Prp40: a Novel Centrin-binding Protein

by

Adalberto Díaz Casas

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
APPLIED CHEMISTRY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2018

Approved by:

_____	_____
Enrique Meléndez, Ph.D. Member, Graduate Committee	Date
_____	_____
Robert Ríos, Ph.D. Member, Graduate Committee	Date
_____	_____
Jorge Ríos, Ph.D. Member, Graduate Committee	Date
_____	_____
Belinda Pastrana, Ph.D. President, Graduate Committee	Date
_____	_____
María Plaza, Ph.D. Representative of Graduate Studies	Date
_____	_____
Enrique Meléndez, Ph.D. Chairperson of the Department	Date

ABSTRACT

Pre-mRNA processing protein 40 (Prp40) is a modular protein that has an essential role in the initiation step of pre-mRNA splicing. In *Homo sapiens*, Prp40 has two putative homologs: Prp40 homolog A (*Hs*Prp40A) and homolog B (*Hs*Prp40B). This splicing factor comprises two WW domains and six tandem FF domains. In addition, Prp40 has two leucine-rich nuclear export signals, but little is known about the function of Prp40 in the export process. Centrin, a calcium-binding protein that is known to have a role in mRNA and protein export, is one of the 350 "eukaryotic signature proteins." We identified a centrin-binding site that is located within the third FF (FF_3) domain of *Hs*Prp40A. This centrin-binding site is highly similar to the first nuclear export signal consensus sequence identified in yeast Prp40. We confirmed the interaction between a *Hs*Prp40A synthetic peptide (*Hs*Prp40Ap) and centrin, using isothermal titration calorimetry and two-dimensional infrared (2D IR) correlation spectroscopy. In addition, spectroscopic and thermodynamic studies carried out on *Hs*Prp40A's full FF_3 domain provided new insights about the stability of this domain. Finally, we proceeded to study the molecular behavior of *Hs*Prp40A's FF_3 domain under thermal stress, using differential scanning calorimetry, circular dichroism, and 2D IR correlation spectroscopies. The comprehensive body of evidence supports a novel centrin target with potential regulatory functions within the nucleus, the scientific implications being that the centrin-Prp40A complex has a functional role in pre-mRNA splicing.

RESUMEN

Pre-mRNA processing protein 40 (Prp40) es una proteína modular que tiene un rol esencial en el paso de iniciación de corte y empalme (*splicing*) del ARNpre-m. En *Homo sapiens*, Prp40 tiene dos homólogos putativos: Prp40 homólogo A (*Hs*Prp40A) y homólogo B (*Hs*Prp40B). Este factor de *splicing* está compuesto por dos dominios WW y seis dominios FF tándem. Además, Prp40 tiene dos señales de exportación nuclear, pero muy poco se conoce sobre la función de Prp40 en el proceso de exportación. Centrin, una proteína que enlaza calcio que es conocida por tener un rol en la exportación de ARNm y proteína, es una de las 350 "proteínas de la firma eucariota". Identificamos un sitio de enlazar centrin que está localizado en el tercer dominio FF (FF_3) de *Hs*Prp40A. Este sitio de enlazar centrin es altamente similar a la secuencia de consenso de la primera señal de exportación nuclear en Prp40 de levadura. Confirmamos la interacción entre un péptido sintético de *Hs*Prp40A (*Hs*Prp40Ap) y centrin, utilizando calorimetría de titulación isotérmica y espectroscopia de correlación infrarroja de dos dimensiones. Además, los estudios espectroscópicos y termodinámicos llevados a cabo en el dominio FF_3 completo de *Hs*Prp40A proveyeron nuevos conocimientos sobre la estabilidad de este dominio. Finalmente, procedimos a estudiar el comportamiento molecular del dominio FF_3 de *Hs*Prp40A bajo estrés térmico, utilizando calorimetría diferencial de barrido, dicroísmo circular y espectroscopia de correlación infrarroja de dos dimensiones. El amplio conjunto de pruebas apoya a *Hs*Prp40A como una proteína objetivo de centrin novedosa con un rol con posibles funciones reguladoras dentro del núcleo. Las implicaciones científicas es que el complejo centrin-Prp40A tiene un rol funcional en el *splicing* del ARNpre-m.

DEDICATION

To my **Lord Jesus Christ**, "who is **God** over all, blessed forever. Amen." (Romans 9:5, Holy Bible).

To **Roselyne**, my beloved wife, for your love and support during my doctoral studies. To my mother, **Leila Margarita**, for her love, care, and prayers through all these years. To my sister and brother-in-law, **Jomary and Eric**, for loving me and supporting me in an unconditional way.

To my **brothers and sisters in Christ**, my spiritual family, for their prayers, their care, shepherding, and fellowship through all these years.

ACKNOWLEDGMENT

The following persons were instrumental in the completion of my Ph.D. degree. Firstly, I would like to express my gratitude to my committee chair, Dr. Belinda Pastrana-Ríos, for her mentoring and guidance in my dissertation project. I admire her passion for her research and her gift for teaching. I would also like to thank her for providing me an opportunity to be part of her research group and carry out my Ph.D. research project. I am also grateful to the rest of my graduate committee: Dr. Enrique Meléndez, Dr. Robert Ríos, and Dr. Jorge Ríos for the input that I receive from them in every Progress Report.

My sincere thanks go to Dr. Walter J. Chazin of Vanderbilt University in Nashville, TN, for his collaboration in the work presented in Díaz Casas et al. 2017. I also thank Dr. Chazin for providing me an opportunity to join his lab as an intern to complete my research project, as a support after the event of hurricane Maria in Puerto Rico. I am also deeply grateful to Dr. Rémy Le Meur of Vanderbilt University for his tremendous assistance and recommendations regarding several experiments during my time at Vanderbilt University; and to Kateryna Ogorodnik of Vanderbilt University for the mass spectrometry analysis of *Hs*Prp40A's FF_3 domain.

In addition, my special thanks go to Dr. William J. Bauer of the BioXFEL program, who allowed me to carry out my 4-month doctorate practicum at the Hauptman-Woodward Institute (Buffalo, NY), and who trained me and provided me guidance on protein

crystallography and X-ray diffraction. I also thank Dr. Edward Snell for giving me access to his laboratory facilities at the Hauptman-Woodward Institute during my doctorate practicum.

I would also like to acknowledge Michael Berne of Tufs University at Medford, MA, for the partial amino acid sequencing analysis of *Hs*Prp40A's FF_3 domain. I also thank Dr. Enrique Melendez and Dr. Juan López-Garriga, for kindly allowing me to use the Jasco J-810 spectropolarimeter and the digital microscope, respectively.

I would like to express special gratitude to my dear friend and colleague, José A. Carmona, for his strong support and assistance throughout all these years, and for accepting the unique opportunity to develop a research project together as part of our graduate courses. I also thank Aslin Rodríguez, Bianca Alamo, Tatiana Garcés, and Adriana Oliveras for helping me survive all the stress from these years and not letting me give up. I also thank all the students who were part of Dr. Pastrana's research group during my graduate studies for their help and collaboration in this research project, especially in the expression and purification of centrins and *Hs*Prp40A's FF_3 domain.

I gratefully acknowledge the following funding sources that made my Ph.D. work possible: the National Institutes of Health Grant SO6GM08103 (B.P.-R.), a Henry Dreyfus Teacher Scholar Award (B.P.-R.), the Alfred P. Sloan Foundation Scholarship Award (A.D.C), the National Institutes of Health-Research Initiative for Scientific Enhancement Scholarship (A.D.C), the National Science Foundation BioXFEL STC Award 1231306 (A.D.C), the

National Institute of Medical Sciences Supplement 5R35GM118089-02 (W.J.C.), and the University of Puerto Rico.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS ABBREVIATIONS

ν_a(N-D)	Guanidinium anti-symmetric stretches
ν_s(N-D)	Guanidinium symmetric stretches
^{13}C-*Cr*cen	^{13}C-labeled *Chlamydomonas reinhardtii* centrin
2D IR	Two-dimensional infrared
At	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
CaM	Calmodulin
CBS	Centrin-binding site
CD	Circular dichroism
Cdc31	*Saccharomyces cerevisiae* centrin homolog
Cr	*Chlamydomonas reinhardtii*
*Cr*cen	*Chlamydomonas reinhardtii* centrin
crn-TRP	Crooked neck-like tetratricopeptide repeat
cryoEM	cryo-electron microscopy
DEAE	Diethylaminoethyl
Dr	*Danio renio*
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
FF_{1-6} domains	First to sixth FF domains

GG-NER	Global genome nucleotide excision repair
H→D exchanged	Hydrogen to deuterium exchanged
HD	Huntington disease
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	High performance liquid chromatography
Hs	*Homo sapiens*
*Hs*cen1-4	*Homo sapiens* centrin1-4
*Hs*Prp40Ap	*Homo sapiens* Prp40 homolog A sythetic peptide
Htt	Huntingtin
IGEPAL	Octylphenoxypolyethoxyethanol
ITC	Isothermal titration calorimetry
K_a	Association constant
K_d	Dissociation constant
MLT	Mellitin
Mm	*Mus musculus*
MS	Mass spectrometry
MTOC	Microtubule organizing center
n	Binding stoichiometry
NES	Nuclear export signal
NMR	Nuclear magnetic resonance
^{Phos}CTD-RNAPII	Phosphorylated C-terminal domain of RNA polymerase II
Pre-mRNA	Precursor messenger RNA

Prp40	Pre-mRNA processing protein 40
Prp40A	Prp40 homolog A
Prp40B	Prp40 homolog B
Rad23B	Rad23 homolog B
RNAPII	RNA polymerase II
Sc	*Saccharomyces cerevisiae*
*Sc*Prp40	*Saccharomyces cerevisiae* Prp40
snRNP	small nuclear ribonucleoprotein
TCEP HCl	tris(2-carboxyethyl)phosphine hydrochloride
T_m	Thermal transition temperature
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
$v(^{13}C{=}O)$	^{13}C-labeled carbonyl stretching vibrations
$v(^{13}COO^-)$	^{13}C-labeled carboxylate vibrational modes
$v(C{=}O)$	Carbonyl stretching vibrations
$v(COO^-)$	Carboxylate vibrational modes
$WW_{1\text{-}2}$ domains	First and second WW domains
Xl	*Xenopus laevis*
XPC	Xeroderma pigmentosum group C
ΔC_p	change in heat capacity at constant pressure
ΔG_B	Change in Gibbs free energy of binding
ΔH_B	Change in enthalpy of binding
ΔH_{calc}	Calculated change in enthalpy

ΔH_D	Change in enthalpy of denaturation
ΔS_B	Change in entropy of binding

CHAPTER I

JUSTIFICATION

Eukaryotic cells are highly compartmentalized, which provides different environmental conditions necessary for the optimal function of each metabolic pathways. Within the nucleus, all the biological networks and pathways are regulated in a pH of ~7.2 (Figure 1) [Casey et. al. 2010]. Among these biological processes, gene expression consists of many steps that begin when a particular segment of DNA is copied into RNA by RNA polymerase II (RNAPII), a process called transcription. The next step is RNA processing, in which a precursor messenger RNA (pre-mRNA) undergoes a series of modifications to become a mature mRNA. One of the most important modifications is pre-mRNA splicing, in which the introns are removed from the primary transcript and exons are joined. This process is carried out by the spliceosome, a megadalton multi-protein complex composed of U1, U2, U4, U5, and U6 small nuclear ribonucleoproteins (snRNPs) and several other proteins [Wahl et al. 2009]. Recently, the structure of the *Schizosaccharomyces pombe* spliceosome at 3.6 Å resolution was determined using cryo-electron microscopy (cryoEM) [Yan et al. 2015]. Although the event of pre-mRNA splicing has been extensively studied over the past years, the mechanisms that governs the early steps of pre-mRNA splicing remain poorly understood.

One of the proteins that is associated with the U1 snRNP and is known to be essential for the first step of pre-mRNA splicing is pre-mRNA processing protein 40 (Prp40), first

identified in *Saccharomyces cerevisiae* [Kao and Siliciano, 1996]. Prp40 is a modular splicing factor comprising of two WW domains and six tandem FF domains, which mediate the interaction between Prp40 and its targets. In higher eukaryotes, Prp40 has two putative homologs: Prp40 homolog A (Prp40A), also known as huntingtin-interacting protein A (HYPA) and formin-binding protein 11 (FBP11); and homolog B (Prp40B), also known as huntingtin-interacting protein C (HYPC) [Chan et al. 1996; Faber et al. 1998; Passani et al. 2000]. A phylogenetic analysis carried out by Becerra and colleagues show that Prp40A is more closely related to *Saccharomyces cerevisiae* Prp40 (*Sc*Prp40) than Prp40B [Becerra et al. 2016]. In addition to the WW and FF domains, *Sc*Prp40 possesses two leucine-rich nuclear export signals (NESs), which are required for proper pre-mRNA splicing [Murphy et al. 2004]. However, little is known about the process of Prp40 export.

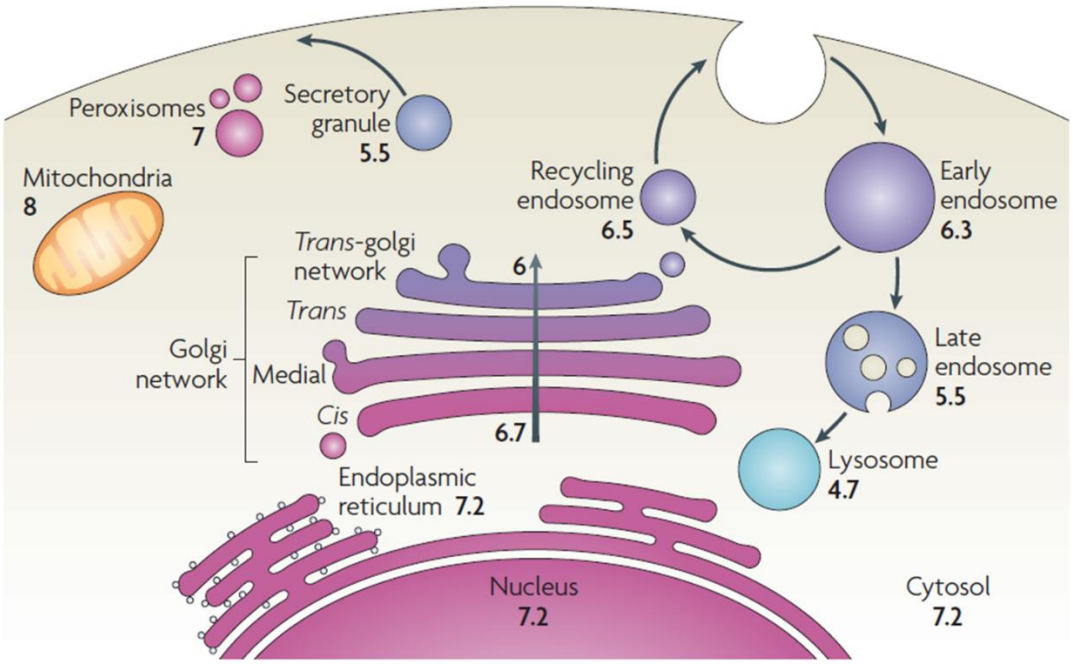


Figure 1. Representation of a eukaryotic cell with the pH of the different compartments. The nucleus is regulated with a pH of ~7.2 [Adapted from Casey et al. 2010 with permission from publisher Springer Nature].

Another protein that is known to have a role in mRNA processing and protein export is centrin [Resendes et al. 2008]. Centrin is a calcium-binding protein that is one of 350 "eukaryotic signature proteins" [Hartman and Fedorov, 2002]. In higher eukaryotes, such as *Homo sapiens*, four isoforms of centrin have been identified: centrin 1 to centrin 4 (*Hs*cen1 to *Hs*cen4). *Hs*cen1 is localized in the sperm, neurons, and highly ciliated cells [Hart et al 1999]. *Hs*cen2 and *Hs*cen3 are found in all somatic cells [Lee and Huang, 1993; Middendorp et al. 1997]. Expression of *Hs*cen4, a pseudogene, is limited to ciliated cells [Gavet et al. 2003]. Interestingly, all four isoforms of centrin are expressed in the photoreceptor cells of the retina [Giessl et al. 2004a; Trojan et al. 2008]. *Hs*cen1 and *Hs*cen2 share approximately 80% sequence identity, but only ~54% with *Hs*cen3. Interestingly, *Hs*cen1 and *Hs*cen2 are closely related to *Chlamydomonas reinhartii* centrin (*Cr*cen), sharing ~70% sequence identity, whereas *Hs*cen3 is closely related to *Saccharomyces cerevisiae* centrin homolog (Cdc31), with ~59% sequence identity [Middendorp et al. 1997]. Thus, this phylogenetic evidence suggests two divergent subfamilies among centrins. Moreover, centrin isoforms exhibit different affinities for calcium (Figure 2) [Pastrana-Ríos et al. 2013]. Centrin has two independent domains, and each one contains two EF-hand motifs, each of which possesses a calcium-binding site [Veeraraghavan et al. 2002]. The binding of calcium induces an open conformation of centrin that allows buried hydrophobic residues to be exposed to the solvent. This mechanism regulates the interaction of centrin with most of its targets.

Among the four human centrin isoforms, *Hs*cen2 is known to be localized in the centriole and the nucleus [Paoletti et al. 1996; Salisbury et al. 2002; Resendes et al. 2008]. In

the centrosome, *Hs*cen2 has many targets, such as Sfi1 and POC5 [Kilmartin, 2003; Li et al. 2006a; Azimzadeh et al. 2009]. Also, *Hs*cen2 interacts with Galectin-3, a protein involved in pre-mRNA splicing, and, like *Sc*Prp40, Galectin-3 possesses a NES [Dagher et al. 1995; Li et al. 2006b; Koch et al. 2010]. Within the nucleus, *Hs*cen2 interacts with xeroderma pigmentosum group C (XPC) and Rad23 homolog B (Rad23B). Thus, *Hs*cen2 is involved in global genome nucleotide excision repair (GG-NER) [Marteijn et al. 2014]. In addition, studies carried out by Jani and colleagues show that *Hs*cen2 and *Hs*cen3 are components of the TREX-2 complex, which is involved in mRNA export [Jani et al. 2012].

Centrin targets possess the consensus sequence $W_1xxL_4xxxL_8$ (x represents any residue), in which the conserved residues are known as the hydrophobic triad $W_1L_4L_8$. This hydrophobic triad has been found in all available centrin-target complex structures and is known to be important for the interaction of centrin with its targets (Figure 3) [Hu and Chazin, 2003; Popescu et al. 2003; Thompson et al. 2006; Li et al. 2006a; Jani et al. 2009; Sosa et al. 2011]. Interestingly, Prp40 was first found to be a potential centrin-binding protein in yeast-two-hybrid screening assays using both *Hs*cen2 and *Cr*cen [Díaz Casas et al. 2017]. In this dissertation, the location of the centrin-binding site (CBS) within *Homo sapiens* Prp40A (*Hs*Prp40A) was identified and validated experimentally. The interaction between centrin and a synthetic peptide of *Hs*Prp40A (*Hs*Prp40Ap) was determined using isothermal titration calorimetry (ITC) and two-dimensional infrared (2D IR) correlation spectroscopy. ITC allows the determination of the thermodynamics governing binding between centrin and *Hs*Prp40Ap, whereas 2D IR correlation spectroscopy determines the molecular requirements of binding and

stability. Finally, we characterized the third FF (FF_3) domain of *Hs*Prp40A, which is the location of the CBS, and its molecular behavior under thermal stress was studied using differential scanning calorimetry (DSC), circular dichroism (CD), and 2D IR correlation spectroscopies. The results indicate that *Hs*Prp40A is a novel centrin-binding protein. Moreover, the interaction between centrin and Prp40A may have a regulatory role in pre-mRNA splicing.

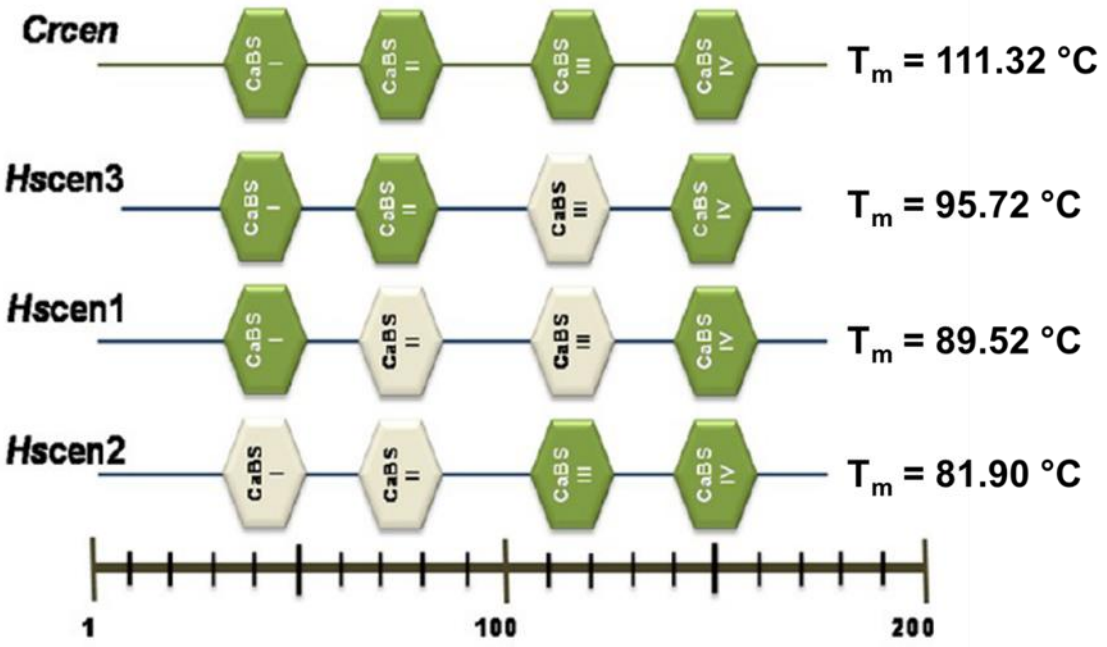


Figure 2. Representation of centrins' calcium-binding sites and affinities for calcium, including their respective thermal transition temperature (T_m) values. Low-affinity and high-affinity calcium-binding sites are colored in white and green, respectively [Adapted from Pastrana-Rios et al. 2013 with permission from publisher American Chemical Society].

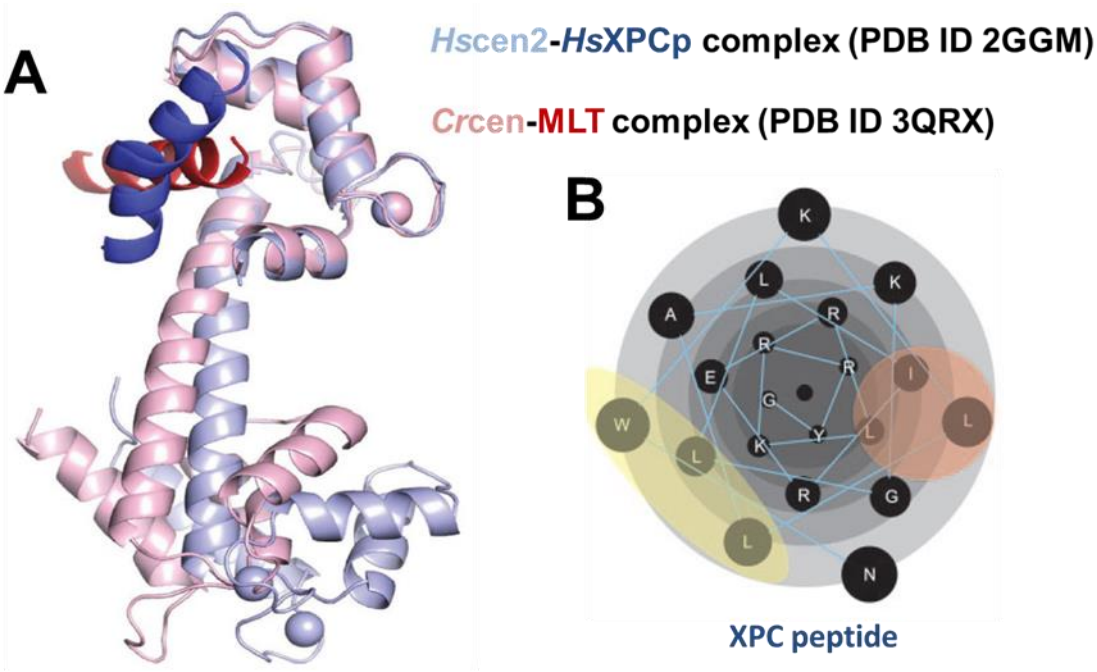


Figure 3. Structures of centrin-target complexes. **(A)** Superposition of the structure of the *Cr*cen-mellitin (MLT) complex **(PDB ID 3QRX)** on the *Hs*cen2-XPC peptide complex **(PDB ID 2GGM)**. **(B)** Helical wheel plot of the centrin-binding site of the XPC peptide. The hydrophobic triad $W_1L_4L_8$ is highlight in yellow [Adapted from Sosa et al. 2011 and Nishi et al. 2005 with permission from publishers John Wiley & Sons, Inc. and the American Society for Microbiology, respectively].

HYPOTHESIS

The hypothesis proposed in this work was the following: *Because centrin's target selectivity is the consensus sequence* $W_1xxL_4xxxL_8$*, we predict that centrin interacts with HsPrp40A, which possesses the hydrophobic triad* $W_1L_4L_8$ *located in its* FF_3 *domain. In addition, centrin interacts with HsPrp40A in its C-terminal domain, and this interaction is dependent of the calcium-binding state.*

OBJECTIVES

In order to validate our hypothesis, we proposed the following objectives:

- To predict the putative CBS within the Prp40 protein using bioinformatics tools.
- To express, isolate, and purify *Hs*Prp40A's FF_3 domain.
- To characterize *Hs*Prp40A's FF_3 domain.
- To determine the thermodynamic parameters governing the binding between centrin and *Hs*Prp40Ap.
- To study the complex dissociation process of ^{13}C-centrin-*Hs*Prp40Ap complex via 2D IR correlation spectroscopy.
- To determine the structure of the centrin-Prp40Ap complex.

CHAPTER II

LITERATURE REVIEW

2.1 Prp40, a nuclear protein with an essential role in the first step of pre-mRNA splicing

Prp40 was first identified in *Saccharomyces cerevisiae* as a protein associated with U1 snRNP with an essential role in the first step of pre-mRNA splicing [Kao and Siliciano, 1996]. The cryoEM structure of *Saccharomyces cerevisiae* U1 snRNP at 3.6 Å was recently reported (Figure 4) [Li et al. 2017]. This nuclear protein comprises 583 amino acids (~69 kDa), with two WW (WW_1 and WW_2) domains and six tandem FF (FF_1–FF_6) domains, which are present in several transcription and splicing factors. Previous studies have shown that the WW domains in *Sc*Prp40 interact with a proline-rich domain of *Sc*Msl5, a branchpoint-binding protein, which also interacts with the splicing factor *Sc*Mud2. *Sc*Prp40 binds to the 5' splice site of the intron, while *Sc*Msl5 and *Sc*Mud2 bind to the branch and the 3' splice sites, respectively. Thus, *Sc*Prp40 is involved in the formation of the cross-intron bridging, a process that occurs in the early steps of spliceosome assembly [Abovich and Rosbach, 1997; Berglund et al. 1997; Bedford et al. 1998; Goldstrohm et al. 2001; Lin et al. 2004; Becerra et al. 2015]. The two WW domains of *Sc*Prp40 also interact with the proline-rich domain of *Sc*Prp8 [Abovich and Rosbach, 1997; Gottschalk et al. 1998]. Moreover, *Sc*Msl5 and *Sc*Prp8 could interact simultaneously with the WW domains of S*c*Prp40 [Wiesner et al. 2002]. The interaction between *Sc*Prp40 and *Sc*Prp8 may help recruit the U5 snRNP to the pre-mRNA [Görnemann

et al. 2011]. Thus, this interaction may also occur during the early steps of spliceosome assembly [Abovich and Rosbach, 1997; Wiesner et al. 2002]. Chung and coworkers showed that the protein *Sc*Clf1 interacts with both *Sc*Prp40 and *Sc*Mud2, acting as a scaffolding protein during spliceosome assembly [Chung et al. 1999; Wang et al. 2003, Gasch et al. 2006]. These findings suggest that Prp40 plays an important structural role during the early steps of spliceosome assembly.

In addition, *Sc*Prp40 interacts with the phosphorylated C-terminal domain of RNA polymerase II (^{Phos}CTD RNAPII). The interaction between *Sc*Prp40 and ^{Phos}CTD RNAP II may facilitate co-transcriptional spliceosome assembly [Morris and Greenleaf, 2000; Goldstrohm et al. 2001; Becerra et al. 2016]. This interaction was also observed in other organisms, such as *Homo sapiens* and *Arabidopsis thaliana* [Allen et al. 2002; Kang et al, 2009]. Unlike *Saccharomyces cerevisiae* and *Homo sapiens*, the three isoforms of *Arabidopsis thaliana* Prp40 (*At*Prp40A, *At*Prp40B, and *At*Prp40C) interact with CTD RNAPII only by their WW domains [Kang et al. 2009]. Ester and Uetz found that *Sc*Prp40 also interacts with the proteins *Sc*Luc7 and *Sc*Snu71 via its FF domains. Like *Sc*Prp40, Luc7 and Snu71 are components of the U1 snRNP [Ester and Uetz, 2008]. The *Sc*Luc7-binding site of Prp40 is located in its FF_1 domain, while the FF_2 and FF_3 domains of *Sc*Prp40 interact with *Sc*Snu71, and this may allow the formation of a ternary complex [Li et al. 2017]. Moreover, most of these interactions between *Sc*Prp40 and its targets have also been found in mammals.

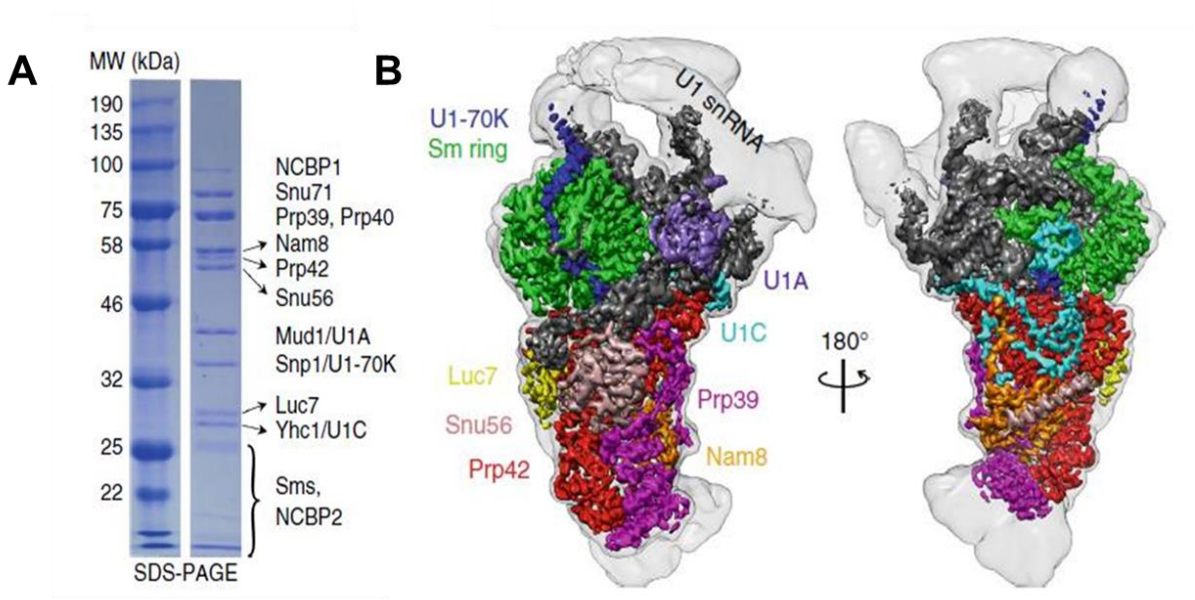


Figure 4. CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP. (A) SDS-PAGE analysis of the components of the U1 snRNP, which includes *Sc*Prp40. (B) Surface representation of the cryoEM structure of U1 snRNP at 3.6 Å resolution [Adapted from Li et al. 2017 with permission from publisher Springer Nature].

2.1.1 Prp40 homolog A and homolog B: the two putative mammalian homologs of ScPrp40

In mammals, such as *Homo sapiens* and *Mus musculus*, *Sc*Prp40 has two putative homologs: Prp40 homolog A (Prp40A), also known as huntingtin-interacting protein A (HYPA) and formin-binding protein 11 (FBP11); and homolog B (Prp40B), also known as huntingtin-interacting protein C (HYPC) [Chan et al. 1996; Faber et al. 1998; Passani et al. 2000]. *Hs*Prp40A and *Hs*Prp40B comprised 953 and 872 amino acids, respectively, and share ~52% sequence identity [Katoh and Katoh, 2003; Becerra et al. 2016]. Interestingly, Prp40A[1] has been found to be more closely related to *Sc*Prp40 than Prp40B[1] is [Becerra et al. 2016]. Like *Sc*Prp40, both Prp40A and Prp40B are involved in pre-mRNA splicing, and they are localized in the nuclear speckles, which are nuclear regions enriched with splicing factors [Bedford et al. 1998; Zhou et al. 2002; Mizutani et al. 2004; Montes et al. 2012]. Although Prp40A is a component of the prespliceosome complex A, unlike *Sc*Prp40, Prp40A was found to associate with the U2 snRNP [Lin et al. 2004; Behzadnia et al. 2007; Makarov et al. 2012]. As modular proteins, *Hs*Prp40A and *Hs*Prp40B have many biological targets in order to carry out their functions.

There are many yeast Prp40-target interactions that also have been found in mammalian organisms. For example, *Hs*Prp40A and *Hs*Prp40B interact with SF1 and U2AF^{65}, the human homologs of *Sc*Msl5 and *Sc*Mud2, respectively [Bedford et al. 1997; Bedford et al. 1998;

[1] In this work, the use of the protein names *Prp40A* and *Prp40B*, in which the species prefix are not included, refer to the two mammalian Prp40 homologs in general.

Goldstrohm et al. 2001; Lin et al. 2004; Becerra et al. 2015]. As in the case of *Saccharomyces cerevisiae*, this interaction may be involved in the cross-intron bridging that occurs in the early steps of spliceosome assembly. In addition, nuclear magnetic resonance (NMR) and ITC studies carried out by Allen and colleagues show that the FF_1 domain of *Hs*Prp40A interacts with ^{Phos}CTD RNAPII. Interestingly, no interaction is observed between *Hs*Prp40A and a non-phosphorylated CTD RNAPII [Allen et al. 2002]. This phosphorylated site is composed of YpSPTpSPS (where pS is phosphoserine) [Morris and Greenleaf, 2000; Goldstrohm et al. 2001; Carty and Greenleaf, 2002]. However, the interaction between *Hs*Prp40B and RNAPII remains unknown [Becerra et al. 2016].

Chan and colleagues found that the WW domains of *Mus musculus* Prp40A (*Mm*Prp40A) interact *in vitro* with the proline-rich region of formins, a family of nuclear proteins involved in murine limb and kidney development. Interestingly, this proline-rich region also corresponds to the binding sites of formin targets that possess an SH3 domain, resulting in competitive binding. Thus, *Mm*Prp40A may play a regulatory role in the interaction of formins with their SH3-domain targets [Chan et al. 1996]. Also, *Mm*Prp40A interacts with the Neural Wiskott-Aldrich syndrome protein (N-WASP), a protein that has a role in actin cytoskeleton reorganization [Mizutani et al. 2004; Pires et al. 2005]. The proline-rich region of N-WASP interacts with the WW domain of *Mm*Prp40A. Both proteins are co-localized in the nucleus of the cell. By this interaction, *Mm*Prp40A suppresses translocation of N-WASP from the nucleus to the cytoplasm. Thus, *Mm*Prp40A inhibits the cytoplasmic function of N-WASP [Mizutani et al. 2004].

2.1.2 Relation of Homo sapiens Prp40 homologs to diseases

During the past few decades, several studies have implicated both human Prp40 homologs in severe neurological disorders. *Hs*Prp40A and *Hs*Prp40B interact with both wild-type and mutant huntingtin protein (Htt), which is responsible for Huntington disease (HD), a neurological disorder that causes brain cells to die in different areas of the brain. This interaction is mediated by the WW domains of *Hs*Prp40A and *Hs*Prp40B that recognize the proline-rich region of Htt located in its N-terminal domain [Faber et al. 1998; Passani et al. 2000]. The poly-glutamine expansion, adjacent to the proline-rich region within the mutant Htt, enhanced the interaction of this protein with *Hs*Prp40A and *Hs*Prp40B [Faber et al. 1998; Jiang et al. 2011]. Mutant Htt sequesters *Hs*Prp40A in the cytosol, thus altering the normal cellular distribution of *Hs*Prp40A. Consequently, this alteration also affects the function of *Hs*Prp40A related to pre-mRNA splicing [Jian et al. 2011]. NAKAP, a nuclear matrix protein, is also an *Hs*Prp40A-binding protein that has a proline-rich region that recognizes the WW domain of *Hs*Prp40A. In HD neurons, Htt and NAKAP are components of nuclear aggregates. Based on this result, it is proposed that the *Hs*Prp40A-NAKAP complex may serve as a nuclear docking site for Htt. This may contribute to the nuclear accumulation of Htt, which is observed in HD. The two WW domains of *Hs*Prp40A may interact simultaneously with NAKAP and Htt, resulting in a ternary complex [Sayer et al. 2005]. Thus, both mammalian Prp40 homologs, *Hs*Prp40A and *Hs*Prp40B, may be involved in HD pathogenesis [Faber et al. 1998; Sayer et al. 2005].

By pull-down and co-immunoprecipitation assays, Bushdorf and Strätling found that the methyl-CpG-binding protein 2 (MeCP2) specifically bound to the WW domains of *Hs*Prp40A and *Hs*Prp40B. The WW domains of both *Hs*Prp40A and *Hs*Prp40B interact in the same proline-rich region of MeCP2. Truncation of this WW domain-binding region in MeCP2 caused a loss of interaction with *Hs*Prp40A and *Hs*Prp40B, which results in the progression of Rett syndrome, another genetic disorder that causes developmental and nervous system problems mostly in girls [Bushdorf and Strätling, 2004].

In addition, *Hs*Prp40A and *Hs*Prp40B interact with CD95 ligand (CD95L) overexpressed in transfected cells and T lymphoblasts. CD95L is a 281-amino acid membrane bound-receptor involved in the regulation of activation-induced cell death. CD95L also functions as co-stimulatory receptor and modulator in T cell activation and C95L-expressing tumor cells. *Hs*Prp40A interacts with the proline-rich region located in the cytosolic tail of CD95L, which recognizes the WW domains of *Hs*Prp40A. Thus, there is a possibility that *Hs*Prp40A is involved in CD95L's regulation and modulation of activation processes in T-cells and in CD95L-expressing tumor cells [Wenzel et al. 2001; Janssen et al. 2003, Becerra et al 2015].

Several studies carried out in both *Homo sapiens* and *Mus musculus* showed that Prp40A is a hypoxia-related gene [Beltran-Parrazal et al. 2010; Rose et al. 2011; Jackson et al. 2012; Oleksiewicz et al. 2017]. In addition, *Hs*Prp40A is associated with obstructive sleep apnea [Perry et al. 2013]. Previous studies showed that knockdown of *Hs*Prp40A inhibits cell

migration in PC3 cells and cytokinesis in HeLa cells. In the case of *Hs*Prp40B, this homolog did not significantly affect cell migration in PC3 cells; however, it led to a reduction in F-actin levels [Bai et al. 2011].

Several studies also have shown that these two Prp40 homologs are implicated in cancer. Wiese and colleagues identified *Hs*Prp40A as one of the 149-gene signature for colorectal cancer [Wiese et al. 2007]. Moreover, *Hs*Prp40A was also identified as one of the genes that is overexpressed in gastric tumor cells and hepatocellular carcinoma [Yang et al. 2007; Suzuki et al. 2009]. In addition, *Hs*Prp40A was found to be overexpressed in Ishikawa endometrial cancer cells after high doses of progesterone [Paulssen et al 2008]. In the case of *Hs*Prp40B, recent studies show that mutations in this homolog are involved in myelodysplasia and pancreatic cancer [Yoshida et al. 2011; Tian et al. 2015]. Interestingly, both *Hs*Prp40A and *Hs*Prp40B are genes differentially expressed in lung adenocarcinoma [Bastide et al. 2010; Imielinski et al. 2012].

2.1.3 Prp40 structure

2.1.3.1 WW domains

In general, WW domains have two conserved tryptophans, which are 20–22 amino acids apart [Bork and Sudol, 1994]. These WW domains are typically 35–40 amino acids in length. In *Homo sapiens*, there are ~50 proteins that contain WW domains [Ingham et al 2005].

These domains typically mediate protein-protein interaction via proline-rich regions [Sudol et al. 1995]. To date, the WW domains have been classified into four groups. Group I is WW domain-containing proteins that recognize a PPxY (PY) motif (where x is any residue). YAP65, dystrophin, and Nedd4 are examples of Group I WW domain proteins. Group II, which includes Prp40A and FE65, recognizes PPLP (PL) motifs. WW domains proteins from Group III, such as FBP30, bind to proline/arginine-containing sequences (PR) motif. Kato and colleagues showed that Group II and Group III can interact with both the PL and PR motifs as well as polyproline stretches. Thus, these two groups are now classified as Group II/III [Kato et al. 2004]. Group IV binds phosphorylated Ser/Thr-Pro (pS/pT-P) motifs and the protein Pin1 belongs to this group [Sudol and Hunter, 2000; Sudol et al. 2001]. Many of these WW domain-containing proteins have been implicated in several human diseases, including Liddle syndrome, Duchenne and Becker muscular dystrophies, Alzheimer's disease, and the previously mentioned HD and Rett syndrome [Sudol, 1996; Faber et al. 1998; Bushdorf and Strätling, 2004].

Prp40 proteins possess two N-terminal WW domains, and the high resolution structures for *Sc*Prp40 and *Hs*Prp40A's domains are presented in Figure 5 [Wiesner et al. 2002; Jiang et al. 2011]. In general, the secondary structure of the WW domains is a triple-stranded curved antiparallel β-sheet [Macias et al. 1996; Wiesner et al. 2002]. Between the WW_1 and WW_2 domains of *Sc*Prp40, there is a linker region comprised of a well ordered helix that possibly appears to impart strict rigidity and fixed orientation upon its tandem WW domains (Figure 5A) [Wiesner et al. 2002; Sudol et al. 2005]. The ligand-binding pockets of the WW domains

face opposite sides in the structure; thus, these domains bind their ligands in an individual, non-cooperative manner [Wiesner et al. 2002]. Interestingly, in the case of *Hs*Prp40A, these ligand-binding pockets point towards each other; and the helical linker is shorter than in *Sc*Prp40 (Figure 5B) [Jiang et al. 2011]. In addition, Pires and colleagues obtained the NMR structure of *Mm*Prp40A's WW_1 domain in complex with a formin peptide of the sequence APPTPPPLPP (Figure 6) [Pires et al. 2005]. This PPLP motif is also present in MeCP2 [Bushdorf and Strätling, 2004].

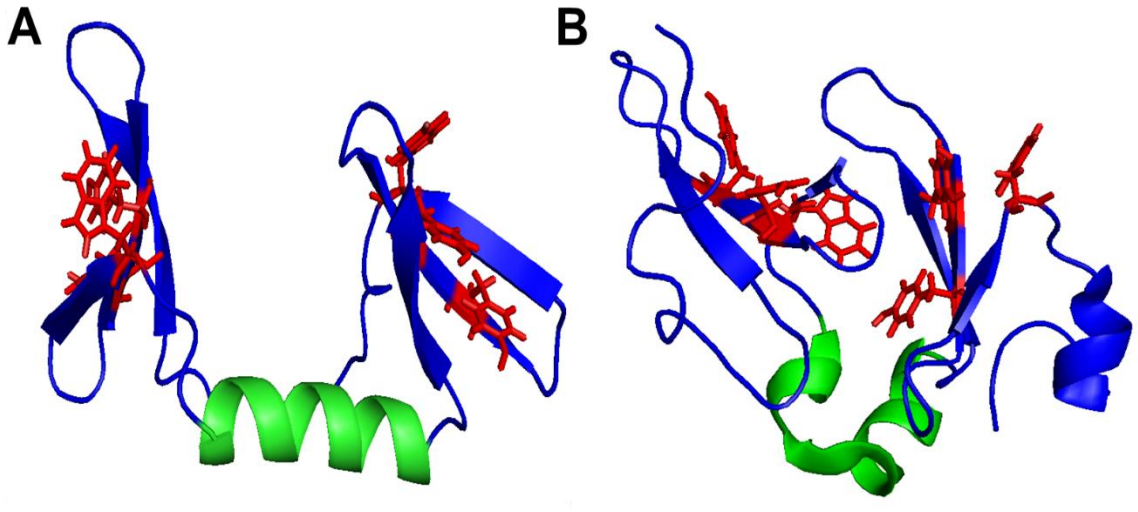


Figure 5. Comparison of the WW domain pair between *Sc*Prp40 and *Hs*Prp40A. Ribbon diagrams of the two tandem WW domain of **(A)** *Sc*Prp40 **(PDB ID 1O6W)** [Wiesner et al. 2002] and **(B)** *Hs*Prp40A **(PDB ID 2L5F)** [Jiang et al. 2011]. The tandem WW domains and the helical linkers are shown in blue and green, respectively; and the side-chains of key residues constituting the ligand-binding pockets are colored in red. This structure was generated using PyMOL version 1.3 from Schrödinger, LLC.

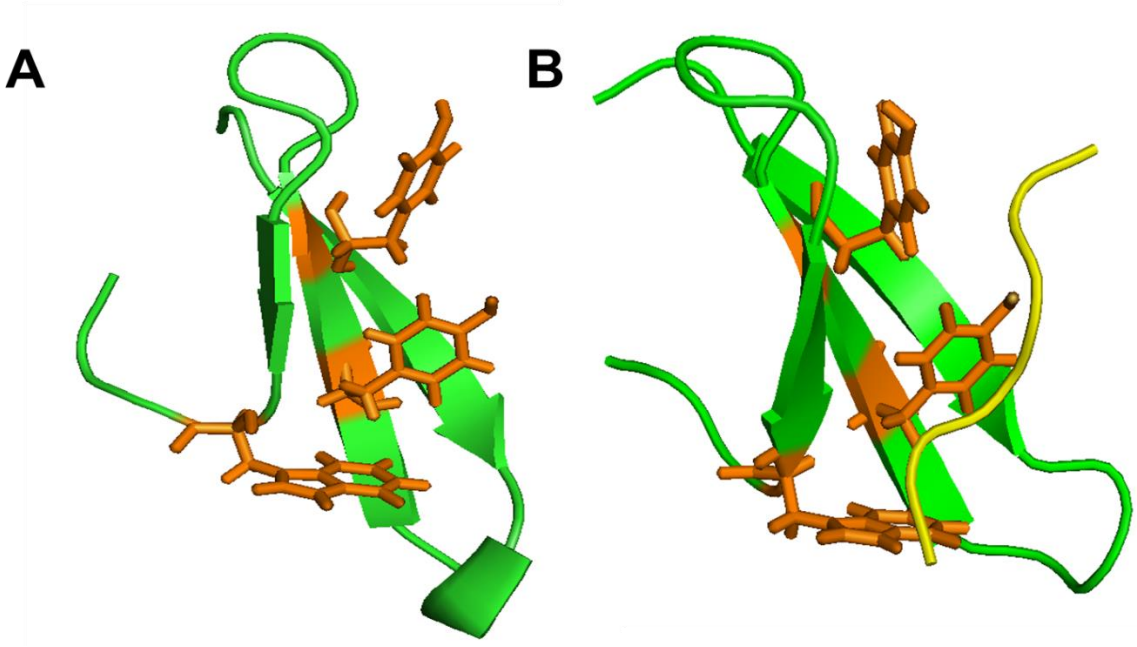


Figure 6. Comparison of the structures of the first WW domain of *Mm*Prp40A and its interaction with a proline-rich ligand. Ribbon diagrams of **(A)** WW_1 domain of *Mm*Prp40A **(PDB ID 1YWJ)** and **(B)** *Mm*Prp40A-formin complex **(PDB ID 1YWI)**. In **(B)**, the WW_1 of *Mm*Prp40A (green) interacts with the proline-rich region of formin (yellow; sequence APPTPPPLPP) [Pires et al. 2005]. In both structures, the side-chains of key residues constituting the ligand-binding pockets are colored in orange. The structure was generated using PyMOL version 1.3 from Schrödinger, LLC.

2.1.3.2 FF domains

In a manner similar to the WW domains, the FF domains are named because of their two conserved phenylalanine residues. These domains are composed of ~60 amino acids and are located towards the C-terminus from the WW domains. FF domains are mostly present in splicing and transcription factors, but they are also found in the Rho GTPase regulatory proteins of the p190 RhoGAP family [Bedford and Leder, 1999]. Prp40 possesses six FF domains, and the high resolution structures of the FF_1 and FF_6 domains of both *Hs*Prp40A and *Sc*Prp40 have been determined using NMR spectrocopy (Figure 7) [Allen et al. 2002; Gasch et al. 2006; Bonet et al. 2008; Bonet et al. 2009a]. The secondary structure of the FF domains consists of three α-helices, with a 3_{10} helix located in the loop that connects the second and third helix. In general, two conserved phenylalanine residues are located in the middle of the first and third α-helices and they are part of the hydrophobic core of the domain [Allen et al. 2002]. Moreover, a FxxLL sequence motif is located in the first α-helix of each FF domain, and was identified as a potential nuclear receptor box [Bedford and Leder, 1999; Allen et al. 2002].

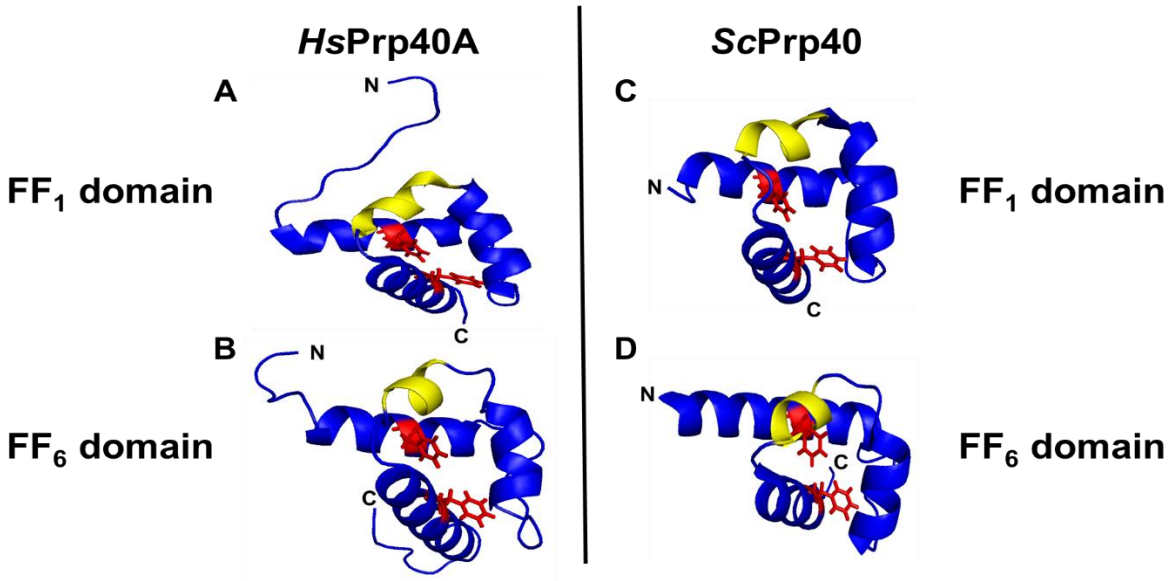


Figure 7. High resolution structures of FF domains in Prp40 proteins. NMR structure of *Hs*Prp40A's **(A)** FF_1 **(PDB ID 1UZC)** and **(B)** FF_6 **(PDB ID 2CQN)** domains, and *Sc*Prp40's **(C)** FF_1 **(PDB ID 2B7E)** and **(D)** FF_6 **(PDB ID 2KFD)** [Allen et al., 2002; Gasch et al. 2006; Bonet et al. 2008; Bonet et al. 2009a]. Typical of a FF domain, these structures consist of three α-helices (colored in blue), with a 3_{10} helix (colored in yellow) located in the loop that connects the second and third helix. The two conserved phenylalanine residues are colored in red. The structures were generated using PyMOL version 1.3 from Schrödinger, LLC.

In addition, crystal structures of three tandem FF domains of the human transcription elongation regulator TCERG1 show that each FF domain is connected to the next one by an α-helix linker (Figure 8) [Lu et al. 2009; Liu et al. 2013]. Despite of the low percent of sequence identity between different FF domains, they possess a high degree of similarity in their structures (Figure 9) [Bedford and Leder, 1999; Bonet et al. 2008]. Interestingly, not all the FF domains in *Sc*Prp40 are functionally equivalent [Gasch et al. 2006]. Results obtained by Ester and Uetz in 2008 found that only FF_1 and FF_2 are essential for *Sc*Prp40, whereas FF_3 and FF_4 may only increase its activity and/or stability [Ester and Uetz, 2008]. Also, *Sc*Prp40's FF_1 domain is negatively charged, in contrast to the FF_1 domain of *Hs*Prp40A, which is predominantly positively charged. Thus, the FF domains can differ in amino acids composition among different organisms. Moreover, *Hs*Prp40A's FF_2, FF_3, and FF_4 domains are negatively charged, whereas its FF_5 and FF_6 domains are comprised primarily of neutral amino acids [Gasch et al. 2006].

Although the WW domains of Prp40 and other proteins have been well studied over the years, the FF domains are still poorly understood. Studies carried out by ITC show that *Hs*Prp40A's FF_1 domain interacts with ^{Phos}CTD RNAPII with a dissociation constant (K_d) of 50 μM [Allen et al. 2002]. Also, the DxRY/F sequence motif located within *Sc*Prp40's FF_1 domain interacts with the N-terminal crooked neck-like tetratricopeptide repeat (crn-TRP) of the splicing factor Clf1 [Chung et al. 1999; Vincent et al. 2003; Gasch et al. 2006]. Previous studies show that the φF/H/LxK/RxG/H/L consensus sequence (where φ represents any hydrophobic residue) of Luc7, and the NDVHY sequence of Snu71, interact with *Sc*Prp40 via

its FF domains [Ester and Uetz, 2008; Li et al. 2017]. Luc7 interacts with the FF_1 domain of *Sc*Prp40, whereas Snu71 interact with *Sc*Prp40's FF_2 and FF_3 domains. Moreover, Li and colleagues co-expressed and purified *Sc*Prp40, Luc7, and Snu71 as a ternary complex, using a protein A-tag on *Sc*Prp40 [Li et al. 2017]. Unlike *Hs*Prp40A, the FF_1 domain of *Sc*Prp40 does not interact with ^{Phos}CTD RNAPII [Gasch et al. 2006]. Taken together, this clearly shows the great variability in the ligand recognition by Prp40's FF domains.

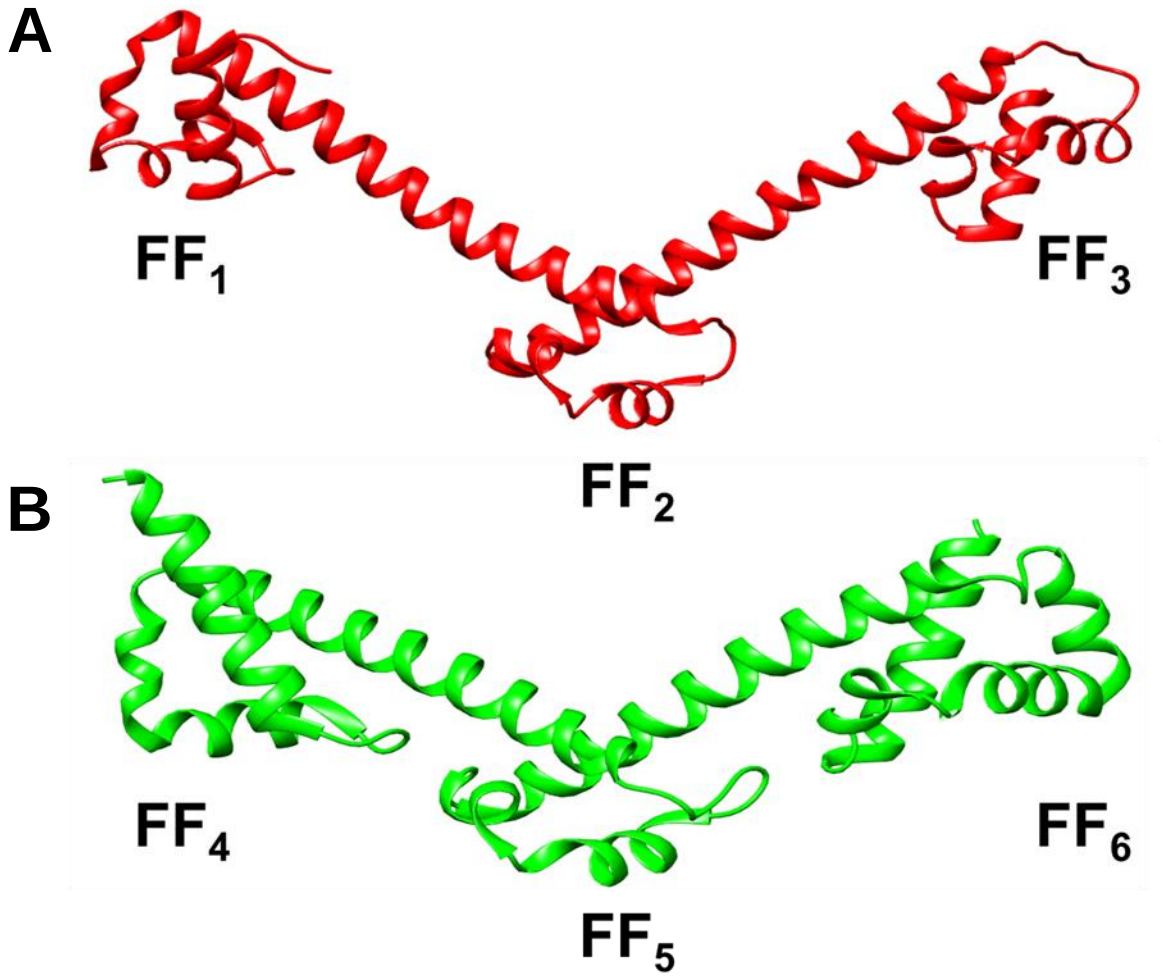


Figure 8. Crystal structures of *Hs*TCERG1's FF domains. **(A)** Ribbon structure of tandem FF_{1-3} **(PDB ID 3HFH)** and **(B)** FF_{4-6} domains **(PDB ID 4FQG)** are colored in red and green, respectively [Lu et al. 2009; Liu et al. 2013]. Each FF domain is connected by an α-helical linker. The structures were generated using UCSF Chimera version 1.9.

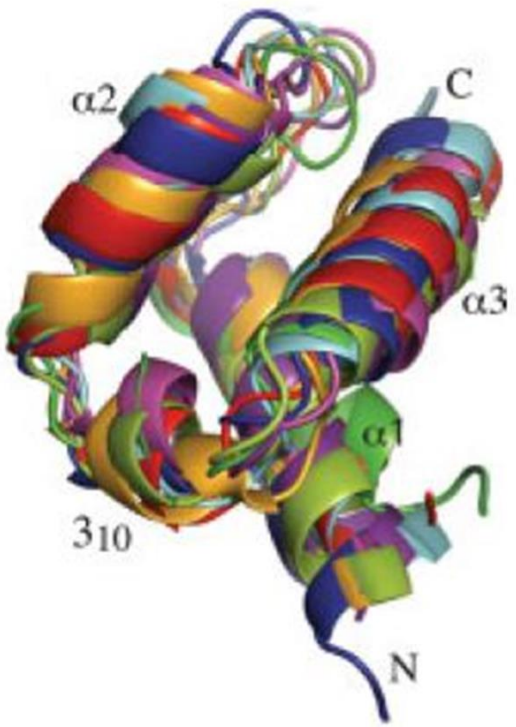


Figure 9. Superposition of different FF domain structures. Shown are the *Sc*Prp40 FF_1 domain (red**, PDB ID 2B7E**), *Hs*Prp40A FF_1 domain (violet, **PDB ID 1UZC**), *Hs*Prp40A FF_6 domain (pale green, **PDB ID 2CQN**), *Hs*CA150 FF_1 domain (light blue, **PDB ID 2DOD**), *Hs*CA150 FF_3 domain (pink, **PDB ID 2DOE**), and *Hs*CA150 FF_4 domain (orange, **PDB ID 2DOF**) [Adapted from Bonet et al. 2008 with permission from publisher John Wiley & Sons].

2.1.3.3. Nuclear export signals

The nuclear export signal (NES) is a sequence that targets the protein for export from the nucleus to the cytoplasm. NES was first found in the human immunodeficiency virus type 1 (HIV-1) Rev protein and cAMP-dependent protein kinase inhibitor (PKI) [Fischer et al. 1995; Wen et al. 1995]. Several proteins that contain the NES have a leucine-rich consensus sequence L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI] (Figure 10) [Bogerd et al. 1996; La Cour et al. 2003]. Several of the NESs are α-helical (Figure 11). In general, the nuclear export signals (NESs) are rich in leucine, glutamate, aspartate and serine residues. Thus, most of the NESs are predominantly negatively charged [La Cour et al. 2004]. In Figure 11, we can see that the conserved hydrophobic residues of the NES are located on one side of the α-helix, thereby allowing for interaction. The polar residues, such as the glutamate residues, are situated on the opposite side of the α-helix, exposed to the solvent environment. The presence of both hydrophobic and polar residues results in an amphipathic α-helical NES. One of the proteins that has been well studied with regard to its export is Crm1 [Fornerod et al. 1997; Fukuda et al. 1997; Stade et al. 1997; Ossareh-Nazari et al. 1997; Ossareh-Nazari et al. 2001]. However, other nuclear export factors have been identified, including the calcium-binding protein centrin [Resendes et al. 2008].

*Sc*Prp40 contains two leucine-rich NESs (NES_1 and NES_2) whose sequences are $_{274}$LKELREYLNGI$_{284}$ and $_{340}$LQNKLNELRL$_{349}$, respectively (Figure 12) [Murphy et al. 2004]. NES_1 and NES_2 are functional for nuclear export through interaction with the nuclear

export factor Crm1. Mutations in NES_1 or both NES_1 and NES_2 caused a defect in splicing at restrictive temperatures *in vivo*. Thus, these two NESs are required for efficient pre-mRNA splicing, in which NES_1 is the most dominant nuclear export signal. Taken together, these results showed that the *Sc*Prp40 export from the nucleus to the cytoplasm is required for pre-mRNA splicing [Murphy et al. 2004]. However, it is possible that another protein has a regulatory role in the *Sc*Prp40 export.

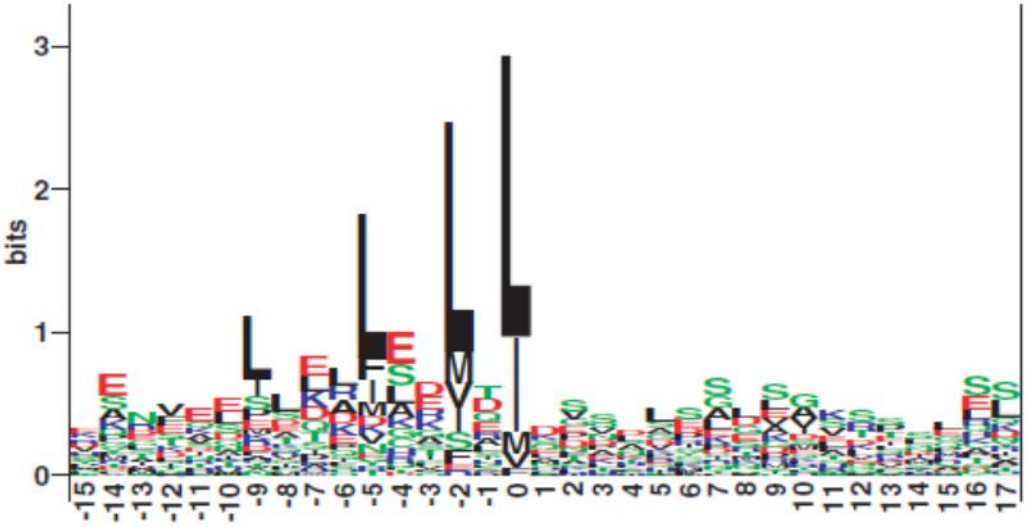


Figure 10. Sequence logos from different NESs show the consensus sequence L-x(2,3)-[LIVFM]-x(2,3)-L-x-[LI] [Adapted from La Cour et al. 2003 with permission from publisher Oxford University Press].

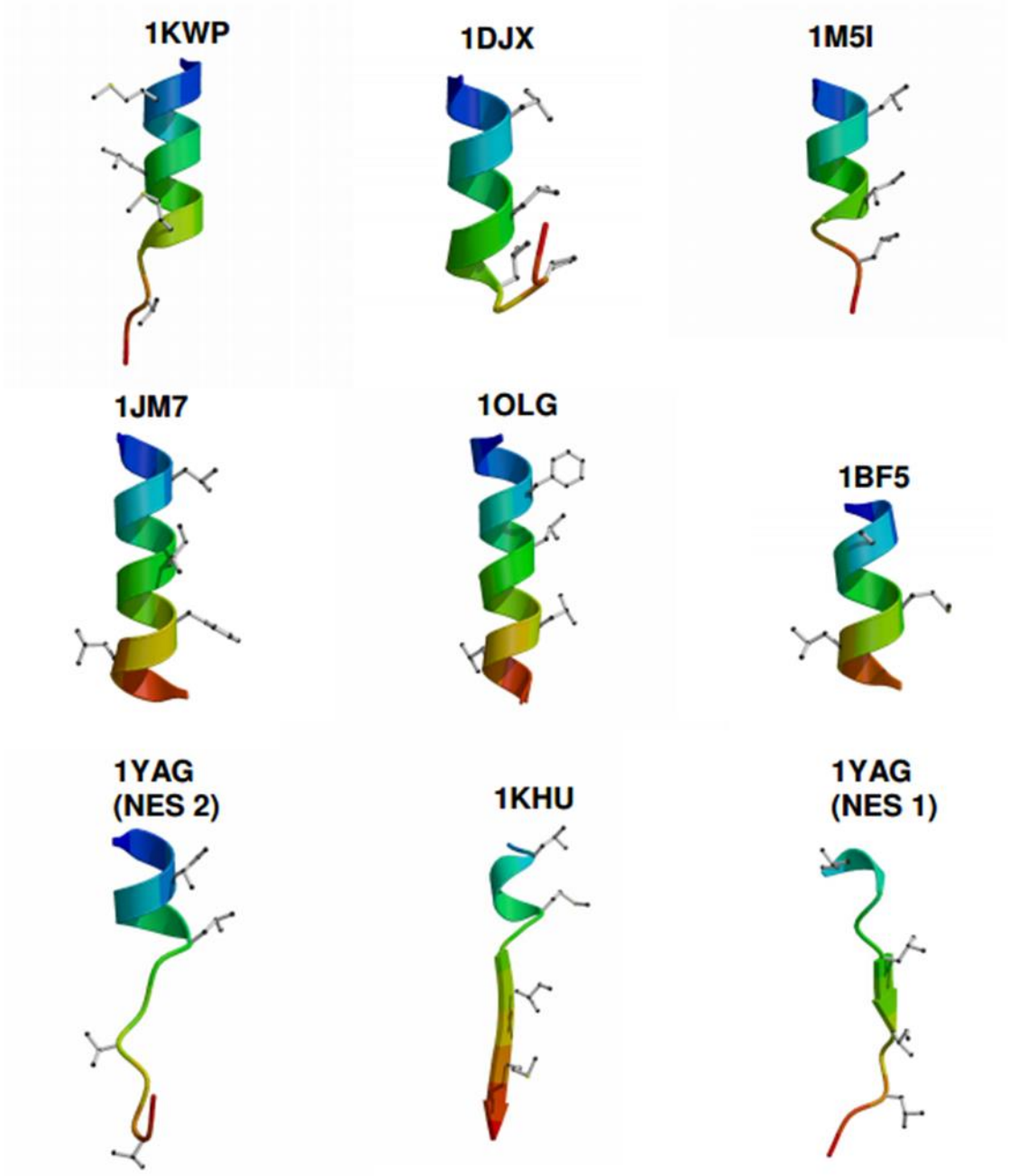


Figure 11. Secondary structures of NESs in different proteins. Most of the NESs are predominantly α-helical and their hydrophobic residues (rendered as balls and stick) face the same side of the helix [Adapted from La Cour et al. 2004 with permission from publisher Oxford University Press]

NES Consensus:	**L** x(2-3)**L** x(2-3)**L** x **L**/**I**
Prp40-NES1:	274-**L** k e - **L** r e y **L** n g **I**-284
mutant nes1:	274-a k e - a r e y a n g **I**-284
Prp40-NES2:	340-**L** q n k **L** n e - **L** r **L**-349
mutant nes2:	340-a q n k a n e - i r **L**-349

Figure 12. Localization of the two NESs in *Sc*Prp40. The two NESs are located at positions 274 to 284 (NES_1) and 340 and 349 (NES_2) [Adapted from Murphy et al. 2004 with permission from publisher Genetics Society of America].

2.2 Centrin, a calcium-binding protein that is involved in protein and mRNA export

Centrin is a calcium-binding protein that has been found to play an essential role in protein and mRNA export [Resendes et al. 2008; Jani et al. 2012; Cunningham et al. 2014]. Centrin is a member of the EF-hand superfamily of calcium-binding proteins [Salisbury, 2007]. Also, it is one of 350 "eukaryotic signature proteins" [Hartman and Federov, 2002; Salisbury, 2007]. Therefore, centrin is critical for the structure and function of the eukaryotic cell [Hartman and Federov, 2002]. This acidic protein was discovered in *Tetraselmis striada,* where it is the major component of the striated flagellar roots [Salisbury et al. 1984]. The striated flagellar roots are organelles that undergo cycles of contraction and extension in the presence of adenosine triphosphate (ATP). Centrin is directly responsible for the contraction of these calcium-sensitive organelles [Salisbury et al. 1984; Baron and Salisbury, 1992]. Centrin is composed of four helix-loop-helix EF-hand consensus motifs, each of which binds a calcium ion. This protein, like the other EF-hand proteins, has two globular domains linked by a tethered helix (Figure 13). Each globular domain is composed of two helix-loop-helix motifs [Hu and Chazin, 2003; Li et al. 2006a; Thompson et al. 2006; Jani et al. 2009; Sosa et al. 2011, Kim et al. 2017]. The tethered helix is located between the second and third EF-hand motifs, resulting in a dumbbell arrangement with two independent domains, the N-terminal domain and the C-terminal domain [Veeraraghavan et al. 2002]. The N-terminal domain imparts stability and directs localization while the C-terminal domain plays a regulatory role that is dependent of the calcium binding state [Matei et al. 2003; Hu and Chazin, 2003; Sheehan et al. 2006; Sosa et al. 2011]. Each EF-hand motif binds calcium with different affinity [Trojan

et al. 2008; Pastrana-Rios et al. 2013] (Figure 2). Studies carried out by our group show that the differences in the stability of centrins may be due to differences in calcium affinity [Pastrana-Ríos et al. 2013]. Centrin's binding to calcium induces an open conformation, allowing buried hydrophobic residues to be exposed to the solvent. Interestingly, centrin is closely related to the calcium sensor calmodulin (CaM), sharing ~50% sequence identity and having a high degree of structural similarity (Figure 14).

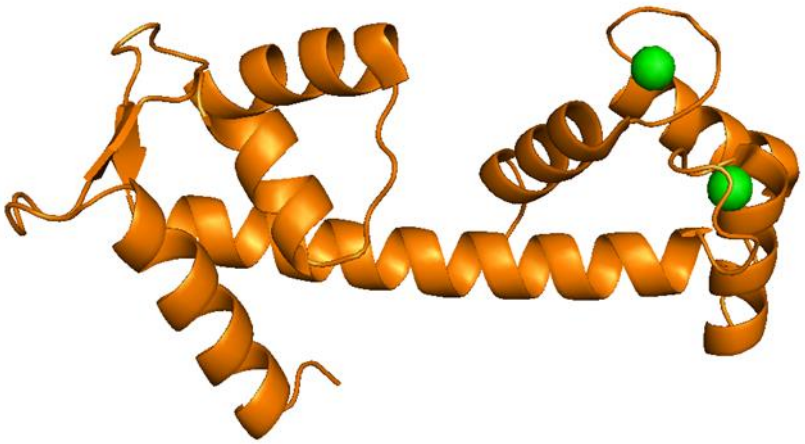


Figure 13. Ribbon diagram of full-length *Hs*cen2 **(PDB ID 2GGM)** [Thompson et al. 2006]. Two calcium ions (green spheres) are bind in helix-loop-helix motifs located in the C-terminal end of *Hs*cen2. This structure was generated using PyMOL version 1.3 from Schrödinger, LLC.

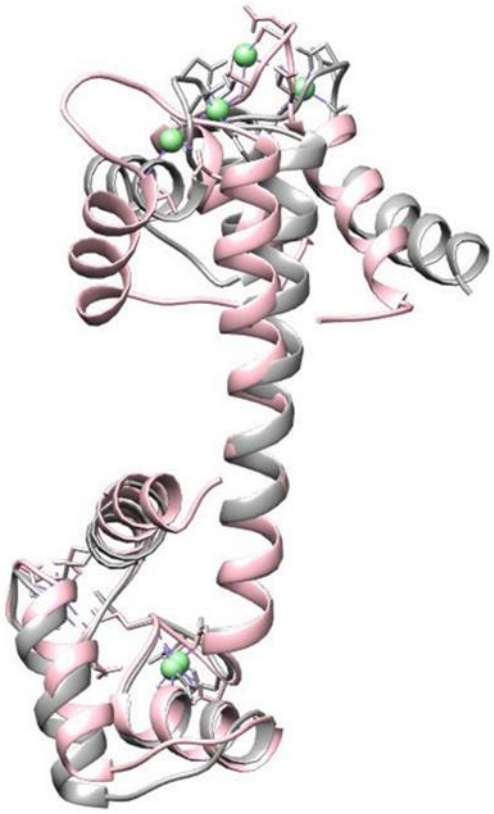


Figure 14. Structural superposition of *Mm*cen1 (pink) **(PDB ID 5D43)** and *Hs*CaM (gray) **(PDB ID 1CLL)** [Chattopadhyaya et al. 1992; Kim et al. 2017]. Calcium ions are colored in light green. The structures were generated using UCSF Chimera version 1.9.

In *Homo sapiens*, there are four isoforms of centrin: centrins 1-3, and the pseudogene centrin 4 (*Hs*cen1-4). *Hs*cen1 and *Hs*cen2 share 84% sequence identity, but they only share 54% sequence identity with *Hs*cen3. Conversely, *Hs*cen3 is more closely related to *Saccharomyces cerevisiae* centrin homolog Cdc31, whereas the other *Homo sapiens* centrin isoforms are more closely related to *Chlamydomonas reinhardtii* centrin (*Cr*cen). Thus, this clearly shows the existence of two different subfamilies among the *Homo sapiens* centrins [Middendorp et al. 1997]. All four isoforms of human centrin are localized in the connecting cilium within the apparatus of the photoreceptor cells, where they interact with the βγ subunit of the G protein transducin. In this interaction, centrins may act as a regulator of the translocation of trandsducin between the photoreceptor outer and inner segments [Giessl et al. 1996; Wolfrum and Salisbury, 1998; Giessl et al. 2004b; Trojan et al. 2008].

*Hs*cen1 is localized at the base of the flagella within the sperm, and is also found in neurons and highly ciliated cells [Hart et al. 1999]. It is an essential component of the sperm connecting piece and is involved in centrosome dynamics during sperm morphogenesis and in zygotes and early embryos during spindle assembly [Moretti et al. 2017]. Previous studies show that abnormal expression or deletion of centrin1 can cause male infertility [Sun et al. 2002; Nakamura et al. 2005; Hinduja et al. 2010; Avasthi et al. 2013]. Moreover, centrin1 interacts with SUMO 2/3 in mouse testis in a calcium-dependent manner. This interaction may play a role in the differentiation of male germ cells [Tanaka et al. 2010]. In addition, previous studies show that *Hs*cen1 is a cancer testis antigen, with up-regulated expression in prostate and pancreatic cancer cells [Kim et al. 2013].

*Hs*cen2 and *Hs*cen3 are ubiquitously expressed in all somatic cells [Lee and Huang, 1993; Middendorp et al. 1997]. *Hs*cen2 is localized in the centriole and in the nucleus [Salisbury et al. 2002; Paoletti et al. 1996; Resendes et al. 2008]. In the centriole, one of its targets is Sfi1, a 1242-amino acid protein with 23 tandem CBSs. The interaction between *Hs*cen2 (or the yeast homolog Cdc31) and Sfi1 was found to be essential for duplication of the centriole in mammals and of the spindle pole body (SPB) in yeast [Kilmartin, 2003; Li et al. 2006a]. *Hs*POC5, another centrin-binding protein, is localized in the distal end of the centriole and has three CBSs. POC5 has been associated with the assembly of the distal half of centrioles and is required for centriole enlongation. GST pull-down experiments show that POC5 interacts with *Hs*cen2 and *Hs*cen3 [Azimzadeh et al. 2009]. Centrin is required for linear POC5 assembly and this event requires calcium binding [Dantas et al. 2015]. Galectin-3 also interacts with centrin2 in the centrosome [Koch et al. 2010]. Like Prp40, Galectin-3 has a NES and is involved in pre-mRNA splicing [Dagher et al. 1995; Li et al. 2006b]. Like centrin2, the deletion of Galectin-3 results in abnormalities at the level of the microtubule and of the primary cilia [Koch et al. 2010]. Studies carried out by Tsang and colleagues show that *Hs*cen2 forms a complex with CP110 and CaM, which plays a role in cytokinesis regulation and genome stability [Tsang et al. 2006]. Moreover, centrin2 regulates primary ciliogenesis by controlling CP110 levels [Prossner and Morrison, 2015]. Strikingly, depletion of centrin2 results in several ciliopathic phenotypes in zebrafish, including enlarged pronephric tubules and pronephric cysts [Denaval et al. 2011]. Also, knockout of *Mm*cen2 leads to olfactory cilia loss [Ying et al. 2014]. Taken together, centrin proteins, and more specifically centrin2, may be involved in ciliopathies, a group of genetic disorders related to defects in the primary cilia [Badano et al.

2006; Bettencourt-Dias et al. 2011 Hildebrandt et al. 2011; Lee and Gleeson, 2011; Novarino et al. 2011; Ostrowski et al. 2011].

Both *Hs*cen3 and its yeast homolog, Cdc31, are required for SPB duplication [Middendorp et at. 1997; Middendorp et al. 2000; Kilmartin, 2003]. Unlike centrin2, which strictly binds calcium ions, centrin3 has an EF-hand that has high affinity to both calcium and magnesium ions [Cox et al. 2005]. Centrin3 acts as an inhibitor of the kinase Msp1, preventing the autophosphorylation of Msp1 and the Msp1-dependent phosphorylation of centrin2. Thus, centrin3 can inhibit the incorporation of centrin2 into centrioles [Sawant et al. 2015]. In *Saccharomyces cerevisiae*, Cdc31 interacts with Vps13, a protein involved in membrane trafficking [Kilmartin, 2003; De et al. 2017; Myers and Payne, 2017]. Interestingly, mutations in *Homo sapiens* Vps13 genes have been related to Huntington-like diseases. The interaction between Cdc31 and Vps13 is required for trans-Golgi network late endosome transport and trans-Golgi homotypic fusion. In both events, Cdc31 is required for Vps13 activity [De et al. 2017].

Centrin4 is restricted to ciliated cells and is a centrin2-related gene [Gavet et al. 2003]. Studies carried out by Hsu and colleagues show that centrin4 interacts with the neuron-specific K channel Eag1 in rat brain and retina. In this interaction, centrin4 may contribute to calcium regulation of Eag1 in neurons [Hsu et al. 2017].

In green algae, *Cr*cen has been found to bind calcium ions in its four EF-hand motifs [Pastrana-Ríos et al. 2013]. *Cr*cen has an essential role in the cell-cycle-dependent duplication and separation of the microtubule organizing center (MTOC). Thus, this protein is involved in the initiation of flagellar excision during the microtubule severing [Taillon et al. 1992; Sanders and Salisbury, 1994]. Our research group has extensively studied *Chlamydomonas* centrin under varying conditions and applying a plethora of biophysical techniques [Pastrana-Ríos et al. 2002; Ortiz et al. 2005; Sanoguet et al. 2006; Pastrana-Ríos, 2006; Meyn et al. 2006; Alfaro et al. 2008; Sosa et al. 2011; Ocaña et al. 2014; Pastrana-Ríos et al. 2015; Díaz Casas et al. 2017]. Our group took advantage of its knowledge of *Cr*cen to establish novel methods for the biophysical characterization of protein-protein interactions. This includes the crystal structure of full-length *Cr*cen in complex with the model peptide MLT (Figure 3A) [Sosa et al. 2011]. Kar1 is another target for centrin, which is an essential component of the SPB [Biggins and Rose, 1994]. A high resolution structure of the C-terminal domain of *Cr*cen in complex with a 19-residue synthetic peptide of Kar1 was obtained by NMR spectroscopy [Hu and Chazin, 2003]. In both complexes, as well as many of the centrin-target complexes reported in the literature [Thompson et al. 2006; Jani et al. 2009], the peptide interacts with the C-terminal end of centrin.

2.2.1 Roles of centrin within the nucleus

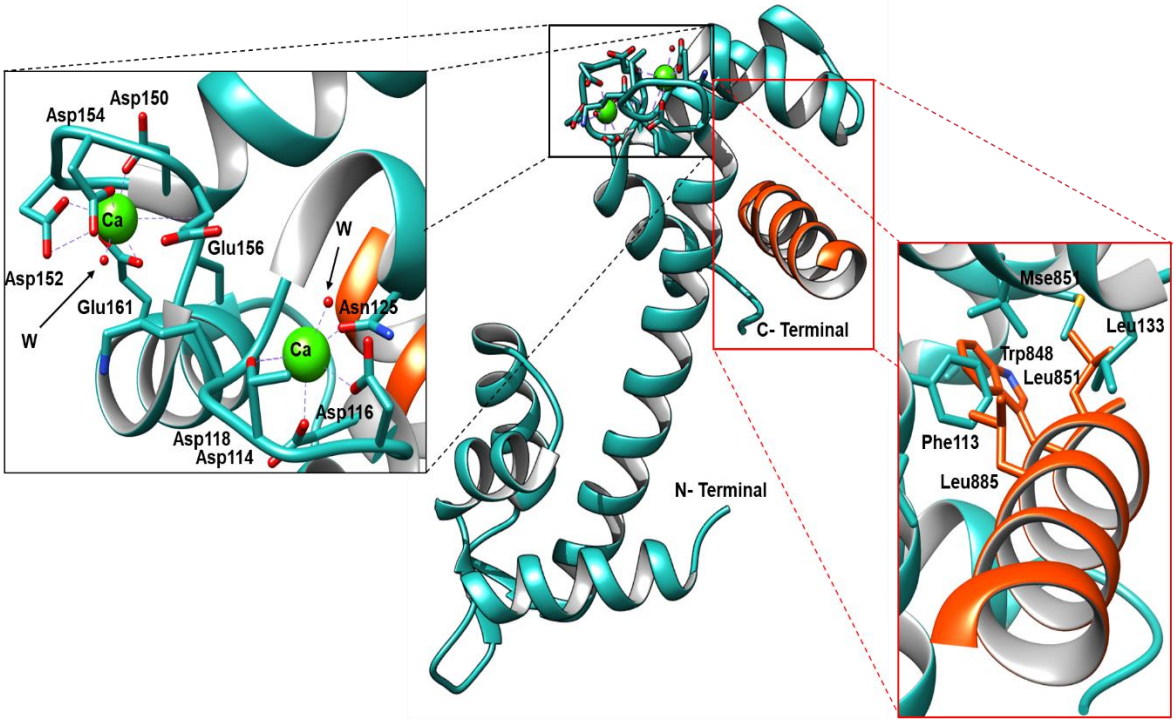
In addition to their essential roles in the centriole, several lines of evidence also show that centrins, and more specifically centrin2 and Cdc31, play different roles within the nucleus.

Centrin2 is a component of the TREX-2 complex, which also includes GANP, DSS1, ENY2, and PCID2. Thus, centrin2 is also involved in mRNA export [Jani et al. 2012]. This complex is also present in *Saccharomyces cerevisiae*, where Cdc31 is the only essential component of the complex [Fischer et al. 2004; González-Aguilera et al. 2008; Jani et al. 2009]. In the TREX-2 complex, Cdc31 interacts with Sac3, the yeast homolog of GANP. However, unlike centrin2, there is no evidence that Cdc31 is involved in protein export [Cunningham et al. 2014]. Studies carried out by Resendes and colleagues show that centrin2 regulates protein export by its association with a NES/nuclear localization signal-bearing construct that contains the NES of HIV protein Rev. In addition, centrin2 localizes in the nuclear pore, where it interacts with the Nup 107-160 complex [Resendes et al. 2008].

In addition to its essential role in mRNA and protein export, centrin2 participates in GG-NER by its interaction with XPC. XPC forms a ternary complex with *Hs*cen2 and RAD23B. This complex functions as a main damage sensor in GG-NER [Araki et al. 2001; Nishi et al. 2005; Marteijn et al. 2014]. In this context, *Hs*cen2 enhances the affinity of XPC with the damaged DNA [Nishi et al. 2005]. Thus, *Hs*cen2 has a crucial role in GG-NER.

The crystal structure of *Hs*cen2 in complex with a synthetic peptide of XPC was obtained by Thompson and colleagues and is presented in Figure 15 [Thompson et al. 2006]. In this complex structure, as well as in all available centrin-target structures [Hu and Chazin, 2003, Li et al. 2006a; Jani et al. 2009; Sosa et al. 2011], centrin's binding site on its targets is the consensus sequence $W_1xxL_4xxxL_8$ (where x represents any residue). These three

hydrophobic residues, which have been identified as important residues for the interaction with centrin, are also known as the hydrophobic triad $W_1L_4L_8$. Among all centrin targets, W_1 is a well conserved residue. L_4 can usually be substituted with another hydrophobic residue such as phenylalanine or tryptophan. L_8 is the least conserved residue in the hydrophobic triad. High resolution structures of *Hs*cen2 in complex with XPC and Sfi1 peptides show that the W_1 of these peptides interacts with a hydrophobic pocket of *Hs*cen2 containing F_{113}, located in its C-terminal domain (Figure 15) [Thompson et al. 2006; Martinez-Sanz et al. 2010]. This is consistent also with the Cdc31-Sac3 peptide complex, where W_1 of Sac3 interacts with F_{105} located in the hydrophobic pocket of the C-terminal domain of Cdc31 [Jani et al. 2009]. Moreover, the crystal structure of *Mus musculus* centrin 1 in a calcium-saturated state was recently obtained at 2.82 Å resolution (Figure 14) [Kim et al. 2017].



Figue 15. Ribbon structure of the *Hs*cen2-XPC peptide complex **(PDB ID 2GGM)**. *Hs*cen2 and the XPC peptide are colored in cyan and orange, respectively. The magnified panel of the right shows the calcium-binding sites of *Hs*cen2, located in the C-terminal end. The calcium ions and the water molecules (W) are represented with green and red spheres, respectively. In the magnified panel of the left, we can see that the hydrophobic triad $W_1L_4L_8$ of the XPC peptide interacts with the hydrophobic pocket of *Hs*cen2, located in the C-terminal end [Thompson et al. 2006]. The structures were generated using UCSF Chimera version 1.9.

CHAPTER III

IDENTIFICATION OF PRP40 HOMOLOG A AS A NOVEL CENTRIN-BINDING PROTEIN

Published in: Díaz Casas, A, Chazin, W.J., Pastrana-Rios, B. (2017) Prp40 homolog A is a novel centrin target. *Biophys. J.* 112, 2529-2539.

3.1 Materials and Methods

3.1.1 Expression, isolation, and purification of recombinant centrins

Expression and purification of unlabeled full-length *Cr*cen and ^{13}C-labeled *Cr*cen (^{13}C-*Cr*cen) were performed as described in Pastrana-Rios et al. 2002 and Sosa et al. 2011, respectively. Full-length *Hs*cen1 and *Hs*cen2 were expressed and purified following the protocol presented in Pastrana Ríos et al. 2013.

3.1.2 HsPrp40A peptide

A synthetic peptide of *Hs*Prp40A (*Hs*Prp40Ap) consisting of the sequence Ac-$_{524}$KQLRKRNWEALKNILDNMANVTYSTTWSEAQQY$_{556}$-CONH$_2$ was purchased from Bio-Synthesis (Lewisville, TX). Mass spectrometry and HPLC analyses were performed by

Bio-Synthesis. To remove the trifluoroacetic acid from *Hs*Prp40Ap, the sample was subjected to repeated lyophilizations in the presence of 0.1 N HCl followed by an exhaustive dialysis against 50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) at pH 7.4, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$. The molar extinction coefficient (ε) of *Hs*Prp40Ap is 13,980 $M^{-1}cm^{-1}$ at 280 nm.

3.1.3 Circular dichroism spectroscopy

*Hs*Prp40Ap (12 μM) in 8 mM HEPES at pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$, and 2 mM $MgCl_2$ was used to acquire far-UV CD spectra on a Jasco J-810 spectropolarimeter (Jasco, Tokyo, Japan). Five scans within the spectral range of 250-195 nm were collected at a scan rate of 200 nm/min, a response time of 2 s, a bandwidth of 2 nm, and a temperature of 25°C. Baseline correction was performed over the spectral range of 250-240 nm. The CD absorbance was converted to mean residue molar ellipticity to analyze the secondary structure of the synthetic peptide. The experiment was carried out in triplicate (n=3).

3.1.4 Isothermal titration calorimetry

ITC experiments were carried out using a VP-ITC microcalorimeter from MicroCal (Northampton, MA). Centrins and *Hs*Prp40Ap samples were exhaustively dialyzed separately against 50 mM HEPES at pH 7.4, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$. The protein and the synthetic peptide concentrations were determined using a Jasco model V-560 UV/Vis

spectrophotometer. The calculated ε for *Hs*cen1, *Hs*cen2, and *Cr*cen is 1,490 M^{-1} cm^{-1}. In a typical experiment, 6-10 μM *Hs*Prp40Ap within the sample cell was titrated with 36-144 μM centrin by automatic injection with volumes ranging from 5 to 10 μL. The first injection was set at 2 μL and was ignored in the final data analysis. All the samples and the reference (buffer) were degassed for at least 10 min before use.

The ITC data were fitted with the One Set of Sites model, using MicroCal Origin software (Northampton, MA). The values for the change in enthalpy of binding (ΔH_B), the change in entropy of binding (ΔS_B), the association constant (K_a), and the stoichiometry of binding (*n*) were determined. The dissociation constant (K_d) was calculated using equation (1):

$$K_d = \frac{1}{K_a} \tag{1}$$

The change in Gibbs free energy of binding (ΔG_B) was calculated using equation (2):

$$\Delta G_B = -RT\ lnK_a \tag{2}$$

For a One Set of Sites model, the shape of the binding isotherm is represented with the *C* value and was calculated using equation (3):

$$C = \frac{[P]_t}{K_d} \tag{3}$$

where $[P]_t$ is the concentration of the peptide located in the cell. The change in heat capacity at constant pressure (ΔC_p) was obtained by the linear best fit for the ΔH_B as a function of temperature. The estimation of ΔH_B at 60°C was calculated from the linear equation of the plot: ΔH_B (60°C) = -0.22 kcal mol^{-1} °C^{-1} (60ºC) - 5.7 kcal mol^{-1}.

We used the structural parameterization of unfolding energetics to estimate the solvent accessibility for the *Cr*cen–*Hs*Prp40Ap complex. The polar and apolar surface areas (ΔASA_{polar} and ΔASA_{apolar}) were estimating using equations (4) and (5), presented by Freire and colleagues [Murphy and Freire, 1992; Xie and Freire, 1994]:

$$\Delta ASA_{polar} = \left[a(60°\mathrm{C})\Delta C_p - a'\Delta H(60°\mathrm{C})\right]/[a(60°\mathrm{C})b' - a'b(60°\mathrm{C})] \quad (4)$$

$$\Delta ASA_{apolar} = \left[b(60°\mathrm{C})\Delta C_p - b'\Delta H(60°\mathrm{C})\right]/[a'b(60°\mathrm{C}) - a(60°\mathrm{C})b'] \quad (5)$$

where a' = 5.37 cal mol^{-1} °C^{-1} Å^{-2}, b'= -3.10 cal mol^{-1} °C $^{-1}$ Å^{-2}, a(60°C) = 31.4 cal mol^{-1} Å^{-2}, and b(60°C) = -8.44 cal mol^{-1} Å^{-2}. These values were obtained from a statistical analysis of a thermodynamic and structural database of proteins [Murphy and Freire, 1992; Xie and Freire, 1994].

3.1.5 FT-IR spectroscopy

^{13}C-*Cr*cen-*Hs*Prp40Ap complex (1:1 molar ratio) was prepared under the desired buffer conditions (25 mM HEPES at pD 6.6, 150 mM NaCl, 2 mM $CaCl_2$, and 2 mM $MgCl_2$) using D_2O after complete hydrogen to deuterium (H→D) exchange via freeze drying as per Sosa et al. 2011. ^{13}C-*Cr*cen-*Hs*Prp40Ap complex (8.25 mg/mL) was deposited onto a round 49 mm x 4 mm custom milled CaF_2 window with a fixed path length of 40 µm from Spectral System LLC (Hopewell Junction, NY); the reference cell contained buffer only. Both cells were set in a custom dual chamber cell holder. The temperature within the cell holder was controlled with a Neslab RTE-740 refrigerated circulating bath (Thermo Electron Corp., Newington, NH). The temperature was monitored with a thermocouple positioned in close contact with the cell. Spectral data acquisition was performed at the desired preset temperature. Once the temperature of the cell was reached, 10 minutes were allowed for thermal equilibrium. The temperature range studied was 5-90°C with 5 °C intervals. The instrument used was a Jasco FT-IR spectrophotometer model 6200 (Jasco Corporation, Tokyo, Japan) equipped with an MCT detector, a sample shuttle, and interface. In this experiment, 640 scans were co-added, apodized with a triangular function, and Fourier transformed to provide a resolution of 4 cm^{-1} with the data encoded every 2 cm^{-1}.

3.1.6 2D IR correlation spectroscopy

This technique was developed by Dr. Isao Noda [Noda, 2004; Noda, 2008] and has been extensively used by our research group [Pastrana-Rios, 2001; Pastrana-Ríos et al. 2002; Ortíz et al. 2005; Pastrana-Ríos, 2006; Sosa et al. 2011; Pastrana-Rios et al. 2013]. We used the FT-IR series of acquired sequential spectra as a function of temperature (5-90 ºC). The spectral data set was subjected to the subtraction of the corresponding spectrum of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex acquired at 5 ºC. This resulted in the generation of the difference spectral data set shown in equation (6):

$$\tilde{A}(v_j, t_k) = \begin{cases} A(v_j, t_k) - \bar{A}(\upsilon_j) & for\, 1 \le k \le m \\ 0 & otherwise \end{cases} \tag{6}$$

Where, $\bar{A}(\upsilon_j)$ is the initial spectrum of the data set to generate the difference spectra. Synchronous 2D correlation intensities that change in phase are defined by equation (7):

$$\Phi(v_1, v_2) = \tilde{A}(v_1, t_j) \cdot \tilde{A}(v_2, t_j) \tag{7}$$

The resulting correlation intensity $\Phi(v_1, v_2)$ as a function of two independent wavenumber axes, v_1 and v_2, is the synchronous plot. This effectively spreads the acquired spectral data in two dimensions.

Asynchronous 2-dimensional correlation peak intensities that change out-of-phase from one another are defined by equation (8):

$$\Psi(v_1, v_2) = \tilde{A}(v_1, t_j) \cdot N_{ij} \tilde{A}(v_2, t_i) \tag{8}$$

The term N_{ij} is the element of the so-called Hilbert-Noda transformation matrix, given by equation (9):

$$N_{ij} = \begin{cases} 0 & for\ \ i = j \\ \dfrac{1}{\pi(j-i)} & otherwise \end{cases} \tag{9}$$

This technique allows for the determination of the changes that occur in the spectral region of interest by the synchronous and the asynchronous plot, and consequently enhances the spectral resolution.

The use of both plots and the application of Noda's rules [Noda, 2015] provides information regarding the sequence of molecular events during the perturbation of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex [Pastrana-Rios, 2001; Pastrana-Rios et al. 2002; Noda, 2004; Noda, 2008; Sosa et al. 2011; Pastrana-Rios et al. 2013]. These plots are symmetrical and, for

this reason, we will always refer to the top left triangle for analysis. We begin with the contour plot that evaluates the out-of-phase peak changes (asynchronous plot):

I. If the asynchronous *cross peak,* ν_2 *is positive, then* ν_2 *is perturbed prior to* ν_1 ($\nu_2 \rightarrow \nu_1$).

II. If the asynchronous *cross peak,* ν_2 *is negative then* ν_2 *is perturbed after* ν_1 ($\nu_2 \leftarrow \nu_1$).

III. If the corresponding synchronous *cross peak is* positive, then the order of the events is established using the asynchronous plot (rules I and II).

IV. However, if the corresponding synchronous *cross peak* is negative and the asynchronous *cross peak* is positive, then the order of event(s) is(are) reversed.

The order of events can be established for each peak observed in the ν_2 axis.

The spectral data were not manipulated, and only baseline correction was performed. The baseline correction and 2D IR plots were generated using the Kinetics program of MATLAB (MathWorks, Natick, MA), which was generously provided by Dr. Erik Goormaghtigh from the Free University of Brussels, Belgium.

3.2 Results

3.2.1 Identification of the centrin-binding site in Prp40A

We analyzed the amino acid sequences of *Sc*Prp40 as well as the two mammalian homologs Prp40A and Prp40B in order to identify a $W_1L_4L_8$ pattern that could act as a potential

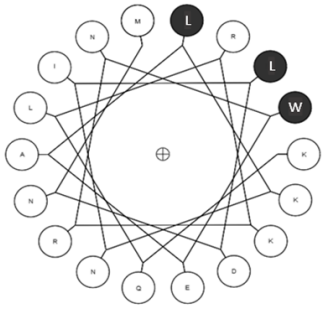
CBS within Prp40. For this, we analyzed the CBSs of known centrin-binding proteins (XPC, Sfi1, POC5, Sac3, and Kar1). The hydrophobic triad $W_1L_4L_8$ was found exclusively in *Hs*Prp40A. Figure 16A shows the CBS for different centrin-binding proteins in which the hydrophobic triad is conserved. The CBS presented has been validated by different molecular biophysical techniques (such as X-ray crystallography and ITC) or *in situ* immunofluorescence studies, where co-localization of the desired proteins was observed [Kilmartin et al. 2003; Thompson et al. 2006; Li et al. 2006a; Azimzadeh et al. 2009; Jani et al. 2009]. Interestingly, the hydrophobic triad of *Hs*Prp40A comprises W_{531}, L_{534}, and L_{538} and is located within the FF_3 domain.

Available crystal and NMR structures of FF domains in different proteins, including *Hs*Prp40A and *Sc*Prp40, reveal a bundle of three α-helices, with a 3_{10}-helix between the second and third α-helices [Allen et al. 2002; Gasch et al. 2006; Bonet et al. 2008, Bonet et al. 2009a; Bonet et al. 2009b; Lu et al. 2009; Liu et al. 2013]. However, the structure of *Hs*Prp40A's FF_3 domain has not been determined to date. Figure 16B shows the helical wheel representation of *Hs*Prp40A, which suggests that the spatial distribution of the hydrophobic triad is limited to one face of the helix. This result is consistent with previous work showing that the hydrophobic triad of other centrin targets is also arranged on a single side of the helix [Hu and Chazin, 2003; Thompson et al. 2006; Li et al. 2006a; Sosa et al. 2011].

A

Protein	aa #								1			4				8				aa #
Hs XPC	841	E	K	R	A	L	G	N	W	K	L	L	A	K	G	L	L	I	R	858
Hs Sfi1	399	A	H	T	V	N	D	K	L	A	R	F	C	F	Y	L	Q	S	H	382
Hs Sfi1	665	V	R	G	Q	Y	T	V	W	A	Q	L	A	R	H	L	V	S	H	648
Hs POC5	173	N	T	K	L	G	S	S	W	L	D	L	V	N	E	M	K	Q	L	156
Hs POC5	262	A	R	V	H	G	I	R	W	H	F	F	T	R	M	L	E	I	K	245
Hs POC5	295	Q	K	Q	V	V	S	R	W	V	K	W	V	K	K	L	L	T	R	278
Sc Sac3	809	A	Q	S	Y	S	A	Q	W	K	E	F	F	K	K	K	L	T	S	792
Sc Kar1	241	R	E	L	I	E	S	K	W	H	R	L	L	F	H	D	K	K	M	258
Hs Prp40A	524	K	Q	L	R	K	R	N	W	E	A	L	K	N	I	L	D	N	M	541

B



C

Protein	aa #								1			4				8				aa #
Hs Prp40A	524	K	Q	L	R	K	R	N	W	E	A	L	K	N	I	L	D	N	M	541
Mm Prp40A	520	K	Q	L	R	K	R	N	W	E	A	L	K	N	I	L	D	N	M	537
Dr Prp40A	418	K	Q	L	R	K	R	N	W	E	A	L	K	N	I	L	D	N	M	435
Xl Prp40A	461	K	Q	L	R	K	R	N	W	E	A	L	K	N	I	L	D	N	M	478
Cr Prp40A	622	A	L	D	K	V	E	R	L	E	V	F	Q	E	Y	M	K	D	M	639
ScPrp40	267	K	K	L	K	T	Q	A	L	K	E	L	R	E	Y	L	N	G	I	284

Figure 16. Amino acid sequence analysis of *Hs*Prp40A. **(A)** Sequence alignment of CBSs of known centrin targets including *Hs*XPC (UniProt ID: Q01831), *Hs*Sfi1 (UniProt ID: A8K8P3), *Hs*POC5 (UniProt ID: Q8NA72), *Sc*Sac3 (UniProt ID: P46674), *Sc*Kar1 (UniProt ID P11927) and the novel centrin-binding protein *Hs*Prp40A (UniProt ID: O75400). *Hs*Sfi1, *Hs*POC5, and *Sc*Sac3 are in a reversed orientation (C-terminal end to N-terminal end) in order to highlight the conserved hydrophobic residues. **(B)** Helical wheel of the CBS within *Hs*Prp40A comprising of the amino acid sequence beginning at Lys_{524} through Met_{541}. The hydrophobic triad $W_1L_4L_8$ is identified in black. The helical wheel representation was generated using an online resource at http://kael.net/helical.htm. **(C)** Sequence alignment of Prp40A in *Homo sapiens*, *Mus musculus* (UniProt ID: Q9R1C7), *Danio renio* (UniProt ID: Q7ZUE4), *Xenopus laevis* (UniProt ID: Q08AZ7), *Chlamydomonas reinhardtii* (available at the Orthologous Matrix database accession number CHLRE00747), and *Sc*Prp40 (UniProt ID: P33203) [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

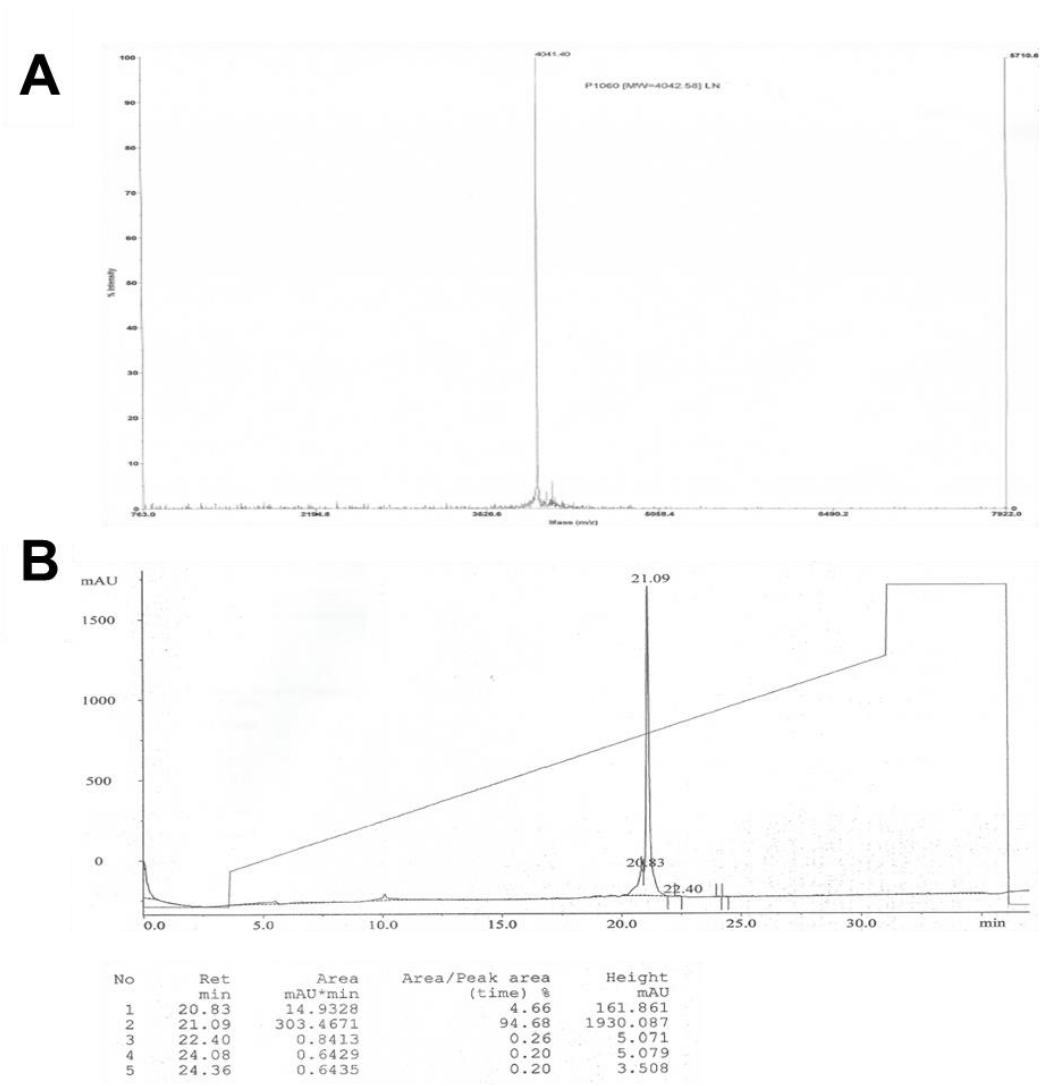
As mentioned above, the potential CBS is localized within the FF_3 domain. Thus, we proceeded to verify the evolutionary conservation of this hydrophobic triad within eukaryotes (Figure 16C). In higher eukaryotes (*Homo sapiens*, *Mus musculus*, *Danio renio*, and *Xenopus laevis*), the CBS was observed to be identical. In *Chlamydomonas reinhardtii* and *Saccharomyces cerevisiae*, the tryptophan in position 1 was substituted with leucine, similar to the ninth CBS within *Hs*Sfi1, whose interaction with centrin was validated by ITC [Martinez-Sanz et al. 2010]. Interestingly, the leucines in positions 1, 4, and 8 of *Sc*Prp40 correspond to the NES_1 as reported by Murphy et al. 2004. Taken together, we found that Prp40A has a potential CBS for the following reasons: 1) the hydrophobic triad found within Prp40A has the well conserved W_1, found in all centrin-binding proteins; 2) the hydrophobic residues in positions 4 and 8 are leucines, which are identical with the hydrophobic residues found in two centrin-binding proteins (XPC and Sfi1); 3) this potential CBS is localized within an FF domain, which is known to mediate protein-protein interactions; 4) the FF_3 domain of Prp40A is highly conserved among different organisms in higher eukaryotes; and 5) this hydrophobic triad is also conserved with the NES_1 of *Sc*Prp40.

3.2.2 Analysis of HsPrp40Ap

The molecular weight and the purity of *Hs*Prp40Ap (Ac-$_{524}$KQLRKRNWEALKNILDNMANVTYSTTWSEAQQY$_{556}$-$CONH_2$) were determined by mass spectrometry (MS) and high performance liquid chromatography (HPLC) analyses, respectively. These analyses were carried-out by Bio-Synthesis after the synthesis of this

peptide. The calculated molecular weight for *Hs*Prp40Ap is 4042.58 Da, while the experimental molecular weight obtained by the MS report was 4041.40 *m/z* (Figure 17A). The purity of *Hs*Prp40Ap obtained by HPLC was >94% (Figure 17B).

In addition, CD spectroscopy was carried-out in order to analyze the secondary structure composition that predominates in *Hs*Pr40Ap (Figure 18). The average crossover point was 201.1 ± 0.9 nm. The average random coil minimum $[\Theta]_{mr}$ was -30,593 ± 781 degrees•cm^2•$dmol^{-1}$ at 205.6 ± 0.2 nm. The average helical contribution $[\Theta]_{mr}$ at 222 nm was 14,759 ± 485 degrees•cm^2•$dmol^{-1}$. Taken together, these results show that the *Hs*Prp40p purchased has a high purity and is predominantly random coil in its free state (as pure component).



No	Ret min	Area mAU*min	Area/Peak area (time) %	Height mAU
1	20.83	14.9328	4.66	161.861
2	21.09	303.4671	94.68	1930.087
3	22.40	0.8413	0.26	5.071
4	24.08	0.6429	0.20	5.079
5	24.36	0.6435	0.20	3.508

Figure 17. Biochemical characterization of *Hs*Prp40Ap. **(A)** Mass spectrometry showed that the molecular weight of *Hs*Prp40Ap was 4041.40 m/z. **(B)** HPLC results confirmed that *Hs*Prp40Ap was obtained with > 94% purity [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

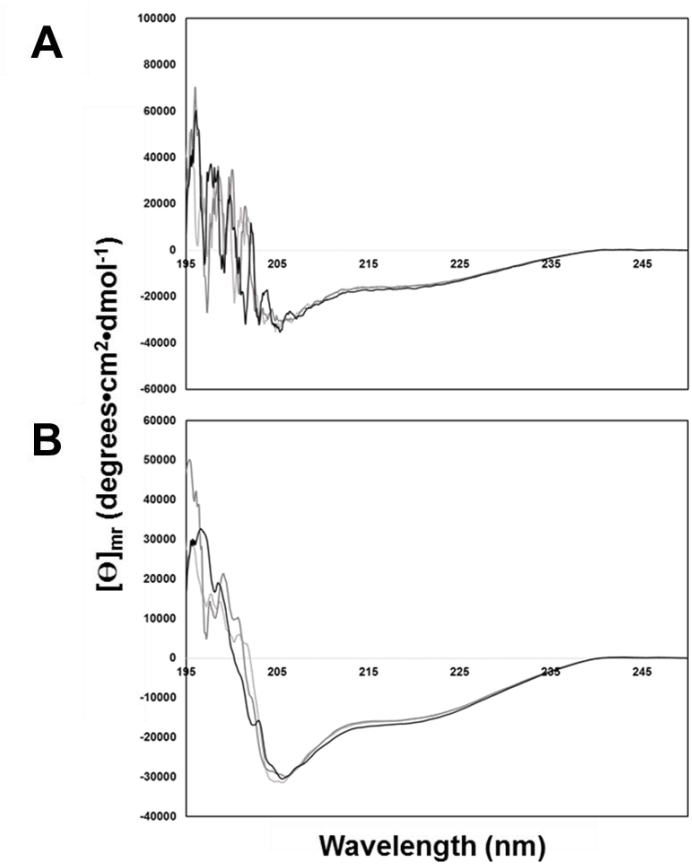


Figure 18. Far-UV CD spectrum of *Hs*Prp40Ap at 25 °C in 8 mM HEPES at pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$, and 2 mM $MgCl_2$. **(A)** Raw data and **(B)** smoothed spectrum using the adaptive smoothing function in the spectral manager software from JASCO. The experiment was carried out in triplicate (n=3) [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

3.2.3 Thermodynamic comparative analysis of the interaction between centrins and Prp40Ap

In order to validate an interaction between centrin and Prp40A in *Homo sapiens*, we first carried out thermodynamic analyses using ITC. All these experiments were performed at pH 7.4, which is in the range of the pH within the nucleus (pH ~7.2) [Casey et al. 2010]. A comparative analysis of the interaction between different centrins and *Hs*Prp40Ap at 25 °C is presented in Figure 19A-C. The power data and the integrated enthalpy clearly show an interaction between full-length *Hs*cen2 and *Hs*Prp40Ap at 25 °C (Figure 19B). The binding stoichiometry (*n)* was close to 0.5. The thermodynamic parameters governing binding for the *Homo sapiens* centrin isoform interaction with *Hs*Prp40Ap are summarized Table 1. However, we did not observe interaction between *Hs*cen1 and *Hs*Prp40Ap at 25 °C and at different molarity ratios (Figure 19C). Among the human centrin isoforms, *Hs*cen2 is localized both in the centriole and in the nucleus. Taken together, we can affirm that *Hs*Prp40Ap interacts selectively with *Hs*cen2.

Many of the centrin-binding proteins interact with the C-terminal domain of centrin in a calcium-dependent manner. In order to determine if the interaction between centrin and *Hs*Prp40Ap is dependent on centrin's affinity for calcium in the C-terminal domain, we performed titrations of *Hs*Prp40A with full-length *C*rcen, which has high affinity for calcium (Figure 2). ITC experiments were carried out at 20, 25, 30 and 35 °C to establish the relative affinity for the target (Figure 20), and the results are summarized in Table 1. The mean

stoichiometry for these experiments was close to ~1.44. In general, the *C* values obtained were ≥10, which suggests that the binding isotherm is sigmoidal and therefore suitable for the determination of the thermodynamic parameters. The exception corresponds to *Cr*cen at 30 °C whose *C* value was determined to be 7. In summary, the binding isotherms were exothermic for all the interactions evaluated for the *Cr*cen-*Hs*Prp40A complex. As the temperature increased, the ΔH also increased. We also estimated the theoretical (calculated) change in enthalphy (ΔH_{calc}) based on structure-based calculations (Table 2) [Murphy and Freire, 1992; Xie and Freire, 1994] and the difference with our experimental ΔH_B was 0-9%. Both the experimental (ΔH_B) and theoretical (ΔH_{calc}) values are consistent, indicating a favorable interaction (such as hydrogen bonding, salt-bridge, hydrophobic interactions, etc.) in the interface between *Cr*cen and *Hs*Prp40A. The K_d was observed to be in the nanomolar range, and thus a high affinity complex was obtained. The highest affinity and stability were determined to be at 35 °C, which is close to physiological temperature. Comparing both complexes at near room temperature (25 °C), the *Hs*cen2-*Hs*Prp40Ap complex exhibited a 2-fold higher affinity (K_d = 278 ± 31 nM versus 588 ± 69 nM) and similar stability (ΔG_B = -8.95 ± 0.07 kcal/mol versus -8.51 ± 0.07 kcal/mol). Also, the $-T\Delta S_B$ for *Hs*cen2 is roughly two times higher than for *Cr*cen. The main contribution of the unfavorable entropy observed during the interaction between centrin and *Hs*Prp40Ap (Table 1) is probably because *Hs*Prp40Ap was undergoing a transition from being flexible (free state) to an ordered structure (bound state), consistent with the CD spectra obtained of free *Hs*Prp40Ap. The relative change in enthalpy plot (Figure 21 and Table 2) resulted in a negative ΔC_p (-0.22 ± 0.09 kcal/mol °C), which suggested an increase in the polar surface between *Cr*cen and *Hs*Prp40Ap during the

interaction, with the estimated binding interface of the *Cr*cen-*Hs*Prp40Ap complex to be more polar (ΔASA_{polar} = -1815 Å^2) than apolar (ΔASA_{apolar} = -1090 Å^2). This result is consistent with the result obtained in the hydrophobicity plot of *Hs*Prp40Ap using the Kyte and Doolittle index [Kyte and Doolittle, 1982], which shows that *Hs*Prp40Ap possesses a higher degree of polar composition (Figure 22).

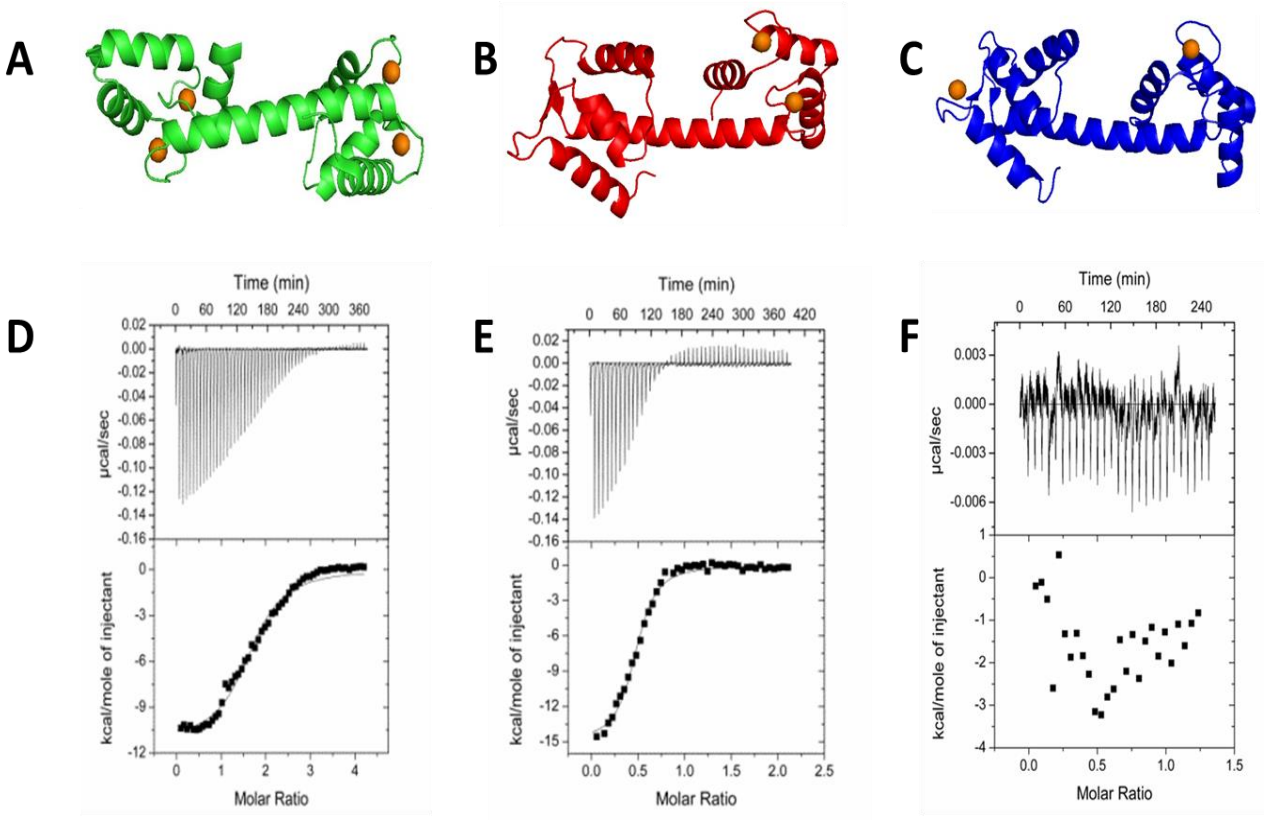


Figure 19. Thermodynamic comparative analysis of the interaction of centrins with *Hs*Prp40Ap at 25 °C. *Cr*cen **(A)** and *Hs*cen2 **(B)** have high affinity for calcium ions in their calcium-binding sites III and IV (located in the C-terminal domain), whereas *Hs*cen1 **(C)** only possesses high affinity for calcium in its calcium-binding site IV (see Figure 2). *Cr*cen and *Hs*cen2 bind to *Hs*Prp40Ap (**D** and **E**, respectively), but not *Hs*cen1 **(F)** [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

Table 1. Summary of the thermodynamic data of the interaction between wild-type centrin homologs and *Hs*Prp40Ap.

Protein	Peptide	Temperature (°C)	K_a (error) (x 10^6 M^{-1})	K_d (error) (nM)	ΔG_B (error) (kcal/mol)	ΔH_B (error) (kcal/mol)	$-T\Delta S_B$ (error) (kcal/mol)	*n* (error)
*Hs*cen1	*Hs*Prp40Ap	25	N.B.[a]	-	-	-	-	-
*Hs*cen2	*Hs*Prp40Ap	25	3.6 (0.4)	278 (31)	-8.95 (0.07)	-15.0 (0.3)	6.1 (0.3)	0.481 (0.007)
*Cr*cen	*Hs*Prp40Ap	20	2.0 (0.2)	500 (50)	-8.45 (0.05)	-9.7 (0.1)	1.2 (0.1)	1.10 (0.01)
*Cr*cen	*Hs*Prp40Ap	25	1.7 (0.2)	588 (69)	-8.51 (0.07)	-11.2 (0.2)	2.7 (0.2)	1.70 (0.02)
*Cr*cen	*Hs*Prp40Ap	30	1.3 (0.1)	769 (59)	-8.48 (0.05)	-13.5 (0.2)	5.0 (0.2)	1.6 (0.2)
*Cr*cen	*Hs*Prp40Ap	35	2.2 (0.3)	455 (62)	-8.95 (0.07)	-12.6 (0.2)	3.6 (0.2)	1.34 (0.02)

[a] N.B., no binding was observed.

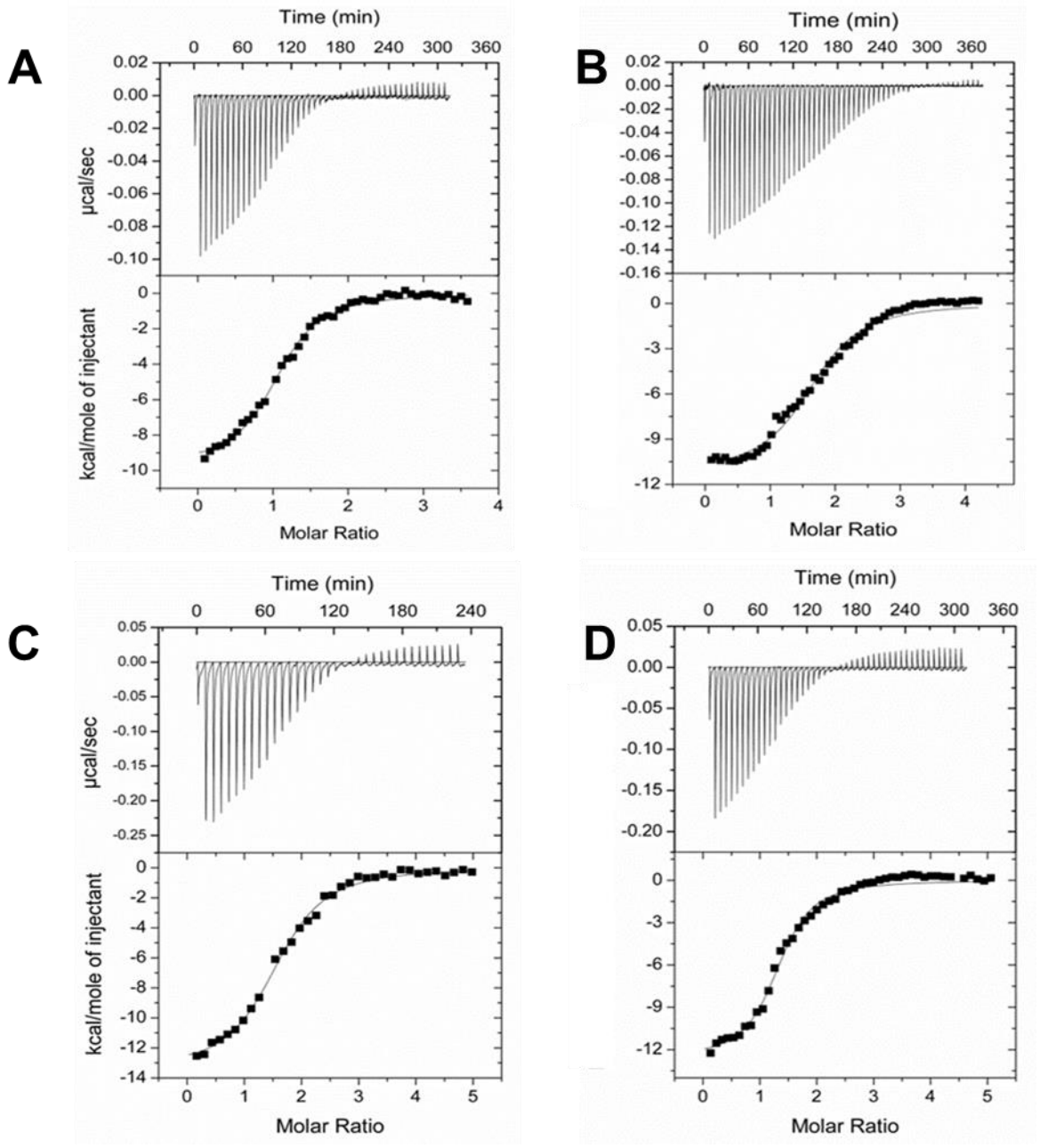


Figure 20. Relative affinities of the interaction between *Cr*cen and *Hs*Prp40Ap at **(A)** 20 °C, **(B)** 25 °C, **(C)** 30 °C and **(D)** 35 °C. The upper panel contains the power data and the lower panel contains the integrated enthalpy fitted to a One Set of Sites model [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

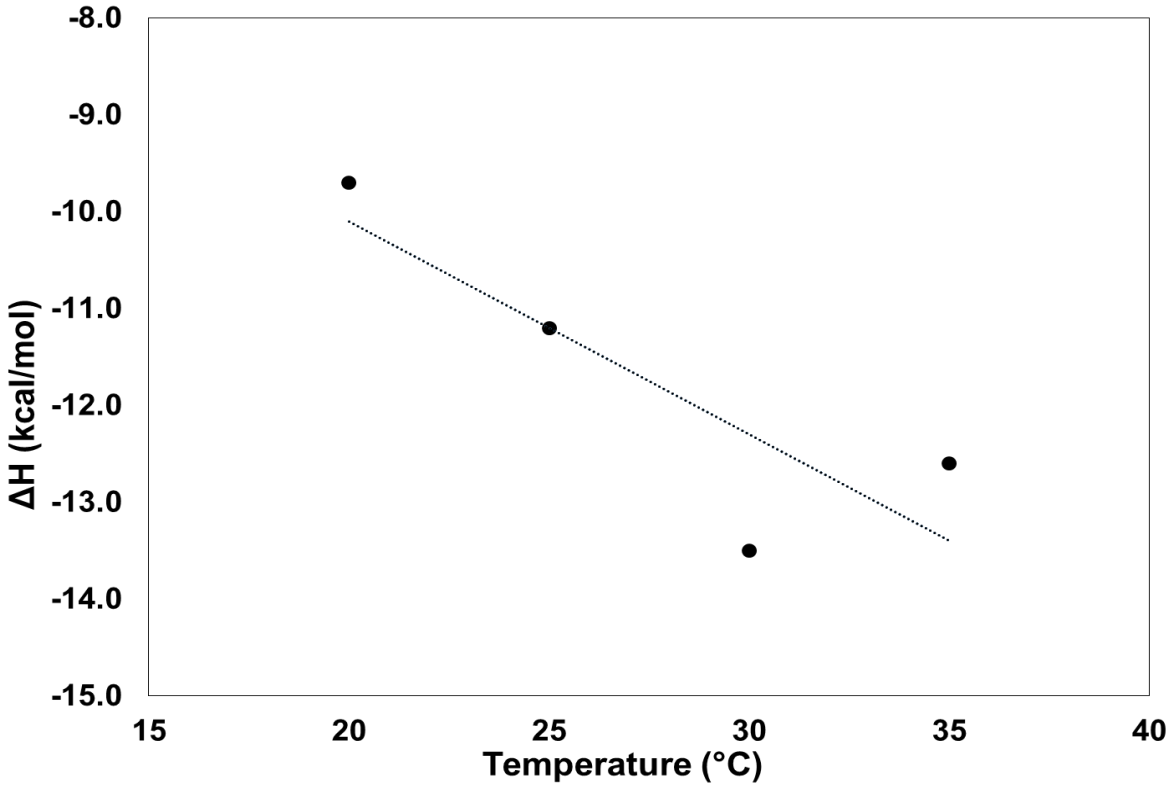


Figure. 21. The relative change in enthalpy for the *Cr*cen-*Hs*Prp40Ap complex. The slope of the graph represents ΔC_p (-0.22 ± 0.09 kcal/mol °C) [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

Table 2. Summary of the estimated thermodynamic parameterization for the *Crcen* and *Hs*Prp40Ap complex interface.

Temperature (°C)	ΔC_p (error)[a] (kcal/mol °C)	ΔH_B (60°C)[b] (kcal/mol)	ΔASA_{polar}[c] (Å^2)	ΔASA_{apolar}[d] (Å^2)	ΔH_{calc}[e] (kcal/mol)
-	-0.22 (0.09)	-18.9	-1815	-1090	-
20	-	-	-	-	-10.1
25	-	-	-	-	-11.2
30	-	-	-	-	-12.3
35	-	-	-	-	-13.5

[a]Determined using the linear slope of the relative ΔH plot (see Figure 21).
[b]Extrapolated using the linear equation for the relative ΔH plot (see Figure 21).
[c]Calculated using Equation 4.
[d]Calculated using Equation 5.
[e]Calculated using equation presented by Xie and Freire (1994).

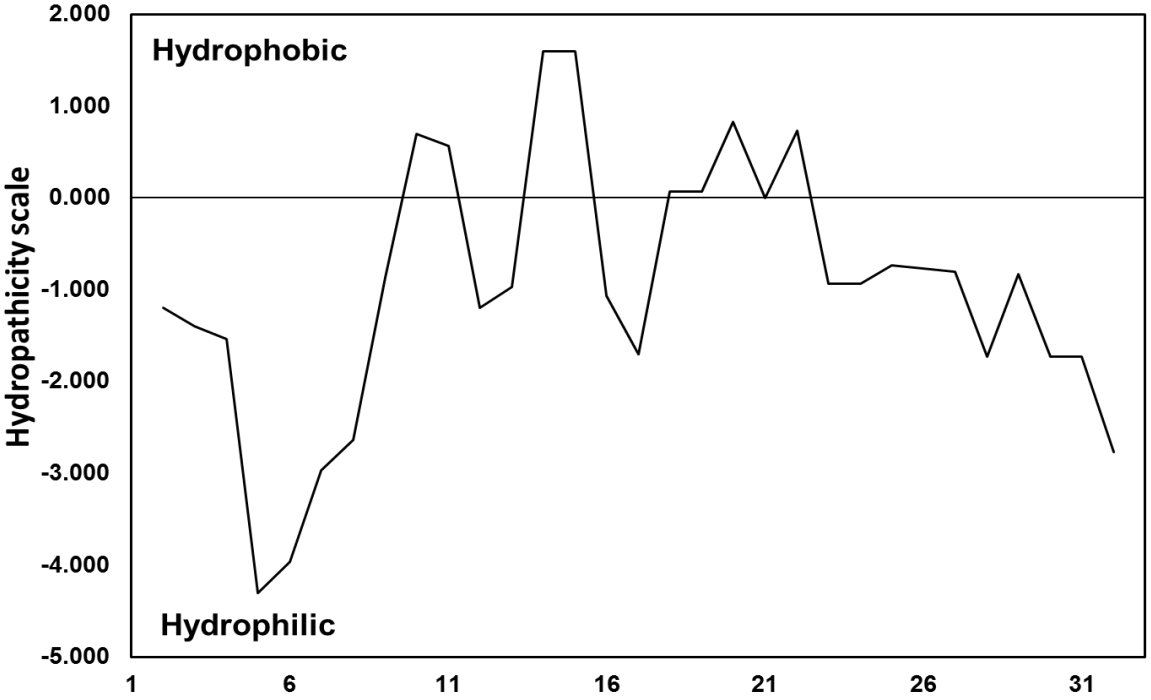


Figure 22. Hydrophobicity plot of *Hs*Prp40Ap. The Kyte-Doolittle hydrophobicity plot [Kyte and Doolittle, 1982] was obtained using the ProScale program of Expasy (http://web.expasy.org/protscale/) [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

3.2.4 FT-IR analysis of the ^{13}C-Crcen-HsPrp40Ap complex

FT-IR was selected as a technique of choice due to its high selectivity and sensitivity to study the conformational dynamics that occur within the *Cr*cen-*Hs*Prp40Ap complex. The spectral region of interest was 1700 – 1500 cm^{-1}. Both the ^{13}C-*Cr*cen and the *Hs*Prp40Ap samples were fully H→D exchanged in order to simplify the amide I' band located typically within 1690 – 1600 cm^{-1} [Venyaminov and Kalnin, 1990; Pastrana-Rios, 2001; Pastrana-Rios et al. 2002]. This band is composed primarily of carbonyl stretching vibrations (ν(C=O)) located within the peptide bond [Bandekar, 1992]. Also, the H→D exchange results in a shift of the amide II band from 1550 cm^{-1} to approximately 1450 cm^{-1}, known as the amide II' band, due to N-D deformation modes within the peptide bond. Also, the buffered solution had a pD of 6.6 which ensures the presence of the deprotonated species of the acidic amino acids (Glu and Asp); evidence to that effect is shown in Figure 23. Thus, the band observed in the 1600 – 1500 cm^{-1} range is attributable to side chains, primarily Arg, Glu^{-}, and Asp^{-} vibrational modes [Chirgadze et al. 1975; Veyaminov and Kalnin, 1990; Bandekar, 1992; Mizuguchi et al. 1997; Pastrana-Rios. 2001; Pastrana-Rios et al. 2002; Pastrana-Rios, 2006; Sosa et al. 2011; Nara et al. 2013]. In the case of *Cr*cen, we homogeneously labeled the protein with ^{13}C. This difference in atomic mass causes the ^{13}C-labeled carbonyl stretching vibrations (ν(^{13}C=O)), also referred to as amide I'*, to shift to lower wavenumbers [Haris et al. 1992; Barth, 2007]. As a result, a simultaneous study of both the target peptide (*Hs*Prp40Ap) and the full-length centrin can be achieved [Sosa et al. 2011]. We also observed the shifts of carboxylate modes (ν(^{13}COO^{-})) within side chains such as Glu^{-} and Asp^{-}. As a result, we were able to distinguish between

vibrational modes pertaining to the protein ($\nu(^{13}C{=}O)$ and $\nu(^{13}COO^-)$) and those of the peptide ($\nu(C{=}O)$ and $\nu(COO^-)$), providing the detailed molecular behavior for each component within the complex. This experimental strategy has proven useful in the past by our group, and the band assignments were made in agreement with previous work [Pastrana-Rios, 2001; Sosa et al. 2011].

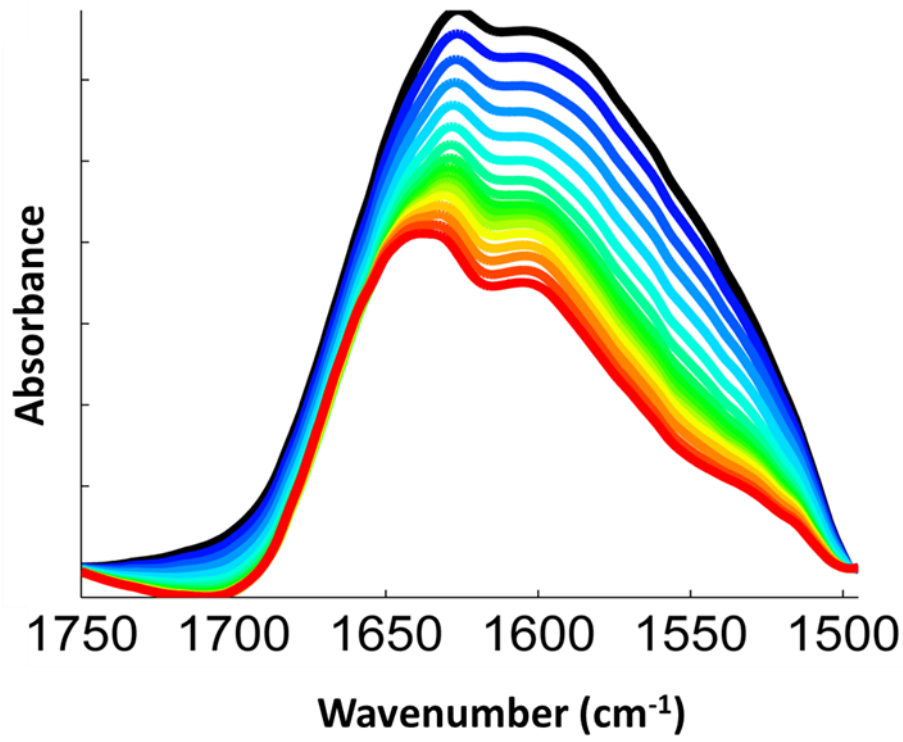


Figure 23. Overlaid FT-IR spectra of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex (1:1, molar ratio) within the spectral region of 1750-1500 cm^{-1} from black (5 °C) to red (90 °C). The absence of the bands of the ν(C=O) vibrational modes of the protonated carboxyl group of Asp and Glu within the spectral regin of 1750-1710 cm^{-1} confirm the existence of only the deprotonated species in the peptide within the complex [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

Figure 24A shows the overlaid FT-IR spectra of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex within the spectral region 1700–1500 cm^{-1} for the temperature range of 5-90 °C. The FT-IR spectral band assignments [Arrondo et al. 1993; Haris and Chapman, 1995; Barth, 2007; Kong and Yu, 2007] for the amide I' band (1690–1600 cm^{-1}) are the following: for *Hs*Prp40Ap, β-turn (1680 cm^{-1}), random coil (1660 cm^{-1}), and α-helix (1649 cm^{-1}). For the amide I'* band in the spectral region of 1620 –1550 cm^{-1}, ^{13}C-*Cr*cen's loop (1638 cm^{-1}), π-helix (1623 cm^{-1}), α-helix (1610 cm^{-1}) and β-sheet (1592 cm^{-1}). Three side chain bands of *Hs*Prp40Ap were observed: two glutamates ν(COO^{-}) (1541 cm^{-1} and 1549 cm^{-1}), and one aspartate ν(COO^{-}) (1567 cm^{-1}) [Mizuguchi et al. 1997 Nara et al. 2013]. These carboxylates within *Hs*Prp40Ap have distinct vibrational modes. We have assigned the 1549 cm^{-1} band to Glu_{532}, which would potentially be participating in an intra-molecular salt-bridge interaction with the positively charged residues located at the N-terminal end of the peptide, once a helical conformation was adopted by the peptide upon complex formation [Pastrana-Rios, 2001]. We assigned the 1567 cm^{-1} band to Asp_{539}, which is located at the center of the peptide, and the 1541 cm^{-1} band to Glu_{552}, which is located near the C-terminal end of the peptide. These residues can serve as probes for the inter-molecular interaction between *Hs*Prp40Ap and ^{13}C-*Cr*cen. ^{13}C-centrin's side chains were assigned as follows: glutamates, 1523 cm^{-1}; arginines' symmetric ν_s(^{13}C-N) and anti-symmetric ν_a(^{13}C-N) stretches, 1557 cm^{-1} and 1577 cm^{-1}, respectively; and aspartates, 1536 cm^{-1}. The ^{13}C-*Cr*cen glutamates and aspartates are located mainly within the calcium-binding site [Sosa et al. 2011].

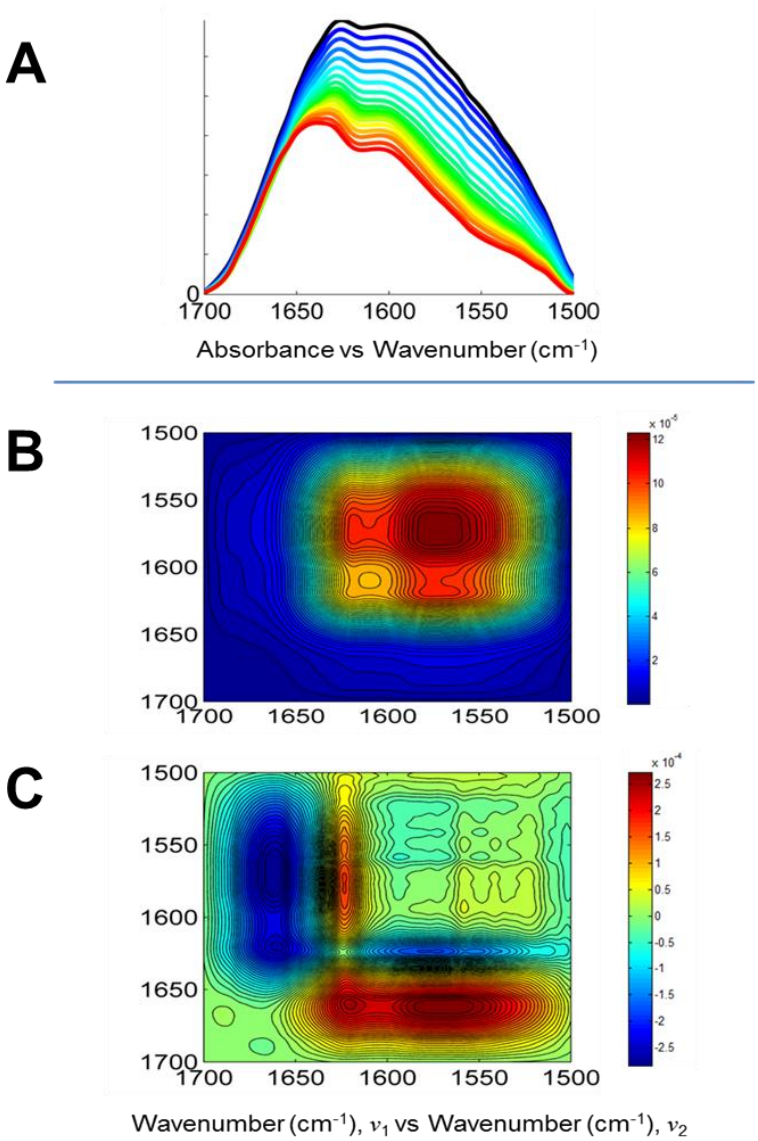


Figure 24. 2D IR correlation spectroscopy of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex (1:1 molar ratio). **(A)** Overlaid FT-IR spectra, **(B)** synchronous, and **(C)** asynchronous plots within the spectral region of 1700-1500 cm^{-1} and over the 5-90 °C temperature range with 5 °C temperature intervals. In the overlaid FT-IR spectra, the black and red lines correspond to 5 °C and 90 °C, respectively [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

2D IR correlation spectroscopy provides the advantage of enhancing the resolution of the spectral region of interest via the asynchronous plot. In Figure 24, B and C, the synchronous and asynchronous plots are shown. Both plots confirm molecular evidence of the interaction between *Cr*cen and *Hs*Prp40Ap. In the synchronous plot (Figure 24B), there is a strong correlation of ^{13}C-*Cr*cen's α-helix (1610 cm^{-1}) with Asp_{539} of *Hs*Prp40Ap (1567 cm^{-1}). In the asynchronous plot (Figure. 24C), we observed that *Hs*Prp40Ap's random coil (1660 cm^{-1}) interacts with ^{13}C-*Cr*cen's β-sheet (1592 cm^{-1}). This random coil of *Hs*Prp40Ap also correlates with the π-helix (1623 cm^{-1}) located at the C-terminal end of ^{13}C-*Cr*cen, suggesting that this interaction occurs at the C-terminal domain of centrin. Also, *Hs*Prp40Ap's Asp_{539} (1567 cm^{-1}) correlates with ^{13}C-*Cr*cen's arginine's ν_a(^{13}C-N) (1577 cm^{-1}), which suggests the existence of an inter-molecular salt bridge.

The synchronous plot contains only positive cross peaks; therefore, the sequential molecular order of events for ^{13}C-*Cr*cen-*Hs*Prp40Ap complex dissociation and unfolding with increasing temperature is defined solely by the asynchronous plot, applying Noda's rules [Noda, 2004; Noda, 2008]. The order of events is summarized in Figure 25 and Table 3. The perturbation begins with ^{13}C-*Cr*cen's C-terminal π-helix (1623 cm^{-1}), followed by centrin's

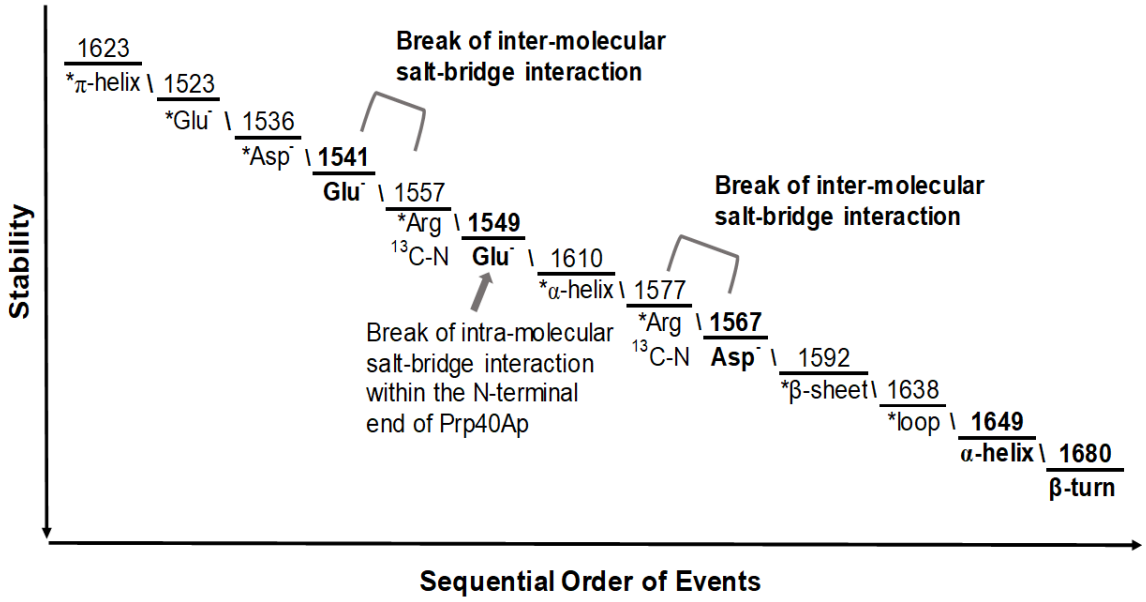


Figure 25. Sequential order of molecular events during the thermal perturbation of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex (1:1 molar ratio) over the temperature range of 5-90 °C. The key to complex formation is the intermolecular salt-bridge interactions between *Hs*Prp40Ap's negatively charged residues and centrin's Arg residues as well as the requirement of the hydrophobic interaction within the binding interface. The text in **bold** corresponds to *Hs*Prp40Ap's vibrational modes [Adapted from Díaz Casas et al. 2017 with permission from publisher Elsevier Inc.].

Table 3. Summary of order of events for ^{13}C-*Cr*cen-*Hs*Prp40Ap complex during thermal stress.

Event	Asynchronous and synchronous plot analysis[a] 5 - 90°C
1	π-helix (1623 cm^{-1}) → Glu^{-} (1523 cm^{-1})
2	π-helix (1623 cm^{-1}) → Asp^{-} (1536 cm^{-1})
3	Asp^{-} (1536 cm^{-1}) → **Glu^{-} (1541 cm^{-1})**
4	Asp^{-} (1536 cm^{-1}) → Arg ^{13}C-N (1557 cm^{-1})
5	Arg ^{13}C-N (1557 cm^{-1}) → **Glu^{-} (1549 cm^{-1})**
6	Arg ^{13}C-N (1557 cm^{-1}) → α-helix (1610 cm^{-1})
7	α-helix (1610 cm^{-1}) → Arg ^{13}C-N (1577 cm^{-1})
8	α-helix (1610 cm^{-1}) → **Asp^{-} (1567 cm^{-1})**
9	**Asp^{-} (1567 cm^{-1})** → β-sheet (1592 cm^{-1})
10	**Asp^{-} (1567 cm^{-1})** → loop (1638 cm^{-1})
11	loop (1638 cm^{-1}) → **α-helix (1649 cm^{-1}),**
12	loop (1638 cm^{-1}) → **β-turn (1680 cm^{-1})**
13	**β-turn (1680 cm^{-1}) → random coil (1660 cm^{-1})**

[a]2D IR correlation peaks corresponding to *Hs*Prp40A are shown in **bold** to distinguish between ^{13}C-*Cr*cen and *Hs*Prp40A cross peaks.

glutamates (1523 cm^{-1}) and aspartates (1536 cm^{-1}), most of which are located within the calcium-binding site of the protein [Sosa et al. 2011]. Increasing temperature then affects the inter-molecular salt-bridge interaction between *Hs*Prp40A's Glu_{552} (1541 cm^{-1}) and ^{13}C-*Cr*cen's arginine $\nu_s(^{13}C\text{-}N)$ (1557 cm^{-1}), followed by *Hs*Prp40Ap's Glu_{532} (1549 cm^{-1}), which is involved in an intramolecular salt bridge interaction with the N-terminal end of the peptide, and lastly ^{13}C-*Cr*cen's α-helix (1610 cm^{-1}). This perturbation within the α-helix of centrin affects the inter-molecular salt bridge interaction between the ^{13}C-*Cr*cen's arginine $\nu_a(^{13}C\text{-}N)$ (1577 cm^{-1}) and *Hs*Prp40Ap's Asp_{539} (1567 cm^{-1}), which is consistent with the cross peaks observed in Figure 24C. Once both inter-molecular salt bridges break, then centrin's EF-hand β-sheet segments (1592 cm^{-1}) and loops (1638 cm^{-1}), located within the calcium-binding site, are perturbed, indicating the unfolding of centrin. Finally, the *Hs*Prp40Ap unfolds, as indicated by the perturbation of its α-helix (1649 cm^{-1}) and then its β-turn (1680 cm^{-1}). These results show that the *Hs*Prp40Ap is more stable than centrin within the complex. The most stable β-turn may correspond to the β-turn located within the loop between the first and second α-helices of the FF domain of Prp40A [Allen et al. 2002; Gasch et al. 2006].

3.3 Discussion

We combined bioinformatics and molecular biophysical (ITC and 2D IR correlation spectroscopy) approaches to determine both the thermodynamics governing binding and the molecular requirements for the interaction between centrin and Prp40Ap. Bioinformatics guided the design of the *Hs*Prp40A fragment-based peptide on the putative CBS containing the

hydrophobic triad ($W_1L_4L_8$) and its convergence with both the NES_1 and the FF_3 domain. Furthermore, we performed a comparative thermodynamic analysis using several centrin homologs (*Hs*cen1, *Hs*cen2, and *Cr*cen), which enabled us to ascertain the requirement of *Hs*Prp40Ap for high Ca^{2+} affinity at the C-terminal end of centrin to allow for formation of the complex. No binding was observed for *Hs*cen1, which, unlike *Hs*cen2 and *Cr*cen, lacks high Ca^{2+} affinity in its third EF-hand. More importantly, the ITC results showed that both *Hs*cen2 and *Cr*cen bound the peptide with K_d values in the nM range and exothermic binding. We then proceeded with establishing the relative affinity for *Cr*cen. Also, the $-T\Delta S_B$ in all cases studied suggested a conformational rearrangement upon formation of the complex. The results for $-T\Delta S_B$ may be due to the transition within *Hs*Prp40Ap from an unordered to a more helical structure as it binds. Also, the thermodynamic parameterization approach confirms the change in polar surface area within the binding interface to be greater than that of the apolar surface area. This is consistent with the combined CD and 2D IR correlation spectroscopy results obtained for *Hs*Prp40Ap, which showed that this peptide is predominantly random coil in the absence of centrin, while in the presence of centrin *Hs*Prp40Ap adopts a primarily helical structure that later was defined exclusively by 2D IR correlation spectroscopy.

2D IR correlation spectroscopy is especially suited for characterizing conformational changes since it enhances the resolution of the spectral region of interest (1700-1500 cm^{-1}) and provides the sequence of molecular events that occur during thermally induced perturbation. The correlation analysis identified the existence of three salt bridge interactions as requirements for formation of the complex, in addition to high Ca^{2+} affinity in the centrin C-

terminal domain and the hydrophobic triad within the target [Pastrana-Rios et al. 2013,Veeraghavan et al. 2002; Hu and Chazin, 2003; Ortiz et al. 2005 , Pastrana-Rios et al. 2002]. The salt bridges are both inter- and intramolecular in nature and involve the three negatively charged residues in *Hs*Prp40Ap. The molecular description of dissociation of the complex provides clues as to the relative orientation of the *Hs*Prp40Ap with respect to centrin. Briefly, centrin's π-helix, located at the C-terminal end, is perturbed initially followed by its glutamates and aspartates; then the first intermolecular salt-bridge interaction involving Glu_{552} of the peptide and an arginine in centrin is perturbed. This Glu_{552} is located at the C-terminal end of the peptide, suggesting that *Hs*Prp40Ap is oriented tail-to-tail with respect to centrin. Glu_{532}, located at the N-terminal end of *Hs*Prp40Ap and adjacent to a stretch of positively charged residues, is perturbed next; presumably, Glu_{532} is involved in an intra-molecular salt-bridge interaction. This event is followed by centrin's α-helical motifs being perturbed, then by the breaking of the second inter-molecular salt bridge involving an arginine from centrin and the single Asp_{539} residue at the center of *Hs*Prp40Ap. Once the salt-bridge interactions between centrin and *Hs*Prp40Ap no longer exist, centrin's short antiparallel β-sheets and associated loops within the Ca^{2+} binding sites are perturbed, presumably within the C-terminal EF-hand. Then and only then does the peptide target begin to unfold.

In summary, our multidisciplinary approach allowed for fragment-based peptide design based on the hydrophobic triad consistently observed in known centrin targets. Like other centrin-target interactions previously studied, the hydrophobic triad $W_1L_4L_8$ in the CBS of *Hs*Prp40A plays a role in the selectivity of the target, along with the high affinity calcium-

binding site within the C-terminal domain [Hu and Chazin, 2003, Thompson et al. 2006, Sosa et al. 2011]. We used full-length centrin and the *Hs*Prp40Ap so as to be representative of the "*in vivo* complex," since the available high resolution structures for other centrin targets require both centrin domains for complex formation. However, the high affinity calcium-binding sites within the C-terminal domain promote the open conformation of the EF-hand domain, exposing the Phe residues that drive the hydrophobic interaction along with the hydrophobic triad found on one face of the peptide's helix, which may be essential for target selectivity. Also crucial for the formation of the complex were the intermolecular salt-bridge interaction between centrin and Prp40Ap. These results provide direct evidence that Prp40 is a novel centrin target.

CHAPTER IV

MOLECULAR BIOPHYSICAL CHARACTERIZATION OF THE THIRD FF DOMAIN OF *HOMO SAPIENS* PRP40 HOMOLOG A

Published in: Díaz Casas, A., Casanova Sepúlveda, G., Sánchez Negrón, O., Caro Muñiz, A.P., Malavé Ramos, S.R., Cebollero López, A.R., Pastrana-Rios, B. (2018) Molecular biophysical characterization of the third FF domain of *Homo sapiens* Prp40 homolog A. *J. Mol. Struct.* 1167, 174-179.

4.1. Materials and Methods

4.1.1. Expression, isolation, and purification of HsPrp40A's FF_3 domain

*Hs*Prp40A's FF_3 domain was commissioned to Blue Heron Biotech LLC (Bothell, WA). *Escherichia coli* BL21 (DE3) cells from New England Biolabs (Ipswich, MA) were transformed with the desired plasmid construct. The newly transformed bacterial cell cultures were grown in 3 L of terrific broth medium (Fluka, St. Louis, MO) at 250 rpm and 37 °C. The bacterial cultures were induced with 0.5 mM isopropyl β-D-thiogalactoside at the onset of the log phase. Cells were harvested by centrifugation once the stationary phase was reached, which was 4 h after induction yielding 23 g of pelleted cells.

The harvested pellet was lysed using cold buffer solution containing 20 mM 2-amino-2-(hydroxymethyl)propane-1,3-diol (Tris) at pH 7.4, 500 mM NaCl, 0.1% octylphenoxypolyethoxyethanol (IGEPAL), and 10 mM dithiothreitol (DTT). To minimize proteolytic cleavage, a Pierce protease inhibitor cocktail was added which contained AEBSF, aprotinin, bestatin, E-64, leupeptin, and pepstatin A (ThermoFischer Scientific, Rockford, IL). The crude lysate was subjected to centrifugation using a JA-14 rotor and a Beckman J2-MC centrifuge at 29,994 x g for 40 min at 4 °C. The supernatant was filtered with a 70 kDa membrane using a Pall Minim II Tangetial Flow Filtration System from Pall Corp. (Ann Arbor, MI). Then, the filtered sample was buffer exchanged and concentrated using a 3 kDa membrane (Pall Corp., Ann Arbor, MI).

For purification, the clarified sample was then subjected to diethylaminoethyl (DEAE) cellulose anion exchange chromatography. The column was equilibrated using 20 mM Tris at pH 7.4, 100 mM NaCl, and 1 mM DTT. The gradient of NaCl used was from 100 mM to 1 M. The eluted fractions were analyzed by SDS-PAGE and fractions containing the *Hs*Prp40A's FF_3 domain were pooled. As a final purification step, the pooled sample was filtrated and then concentrated using a 50 and 5 kDa membranes, respectively, using an Amicon stirred ultrafiltration apparatus from EMD Millipore (Darmstadt, Germany). During this step, the buffer was exchanged to 50 mM HEPES at pH 7.4, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$. The concentration of the FF_3 domain was determined using a Jasco model V-560 UV/Vis spectrophotometer (Jasco Corporation, Tokyo, Japan). The calculated molar extinction

coefficient for *Hs*Prp40A's FF_3 domain was $\varepsilon = 13{,}980\ M^{-1}\ cm^{-1}$ at 280 nm. The FF_3 domain was subsequently sent for partial amino acid sequencing and MS analysis.

4.1.2 Differential scanning calorimetry

*Hs*Prp40A's FF_3 domain (0.06 mM) in 50 mM HEPES at pH 7.4, 150 mM NaCl, 4 mM $CaCl_2$, 4 mM $MgCl_2$, and 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP HCl) was analyzed with a VP-DSC microcalorimeter from MicroCal LLC (Northampton, MA). Thermograms were collected at 25 psi and a scan rate of 60 °C/h over a temperature range of 10-95 °C, with a 16 s filtering period. The data analysis was performed using Origin from MicroCal. The sample thermogram was reference subtracted and a progressive baseline was obtained to allow for accurate determination of the change in enthalpy of denaturation (ΔH_D) and the T_m.

4.1.3 Circular dichroism spectroscopy

Far-UV CD spectra were acquired using a Jasco J-810 spectropolarimeter (Jasco Corporation, Tokyo, Japan) for *Hs*Prp40A's FF_3 domain (6 μM) in 8 mM HEPES at pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$, and 2 mM $MgCl_2$. Six scans were co-added within the spectral range of 250-190 nm at a scan rate of 20 nm/min, a response time of 2 s, and a bandwidth of 2 nm at different equilibrium temperatures: 5, 25, and 85 °C. In addition, a temperature dependence spectrum was also collected from 0 to 85 °C at a rate of 1 °C/min, while monitoring

at 222 nm. Baseline correction was performed over the spectral range of 250-240 nm. The CD absorbance was converted to mean residue molar ellipticity units to analyze the helical content of the protein.

FT-IR spectroscopy

*Hs*Prp40A's FF_3 domain was fully H→D exchanged following a procedure well established by our laboratory [Pastrana-Rios, 2001; Pastrana-Rios et al. 2002; Ortíz et al. 2005; Pastrana-Rios, 2006; Sosa et al. 2011; Pastrana-Rios et al. 2013; Díaz Casas et al. 2017]. The recombinant FF_3 domain at a final concentration of 120 mg/mL in 50 mM HEPES, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$ at pD 6.6 was deposited onto a 49 mm x 4 mm custom-milled CaF_2 window with a fixed path length of 40 µm from Spectral System LLC (Hopewell Junction, NY). Similar to the sample, 80 µL of the buffer was deposited onto reference cell. Both cells were set in a custom dual chamber cell holder. The temperature within the cell holder was controlled with a Neslab RTE-740 refrigerated circulating bath (Thermo Electron Corp., Newington, NH). The temperature was monitored with a thermocouple positioned in close contact with the cell. Spectral data was acquired at the desired preset temperature only after 10 min equilibration time was achieved. The temperature range studied was 5-90 °C with 5 °C intervals. The instrument used was a Jasco FT-IR spectrophotometer model 6200 (Jasco Corporation, Tokyo, Japan) equipped with a MCT detector, sample shuttle, and interface. In this experiment, 512 scans were co-added, apodized with a triangular function, and Fourier transformed to provide a resolution of 4 cm^{-1} with the data encoded every 2 cm^{-1}.

4.1.4 2D IR correlation spectroscopy

For the generation of the 2D IR correlation spectra and the determination of the molecular sequence of events of *Hs*Prp40A's FF_3 domain, we followed the protocol presented in Chapter III.

4.2 Results

4.2.1 Biochemical characterization of HsPrp40A's FF_3 domain

A representation of the full-length *Hs*Prp40A, which includes its available structural data, is presented in Figure 26A. In this work, the design construct of *Hs*Prp40A's FF_3 domain is comprised of the sequence from Ser_{516} to Glu_{593} (Figure 26B). Once the FF_3 domain was purified, the sample was sent for a partial amino acid sequencing and the results are shown in Figure 27A. The loss of methionine at the N-terminal end coincides with the expected amino acid sequence as presented in the Uniprot accession number O75400. Figure 27B shows the SDS-PAGE analysis of the purified FF_3 domain. A band near the expected molecular weight (approximately 9 kDa) was observed. *Hs*Prp40A's FF_3 domain was determined to be in its monomeric form, with an empirical *m/z* value of 9337.00 (Figure 27C). The calculated molecular weight for *Hs*Prp40A's FF_3 domain is 9338.35 Da. The combined sequencing data

along with the single SDS-PAGE band observed with the expected molecular weight and the MS result suggests purity levels to be greater than 98%.

Circular dichroism spectroscopy

CD spectroscopy was performed to determine the extent of helical content of the FF_3 domain. The CD spectra acquired (Figure 28) suggest that the secondary structure of the FF_3 domain in solution is predominantly α-helical. These results are consistent with available high resolution structural data on FF domains, which show that the structure of these domains has a high percentage of helical content [Allen et al. 2002; Gasch et al. 2006; Bonet et al. 2008; Bonet et al. 2009a; Bonet et al. 2009b; Lu et al. 2009; Liu et al. 2013]. The mean residue molar ellipticity $[\Theta]_{mr}$ at 222 nm was determined to be -22,689.7 (5 $^{\circ}$C), -21,163.6 (25 $^{\circ}$C), and -9,736.6 (85 $^{\circ}$C) degrees•cm^2•$dmol^{-1}$, suggesting a loss in helical contribution as a function of increasing temperature. The helical content of the FF_3 domain at 25 $^{\circ}$C was about 63.3%.

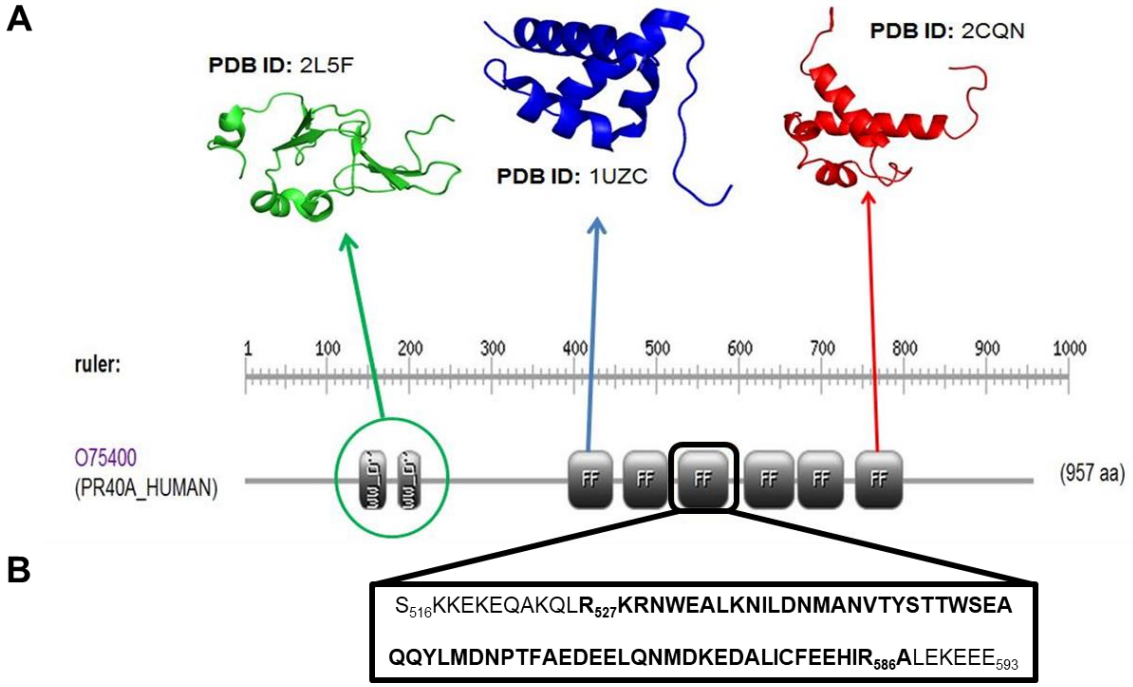
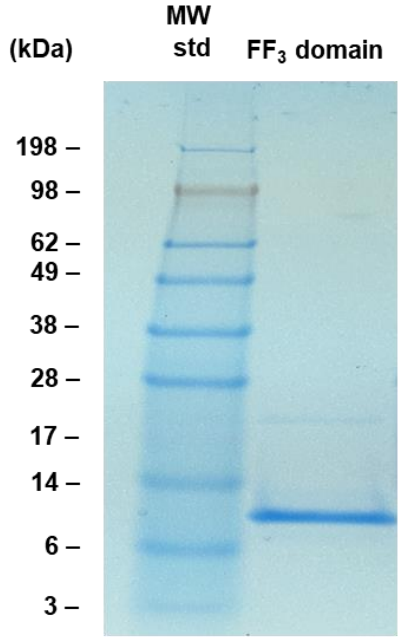


Figure 26. Representation of *Hs*Prp40A with its domains. **(A)** NMR ribbon structures of the two tandem WW domains **(PDB: 2L5F)**, FF_1 domain **(PDB: 1UZC)**, and FF_6 domain **(PDB: 2CQN)** are presented in green, blue and red colors, respectively [Jiang et al. 2011; Allen et al. 2002, Bonet et al. 2008]. **(B)** The amino acid sequence of the FF_3 domain of *Hs*Prp40A is presented in the black box. The amino acids corresponding to the structural FF_3 domain are in **bold**.

A

Protein	Uniprot Accession Number	1				5					10
Hs Prp40A FF3 Domain	O75400	M	S	K	K	E	K	E	Q	A	K	Q
			S	K	K	E	K	E	Q	A	K	Q

B



C

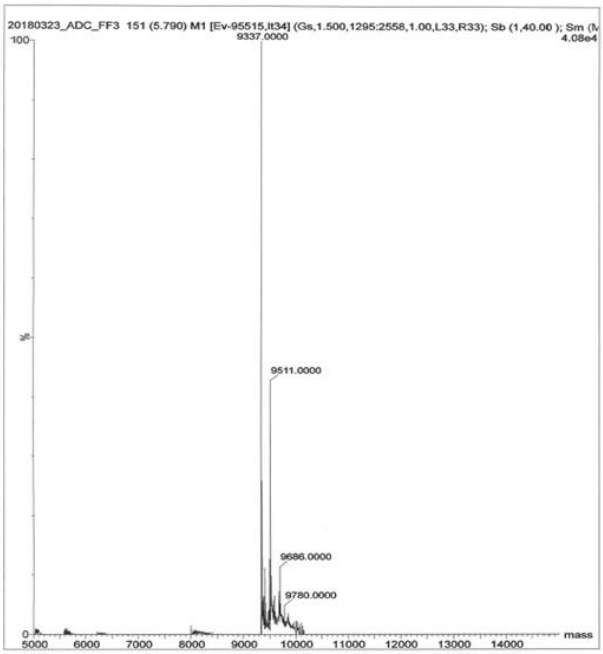


Figure 27. Biochemical characterization of *Hs*Prp40A's FF_3 domain. **(A)** Partial amino acid sequencing of the N-terminal of FF_3 domain, validating its identity and the loss of M_1. **(B)** SDS-PAGE of the purified FF_3 domain. The molecular weight standard (MW std) used was SeeBlue Plus 2 Prestained Standard from Novex (Carlsbad, CA). **(C)** MS analysis showed that the molecular weight of *Hs*Prp40A's FF_3 domain was 9337.0 *m/z* [Adapted from Díaz Casas et al. 2018 with permission from publisher Elsevier Inc.].

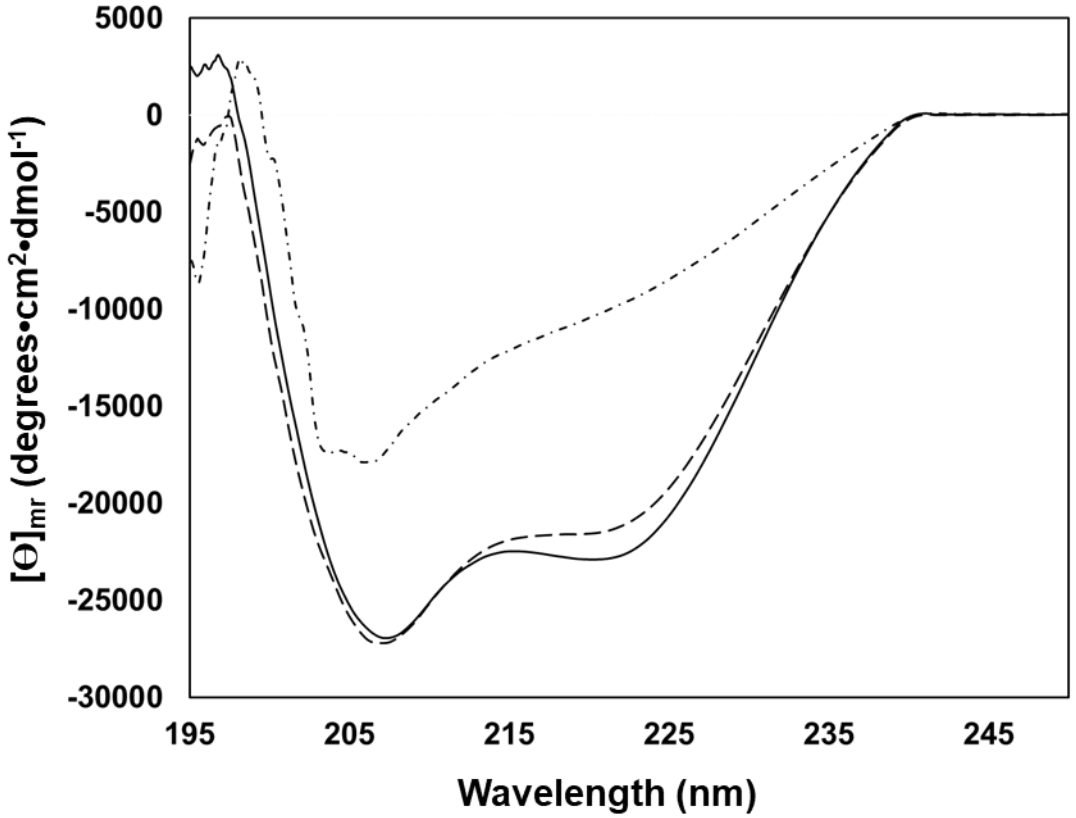


Figure 28. Far-UV CD overlaid spectra of *Hs*Prp40A's FF_3 domain at 5 (–), 25 (– – –), and 85 °C (– · –) within the spectral region of 195–250 nm. Typical helical spectral features are observed [Adapted from Díaz Casas et al. 2018 with permission from publisher Elsevier Inc.].

4.2.2 FT-IR spectroscopy

FT-IR spectra of *Hs*Prp40A's fully H→D exchanged FF_3 domain were analyzed within the spectral region of 1735-1500 cm^{-1},contains the amide I' and side chain bands. The amide I' band observed in the spectral region of 1735-1600 cm^{-1} is composed primarily of carbonyl stretching vibrations (ν(C=O)) associated with the backbone of the domain and is highly sensitive to conformational changes [Bandekar. 1992]. whereas the spectral region of 1600-1500 cm^{-1} exhibits side chain vibrational modes which are sensitive to weak interactions. The side chain modes are observed in this spectral region due to the H→D exchange of the amides within the FF_3 domain, resulting in a shift of the amide II band from 1550 cm^{-1} to approximately 1450 cm^{-1}. As a result, the N-D deformation modes within the peptide bond do not overlap with the side chain modes, allowing for an in-depth analysis of the effects of thermal stress on the side chain modes of the FF_3 domain.

For the band assignments, we used the asynchronous plot, which provides enhanced spectral resolution (Figure 29A-C), resulting in the following assignments for the amide I' band: β-turn (1674 cm^{-1}), 3_{10}-helix (1664 cm^{-1}), and α-helix (1643 cm^{-1}). These band assignments are consistent with the known high resolution structures available for other FF domains [Allen et al. 2002; Gasch et al. 2006; Bonet et al. 2008; Bonet et al. 2009a] and are in agreement with band assignments made for proteins in general [Chirgadze et al. 1975; Venyaminov and Kalnin, 1990; Bandekar, 1992; Arrondo et al. 1993; Haris and Chapman, 1995; Barth, 2007]. This FF_3 domain also contains secondary structure contributions from β-

sheet at 1630 cm^{-1}, which may be an artifact of the additional sequence included in this fragment to ensure the FF_3 domain's bacterial expression.

Five side-chain bands were observed in the spectral region of 1600-1500 cm^{-1}: arginine's guanidinium v_a(N–D) (1615 cm^{-1}), aspartates $v(COO^-)$ (1579 cm^{-1}), glutamates $v(COO^-)$ (1558 and 1541 cm^{-1}) and a tyrosine v(CC) (1520 cm^{-1}). The carboxylates of the glutamates have two distinct vibrational modes, suggesting they are located in two different chemical environments. The glutamates from the N-terminal end are located in a predominantly positively charged area within the first α-helix, whereas the glutamates from the C-terminal end are located in a predominantly negatively charged area within the third α-helix.

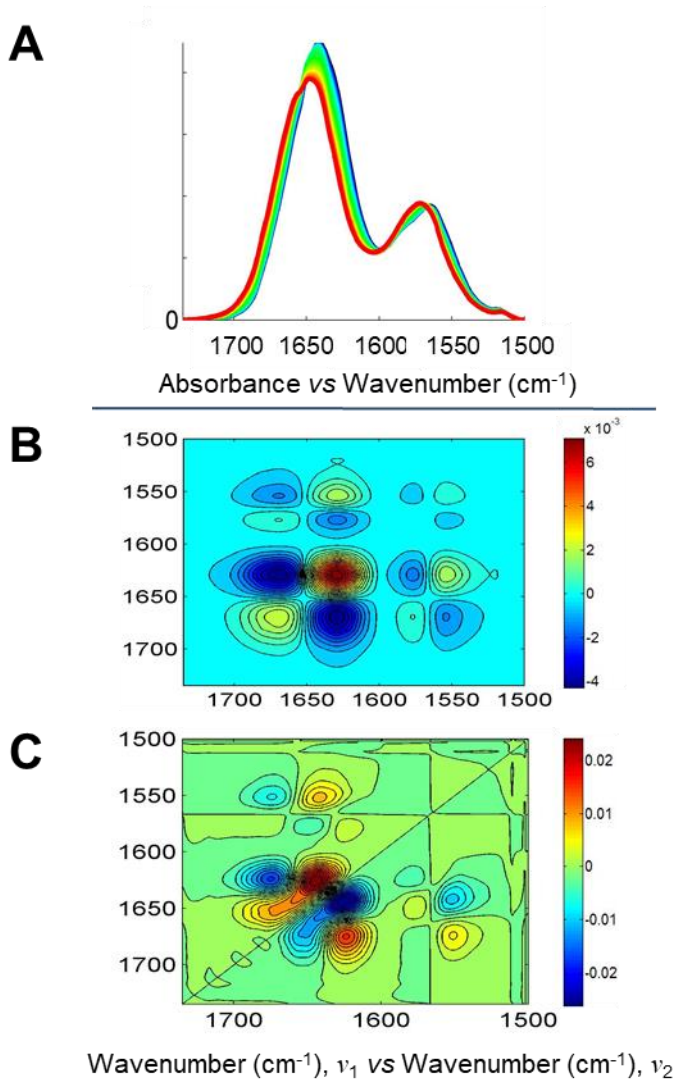


Figure 29. 2D IR correlation spectroscopy of *Hs*Prp40A's FF_3 domain. **(A)** Overlaid FT-IR spectra, **(B)** synchronous, and **(C)** asynchronous plots within the spectral region of 1735–1500 cm^{-1} and over the 5-90 °C temperature range and 5 °C intervals. In the overlaid FT-IR spectra, the black and red lines correspond to 5 °C and 90 °C, respectively [Adapted from Díaz Casas et al. 2018 with permission from publisher Elsevier Inc.].

4.2.3 Thermal dependence study of HsPrp40A's FF_3 domain

The thermal unfolding process for *Hs*Prp40A's FF_3 domain was analyzed by DSC and the result is shown in Figure 30A. The unfolding process for *Hs*Prp40A's FF3 domain was determined to be endothermic (ΔH_D = 54.6 kcal/mol), and the T_m obtained was 50.1 °C. In addition, the helical content of *Hs*Prp40A's FF_3 domain was monitored using CD spectroscopy, while FT-IR provided an in-depth analysis as to the entire structural rearrangement of the domain as a function of temperature by simply monitoring the amide I' band. The thermal dependence plots are shown in Figure 30B. These plots are consistent in that a pre-transition is observed, but while the thermal dependence in the CD spectrum is observed to plateau, the FT-IR amide I' band is observed to continue to transition, suggesting other secondary structural contributions are still transitioning into the random coil state. In both cases, a sigmoidal curve is observed suggesting that cooperativity defines the thermal unfolding process. Moreover, the inflexion points in both the CD and FT-IR plots are shown to be consistent with the thermal transition temperature (T_m) observed in the DSC analysis. FT-IR provides greater selectivity and sensitivity than CD to define this cooperative process at the molecular level.

4.2.4 2D IR correlation spectroscopy

2D IR correlation spectroscopy has proven essential to understanding the flexibility and dynamics of the FF_3 domain under thermal stress. The synchronous and asynchronous

plots are shown in Figure 29B and C. By applying Noda's rules [Noda, 2015], we were able to determine the sequential order of molecular events for *Hs*Prp40A's FF_3 domain (Figure 31).

The least stable component is the β-sheet (1630 cm^{-1}), which may be located at either the N- or C-terminal end due to the additional residues contained in the FF_3 domain. Then the tyrosines (1520 cm^{-1}) are perturbed, followed by the glutamates (1541 cm^{-1}) and arginine ν_a(N–D) (1615 cm^{-1}), suggesting disruption of the intramolecular salt-bridge [Pastrana-Ríos et al. 2001] that is most likely located in the C-terminal end, within the last α-helix of the domain. Presumably, the arginine at position 586 is the residue involved in this interaction, as it is surrounded by a glutamate-rich region. Then, other glutamates (1558 cm^{-1}) are perturbed, followed by the α-helix (1643 cm^{-1}). The perturbation of the α-helix affects the β-turn (1674 cm^{-1}) stability, followed by the 3_{10}-helix (1664 cm^{-1}). This β-turn may be located between the second α-helix and the 3_{10}-helix, where a proline residue is located. Finally, the aspartates (1579 cm^{-1}) are perturbed. After these events occur, the FF_3 domain proceeds to adopt a more disordered conformation during the thermal stress leading up to the denatured state.

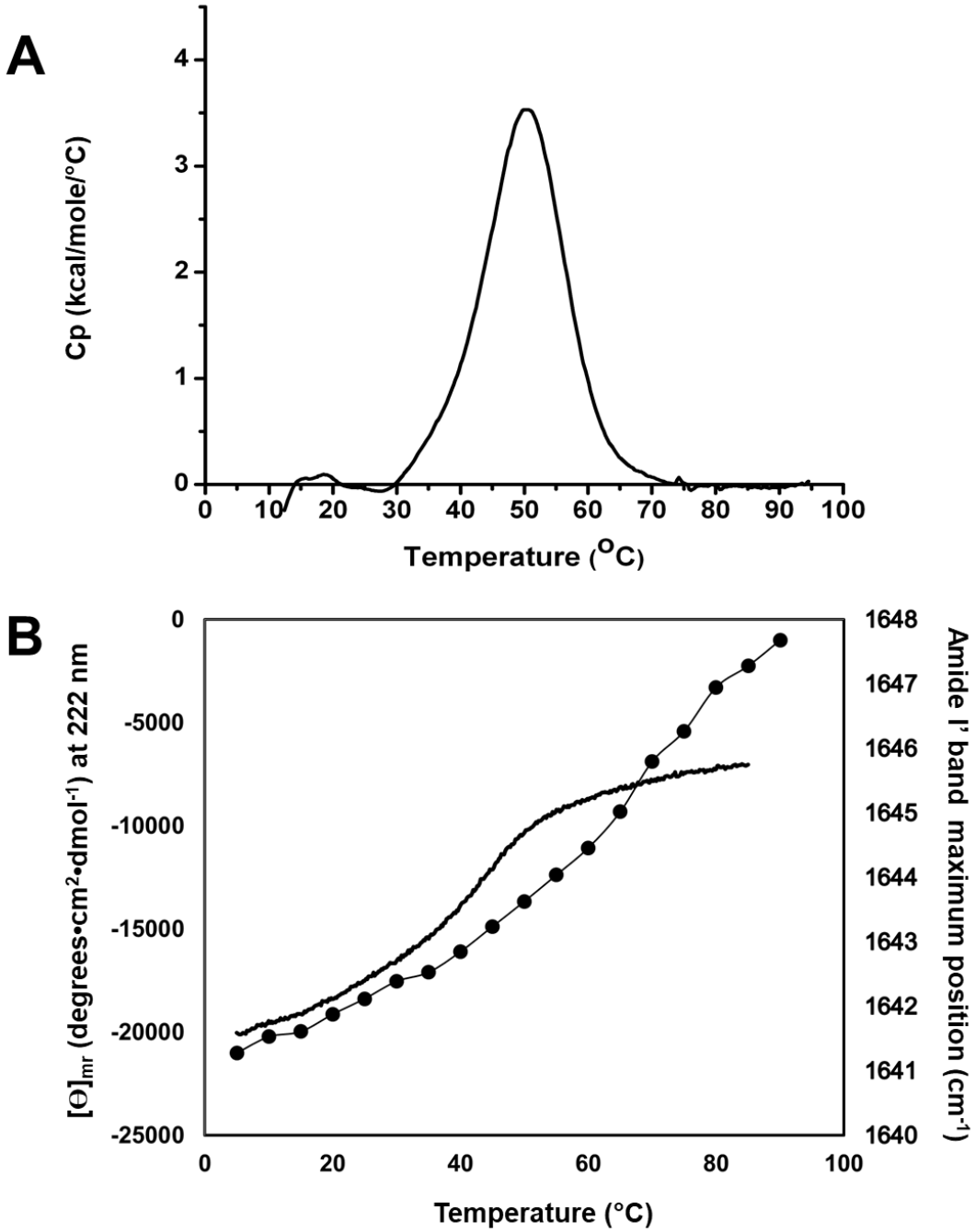


Figure 30. Thermal dependence plots of *Hs*Prp40A's FF_3 domain. **(A)** DSC thermogram where the T_m value is 50.1 °C. **(B)** Thermal dependence plots for CD and FT-IR, using the $[\Theta]_{mr}$ at 222 nm (thick line), and the amide I' band maximum position (line and solid circles), respectively [Adapted from Díaz Casas et al. 2018 with permission from publisher Elsevier Inc.].

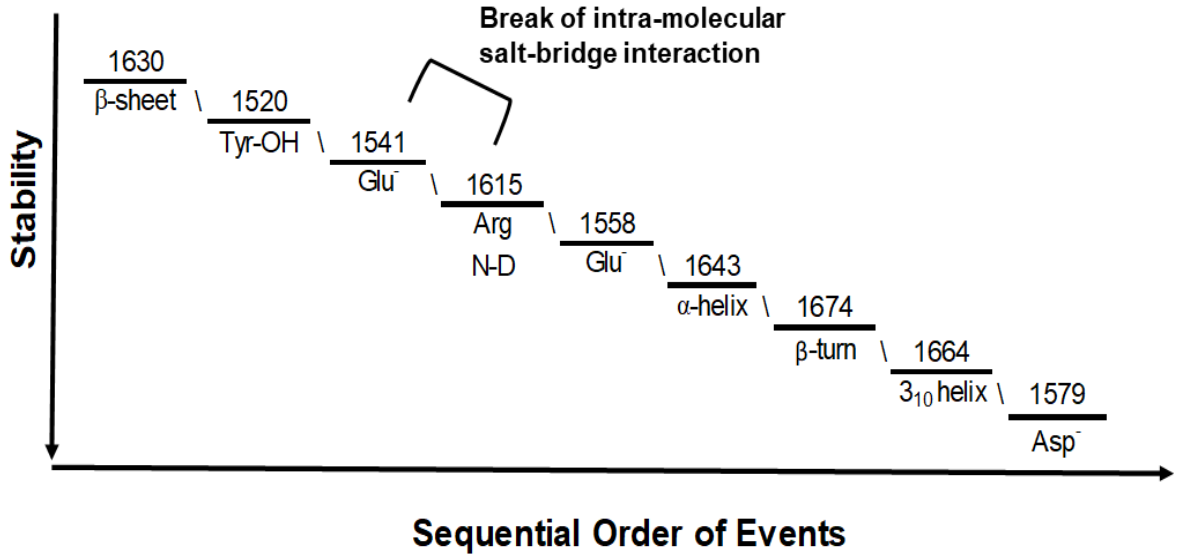


Figure 31. Sequential order of molecular events during the thermal perturbation of *Hs*Prp40A's FF_3 domain over the temperature range of 5-90 °C [Adapted from Díaz Casas et al. 2018 with permission from publisher Elsevier Inc.].

CHAPTER V

CONCLUSION

The main objective of this dissertation was to validate, through different molecular biophysical techniques, that Prp40 is a novel centrin-binding protein. First, we were able to identify a potential CBS within *Hs*Prp40, which has the hydrophobic triad $W_1L_4L_8$ consensus sequence, known to be important for centrin selectivity for its targets. Interestingly, this CBS is highly conserved among higher eukaryotes (i.e. mammalian). In addition, it is localized within the FF_3 domain and is conserved with the first NES of *Sc*Prp40.

The results found by bioinformatics guided us to design a synthetic peptide of the putative CBS containing the hydrophobic triad $W_1L_4L_8$. The interaction between centrin and *Hs*Prp40Ap was validated *in vitro*, using ITC and FT-IR as two orthogonal techniques. By ITC, the thermodynamics governing binding between centrin and *Hs*Prp40A were determined; these showed a high affinity interaction between centrin and *Hs*Prp40A. The comparative thermodynamic analysis using full-length *Hs*cen1, *Hs*cen2, and *Cr*cen showed that *Hs*Prp40 interacts with centrin in its C-terminal domain and that this interaction is dependent on calcium affinity. This is consistent with the mode of interaction of other centrin-binding proteins, such as Kar1. Centrin's affinity for calcium in its C-terminal domain results in an open conformation that allows the hydrophobic triad of *Hs*Prp40Ap to interact with the exposed hydrophobic surface of centrin. In addition, by 2D IR correlation spectroscopy, we were able to determine

the existence of two intermolecular salt-bridge interactions that were also crucial for the formation of the complex. Moreover, 2D IR correlation spectroscopy allowed us to describe the molecular sequence of events for the dissociation of the ^{13}C-*Cr*cen-*Hs*Prp40Ap complex as well as the unfolding of the FF_3 domain under thermal stress. In both cases, 2D IR correlation spectroscopy was found to be an excellent technique to study the molecular behavior of proteins. In conclusion, the use of multidisciplinary approaches presented in this work allowed us to provide evidence that *Hs*Prp40A is a centrin-binding protein.

.

CHAPTER VI

FUTURE WORK

As our next step, we are interested in determining high resolution structures of *Hs*Prp40A's FF_3 domain and the *Hs*cen2-*Hs*Prp40Ap complex. Further investigation of the interaction between *Hs*cen2 mutants and *Hs*Pr40Ap by ITC will be very useful in order to determine key amino acids that affect the binding of *Hs*cen2 to *Hs*Prp40Ap. We will validate the key residues involved in the interaction interface and the critical salt-bridge interaction for complex formation. In addition, it is of great interest to validate the existence of the interaction between *Hs*cen3 and *Hs*Prp40Ap.

The use of bioinformatics tools allowed us to define the hydrophobic triad $W_1L_4L_8$ of *Hs*Prp40Ap is highly conserved with the NES_1 of *Sc*Prp40. Interestingly, Resendes and colleagues showed that *Hs*cen2 interacts with the NES sequence of HIV Rev protein. Moreover, *Sc*Prp40 NES_1 is essential for proper pre-mRNA splicing. Thus, it will be very important to validate if Cdc31 could bind to the leucine-rich NES of *Sc*Prp40.

Based on this line of evidence and on the results obtained in this dissertation, there is a question still needing to be answered: Is centrin playing a regulatory role in Prp40 export, and consequently, in pre-mRNA splicing? To answer this question, *in vivo* splicing assays and

other cellular studies are required in order to study the biological implications of the interaction between centrin and Prp40.

REFERENCES

In alphabetical order

1. Abovich, N. and Rosbach, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*. 89, 403-412.

2. Alfaro, E., Sosa, L del V., Sanoguet, Z., Pastrana-Rios, B., and Schreiter, E.R. (2008) Crystallization and preliminary X-ray characterization of full-lenght *Chlamydomonas reinhardtii* centrin. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 64, 402-404.

3. Allen, M., Friedler, A., Schon, O., and Bycroft, M. (2002) The structure of a FF domain from human HYPA/FBP11. *J. Mol. Biol.* 323, 411-416.

4. Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001) Centrosome protein centrin 2/caltractin is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J. Biol. Chem.* 276, 18665-18672.

5. Arrondo, J.L., Muga, A., Castresana, J., and Goñi, F.M. (1993) Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* 59, 23-56.

6. Avasthi, P., Scheel, J.F., Ying, G., Frederick, J.M., Baehr, W., and Wolfrum, U. (2013) Germline deletion of Cetn1 causes infertility in mala mice. *J. Cell. Sci.* 126, 3204-3213.

7. Azimzadeh, J., Hergert, P., Delouvée, A., Euteneuer, U., Formstecher, E., Khodjakov, A., and Bornes, M. (2009) hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. *J. Cell. Biol.* 185, 101-114.

8. Badano, J.L., Mitsuma, N., Beales, P.L., and Katsanis, N. (2006) The ciliopathies: an emerging class of human genetic disorders. *Annu. Rev. Genomics Hum. Genet.* 7, 125-148.

9. Bai, S.W., Herrera-Abreu, M.T., Rohn, J.L., Racine, V., Tajadura, V., Suryavanshi, N., Bechtel, S., Wiemann, S., Baum, B., and Ridley, A.J. (2011) Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* 9:54.

10. Bandekar, J. (1992) Amide modes and protein conformation. *Biochim. Biophys. Acta.* 1120, 123-143.

11. Baron, A.T. and Salisbury, J.L. (1992) Role of centrin in spindle pole dynamics. In: Kalnins, V.I. (ed.). The centrosome. Academic Press, Inc. pp. 167-195.

12. Barth, A. (2007) Infrared spectroscopy of proteins. *Biochim. Biophys. Acta.* 176, 1073-1101.

13. Bastide, K., Ugolin, N., Levalois, C., Bernaudin, J.F., and Chevillard, S. (2010) Are adenosquamous lung carcinomas a simple mix of adenocarcinomas and squamous cells carcinomas, or more complex at the molecular level? *Lung Cancer*. 68, 1-9.

14. Becerra, S., Andrés-León, E., Prieto-Sánchez, S., Hernández-Munain, C., and Suñé, C. (2016) Prp40 and early events in splice site definition. *Wiley Interdiscip. Rev. RNA.* 7, 17-32.

15. Becerra, S., Montes, M., Hernández-Munain, C., and Suñé, C. (2015) Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternatively pre-mRNA splcing in vivo. *RNA*. 21, 438-457.

16. Bedford, M. T. and Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci*. 24, 264–265.

17. Bedford, M.T., Chan, D.C., and Leder, P. (1997) FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *EMBO J.* 16, 2376-2383.

18. Bedford, M.T., Reed, R., and Leder, P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. *Proc. Natl. Acad. Sci. USA.* 95, 10602-10607.

19. Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007) Composition of the three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* 26, 1737-1748.

20. Beltran-Parrazal, L., Acuna, D., Ngan, A.M., Kim, E., Ngan, A., Kawakami, K., Edmond, J., and Lopez, I.A. (2010) Neuglobin, cytoglobin, and transcriptional profiling of hypoxia-related genes in the rat cerebellum after prenatal chronic very mild carbon monoxide exposure (25 ppm). *Brain Res.* 1330, 61-71.

21. Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbach M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell.* 89, 781-787.

22. Bettencourt-Dias, M., Hildebrant, F., Pellman, D., Woods, G., and Godinho, S.A. (2011) Centrosomes and cilia in human disease. *Trends Genet.* 27, 307-315.

23. Biggins, S. and Rose, M.D. (1994) Direct interaction between yeast spindle pole body components: Kar1p is required for Cdc31p localization to the spindle pole body. *J. Cell Biol.* 125, 843-852.

24. Bogerd, H.P., Fridell, R.A., Benson R.E., Hua, J., and Cullen, B.R. (1996) Protein sequence requirement for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay. *Mol. Cell Biol.* 16, 4207-4214.

25. Bonet, R., Ramirez-Espain, X., and Macias, M.J. (2008) Solution structure of the yeast URN1 splicing factor FF domain: comparative analysis of charge distribution in FF domain structures-FFs and SURPs, two domains with similar fold. *Proteins.* 73, 1001-1009.

26. Bonet, R., Ruiz, L., Aragón, E., Maltín-Malpartida, P., and Macias, M.J. (2009b) NMR structural studies on human p190 RhoGAPFF1 revealed that domains phosphorylation by the PDGF-receptor alpha requires its previous unfolding. *J. Mol. Biol.* 389, 230-237.

27. Bonet, R., Ruiz, L., Morales, B., and Macias, M.J. (2009a) Solution structure of the fourth FF domain of yeast Prp40 splicing factor. *Proteins.* 77, 1000-1003.

28. Bork, P. and Sudol, M. (1994) The WW domain: a signaling site in dystrophin? *Trends Biochem. Sci.* 19, 531-533

29. Bushdord, J.P. and Strätling, W.H. (2004) A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J. Mol. Med.* 82, 135-143.

30. Carty, S.M. and Greenleaf, AL. (2002) Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/Chromatin modification and RNA processing. *Mol. Cell Proteomics.* 1, 598-610.

31. Casey, J.R., Grinstein, S., and Orlowski, J. (2010) Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell. Biol.* 11, 50-61.

32. Chan, D.C., Bedford, M.T., and Leder, P. (1996) Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. *EMBO J.* 15, 1045-1054.

33. Chattopadhyaya, R., Meador, W.E., Means, A.R., and Quiocho, F.A. (1992) Calmodulin structure refined at 1.7 Å resoluction. *J. Mol. Biol.* 228, 1177-1192.

34. Chirgadze, Y.N., Fedorov, O.V., and Trushina, N.P. (1975) Estimation of amino acid residue side-chain absorption in the infrared spectra of protein solutions in heavy water. *Biopolymers*. 14, 679-694.

35. Chung, S., McLean, M.R., and Rymond, B.C. (1999) Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA*. 5, 1042-1054.

36. Cox, J.A., Tirone, F., Durussel, I., Firanescu, C., Blouquit, Y., Duchambon, P., and Craescu, C.T. (2005) Calcium and magnesium binding to human centrin 3 and interaction with target peptides. *Biochemistry*. 44, 840-850.

37. Cunningham, C.N., Schmidt, C.A., Schramm, N.J., Gaylord, M.R., and Resendes, K.K. (2014) Human TREX2 components PCID2 and centrin 2, but not ENY2, have distinct

functions in protein export and co-localize to the centrosome. *Exp. Cell Res.* 320, 209-218.

38. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995) Identification of galectin-3 as a factor in pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA*. 92, 1213-1217.

39. Dantas, T.J., Daly, O.M., Convoy, P.C., Tomas, M., Wang, Y., Lalor, P., Dockery, P., Ferrando-May, E., and Morrison, C.G. (2015) Calcium-binding capacity of centrin2 is required for linear POC5 assembly but not nucleotide excision repair. *PLoS One*. 8, e68487.

40. De, M., Oleskie, A.N., Ayyash, M., Dutta, S., Mancour, L., Abazeed, M.E., Brace, E.J., Skiniotis, G., and Fuller, R.S. (2017) The Vps13p-Cdc31p complex is directly required for TGN late endosome transport and TGN homotypic fusion. *J. Cell Biol.* 216, 425-439.

41. Denaval, B., Covassin, L., Lawson, N.D., and Doxsey, S. (2011) Centrin depletion causes cyst formation and other ciliopathy-related phenotypes in zebrafish *Cell Cycle*. 10, 3964-3972.

42. Díaz Casas, A, Chazin, W.J., and Pastrana-Rios, B. (2017) Prp40 homolog A is a novel centrin target. *Biophys. J.* 112, 2529-2539.

43. Díaz Casas, A., Casanova Sepúlveda, G., Sánchez Negrón, O., Caro Muñiz, A.P., Malavé Ramos, S.R., Cebollero López, A.R., and Pastrana-Rios, B. (2018) Molecular biophysical characterization of the third FF domain of *Homo sapiens* Prp40 homolog A. *J. Mol. Struct.* 1167, 174-179.

44. Ester, C. and Uetz, P. (2008) The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochemistry*. 9, 1-11.

45. Faber, P.W., Barnes, G.T., Srinidhi, J., Chen, J., Gusella, J.F., and MacDonald, M.E. (1998) Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* 7, 1463-1474.

46. Fischer, T., Rodríguez-Navarro, S., Rácz, A., Schiebel, E., and Hurt, E. (2004) Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery. *Nat. Cell Biol.* 6, 840-848.

47. Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W., and Lührmann, R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*. 82, 475-483.

48. Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*. 90, 1051-1060.

49. Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. (1997) CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature*. 390, 308-311.

50. Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L., and Macias, M.J. (2006) The structure of Prp40 FF1 domain and its interaction with the crn-TRP1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem.* 281, 356-364.

51. Gavet, O., Alvarez, C., Gaspar, P., and Bornes, M. (2003) Centrin4p, a novel mammalian centrin specifically expressed in ciliated cells. *Mol. Biol. Cell.* 14, 1818-1834.

52. Giessl, A., Pulvermüller, A., Trojan, P., Park, J.H., Choe, H.W., Ernst, O.P., Hofmann, K.P., and Wolfrum, U. (2004a) Differential expression and interaction with the visual G-protein transducing of centrin isoforms in mammalian photoreceptor cells. *J. Biol. Chem.* 279, 51472-51481.

53. Giessl, A., Trojan, P., Puvermüller, A., and Wolfrum, U. (2004b) Centrins, potential regulators of transducin translocation in photoreceptor cells. In D.S. Williams (Ed.), *Cell biology andrelated disease of the outer retina* (pp. 122-195). Singapore: World Scientific Publishing Company Pte. Ltd.

54. Giessl, A., Trojan, P., Rausch, S., Pulvermüller, A., and Wolfrum, U. (1996) Centrins, gatekeepers for the light-dependent translocation of transducing through the photoreceptor cell connecting cilium. *Vision Res.* 46, 4502-4509.

55. Goldstrohm, A.C., Albrecht, T.R., Suñé, C., Bedford, M.T., and Garcia-Blanco, M.A. (2001) The transcription elongation factor CA150 interacts with RNA Polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* 21, 7617-7628.

56. González-Aguilera, C., Tous, C., Gómez-González, B., Huertas, P., Luna, R., and Aguilera, A. (2008) The THP1-SAC3-SUS1-CDC31 complex Works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol. Biol. Cell*. 19, 4310-4318.

57. Görnemann, J., Barrandon, C., Hujer, K., Rutz, B., Rigaut, G., Kotovic, K.M., Faux, C., Neugebauer, K.M., and Séraphin, B. (2011) Cotranscriptional spliceosome assembly and slicing are independent of the Prp40 WW domain. *RNA.* 17, 2119-2129.

58. Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer G., Colot, H.V., Mann, M., Séraphin, B., Rosbach, M., Lührmann, R., and Fabrizio, P. (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*. 4, 374-393.

59. Haris, P.I. and Chapman, D. (1995) The conformational analysis of peptides using Fourier transform IR spectroscopy. *Biopolymers.* 37, 251-263.

60. Haris, P.I., Robillard, G.T., van Dijk, A.A., and Chapman, D. (1992) Potential of ^{13}C and ^{15}N labeling for studying protein-protein interactions using Fourier transform infrared spectroscopy. *Biochemistry*. 31, 6279-6284.

61. Hart, P.E., Glantz, J.N., Orth, J.D., Poynter, G.M., and Salisbury, J.L. (1999) Testis-specific murine centrin, Cetn1: genomic characterization and evidence for retroposition of a gene encoding a centrosome protein. *Genomics.* 60, 111-120.

62. Hartman, H. and Federov, A. (2002) The origin of the eukaryotic cell: A genomic investigation. *Proc. Natl. Acad. Sci. USA.* 99, 1420-1425.

63. Hildebrandt, F., Benzing, T., and Katsanis, N. (2011) Ciliopathies. *N. Engl. J. Med.* 364, 1533-1543.

64. Hinduja, I., Baliga, N.B., and Zaveri, K. (2010) Correlation of human sperm centrosomal proteins with fertility. *J. Hum. Reprod. Sci.* 3, 95-101.

65. Hsu, P.H., Chiu, Y.C., Lin, T.F., and Jeng, C.J. (2017) Ca^{2+}-binding protein centrin 4 is a novel binding partner of rat Eag1 K^+ channels. *FEBS Open Bio.* 6, 349-357.

66. Hu, H. and Chazin, W.J. (2003) Unique features in the C-terminal domain provide caltractin with target specificity. *J. Mol. Biol.* 330, 473-484.

67. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., Sougnez, C., Auclair, D., Lawrence, M.S., Stojanov, P., Cibulski, K., Choi, K., de Waal, L., Sharifnia, T., Brooks, A., Greulich, H., Banerji, S., Zander, T., Seidel, D., Leenders, F., Ansén, S., Ludwig, C., Engel-Riedel, W., Stoelben, E., Wolf, J., Goparju, C., Thompson, K., Winckler, W., Kwiatkowski, D., Johnson, B.E., Jänne, P.A., Miller, V.A., Pao, W., Travis, W.D., Pass, H.I., Gabriel, S.B., Lander, E.S., Thomas, R.K., Garraway, L.A., Getz, G., and Meyerson, M. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 150, 1107-1120.

68. Ingham, R.J., Colwill, K., Howard, C., Dettwiler, S., Lim, C.S., Yu, J., Hersi, K., Raaijmakers, J., Gish, G., Mbamalu, G., Taylor, L., Yeung, B., Vassilovski, G., Amin, M., Chen, F., Matskova, L., Winberg, G., Ernberg, I., Linding, R., O'donnel, P., Starostine, A., Keller, W., Metalnikov, P., Stark. C., and Pawson, T. (2005) WW domains provide a platform for the assembly of multiprotein networks. *Mol. Cell. Biol.* 25, 7092-7106.

69. Jackson, I.L., Zhang, X.; Hadley, C., Rabbani, Z.N., Zhang, Y., Marks, S., and Vujaskovic, Z. (2012) Temporal expression of hypoxia-regulated genes is associated

with early changes in redox status in irradiated lung. *Free Radic. Biol. Med.* 53, 337-346.

70. Jani, D., Lutz, S, Marshall, N.J., Fischer, T., Köhler, A., Ellisdon, A.M., Hurt, E., and Stewart, M. (2009) Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. *Mol. Cell.* 33, 727-737.

71. Jani, D., Lutz, S., Hurt, E., Laskey, R.A., Stewart, M., and Wickramasinghe, V.O. (2012) Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acids Res.* 40, 4562-4573.

72. Janssen, O., Qian, J., Linkermann, A., and Kabelitz, D. (2003) CD95 ligand – death factor and costimulatory molecule? *Cell Death Differ.* 10, 1215-1225.

73. Jiang, Y.J., Che, M.X., Yuan, J.Q., Xie, Y.Y., Yan, X.Z., and Hu, H.Y. (2011) Interaction with polyglutamine-expanded huntingtin alters cellular distribution and RNA processing of huntingtin yeast two-hybrid protein A (HYPA). *J. Biol. Chem.* 286, 25236-25245.

74. Kang, C.H., Feng, Y., Vikram, M., Jeong, I.S., Lee, J.R., Bahk, J.D., Yun, D., Lee, S.Y., and Koiwa, H. (2009) *Arabidopsis thaliana* Prp40s are RNA polymerase II C-terminal domain-associating proteins. *Arch. Biochem. Biophys.* 484, 30-38.

75. Kao, H-Y and Siliciano, P.G. (1996) Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 16, 960-967.

76. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J.J., Sudol, M., and Tanokura, M. (2004) Common mechanism of ligand recognition by group II/III WW domains. *J. Biol. Chem.* 279, 31833-31841.

77. Katoh, M. and Katoh, M. (2003) Identification and characterization of human FNBP3 gene in silico. *Int. J. Mol. Med.* 12, 651-656.

78. Kilmartin, J.V. (2003) Sfi1p has conserved centrin-binding sites and essential function in budding yeast spindle pole body duplication. *J. Cell Biol.* 162, 1211-1221.

79. Kim, J.J., Rajagopalan, K., Hussain, B., Williams, B.H., Kulkarni, P., and Mooney, S.M. (2013) CETN1 is a cancer testis antigen with expression in prostate and pancreatic cancers. *Biomark. Res.* 1, 22.

80. Kim, S.Y., Kim, D.S., Hong, J.E., and Park, J.H. (2017) Crystal structure of wild-type centrin1 from Mus musculus occupied by Ca^{2+}. *Biochemistry (Mosc).* 82, 1129-1139.

81. Koch, A., Poirier, F., Jacob, R., and Delacour, D. (2010) Galectin-3, a novel centrosome-associated protein, required for epithelial morphogenesis. *Mol. Biol. Cell.* 21, 219-231.

82. Kong, J. and Yu, S. (2007) Fourier transform infrared spectroscopic analysis of protein secondary sturctures. *Acta Biochim. Biophys. Sin. (Shanghai)*. 39, 549-559.

83. Krimm, S. and Bandekar, J. (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* 38, 181-364.

84. Kyte, J. and Doottle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.

85. La Cour, T., Gupta,R., Rapacki, K., Skriver, K., Poulsen, F.M., and Brunak, S. (2003) NESbase version 1.0: a database of nuclear export signals. *Nucleic Acids Res.* 31, 393-396.

86. La Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* 17, 527-536.

87. Lee, J.E. and Gleeson, J.G. (2011) A system-biology approach to understanding the cioliopathy disorders. *Genome Med.* 3, 59

88. Lee, V.D. and Huang, B. (1993) Molecular cloning and centrosomal localization of human caltractin. *Proc. Natl. Acad. Sci. USA.* 90, 11039-11043.

89. Li, S., Sandercock, A.M., Conduit, P., Robinson, C.V., Williams, R.L., and Kilmartin, J.V. (2006a) Structural role of Sfi1p-centrin filaments in budding yeast spindle pole body duplication. *J. Cell Biol.* 173, 867-877.

90. Li, S.Y., Davidson, P.J., Lin, N.Y., Patterson, R.J., Wang, J.L., and Arnoys, E.J. (2006b) Transport of galectin-3 between the nucleus and cytoplasm. II. Identification of the signal for nuclear export. *Glycobiology*. 16, 612-622.

91. Li, X., Liu, S., Jiang, J., Zhang, L., Espinosa, S., Hill, R.C., Hansen, K.C., Zhou, Z.H., and Zhao, R. (2017) CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat. Commun.* 8, 1035.

92. Lin, K.T., Lu, R.M., and Tarn, W.Y. (2004) The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing in vivo. *Mol. Cell. Biol.* 24, 9176-9185.

93. Liu, J., Fan, S., Lee, C.J., Greenleaf, A.L., and Zhou, P. (2013) Specific interaction of the transcription elongation regulator TCERG1 with RNA polymerase II requires simultaneous phosphorylation at Ser2, Ser5, and Ser7 within the carboxyl-terminal domain repeat. *J. Biol. Chem.* 288, 10890-10901.

94. Lu, M., Yang, J., Ren, Z., Sabui, S., Espejo, A., Bedford, M.T., Jacobson, R.H., Jeruzalmi, D., McMurray, J.S., and Chen, X. (2009) Crystal structure of the three tandem FF domains of the transcription elongation regulator CA150. *J. Mol. Biol.* 393, 397-408.

95. Macias, M.J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) Structure of the WW domain of kinase-associated protein complexed with a proline-rich peptide. *Nature*. 382, 646-649.

96. Makarov, E.M., Owen, N., Bottrill, A., and Makarova, O. (2012) Functional mammalian spliceosomal complex E contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acid Res.* 40, 2639-2652.

97. Marteijn, J.A., Lans, H., Vermeulen, W., and Hoeijmakers, J.H. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell. Biol.* 15, 465-481.

98. Martinez-Sanz, J., Kateb, F., Assairi, L., Blouquit, Y., Bodenhausen, G., Abergel, D., Mouawad, L., and Craescu, C.T. (2010) Structure, dynamics and thermodynamics of the human centrin 2/hSfi1 complex. *J. Mol. Biol.* 395, 191-204.

99. Matei, E., Miron, S., Blouquit, Y., Duchambon, P., Durussel, I., Cox, J.A., and Craescu, C.T. (2003) C-terminal half of human centrin 2 behaves like a regulatory EF-hand domain. *Biochemistry.* 42, 1439-1450.

100. Meyn, S.M., Seda, C., Campbell, M., Weiss, K.L., Hu, H., Pastrana-Rios, B., and Chazin, W.J. (2006) The biochemical effect of Ser167 phosphorylation on *Chlamydomonas reinhardtii* centrin. *Biochem. Biophys. Res. Commun.* 342, 342-348.

101. Middendorp, S., Küntziger, T., Abraham, Y., Holmes, S., Bordes, N., Paintrand, M., Paoletti, A., and Bornes, M. (2000) A role for centrin 3 in centrosome reproduction. *J. Cell. Biol.* 148, 405-416.

102. Middendorp, S., Paoletti, A., Schiebel, E., and Bornes, M. (1997) Identification of a new mammalian centrin gene, more closely related to Saccharomyces cerevisiae CDC31 gene. *Proc. Natl. Acad. Sci. USA.* 94, 9141-9146.

103. Mizuguchi, M., Nara, M., Kawano, K., and Nitta, K. (1997) FT-IR study of the Ca^{2+}-binding to bovine alpha-lactalbumin. Relationships between the type of coordination and characteristics of the bands due to the Asp COO^- groups in the Ca^{2+}-binding site.. *FEBS Lett.* 417, 153-156.

104. Mizutani, K., Suetsugu, S., and Takenawa, T. (2004) FBP11 regulates nuclear localization of N-WASP and inhibits N-WASP-dependent microspike formation. *Biochem. Biophys. Res. Commun.* 313, 468-474.

105. Montes, M., Becerra, Sánchez-Álvarez, M., and Suñé, C. (2012) Functional coupling of transcription and splicing. *Gene*. 501, 104-117.

106. Moretti, E., Pascarelli, N.A., Belmonte, G., Renieri, T., and Collodel, G. (2017) Sperm with fibrous sheath dysplasia and anomalies in head-neck junction: focus on centriole and centrin 1. *Andrologia*. 49(7).

107. Morris, D.P. and Greenleaf, A.L. (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 275, 39935-39943.

108. Murphy, K.P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* 43, 313-361.

109. Murphy, M.W., Olson, B.L., and Siliciano, P.G. (2004) The yeast splicing factor Prp40p contains functional leucine-rich nuclear export signals that are essential for splicing. *Genetics*. 166, 53-65.

110. Myers, M.D. and Payne, G.S. (2017) Vps13 and Cdc31/centrinL Puzzling partners in membrane traffic. *J. Cell Biol.* 216, 299-301.

111. Nakamura, S., Terada, Y., Rawe, V.Y., Uehara, S., Morito, Y., Yoshimoto, T., Tachibana, M., Murakami, T., Yaegashi, N., and Okamura, K. (2005) A trial to restore defective human sperm centrosomal function. *Humm. Reprod.* 20, 1933-1937.

112. Nara, M., Morii, H., and Tanokura, M. (2013) Coordination to divalent cations by calcium-binding proteins studied by FTIR spectroscopy. *Biochim. Biophys. Acta.* 1828, 2319-2327.

113. Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005) Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol. Cell Biol.* 25, 5664-5674.

114. Noda, I. (2004) Advances in two-dimensional correlation spectroscopy. *Vibrat. Spectrosc.* 36, 143-165.

115. Noda, I. (2008) Recent advancement in the field of two-dimensional correlation spectroscopy. *J. Mol. Struct.* 883, 2-26.

116. Noda, I. (2015) Techniques of two-dimensional (2D) correlation spectroscopy useful in life science research. *Biomed. Spectrosc. Imaging*. 4, 109-127.

117. Novarino, G., Akizu, N., and Gleeson, J.G. (2011) Modeling human disease in humans: the ciliopathies. *Cell*. 147, 70-79.

118. Ocaña, W. and Pastrana-Ríos, B. (2014) Calcium titration of Chlamydomonas reinhardtii centrin and its structural changes. *J. Mol. Struct.* 1069, 73-78.

119. Oleksiewicz, U., Liloglou, T., Tasopoulou, K.M., Daskoulidou, N., Gosney, J.R., Field, J.K., and Xinarianos, G. (2017) COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. *J. Cancer Res. Clin. Oncol.* 143, 1133-1141.

120. Ortíz, M., Sanoguet, Z., Hu, H., Chazin, W.J., McMurray, C.T., Salisbury, J.L., and Pastrana-Rios, B. (2005) Dynamics of hydrogen-deuterium exchange in *Chlamydomonas* centrin. *Biochemistry.* 44, 2409-2418.

121. Ossareh-Nazari, B., Bachalerie, F., and Dargemont, C. (1997) Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science*. 278, 141-144.

122. Ossareh-Nazari, B., Gwizdek, C., and Dargemont, C. (2001) Protein export from the nucleus. *Traffic*. 2, 684-689.

123. Ostrowski, L.E., Dutcher, S.K., and Lo, C.W. (2011) Cilia and models for studying structure and function. *Proc. Am. Thorac. Soc.* 8, 423-429.

124. Paoletti,A., Moudjou, M., Paintrand, M., Salisbury, J.L., and Bornens, M. (1996) Most of centrin in animal cells is not centrosome-associated and centrosomal centrin is confined to the distal lumen of centrioles. *J. Cell Sci.* 109, 3089-3102.

125. Passani, L.A., Bedford, M.T., Faber, P.W., McGinnis, K.M., Sharp, A.H., Gusella, J.F., Vonsattel, J.P., and MacDonald, M.E. (2000) Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease. *Hum. Mol. Genet.* 9, 2175-2182.

126. Pastrana-Ríos, B, Del Valle Sosa, L., and Santiago, J. (2015) Trifluoroacetic acid as excipient destabilizes melittin causing the selective aggregation of melittin within the centrin-melittin-trifluoroacetic acid complex. *Struct Dyn.* 2, 041711.

127. Pastrana-Rios, B. (2001) Understanding the mechanism of peptide unfolding. *Biochemistry.* 40, 9074-9081.

128. Pastrana-Ríos, B. (2006) Simulation of FT-IR spectra and 2D-COS analysis for the thermal perturbation of apo-centrin. *J. Mol. Struct.* 799, 163-167.

129. Pastrana-Ríos, B., De Orbeta, J., Meza, V., Reyes, M., Narváez, D., Gómez, A.M., Rodríguez-Nassif, A., Almodovar, R., Díaz Casas, A., Robles, J., Ortíz, A.M., Irizarry, L., Campbell, M., and Colón M. (2013) Relative stability of human centrin's and its relationship with their calcium binding sites. *Biochemistry.* 52, 1236-1248.

130. Pastrana-Ríos, B., Ocaña, W., Rios, M., Vargas, G.L., Ysa, G., Poynter, G., Tapia, J., and Salisbury, J.L. (2002) Centrin: its secondary structure in the presence and absence of cations. *Biochemistry*. 41, 6911-6919.

131. Paulssen, R.H., Moe, B., Grønaas, H., and Orbo, A. (2008) Gene expression in endometrial cancer cells (Ishikawa) after short time high dose exposure to progesterone. *Steroids*. 73, 116-128.

132. Perry. J.C., Guindalini, C., Bittencourt, L., Garbuio, S., Mazzotti, D.R., and Tufik, S. (2013) Whole blood hypoxia-related gene expression reveals novel pathways to obstructive sleep apnea in humans. *Respir. Physiol. Neurobiol.* 189, 649-654.

133. Pires, J.R., Parthier, C., Aido-Machado, R., Wiedemann, U., Otte, L., Rudolph, R., and Oschkinat, H. (2005) Structural basis for APPTPPPLPP peptide recognition by the FBP11WW1 domain. *J. Mol. Biol.* 348, 399-408.

134. Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P., and Craescu, C.T. (2003) Xeroderma pigmentosum group C possesses a high affinity binding site to human centrin 2 and calmodulin. *J. Biol. Chem.* 278

135. Prossner, S.L. and Morrison, C.G. (2015) Centrin2 regulates CP110 removal in primary cilium formation. *J. Cell. Biol.* 208, 693-701.

136. Pulvermüller, A., Giessl, A., Heck, M., Wottrich, R., Schmitt, A., Ernst, O.P., Choe, H.W., Hofmann, K.P., and Wolfrum, U. (2002) Calcium-dependent assembly of centrin-G-protein complex in photoreceptor cells. *Mol. Cell. Biol.* 22, 2194, 2203.

137. Resendes, K.K., Rasala, B.A., and Forbes, D.J. (2008) Centrin 2 localizes to the vertebrate nuclear pore and plays a role in mRNA and protein export. *Mol. Cell. Biol.* 28, 1755-1769.

138. Rose, J.B., Naydenova, Z., Bang, A., Ramadan, A., Klawitter, J., Schram, K., Sweeney, G., Grenz, A., Eitzschig, H. Hammond, J., Choi, D.S., and Coe, I.R. (2011) Absence of

equilibrative nucleoside transporter 1 in ENT1 knockout mice leads to altered nucleoside levels following hypoxic challenge. *Life Sci*. 89, 621-630.

139. Salisbury, J.L. (2007) A mechanistic view on the evolutionary origin of centrin-based control of centriole duplication. *J. Cell. Physiol.* 213, 420-428.

140. Salisbury, J.L., Baron, A., Surek, B., and Melkonian, M. (1984) Striated flagellar roots: Isolation and partial characterization of a calcium-modulated contractile organelle. *J. Cell Biol.* 99, 962-970.

141. Salisbury, J.L., Suino, K.M., Busby, R., and Springett, M. (2002) Centrin-2 is required foor centriole duplication in mammalian cells. *Curr. Biol.* 12, 1287-1292.

142. Sanders, M.A. and Salisbury, J.L. (1994) Centrin plays an essential role in microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. *J. Cell Biol.* 124, 795-805.

143. Sanoguet, Z., Campbell, M., Ramos, S., Seda, C., Pérez Moreno, L., and Pastrana-Ríos, B. (2006) Effects of phosphorylation in *Chlamydomonas* centrin Ser 167. *Calcium Bind. Proteins.* 1, 108-114.

144. Sawant, D.B., Majumder, S., Perkins, J.L., Yang, C.H., Eyers, P.A., and Fisk, H.A. (2015) Centrin 3 is an inhibitor of centrosomal Mps1 and antagonizes centrin 2 function. *Mol. Biol. Cell.* 26, 3741-3753.

145. Sayer, J.A., Manczak, M., Akileswaran, L., Reddy, P.H., and Coghlan, V.M. (2005) Interaction of the nuclear matrix protein NAKAP with HypA and Huntingtin. *Neuromolecular Med.* 7, 297-310.

146. Sheehan, J.H., Bunick, C.G., Hu, H., Fagan, P.A., Meyn, S.M., and Chazin, W.J. (2006) Structure of the N-terminal calcium sensor domain of centrin reveals the biochemical basis for domain-specific function. *J. Biol. Chem.* 281, 2876-2881.

147. Sosa, L del V., Alfaro, E., Santiago, J., Narváez, D., Rosado, M.C., Rodríguez, A., Gómez, A.M., Schreiter, E.R., and Pastrana-Rios, B. (2011) The structure, molecular dynamics and energetics of centrin-melittin complex. *Proteins.* 79, 3132-3143.

148. Stade, K., Ford, C., Guthrie, C., and Weis, K. (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell*. 1041-1050.

149. Sudol, M. (1996) Structure and function of the WW domains. *Prog. Biophys. Mol. Biol.* 65, 113-132.

150. Sudol, M., Chen, H.I., Bougeret, C., Einbond, A., and Bork, P. (1995) Characterization of a novel protein-binding module — the WW domain. *FEBS Lett.* 369, 67-71.

151. Sudol, M. and Hunter, T. (2000) NeW Wrinkles for an old domain. *Cell*. 103, 1001-1004.

152. Sudol, M., Recinos, C.C., Abraczinskas, J., Humbert, J., and Farooq, A. (2005) WW or WoW: The WW domains in a union of bliss. *IUBMB Life*. 57, 773-778.

153. Sudol, M., Sliwa, K., and Russo, T. (2001) Functions of WW domains in the nucleus.. *FEBS Lett.* 490, 190-195.

154. Sun, X., Ma, J., Ge, Y., Li, S., Yu, Z., Xue, S., and Han, D. (2002) Abnormal expression of centrosome protein (centrin) in spematozoa of male human infertility. *Chin. Sci. Bull.* 47, 822-823.

155. Suzuki, S., Takeshita, K., Asamoto, M., Takahashi, S., Kandori, H., Tsujimura, K., Saito, F., Masuko, K., and Shirai, T. (2009) High mobility group box associated with cell proliferation appears to play an important role in hepatocellular carcinogenesis in rats and humans. *Toxicology.* 255, 160-170.

156. Taillon, B.E., Adler, S.A., Suhan, J.P., and Jarvik, J.W. (1992) Mutational analysis of centrin: an EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of *Chlamydomonas*. *J. Cell Biol.* 119, 1613-1624.

157. Tanaka, N., Goto, M., Kawasaki, A., Sasano, T., Eto, K., Nishi, R., Sugasawa, K., Abe, S., and Saitoh, H. (2010) An EF-hand protein, centrin-1, is an EGTA-sensitive SUMO-interacting protein in mouse testis. *Cel. Biochem. Funct.* 28, 604-612.

158. Thompson, J.R., Ryan, Z.C., Salisbury, J.L., and Kumar, R. (2006) The structure of the human centrin 2 – xeroderma pigmentosum group C protein complex. *J. Biol. Chem.* 281, 18746-18752.

159. Tian, J., Liu, Y., Zhu, B., Tian, Y., Zhong, R., Chen, W., Lu, X., Zou, L., Shen, N., Qian, J., Li, H., Miao, X., and Wang, L. (2015) SF3A1 and pancreatic cancer: new evidence for the association of the spliceosome and cancer. *Oncotarget*. 6, 37750-37757.

160. Trojan, P., Krauss, N., Choe, H., Gieβl, A., Pulvermüller, A., and Wolfrum U. (2008) Centrins in retinal photoreceptor cells: Regulators in the connecting cilium. *Prog Retin. Eye Res.* 27, 237-259.

161. Tsang, W.Y., Spektor, A., Luciano, D.J., Indjeian, V.B., Chen, Z., Salisbury, J.L., Sánchez, I., and Dynlacht, B.D. (2006) CP110 cooperates with two calcium-binding proteins to regulate cytokinesis and genome stability. *Mol. Biol. Cell.* 17, 3423-3434.

162. Veeraraghavan, S., Fagan, P.A., Hu, H., Lee, V., Harper, J.F., Huang, B., and Chazin, W.J. (2002) Structural independence of the EF-hand domains of caltractin. *J. Biol. Chem.* 277, 28564-28571.

163. Venyaminov, S.Y. and Kalnin, N.N. (1990) Quantitative IR spectrophotometry of peptide compunds in water (H_2O) solutions. I. Spectral Parameters of amino acid residue absorption bands. *Biopolymers.* 30, 1243-1257.

164. Vincent, K., Wang, Q., Jay, S., Hobbs, K., and Rymond, B.C. (2003) Genetic interactions with CLF1 identify additional pre-mRNA splicing. *Genetics*. 164, 895-907.

165. Wahl, M.C., Will, C.L., and Lührmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*. 136, 701-718.

166. Wang, Q. Hobbs, K., Lynn, B., and Rymond, B.C. (2003) The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. *J. Biol. Chem.* 278, 7875-7883.

167. Wen, W., Meinkoth, J.L., Tsien, R.Y., and Taylor, S.S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*. 82, 463-473.

168. Wenzel, J. Sanzenbacher, R., Ghadimi, M., Lewitzky, M.,, Zhoy, Q., Kaplan, D.R., Kabelitz, D., Feller, S.M., and Janssen, O. (2001) Multiple interactions of the cytosolic polyproline region of the CD95 ligand: hints for the reverse signal transduction capacity of a death factor. *FEBS Lett.* 509, 255-262.

169. Wiese, A.H., Auer, J., Lassmann, S., Nährig, J., Rosenberg, R., Höfler, H., Rüger, R., and Werner, M. (2007) Identification of gene signatures for invasive colorectal tumor cells. *Cancer Detect Prev.* 31, 282-295.

170. Wiesner, S., Stier, G., Sattler, M., and Macias, M.J. (2002) Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.* 324, 807-822.

171. Wolfrum, U. and Salisbury, J.L. (1998) Expression of centrin isoforms in the mammalian retina. *Exp. Cell. Res.* 242, 10-17.

172. Xie, D. and Freire, E. (1994) Structure based prediction of protein folding intermediates. *J. Mol. Biol.* 242, 62-80.

173. Yan, C., Hang, J., Wan, R., Huan, M., Wong, C.C., and Shi, Y. (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*. 349, 1182-1191.

174. Yang, S., Jeung, H.C., Jeong, H.J., Choi, Y.H., Kim, J.E., Jung, J.J., Rha, S.Y., Yang, W.I., and Chung, H.C. (2007) Identification of genes with correlated patterns of variations in DNA copy number and gene expression level in gastric cancer.

175. Ying, G., Avasthi, P., Irwin, M., Gerstner, C.D., Frederick, J.M., Lucero, M.T., and Baehr, W. (2014) Centrin 2 is required for mouse olfactory ciliary trafficking and development of ependymal cilia planar polarity. *J. Neurosci*. 34, 6377-6388.

176. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W.K., Miyawaki, S., Sugano, S., Haferlach, C.,

Koeffler, H.P., Shih, L.Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S., and Ogawa, S. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 478, 64-69.

177. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature.* 419, 182-185.

APPENDIX A

LIST OF PUBLICATIONS AND PRESENTATIONS

Publications (in chronological order):

1. Pastrana-Ríos, B., De Orbeta, J., Meza, V., Reyes, M., Narváez, D., Gómez, A.M., Rodríguez-Nassif, A., Almodovar, R., **Díaz Casas, A**., Robles, J., Ortíz, A.M., Irizarry, L., Campbell, M., and Colón M. (2013) Relative stability of human centrin's and its relationship with their calcium binding sites. *Biochemistry.* 52, 1236-1248.
2. **Díaz Casas, A**, Chazin, W.J., Pastrana-Rios, B. (2017) Prp40 homolog A is a novel centrin target. *Biophys. J.* 112, 2529-2539.
3. **Díaz Casas, A**., Casanova Sepúlveda, G., Sánchez Negrón, O., Caro Muñiz, A.P., Malavé Ramos, S.R., Cebollero López, A.R., Pastrana-Rios, B. (2018) Molecular biophysical characterization of the third FF domain of *Homo sapiens* Prp40 homolog A. *J. Mol. Struct.* 1167, 174-179.

Presentations:

Oral Presentation

1. **Díaz Casas, A.** and Pastrana-Rios, B. Thermodynamic Governing Binding of Centrin–Target Peptide Complexes. *37th ACS Senior Technical Meeting*. Aguadilla, Puerto Rico. November 8, 2013.

Poster Presentations

1. **Díaz Casas, A.** and Pastrana-Rios, B. Thermodynamic Governing Binding of Centrin-Sfi1p$_{21}$ Complex. *2012 Lilly Academy Technical Forum*. San Juan, Puerto Rico. April 13, 2012.
2. **Díaz Casas, A.** and Pastrana-Rios, B. Thermodynamic Governing Binding of Centrin-Sfi1p$_{21}$ Complex. *IFPAC/PAT Summer Summit*. San Juan, Puerto Rico. June 12-13, 2012.
3. **Díaz Casas, A.** and Pastrana-Rios, B. Comparative thermodynamic analysis of *Homo sapiens* Centrins – *Homo sapiens* Sfi1p$_{21}$ Complexes. *2014 Lilly Academy Technical Forum*. San Juan, Puerto Rico. March 28, 2014.
4. **Díaz Casas, A.** and Pastrana-Rios, B. Thermodynamics of Novel Centrin Biological Target Interactions: Prp40A. *38th ACS Senior Technical Meeting*. San Juan, Puerto Rico. November 7, 2014.
5. **Díaz Casas, A.** and Pastrana-Rios, B. Molecular Biophysical Studies that confirm *Homo sapiens* Prp40 Homolog A as a Centrin Biological Target. *BioXFEL 2nd Annual Conference*. Ponce, Puerto Rico. January 14, 2015.
6. **Díaz Casas, A.** and Pastrana-Rios, B. Molecular Biophysical Studies that confirm *Homo sapiens* Prp40 Homolog A as a Centrin Binding Protein. *14th Annual PepTalk: The Protein Science Week.* San Diego, California. January 19-23, 2015.
7. **Díaz Casas, A.**, Bauer, W.J., and Pastrana-Rios, B. Molecular Biophysical Studies and Crystallization Screening of Centrin-Prp40Ap Complex. *BioXFEL 3rd Annual Conference*. San Juan, Puerto Rico. January 13-15, 2016.

8. **Díaz Casas, A.**, Bauer, W.J., and Pastrana-Rios, B. Molecular Biophysical Studies and Crystallization Screening of Centrin-Prp40Ap Complex. *BioXFEL 4th Annual Conference.* Las Vegas, Nevada. January 9-12, 2017.
9. **Díaz Casas, A.**, Bauer, W.J., and Pastrana-Rios, B. Molecular Biophysical Studies and Crystallization Screening of Centrin-Prp40Ap Complex. *2017 Lilly Academy Technical Forum.* San Juan, Puerto Rico. March 24, 2017.
10. **Díaz Casas, A**. and Pastrana-Ríos, B. Molecular Biophysical Characterization of FF_3 domain of *Homo sapiens* Prp40 Homolog A. Victoria, Canada. June 7-10, 2017.
11. **Díaz Casas, A.**, Casanova Sepúlveda, G., and Pastrana-Rios, B. Spectroscopic Studies and Crystallization Screening of the third FF domain of *Homo sapiens* Prp40 Homolog A. *BioXFEL 5th Annual Conference*. New Orleans, LA, February 13-15, 2018.