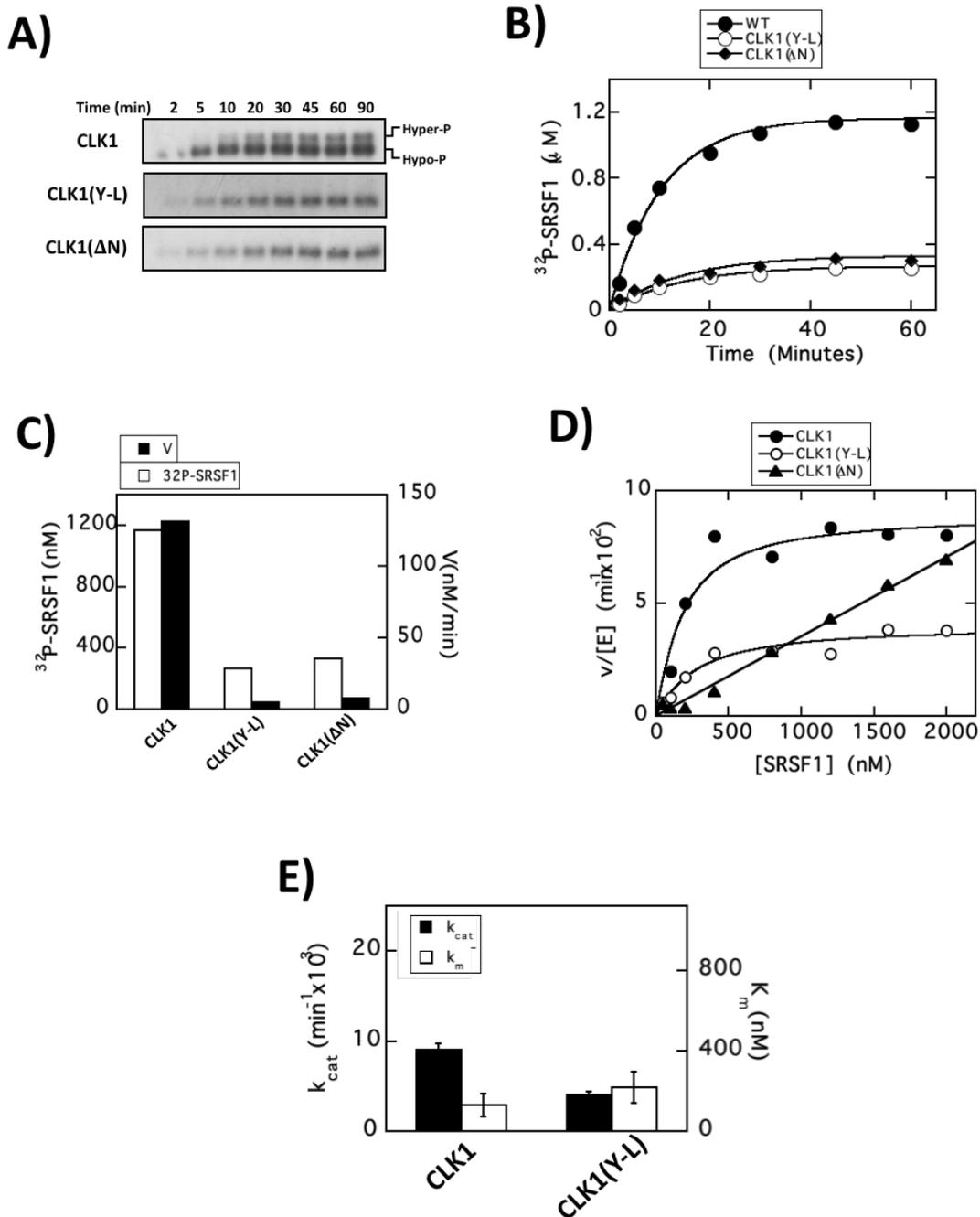
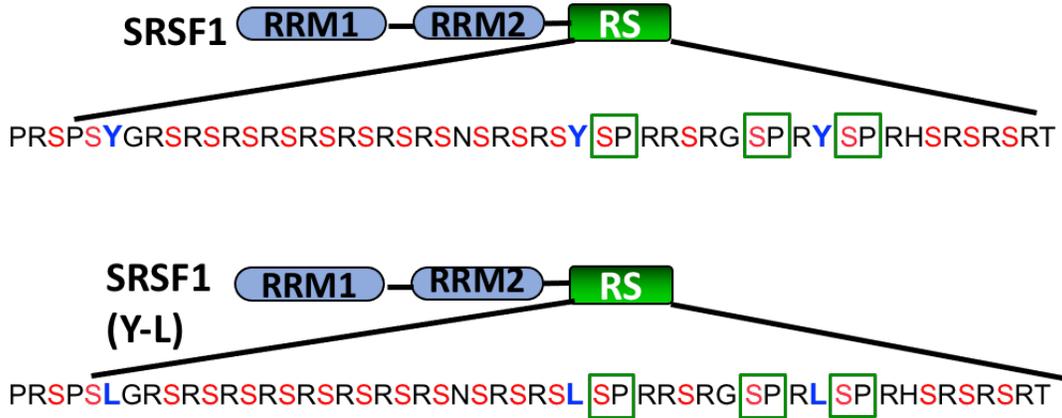


Figure 5.8. Hyper-phosphorylation is impaired in CLK1(Y-L). **(A)** ^{32}P -Autoradiograms of single turnover assays with CLK1(Y-L) **(B)** Single turnover progress curves of CLK1, CLK1(Y-L) and CLK1(ΔN) **(C)** Velocities and maximum amplitudes from **(D)** after fitting to single exponential function **(D)** Velocity vs [SRSF1] curves to determine K_m **(E)** k_{cat} and K_m for CLK1(Y-L) mutant.

A)



B)

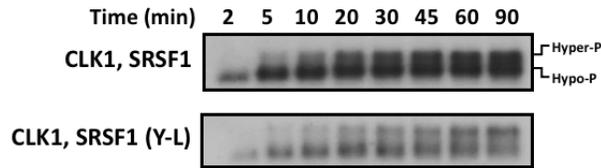


Figure 5.9. SRSF1 hyper-phosphorylation is not mediated by π - π stacking interactions between the enzyme and the substrate. **(A)** RS domain sequence of SRSF1 (Y-L) mutant. **(B)** Autoradiograms of SRSF1(Y-L) phosphorylation by CLK1.

SRSF1 RS domain and CLK1 N-terminus. The RS domain in SRSF1 has 3 tyrosine residues, with two of them placed right before the Ser-Pro dipeptides. We mutated the YSP to LSP in the SRSF1(Y-L) mutant and tested whether CLK1 could induce a gel-shift in the autoradiograms. Interestingly, CLK1 induced hyper-phosphorylation in SRSF1(Y-L) based on the observed gel shift (Fig.5.9B). Such findings suggest that the ability of CLK1 to induce Ser-Pro phosphorylation in SRSF1 is not likely to be due to π - π stacking interactions between the CLK1 N-terminus and the SRSF1 RS domain.

5.3. Discussion

Overall, the results presented here suggest that there is a strong correlation between the quaternary structure of CLK1 and hyper-phosphorylation. Enhancing the oligomerization of CLK1 by bringing more CLK1 units together through the addition of SRPK1 or by lowering the ionic strength of the reaction mixture resulted in an enhanced hyper-phosphorylation. Weakening the oligomerization by removal of aromatic residues on the N-terminus results in complete loss of hyper-phosphorylation activity. This correlation, in fact, helps in explaining how oligomeric structure provides CLK1 with an optimal catalytic efficiency in the cellular context. Since CLK1 and its substrate SRSF1 are found in nuclear speckles which are biomolecular condensates, the high concentration of CLK1 in speckles allow for tighter binding, oligomerization, and faster hyper-phosphorylation.

Unlike hyper-phosphorylation, SRSF1 binding is solely dependent on the length of the N-terminus. The SRSF1 K_m for all single block deletions is comparable to that of CLK1. The K_m for the two-block deletion constructs, CLK1(Δ 12) and CLK1(Δ 13), on the other hand, was five-fold higher than that of CLK1, implying weakened binding. CLK1(Y-L) mutant also showed a similar K_m to that of CLK1, implying that tyrosines are not necessary for SRSF1 binding. Previous pull-

down assays have shown that SRSF1 binding affinity decreases progressively with decrease in length of the N-terminus (22), also suggesting that the SRSF1 binding is dependent on the length of the N-terminus rather than being mediated by sequence-specific interactions. Since the nuclear import of CLK1 is mediated by piggybacking with SRSF1, a direct correlation between the length of the N-terminus and nuclear localization was also observed with the nuclear import of these deletion constructs in chapter 3. Two-block deletions with a five-fold higher K_m expressed in the nucleus and in the cytoplasm, whereas all single block deletion constructs with unimpaired SRSF1 K_m expressed exclusively in the nucleus. Thus, while tyrosines are important for hyper-phosphorylation, interactions between the N-terminus and RS domain are vital for CLK1 nuclear localization.

Among all the deletion constructs used in this study, the hyper-phosphorylation kinetics of CLK1($\Delta 2$) is remarkably faster. CLK1($\Delta 2$) could also have an altered quaternary structure where the kinase domains are pulled closer together due to the N-K interactions from block 1 on an N-terminus that is about 50 residues shorter than CLK1. While more structural studies have to be done to explain why CLK1($\Delta 2$) is a more efficient kinase, based on our understanding so far, block 2 is most likely inhibitive towards hyper-phosphorylation. This makes the question of the necessity of block 2 very intriguing.

Hyper-phosphorylation at Ser-Pro dipeptides in SR proteins is a feature unique to CLK1. The functional consequences of Ser-Pro phosphorylation vary immensely and ranges from dissociation of SRSF1 from speckles to enhanced RNA binding. As discussed in this chapter, the quaternary structure of CLK1 has a heavy impact on SRSF1 hyper-phosphorylation kinetics. The correlation between CLK1 quaternary structure and hyper-phosphorylation kinetics is shown in Fig. 5.10. CLK1(Y-L) mutant, with minimal N-N interactions, completely lost its ability to

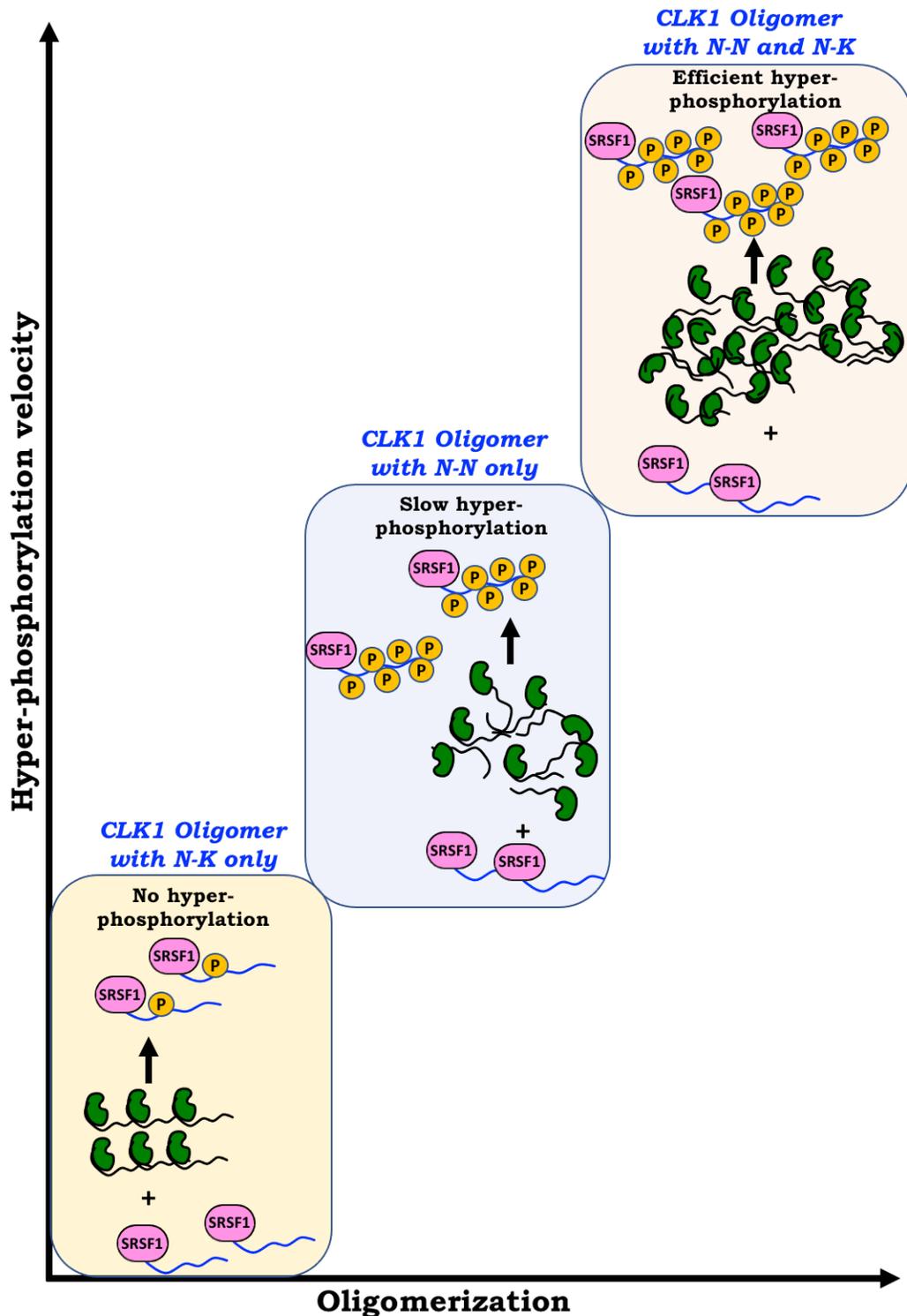


Figure 5.10. Summary of the correlation between quaternary structure and phosphorylation kinetics of CLK1. N-N and N-K interactions lead are important for efficient oligomerization and activity of CLK1. The absence of N-N interaction completely abolishes hyper-phosphorylation while the absence of N-K interactions slows down the velocity of hyper-phosphorylation.

hyper-phosphorylate SRSF1, suggesting that the tyrosines on the N-terminus are essential for Ser-Pro phosphorylation. Since the SRSF1 K_m for CLK1(Y-L) is not very different from that of CLK1, substrate binding is mostly unaffected. Since CLK1(Y-L) forms oligomers that are solely mediated by N-K interactions, we know that such an oligomer is rendered ineffective in hyper-phosphorylation. Next, we considered the impact of N-K interactions on hyper-phosphorylation of CLK1. Pulldown assays from chapter 3 show that block 1 and block 3 bind the kinase domain. However, the interaction of block 3 with the kinase domain is most likely an intrasteric interaction because CLK1(Δ 12) is a monomer. This implies that the majority of the N-K interactions happen between the stretch of K/R residues on block 1 and the kinase domain (Fig. 4.8). CLK1(Δ 1) is, therefore, a construct that lacks strong N-K interactions, and hence, the functional role of N-K interactions should be impaired in the phosphorylation kinetics of CLK1(Δ 1). Although the total number of sites phosphorylated was unaffected, we found that the velocity of hyper-phosphorylation was slightly compromised in CLK1(Δ 1) under single turnover conditions. When both N-N and N-K interactions are present, CLK1 oligomerization is the most efficient, which inevitably results in faster hyper-phosphorylation kinetics.

Based on our understanding of the CLK1 N-terminus and its role in regulating SRSF1 hyper-phosphorylation gathered so far, an interesting question that arises here is the need for having a 150 residue-long N-terminus when hyper-phosphorylation was not impaired even in two-block deletions. Our studies in this chapter were confined to SRSF1 hyper-phosphorylation. In a physiological context, the substrates of CLK1 and its binding partners are not confined to SRSF1 alone. For example, recent studies from our lab have shown that CLK1 phosphorylates U1-70K. N-terminus has also been shown to be important for hyper-phosphorylation of several other SR

proteins and other proteins with RS-like domains like Tra2 β (18). Hence, the need for having a long N-terminus maybe in regulating the phosphorylation of other physiological substrates.

5.4. Conclusions

Studies presented in this chapter suggest that there is a strong correlation between the interactions forming the quaternary structure of CLK1 and hyper-phosphorylation kinetics. An efficient oligomer formed from both N-N and N-K interactions are necessary for fast and efficient hyper-phosphorylation. Enhancing the self-association of CLK1 by the addition of SRPK1 or by lowering the ionic strength of the solution enhances the hyper-phosphorylation of CLK1. Lowering the affinity of CLK1 to self-associate by removing tyrosines results in complete loss of hyper-phosphorylation activity.

Chapter 5 is currently being prepared for submission for publication of the material: George A, Aubol BE, Adams JA. The dissertation author was a primary investigator and author of this paper.

References

1. Prasad J, Colwill K, Pawson T, Manley JL. The Protein Kinase Clk/Sty Directly Modulates SR Protein Activity: Both Hyper- and Hypophosphorylation Inhibit Splicing. *Mol Cell Biol.* 1999;19(10):6991–7000.
2. Keshwani MM, Aubol BE, Fattet L, Ma CT, Qiu J, Jennings PA, Fu XD, Adams JA . Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochem J.* 2015;466:311–22.
3. Ngo JCK, Gullingsrud J, Giang K, Yeh MJ, Fu XD, Adams JA, McCammon JA, Ghosh G. SR Protein Kinase 1 Is Resilient to Inactivation. *Structure.* 2007;15(1):123–33.
4. Bullock AN, Das S, Debreczeni JÉ, Rellos P, Fedorov O, Niesen FH, Guo K, Papagrigoriou E, Amos AL, Cho S, Turk BE, Ghosh G, Knapp S. Kinase Domain Insertions Define Distinct Roles of CLK Kinases in SR Protein Phosphorylation. *Structure.* 2009;17(3):352–62.
5. Colwill K, Feng LL, Yeakley JM, Gish GD, Ca JF, Pawson T, Fu XD. SRPK1 and Clk / Sty Protein Kinases Show Distinct Substrate Specificities for Serine / Arginine-rich Splicing Factors *. 1996;271(40):24569–75.
6. Aubol BE, Plocinik RM, Hagopian JC, Ma CT, McGlone ML, Bandyopadhyay R, Fu XD, Adams JA. Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *J Mol Biol.* 2013;425(16):2894–909.
7. Ma CT, Velazquez-Dones A, Hagopian JC, Ghosh G, Fu XD, Adams JA. Ordered Multi-site Phosphorylation of the Splicing Factor ASF/SF2 By SRPK1. *J Mol Biol.* 2008;376(1):55–68.
8. Velazquez-Dones A, Hagopian JC, Ma CT, Zhong XY, Zhou H, Ghosh G, Fu XD, Adams JA. Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *J Biol Chem.* 2005;280(50):41761–8.
9. Cho S, Hoang A, Sinha R, Zhong XY, Fu XD, Krainer AR, et al. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A.* 2011;108(20):8233–8.
10. Aubol BE, Chakrabarti S, Ngo J, Shaffer J, Nolen B, Fu XD, et al. Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc Natl Acad Sci U S A.* 2003;100(22):12601–6.
11. Ma CT, Ghosh G, Fu XD, Adams JA. Mechanism of Dephosphorylation of the SR Protein ASF/SF2 by Protein Phosphatase 1. *J Mol Biol.* 2010;403(3):386–404.
12. Aubol BE, Adams JA. Applying the brakes to multisite SR protein phosphorylation:

- Substrate-induced effects on the splicing kinase SRPK1. *Biochemistry*. 2011;50(32):6888–900.
13. Aubol BE, Plocinik RM, McGlone ML, Adams JA. Nucleotide release sequences in the protein kinase SRPK1 accelerate substrate phosphorylation. *Biochemistry*. 2012;51(33):6584–94.
 14. Ngo JCK, Chakrabarti S, Ding JH, Velazquez-Dones A, Nolen B, Aubol BE, Adams JA, Fu XD, Ghosh G. 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*. 2005;20(1):77–89.
 15. Ngo JCK, Giang K, Chakrabarti S, Ma CT, Huynh N, Hagopian JC, Dorrestein PC, Fu XD, Adams JA, Ghosh G. A Sliding Docking Interaction Is Essential for Sequential and Processive Phosphorylation of an SR Protein by SRPK1. *Mol Cell*. 2008;29(5):563–76.
 16. Ghosh G, Adams JA. Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J*. 2011;278(4):587–97.
 17. Aubol BE, Wu G, Keshwani MM, Movassat M, Fattet L, Hertel KJ, Fu XD, Adams JA. Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol Cell*. 2016;63(2):218–28.
 18. Aubol BE, Plocinik RM, Keshwani MM, Mcglone ML, Hagopian JC, Ghosh G, Fu XD, Adams JA. N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. *Biochem J*. 2014;462(1):143–52.
 19. Keshwani MM, Hailey KL, Aubol BE, Fattet L, McGlone ML, Jennings PA, Adams JA. Nuclear protein kinase CLK1 uses a non-Traditional docking mechanism to select physiological substrates. *Biochem J*. 2015;472(3):329–38.
 20. Aubol BE, Fattet L, Adams JA. A conserved sequence motif bridges two protein kinases for enhanced phosphorylation and nuclear function of a splicing factor. *FEBS J*. 2020;1:1–16.
 21. Aubol BE, Keshwani MM, Fattet L, Adams JA. Mobilization of a splicing factor through a nuclear kinase–kinase complex. *Biochem J*. 2018;475(3):677–90.
 22. George A, Aubol BE, Fattet L, Adams JA. Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate. *J Biol Chem*. 2019;294(24):9631–41.

Chapter 6

Conclusions

6.1 Conclusions

Our studies illustrate how the disordered N-terminus regulates various functions of CLK1. Although CLK1 was always known to be a strictly nuclear kinase, the pertinent question of how the kinase enters the nucleus had long gone unanswered. Although early studies had speculated that the nuclear import of CLK/STY, the mouse form of CLK1, was mediated by a classical NLS on its N-terminus (1), we found that the nuclear import of CLK1 was not mediated by such a classical system. Instead, TRN-SR2, a karyopherin known for transporting SR proteins into the nucleus, is also responsible for bringing CLK1 into the nucleus (2). However, we found that TRN-SR2 cannot bind to CLK1 directly but instead interacts indirectly with CLK1 through its phosphorylated SR protein substrate, SRSF1. We found that the N-terminus of CLK1 makes high-affinity interactions with the RS domain of SRSF1 facilitating a piggyback mechanism where CLK1 transits from the cytoplasm to the nucleus attached to its physiological substrate which contains its own NLS, the phospho-RS domain. Although nuclear import by piggybacking has been reported previously in some rare cases (3), CLK1 nuclear import is unique since the binding domains responsible for this phenomenon are all intrinsically disordered.

Previous findings from our laboratory had shown that CLK1 undergoes self-association mediated by its N-terminus and that this self-association serves as a substrate selection mechanism allowing CLK1 to discriminate between physiological and non-physiological substrates (4). However, investigating the structural aspects of oligomerization was a daunting task owing to the aggregation-prone nature of the protein and its lack of residue conservation. Through extensive mutagenesis studies, we deciphered that CLK1 oligomerization is not only mediated by N-N interactions, but also N-K interactions. The N-N interactions are mediated by tyrosine residues in the N-terminus and the N-K interactions are mediated by electrostatic interactions between the N-

terminus and the kinase domain. While aromatic residues in intrinsically disordered regions have been shown previously to drive protein self-association (5,6), it had been suspected that the RS-like nature of the CLK1 N-terminus would be the major driver of oligomerization, similar to RS domain-driven protein-protein interactions within the SR protein family. In contrast to these prevailing ideas, we found that tyrosines rather than arginine-serine dipeptides facilitate N-N contacts in CLK1. Most importantly, we found that there is a direct correlation between the quaternary structure of the CLK1 oligomer and the phosphorylation kinetics. We found that oligomerization was necessary for fast and efficient serine-proline dipeptide hyper-phosphorylation. Strengthening the oligomerization by lowering the ionic strength of the solution or adding SRPK1 results in an enhanced hyper-phosphorylation. Weakening the oligomerization by removing N-N interactions completely abolishes hyper-phosphorylation. Hopefully, knowledge of how the N-terminus mediates this plethora of CLK functions can be used in the future for manipulating CLK function in desirable ways.

6.2 Future directions

Our findings on how the disordered N-terminus modulates the binding and hyper-phosphorylation of SRSF1 may be of immense therapeutic potential in the future. Inhibition of CLK1 can open doors to the treatment of several spliceopathies and cancers. In fact, a high-throughput screen of small molecules to correct the cryptic splicing of lamin A/C (LMNA) gene causing the Hutchinson-Gilford progeria syndrome (HGPS) led to the discovery of new CLK4 inhibitors (7). While most kinase inhibitors target the ATP-binding pocket on the kinase domains, enhancing inhibitor specificity and selectivity has always been a challenging task owing to the structural similarity in the ATP binding pockets of canonical kinases. Allosteric inhibition by designing drugs that bind at sites away from the substrate binding pockets of kinases has attracted

much attention in the past decade. Most CLK inhibitors known today also target the ATP binding pocket. However, as we deciphered the nature of interactions driving self-association of CLK1 that activates the kinase towards hyper-phosphorylation of SR proteins, new strategies targeting the N-terminus for CLK1 inhibition can be envisioned. Although initial ideas on drugs targeting intrinsically disordered domains were received with skepticism, recent studies have had successes in designing drugs that bind unstructured regions. For example, in a recent study, a small molecule drug called trifluoperazine dihydrochloride was identified to bind the intrinsically disordered protein nuclear protein 1 (NUPR1) known to be involved in causing pancreatic ductal adenocarcinoma (PDAC) (8). Injection of this molecule was shown to completely arrest tumor development on xenografted PDAC derived cells on mice (7). A remarkably high proportion of proteins regulating diseases like cancers, neurodegenerative diseases and viral infections are intrinsically disordered, making ‘drugging the disorderome’ an enticing, but challenging idea (9). Small molecule CLK1 inhibitors, including the conventional CLK1 inhibitor TG003, have been shown to be effective in correcting the cryptic exon inclusion in the dystrophin gene that generates dysfunctional dystrophin protein causing duchenne muscular dystrophy (DMD) (10,11). Since oligomerization mediated by the disordered N-terminus is an activation and substrate selection mechanism for CLK1, inhibiting N-N interactions may be a new strategy for regulating hyper-phosphorylation of SR proteins. Thus, the findings presented in this dissertation may open doors to new strategies for CLK1 inhibition.

References

1. Duncan PI, Howell BW, Marius RM, Drmanic S, Douville EMJ, Bell JC. Alternative splicing of STY, a nuclear dual specificity kinase. *J Biol Chem.* 1995;270(37):21524–31.
2. George A, Aubol BE, Fattet L, Adams JA. Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate. *J Biol Chem.* 2019;294(24):9631–41.
3. Tessier TM, Macneil KM, Mymryk JS. Piggybacking on classical import and other non-classical mechanisms of nuclear import appear highly prevalent within the human proteome. *Biology (Basel).* 2020;9(8):1–20.
4. Keshwani MM, Hailey KL, Aubol BE, Fattet L, McGlone ML, Jennings PA, Adams JA. Nuclear protein kinase CLK1 uses a non-Traditional docking mechanism to select physiological substrates. *Biochem J.* 2015;472(3):329–38.
5. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science (80-).* 2020;367(6478):694–9.
6. Song J, Ng SC, Tompa P, Lee KAW, Chan HS. Polycation- π Interactions Are a Driving Force for Molecular Recognition by an Intrinsically Disordered Oncoprotein Family. *PLoS Comput Biol.* 2013;9(9).
7. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med.* 2012;18(8):472–82.
8. Neira JL, Bintz J, Arruebo M, Rizzuti B, Bonacci T, Vega S, Lanas A, CampoyAV, Iovanna JL, Abian O. Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. *Sci Rep.* 2017;7(January):1–15.
9. Fuertes G, Nevola L, Esteban-Martín S. Perspectives on drug discovery strategies based on IDPs. *Intrinsically Disordered Proteins.* Elsevier Inc.; 2019. 275–327 p.
10. Sako Y, Ninomiya K, Okuno Y, Toyomoto M, Nishida A, Koike Y, Ohe K, Kii Isao, Yoshida S, Hashimoto N, Hosoya T, Matsuo M, Hagiwara M. Development of an orally available inhibitor of CLK1 for skipping a mutated dystrophin exon in Duchenne muscular dystrophy. *Sci Rep.* 2017;7(April 2016):6–7.
11. Nishida A, Kataoka N, Takeshima Y, Yagi M, Awano H, Ota M, et al. Chemical treatment enhances skipping of a mutated exon in the dystrophin gene. *Nat Commun.* 2011;2(1).