MOLECULAR BIOPHYSICAL CHARACTERIZATION OF A NOVEL CENTRIN-Krr1 COMPLEX

By:

Aslin Marie Rodríguez Nassif

A dissertation submitted in partial fulfilment 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
_____	_____
Jorge Ríos, Ph. D. Member, Graduate Committee	Date
_____	_____
Marisol Vera, Ph. D. Member, Graduate Committee	Date
_____	_____
Belinda Pastrana-Ríos, Ph. D. President, Graduate Committee	Date
_____	_____
Ivette Cruzado, Ph. D. Representative of Graduate Studies	Date
_____	_____
Enrique Meléndez, Ph. D. Chairperson of the Department	Date

ABSTRACT

The nucleus is the natural compartment for DNA in eukaryotes. This compartment is also the setting for genetic mutations, nucleotide excision repair (NER), and ribosome biogenesis. The study of protein-protein interactions (PPIs) and protein-DNA interactions is the key to understanding these processes in the cell. Abnormalities in ribosome biogenesis cause specific clinical syndromes and recently have been associated with tumorigenesis. One example involves the KH domain of Krr1, which has been identified as a single-stranded DNA (ssDNA) recognition motif. Centrin, an essential calcium binding protein, has been found to regulate NER within the nucleus. Krr1, a novel centrin target, is required for ribosome biogenesis. The structure of *Homo sapiens* Krr1 (*Hs*Krr1) has not been determined to date. We describe the spectroscopic characterization of the GXXG loop peptide (GXXGlp), which is present in KH domain-containing proteins. The sequence of *Hs*Krr1's GXXGlp is evolutionarily conserved and has been associated with ssDNA interaction and ribosome biogenesis. 2D IR correlation spectroscopy was used to determine the stability of GXXGlp, as a synthetic peptide, and the optimal formulation conditions for the use of the peptide. We observed differences in the molecular behavior of GXXGlp in the presence and absence of trifluoroacetate at various peptide concentrations. 2D IR correlation spectroscopy was also used for the elucidation of the unfolding process, the mechanism and extent of peptide aggregation, and the effect of TFA on the stability of the peptide. This spectroscopic method can be applied to the characterization of any synthetic peptide. Furthermore, our findings indicate a GXXGlp-ssDNA interaction. We used the same technique to assess the stability of the complex via

thermal perturbation. Preliminary studies identified a *Hs*centrin2-GXXGlp interaction using NMR, indicating that the interaction occurs in the C-terminal domain of *Hs*centrin, the Ca^{2+}-dependent domain. New signals in the 2D HSQC centrin-GXXGlp titration experiments suggest that centrin adopts a new conformation upon binding to GXXGlp, exposing amino acids located in the EF hands of the C-terminal domain. Moreover, these studies identified a novel interaction between *Hs*Krr1 and *Hs*centrin2's C-terminal domain.

RESUMEN

El núcleo es un compartimiento natural para el ADN en las células eucariotas. Este compartimento también es el escenario de mutaciones genéticas, reparación de escisión nuclear (REN) y del comienzo de la generación de ribosomas. El estudio de las interacciones proteína-proteína (IPP) y las interacciones proteína-ADN (IPA) es la clave para entender muchos de los procesos regulatorios en la célula. Anomalías en la generación de los ribosomas causan síndromes clínicos específicos y recientemente se han asociado con la creación de tumores. Un ejemplo envuelve el dominio homólogo K (KH) de Krr1 que se ha identificado que reconoce ácidos nucleicos de hebra sencilla (ADNhs). Centrin, una proteína esencial que enlaza a calcio forma parte de las proteínas que regulan REN en el núcleo mientras que Krr1, un nuevo blanco de centrin, es necesario para la creación de ribosomas. La estructura de Homo sapiens Krr1 (*Hs*Krr1) no se ha resuelto aún. Se describió la caracterización espectroscópica del péptido GXXG (GXXGlp), que está presente en las proteínas que contienen el dominio KH. La secuencia de HsKrr1 GXXGlp se conserva evolutivamente y se ha asociado con la interacción ADNhs y la creación de ribosomas. La espectroscopía de correlación en infrarrojo de dos dimensiones (IR 2D) se utilizó para determinar la estabilidad del GXXGlp, como péptido sintético, y las condiciones de formulación óptimas para el uso del péptido. Se observó diferencias en el comportamiento molecular del GXXGlp en presencia y ausencia de ácido trifluoroacético (ATF) a diversas concentraciones de péptido. Con la espectroscopía de correlación IR 2D se determinó el proceso de despliegue, el mecanismo, la extensión de la agregación de péptidos, y el efecto de ATF sobre la estabilidad del

péptido, demostrando que este método espectroscópico se puede aplicar a la caracterización de cualquier péptido sintético. Además, los resultados han identificado la interacción GXXGlp-ADNhs y la estabilidad del complejo con perturbación térmica utilizando la misma técnica. La interacción *Hs*centrin-GXXGlp se identificó utilizando estudios en dos dimensiones de resonancia magnética nuclear (RMN), lo que indica que la interacción tiene lugar en el dominio C-terminal de Centrin, el dominio dependiente de calcio. Las nuevas señales en los resultados de los experimentos de RMN sugieren que centrin adopta una nueva conformación exponiendo los aminoácidos localizados en su dominio del C-terminal.

DEDICATION

Each important step requires your own effort and unconditional support from the ones you love.

I dedicate this work to my family, especially my beloved husband, Luis, and children, Adrian and Damian, for your love, motivation, and unconditional support to achieve my goals throughout my scientific career.

ACKNOWLEDGMENTS

Thanks to the Chemistry Department of the University of Puerto Rico at Mayagüez for accepting me and providing financial support. I would like to express my gratitude to Dr. Belinda Pastrana, my research advisor, for her valuable guidance and constant encouragement with my research. She has guided me all these years to complete the work in a successful manner and has been a constant support to face every difficult step. I would like to thank my thesis committee, Dr. Enrique Melendez, Dr. Jorge Rios, and Dr. Marisol Vera, for their support and suggestions for my research. My sincere gratitude to Lilly del Caribe for giving me the opportunity of an internship and inspiring me to pursue an industrial professional career. Special thanks to the research laboratory members including technical assistants, undergraduate students, and graduate colleagues. A special thanks to Adalberto Diaz, my coworker and friend, for his advice, help, and support in every step of my research. My deepest gratitude to my friends, specially to Bianca Alamo, Tatiana Garces, Leishla Cruz, and Jose Carmona, for their moral support in difficult times and their encouragement to continue pursuing my academic goals. I will always be thankful for that. I am most grateful to the laboratory collaborators, Dr. Walter J. Chazin and Dr. William Bauer for lending me their expertise to my scientific and technical problems. I would like to express my gratitude to my parents and sisters, Juan Rodriguez, Ruth Nassif, Karen Rodriguez, and Ruth Rodriguez for their unfailing emotional support, unconditional trust, and endless patience. Finally, I thank with love to Luis Lafontaine, Adrian, and Damian,

my husband and children for their company, love, support, encourage, and invaluable help through this period in the most positive way.

TABLE OF CONTENTS

ABSTRACT ..... **ii**

RESUMEN ..... **iv**

DEDICATION ..... **vii**

ACKOWLEDGMENTS ..... **viii**

TABLE OF CONTENTS ..... **x**

LIST OF TABLES ..... **xii**

LIST OF FIGURES ..... **xiii**

CHAPTER I: JUSTIFICATION **1**

OBJECTIVES 6
HYPOTHESIS 6

CHAPTER II: PREVIOUS WORK **7**

CENTRIN 7
CENTRIN TARGETS 11
Prp40 ..... *11*
Sf1 ..... *14*
Melittin *16*
XPC *18*
Kar1 ..... *21*
KRR1, A NOVEL CENTRIN TARGET 23
K Homology Domain *28*
GXXG Loop *28*

CHAPTER III: MATERIALS AND METHODS **30**

PROTEIN EXPRESSION AND PURIFICATION 30
Hscen2 ..... *33*
Hscen2 C-terminal *34*
Hscen2 N-terminal *35*
HsKrr1 ..... *36*
PEPTIDE PREPARATION 37
ssDNA SYNTHESIS AND PURIFICATION 38
FT-IR SPECTROSCOPY 39
2D IR CORRELATION SPECTROSCOPY AND SPECTRAL ANALYSIS 40
CIRCULAR DICHROISM 41
ISOTHERMAL TITRATION CALORIMETRY 42
NUCLEAR MAGNETIC RESONANCE 43

Crystallization Tray 44
Small Angle X-ray Scattering and Modeling 45

CHAPTER IV: RESULTS AND DISCUSSION 46

Bacterial Protein Expression and Purification 46
Hscen2 46
Hscen2 C-terminal 48
Hscen2 N-terminal 50
HsKrr1 52
Analysis of GXXG Loop Peptide 54
Analysis of ssDNA, a GXXG Loop Target 56
Analysis of secondary structure and Stability Studies of HsKrr1 and its GXXG Loop 59
Preliminary Studies of the Interaction Between HsKrr1 and Hscentrin2 C-terminal 63
Nuclear Magnetic Resonance Spectroscopy Studies of the Hscen2/Crcen and HsKrr1 GXXG Loop Complex 66
FT-IR Spectroscopy Analysis of the GXXG Loop and its Complex with ssDNA 69
2D IR Correlation Spectroscopy in the Determination of Aggregation and Stability of the GXXG Loop 72
Small Angle X-ray Scattering, Modeling, and Preliminary Crystallization Tray of HsKrr1 80

CHAPTER V: CONCLUSIONS AND FUTURE WORK 83

REFERENCES 85

APPENDIX A 93

APPENDIX B 94

APPENDIX C 95

APPENDIX D 96

APPENDIX E 97

LIST OF TABLES

TABLE 1. SEQUENCE DETAILS OF HSCEN2, HSKRR1, HSCEN2 C-TERMINAL, AND HSCEN2 N-TERMINAL 32

TABLE 2. THERMODYNAMIC DATA OF THE INTERACTION BETWEEN HSKRR1 AND HSCEN2 C-TERMINAL DOMAIN AT 20 °C 65

TABLE 3. SUMMARY OF PEAK ASSIGNMENTS FOR THE GXXGLP IN D2O IN THE SPECTRAL REGION OF 1750-1500 CM^{-1} 76

LIST OF FIGURES

FIGURE 1. RIBBON MODEL OF KH DOMAIN 3

FIGURE 2. RIBBON DIAGRAM OF THE CRYSTAL STRUCTURE OF *CHLAMYDOMONAS REINHARDTII* CENTRIN 5

FIGURE 3. SCHEMATIC REPRESENTATION OF THE EF-HAND CONSENSUS SEQUENCE 9

FIGURE 4. HOMOLOGY RIBBON MODELS OF HUMAN CENTRINS 10

FIGURE 5. COMPARISON OF PRP40 AND ITS PROBABLE ORTHOLOGS' WW DOMAINS AND FF REGIONS 12

FIGURE 6. SCHEMATIC OVERVIEW OF PRP40 FF DOMAIN INTERACTIONS 13

FIGURE 7. MODEL OF THE MULTIMERIC ASSEMBLY OF SFI1-CENTRIN 15

FIGURE 8. SUPERIMPOSITION OF THE CRCEN/MELITTIN COMPLEX AND THE HSCEN2/XPC COMPLEX 17

FIGURE 9. CD AND NMR ANALYSIS OF HSCEN2 AND XPC INTERACTION 19

FIGURE 10. STRUCTURE OF HSCEN2 BOUND TO XPC PEPTIDE.....20

FIGURE 11. STRUCTURAL CHARACTERIZATION OF THE COMPLEX OF CDC31 AND KAR1 PEPTIDE.....22

FIGURE 12. THE SEQUENCE OF HSKRR1 PROTEIN 25

FIGURE 13. HSKRR1 EXPRESSION LEVELS IN DIFFERENT HUMAN TISSUES.....26

FIGURE 14. STRUCTURE OF SCKRR1-SCFAF1 COMPLEX.....27

FIGURE 15. THE GXXG LOOP IN KH DOMAIN PROTEINS.....29

FIGURE 16. 4-20% BIS-TRIS GRADIENT SDS-PAGE FOR THE PURIFICATION OF HSCEN2 47

FIGURE 17. EXPRESSION AND PURIFICATION OF HSCEN2 C-TERMINAL 49

FIGURE 18. EXPRESSION AND PURIFICATION OF HSCEN2 N-TERMINAL 51

FIGURE 19. PURIFICATION OF HSKRR1 53

FIGURE 20. PEPTIDE PURIFICATION 55

FIGURE 21. ANION EXCHANGE CHROMATOGRAPHY FOR OLIGONUCLEOTIDE PURIFICATION 57

FIGURE 22. HPLC ANALYSIS TO VALIDATE THE PURITY OF THE OLIGONUCLEOTIDES.... 58

FIGURE 23. CD ANALYSIS OF GXXGLP AND HSKRR1 60

FIGURE 24. THERMAL DEPENDENCE PLOTS OF GXXGLP AND GXXGLP/SSDNA13MER COMPLEX 61

FIGURE 25. THERMAL DEPENDENCE PLOT OF HSKRR1 62

FIGURE 26. ITC ISOTHERM OF THE INTERACTION BETWEEN HSKKR1 AND HSCEN2 C-TERMINAL AT 20 °C 64

FIGURE 27. HSQC SPECTRA OF UNIFORMLY 15N-LABELED HSCEN2 67

FIGURE 28. HSQC SPECTRA OF UNIFORMLY 15N-LABELED CRCEN 68

FIGURE 29. FT-IR SPECTRAL OVERLAY OF GXXGLP 71

FIGURE 30. 2D IR CORRELATION SPECTROSCOPY DURING THERMAL PERTURBATION OF THE GXXGLP 77

FIGURE 31. SEQUENTIAL ORDER OF MOLECULAR EVENTS FOR ALL GXXGLP SAMPLES 78

FIGURE 32. MODEL OF THE INTERMOLECULAR SALT-BRIDGE INTERACTION THAT MEDIATES THE SELF-ASSOCIATION OF THE GXXGLP 79

FIGURE 33. HSKKR1 SMALL ANGLE X-RAY SCATTERING MODEL 81

FIGURE 34. MICROCRYSTAL OF HSKRR1 82

CHAPTER I
JUSTIFICATION

Previous studies have been directed towards the recognition of protein-protein interactions (PPIs) to understand biological processes and develop novel therapeutics. PPIs occur when two or more proteins interact to induce a cellular response *via* a specific biological function. DNA replication and ribosome biogenesis are examples of molecular processes that depend on PPIs. Information about PPIs improves our understanding of the normal and aberrant cellular processes at the molecular level, elucidating the molecular mechanisms of a disease and therefore providing the basis for new therapeutic approaches.

One mechanism of diseases such as cancer is through a genetic mutation, which may be translated into the generation of protein variants and affect numerous PPIs, whereby the disease state becomes manifest. A second mechanism of disease is unusual ribosome biogenesis, caused by genetic mutations that generate mutant proteins. This process is required for protein synthesis. Abnormalities in ribosome biogenesis cause specific clinical syndromes and recently have been associated with tumorgenesis.[1–5] One example involves the K homology (KH) domain (Figure 1) of Krr1, which has been associated with fragile X syndrome, causing mental retardation.[6–8] This form of inherited mental impairment is the most common type seen in humans.[9,10] Moreover, symptoms of fragile X syndrome can range from learning disabilities to more severe cognitive or intellectual disabilities. This genetic disease is the most common known cause of autism or "autistic-like" behaviors. Approximately

one in 3,600 to 4,000 males and one in 4,000 to 6,000 females have fragile X syndrome.

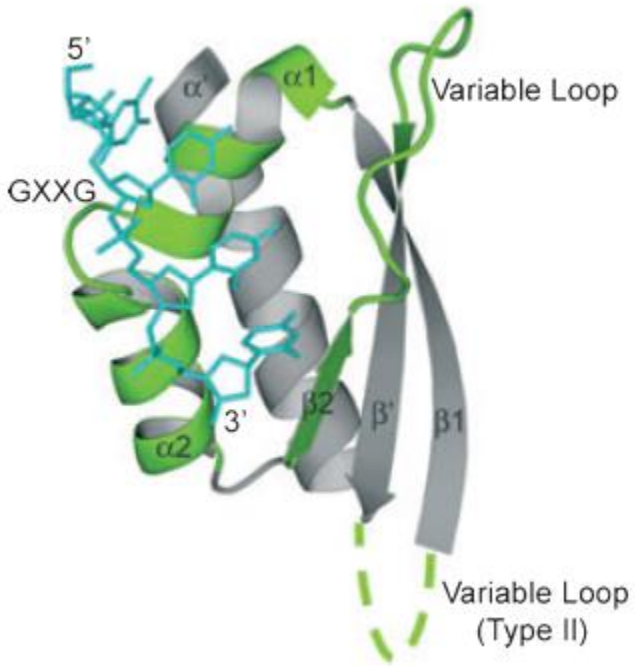


Figure 1. Ribbon model of type I KH domain. Secondary structural elements (colored in green) include the α_1-helix, GXXG loop, α_2-helix, β_2-strand, and variable loop. Amino acids rendered as cyan stick models play an important role in the recognition of four nucleotides. The green dotted line represents the location of a second variable loop found in Type II KH domains. Typically Van der Waals forces, hydrophobic interactions, and electrostatic interactions contribute to the nucleic acid binding affinity. *Adapted with permission from John Wiley and Sons, Valverde, 2008.*[6]

Symptoms can also include characteristic physical and behavioral features, along with delays in speech and language development.[6–9]

Centrin, found only in eukaryotes, is an essential calcium binding protein (CaBP) that has been found to regulate both nucleotide excision repair (NER) within the nucleus and mRNA export from the nucleus. Centrins are a major class of Ca^{2+} sensor proteins that play a crucial role in various cellular signaling cascades through regulation of numerous target proteins in a Ca^{2+}-dependent manner. Centrin, like calmodulin, contains four EF-hand motifs with highly conserved amino acid sequences. The EF-hand is a helix-loop-helix motif, which contains a canonical sequence within the loop that binds Ca^{2+}. Also, a tethered helix is found between the second and third EF-hands, thus generating a dumbbell arrangement with two independent domains (Figure 2).[11–18]

Krr1, a novel centrin target, is required for proper processing of pre-rRNA, for the synthesis of 18S rRNA, and for the assembly of the 40S subunit. Abnormalities in ribosome biogenesis lead to defects in protein synthesis, affect PPIs, and cause diseases. The structure of human Krr1 has not been determined to date, although this target protein does contain a KH domain [6] which has been identified as a DNA recognition motif.

The goal of this research project involved the molecular and thermodynamic characterization of the binary complex between *Homo sapiens* centrin 2 and Krr1 (*Hs*cen2-*Hs*Krr1) and the ternary *Hs*cen2-*Hs*Krr1-DNA complex.

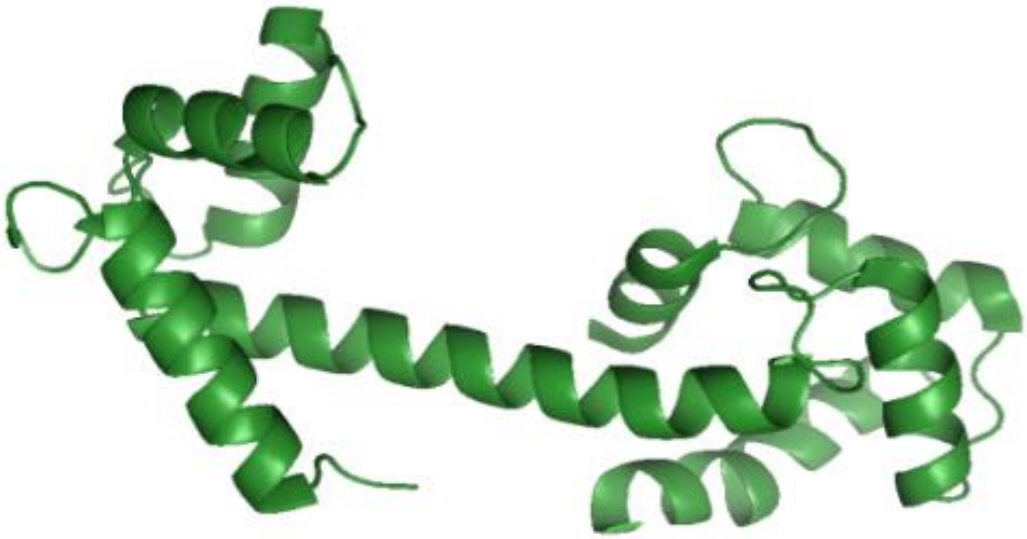


Figure 2. Ribbon diagram of the crystal structure of *Chlamydomonas reinhardtii* centrin (PDB ID: 3QRX). *Adapted with permission from John Wiley and Sons, del Valle Sosa, 2011*.[17]

OBJECTIVES

The following objectives were proposed for this study:

- To identify the binding site of Krr1 in *Hs*cen2.
- To determine the thermodynamics of binding of the *Hs*cen2-*Hs*Krr1 complex by molecular characterization.
- To determine the binding affinity of Krr1 for centrin and ssDNA.
- To determine the molecular mechanism of complex formation and the secondary structure of the complex.

HYPOTHESIS

The origins of many diseases have been linked to genetic mutations, defects in PPIs, and unusual cellular functions. The PPIs between centrin 2 and Krr1, a novel biological centrin target associated with mental impairment, is one example of the role of PPIs in the molecular mechanism of a disease. Many syndromes are also related to defects in interactions between proteins and DNA. The study of the interaction between Krr1 and centrin 2 and DNA will improve the understanding of Krr1-mediated syndromes and provide the basis for the development of therapeutic drugs. If the binary complex of *Hs*Krr1-DNA occurs, then the interaction of *Hs*cen2 with that complex may affect the regulation of ribosome biogenesis.

CHAPTER II
PREVIOUS WORK

Centrin

Centrin is a highly conserved CaBP that acts as a Ca^{2+} sensor and regulator for many cellular processes. These include centriole duplication and separation, assembly of the basal bodies and cilia, the initiation and regulation of NER, and mRNA nuclear export. Centrins are also related to the proteasome and implicated in protein degradation.[17,19–23] Centrin is a member of the EF-hand superfamily and is composed of two independent domains.[17,20–24] The name of EF-hand was given to the helix-loop-helix motif containing a Ca^{2+} binding site (see Figure 3). The EF-hand's role is to translate a signal into a functional response.[25] Each of centrin's terminal domains contains two EF-hand motifs.[24] The C-terminal and N-terminal domains of centrin have similar structure but have different affinity for Ca^{2+} and for its target proteins, suggesting these domains perform different functions within the protein.[25–27]

Four centrin isoforms have been identified in mammals (see Figure 4).[27–29] Their genes have been identified as *CETN1*, *CETN2*, *CETN3*, and pseudogene *CETN4*, located in chromosomes 18p11.32, Xq28, 5q14.3, and 4q27, respectively.[30,31] These CaBPs have been localized in specific sites suggesting different functions: ***Hs*cen1**, to the base of the flagella in the sperm for motility;[28,31–33] ***Hs*cen2**, to the centriole, basal bodies, ciliary structure, and nucleus, having roles in centriole separation and duplication, basal body and cilia assembly, and DNA excision repair and mRNA export;[31,33] ***Hs*cen3**, to the spindle poles, in association with mitosis;[30,31,33] and

***Hs*cen4**, to neurons in association with neuronal elongation. In addition, all four centrin isoforms have been localized to the connecting cilium in the retina, where they are known to be associated with G-coupled signal transduction.[31,33]

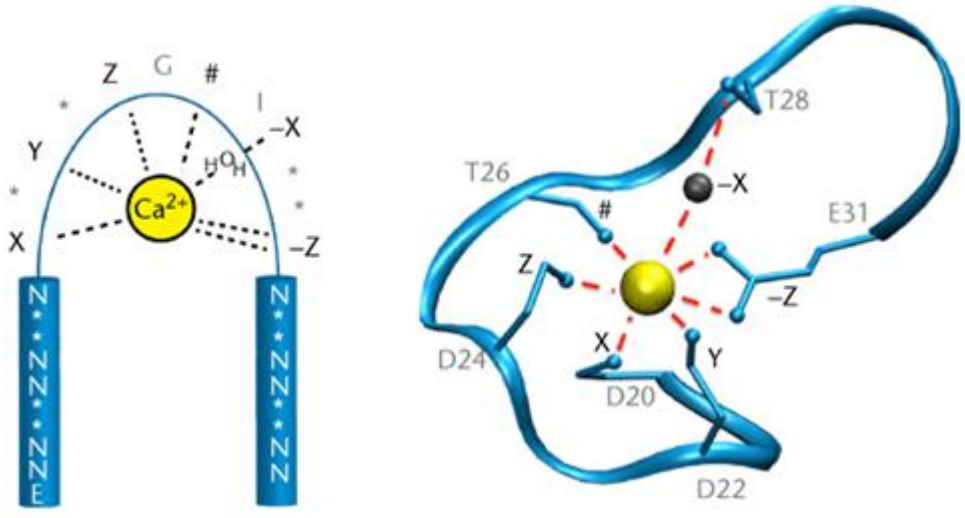


Figure 3. Schematic representation of the EF-hand consensus sequence. The helix–loop–helix structural motif represented in blue coordinates to Ca^{2+} in the following positions: X, Y, Z, and −Z. The Ca^{2+} coordination is shown with a ribbon representation of the loop from the first EF-hand of calmodulin (PDB ID: 1CLL). The ligating oxygen atoms of amino acids are shown in a ball and stick representation. *Adapted with permission from John and Wiley Sons, Johnson, 2014*. [35]

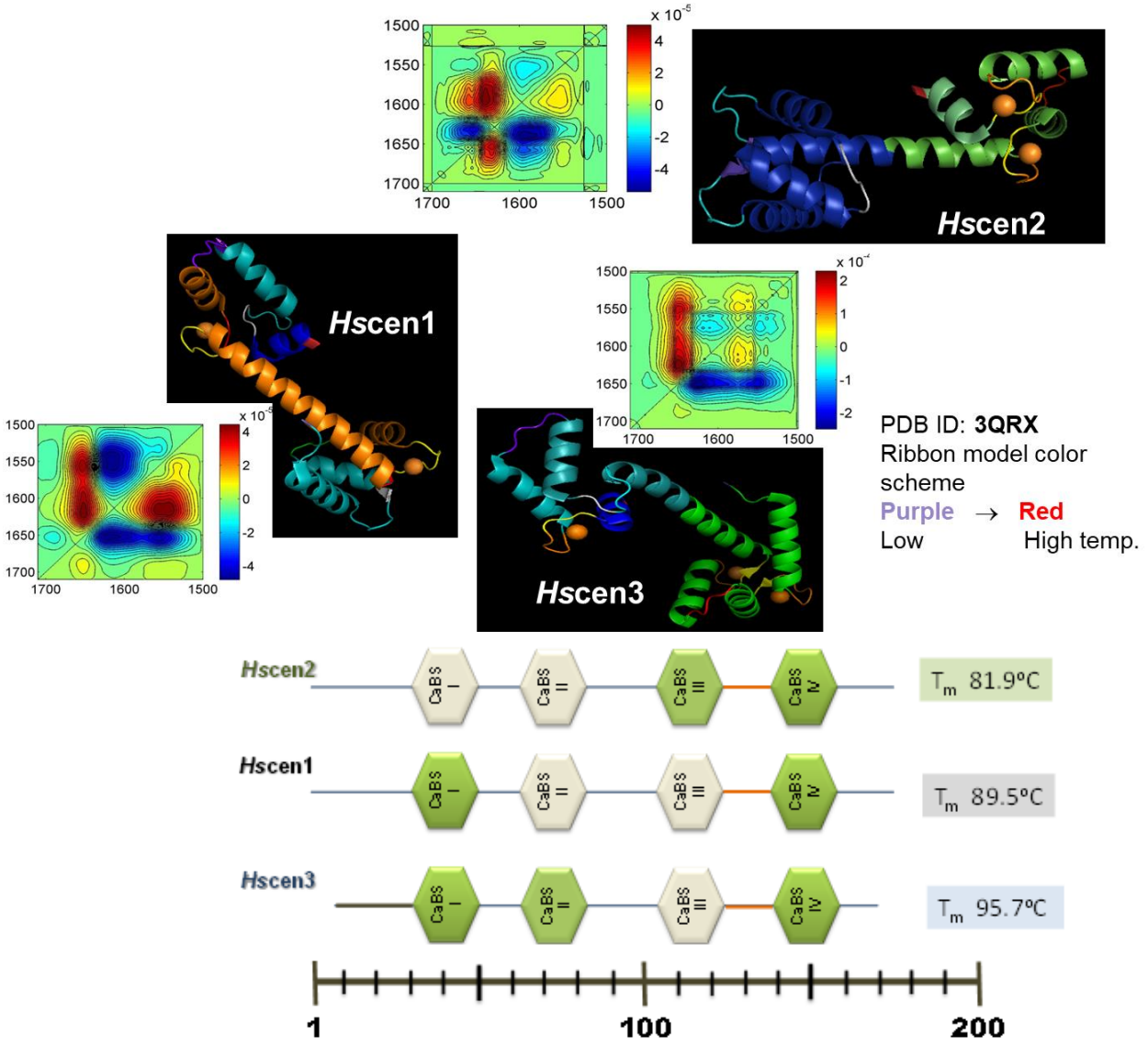


Figure 4. Homology ribbon models of human centrins. Top: Homology models of *Hs*cen1, 2, and 3 were created based on the available full-length *Cr*centrin structure (PDB ID: 3QRX) using pyMOL from Schrödinger, LLC. Ca^{2+} is represented as orange spheres and the structural stability is represented by schematic color legend, defining purple as the least stable part and red as the most stable part using thermal perturbation. Bottom: Schematic diagram of the predicted low-affinity (cream) and high-affinity (green) domains of the *Hs*centrins based on sequence analysis. *Adapted with permission from American Chemical Society, Pastrana-Rios, 2013* [18]

Centrin Targets

Prp40

Prp40 is a ribonucleoprotein required for pre-spliceosome formation where the pre-RNA goes through a process that removes introns, leaving the exons aligned in the messenger (m)RNA. Prp40 is an essential protein that contains two WW domains, which regulate the RNA metabolism and transcription processes, followed by six FF repeats (Figure 5.) The first two FF domains are the most important domains to Prp40.[16,36–40] One of the functions inside the pre-spliceosome complex is to define a bridging interaction that links both terminals of the introns. Prp40 interacts with several proteins forming the pre-spliceosome complex including SF1 and U2AF, and the U1 complex (Figure 6.) The interaction of Prp40 with centrin was validated by performing yeast two hybrid assays, isothermal titration calorimetry (ITC), and two-dimensional correlation spectroscopy. The interaction between the nuclear proteins is Ca^{2+}-dependent, and Prp40 adopts a primarily helical structure.[16]

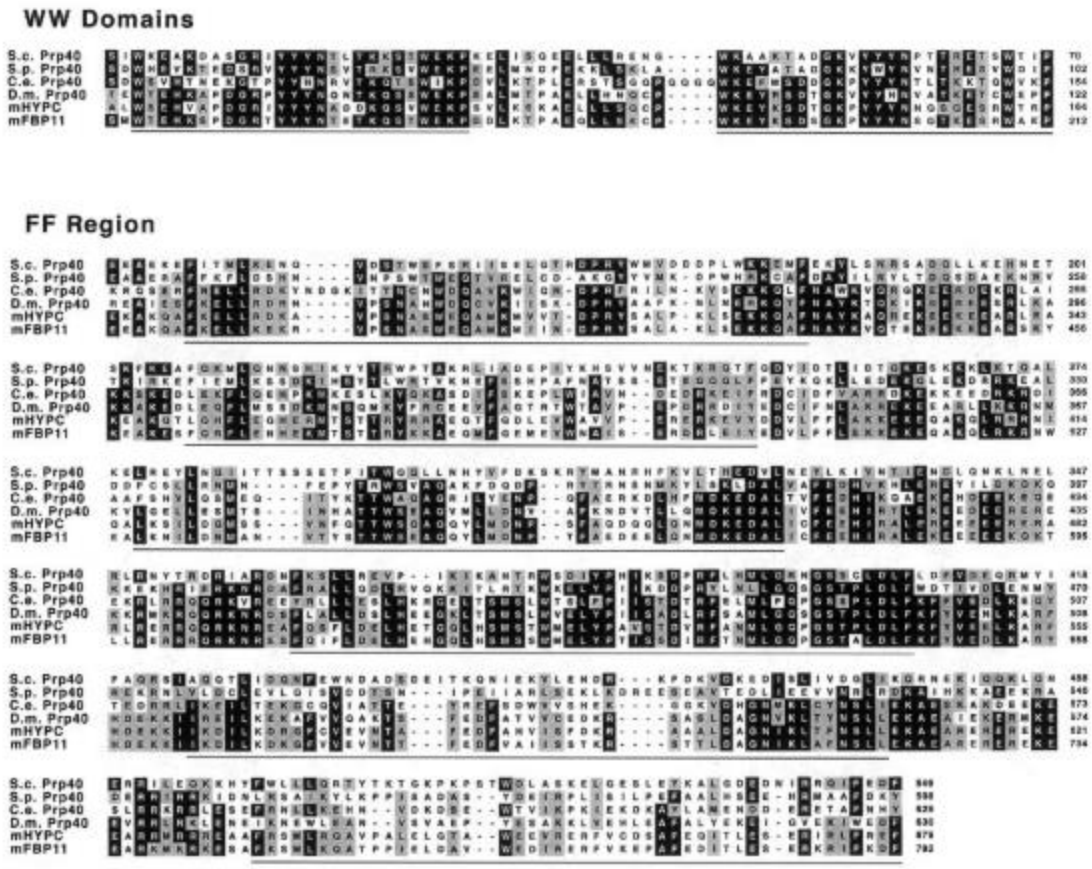


Figure 5. Comparison of Prp40 and its probable orthologs' WW domains and FF regions. Sequence alignment of Prp40 in *Saccharomyces cerevisiae* (*Sc*), *S. pombe* (*Sp*), *C. elegans* (*Ce*), *D. melanogaster* (*Dm*), as well as mHYPC and mFBP11. The WW domains (underlined) show a very high level of conservation. The alignment of the FF region indicates there are six repeating units. *Adapted with permission from American Society for Biochemistry and Molecular Biology, Morris, 2000.* [39]

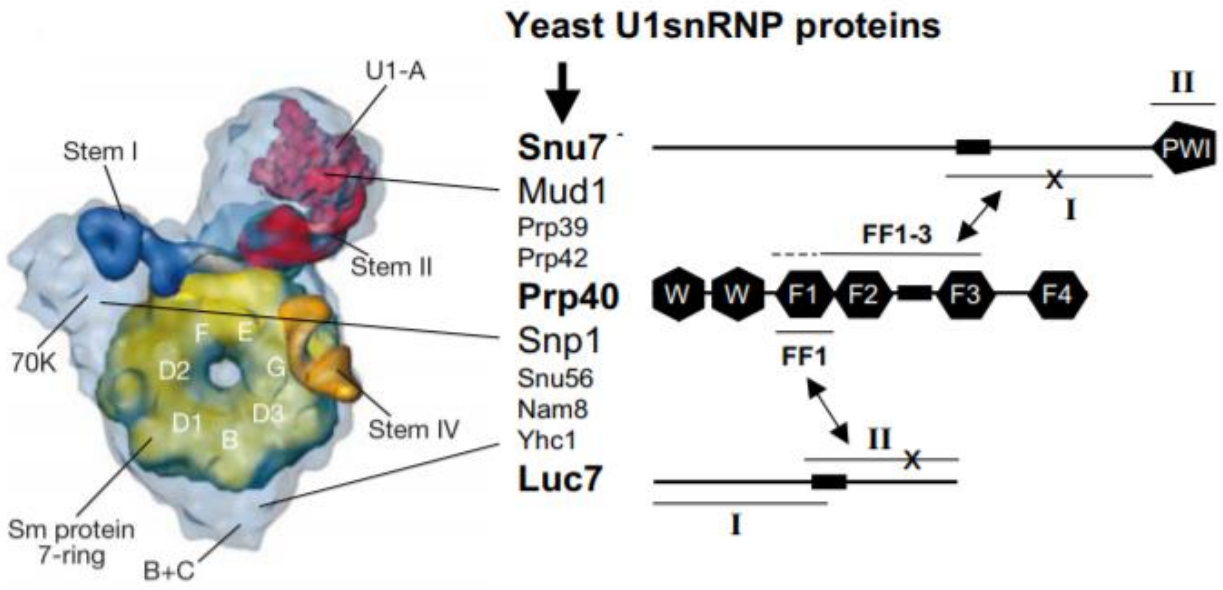


Figure 6. Schematic overview of Prp40 FF domain interactions. Prp40 interacts with Snu71 and Luc7 in the U1snRNP protein complex. At left is shown a structural homology model of the human U1snRNP. *Adapted with permission from BioMed Central, Ester, 2008*. [40]

Sfi1

Splicing factor 1 (Sfi1) is a centrosome protein with homologs in other eukaryotic organisms. Sfi1 plays an important role in pre-mRNA splicing in the recognition of the introns at the 3' splice site. The human protein contains in its C-terminal region a proline- rich domain; in its N-terminal region a ligand motif for U2 auxiliary factor, which is required in the early stages of spliceosome genesis; and a central KH domain, which binds to the intron segment of the DNA. *In vitro* results show the interaction of Sfi1 with *Hs*cen2. Craescu et al. 2006 characterized the complex between centrin and Sfi1 with different repeat segments of Sf1 and *Hs*cen2 (Figure 7). The interaction between the Sfi1 segments and *Hs*cen2 is moderately dependent on Ca^{2+}, which shows that the binding site in *Hs*cen2 is in the C-terminus, which is the Ca^{2+} dependent terminus of the protein and has an affinity of ~ $10^{7}{}_{M}{}^{-1}$. On the other hand, the *Hs*cen2 N-terminal domain does not play a significant role among the protein-protein interaction (PPI).[41–43]

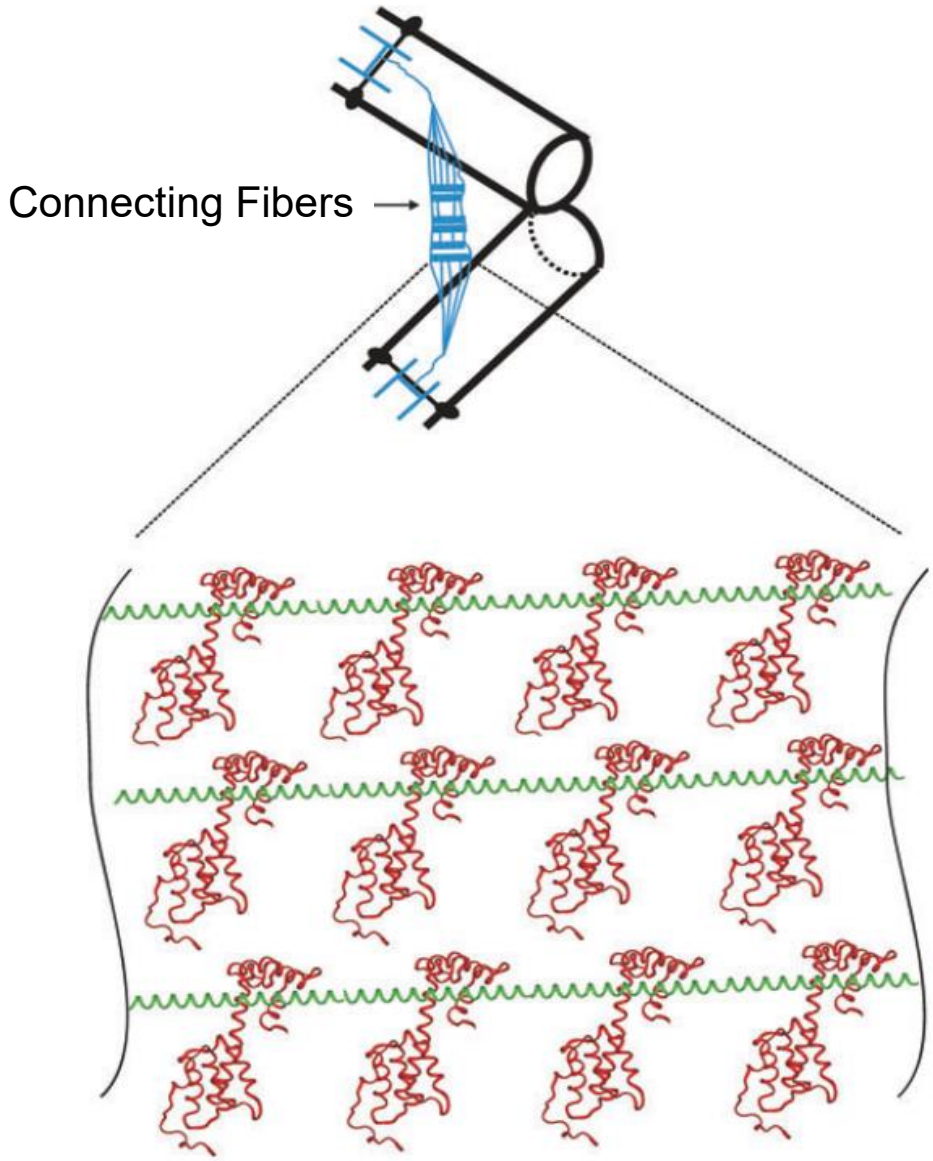


Figure 7. Model of the multimeric assembly of Sfi1–centrin. Assembly model of *Hs*Sfi1 and *Hs*centrin2 in the basal body connecting fiber structure. Centrin is represented in a red ribbon diagram bound to segments of Sfi1 (green) through its C-terminal domain. *Adapted with permission from John Wiley and Sons, Martinez-Sanz, 2006.* [43]

Melittin

Melittin, a peptide, is the major component of bee venom. Because it is associated with cell lysis, it is a potential candidate for cancer therapy. Clinical treatments with this peptide have shown numerous anti-cancer effects including the inhibition of the proliferation of certain cancer cells.[44,45] Pastrana-Rios et al. 2011 characterized with biophysical techniques the complex between melittin (26 aa) and *Chlamydomonas reinhardtii* centrin (*Cr*cen) (Figure 8). Two-dimensional infrared correlation spectroscopy was used to study ^{13}C-*Cr*cen and melittin, showing that melittin's helicity increased and *Cr*cen was stabilized during the interaction.[17]

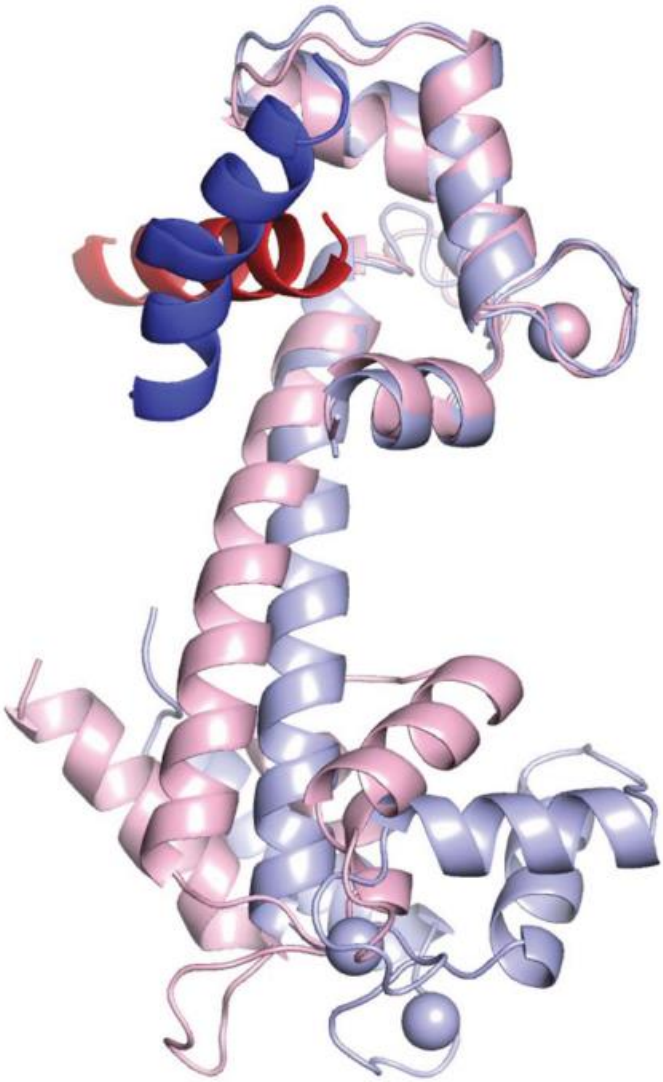


Figure 8. Superimposition of the *Cr*cen/melittin complex and the *Hs*cen2/XPC complex. *Cr*cen (light blue) is bound to melittin (blue), and *Hs*cen2 (light pink) is bound to XPC peptide (red). The C-termini of the CaBPs were aligned to show the distinct interaction with the respective peptides. *Adapted with permission from John Wiley and Sons, del Valle Sosa, 2011.* [17]

XPC

Xeroderma pigmentosum group C (XPC) is a protein involved in NER, detecting damage in DNA. To repair the DNA, XPC binds to a complex including RAD23A, RAD23B, XPA, RPA, TFIIH complex, and a short ssDNA segment. It recognizes the damaged DNA *via* distortions in the DNA's helical structure. XPC also recognizes and interacts with undamaged unpaired bases in ssDNA.[11] Nishi et al. 2005 [46] showed that *Saccharomyces cerevisiae* centrin is part of NER and forms a complex with XPC. In mutation analysis experiments the authors found the co-overexpression of XPC and the C-terminal domain of centrin, suggesting an interaction with that domain. The N-terminal domain of centrin also plays an important role in the recognition of damaged DNA by XPC, suggesting that centrin acts like a bridge between XPC and part of the NER complex.[46,47] *In vitro* studies by Craescu et al. 2006 identified two putative centrin binding sites on XPC, which were confirmed by experiments to evaluate binding properties of the complex using ITC. Only one of the XPC peptides interacted strongly with the C-terminal domain of centrin in a Ca^{2+}-dependent manner ($K_a = 1.1 \times 10^8 M^{-1}$). Circular dichroism and nuclear magnetic resonance were used to characterize the complex (Figure 9).[48] In 2006 and 2007, Thompson and Craescu characterized the crystal structure of the complex to understand the electrostatic and non-polar intermolecular interactions between centrin and XPC. The crystal structure shows that the C-terminal domain of centrin binds to both XPC and two Ca^{2+} atoms, whereas the N-terminal domain remains in a closed conformation (Figure 10).[12,13,15]

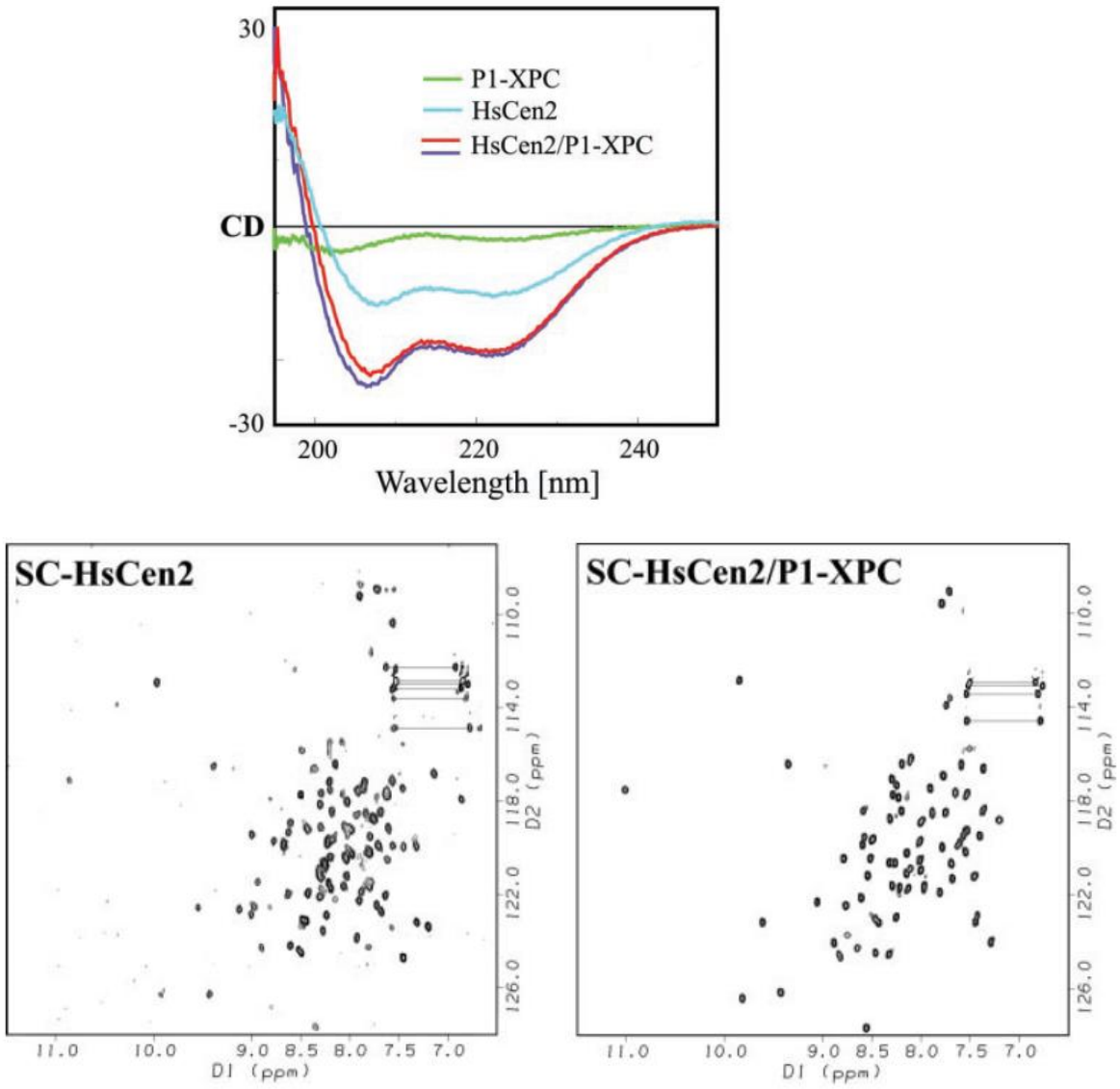


Figure 9. CD and NMR analysis of *Hs*cen2 and XPC interaction. Top: Far-UV CD spectra of XPC (30 µM), *Hs*cen2 (30 µM), and the complex at 22 °C show that *Hs*cen2 adopts a more helical secondary structure in the presence of the XPC peptide (1:1 and 2:1 peptide-to-protein ratio in red and magenta, respectively). Bottom: HSQC spectra at 500 MHz of *Sc*cen alone (left) and in the presence of the XPC peptide (right) at equimolar concentration. Spectra were recorded at 308 K. Shifting and fading of the signals suggest an interaction between the protein and the peptide. *Adapted with permission from Elsiever, Popescu, 2003.* [50]

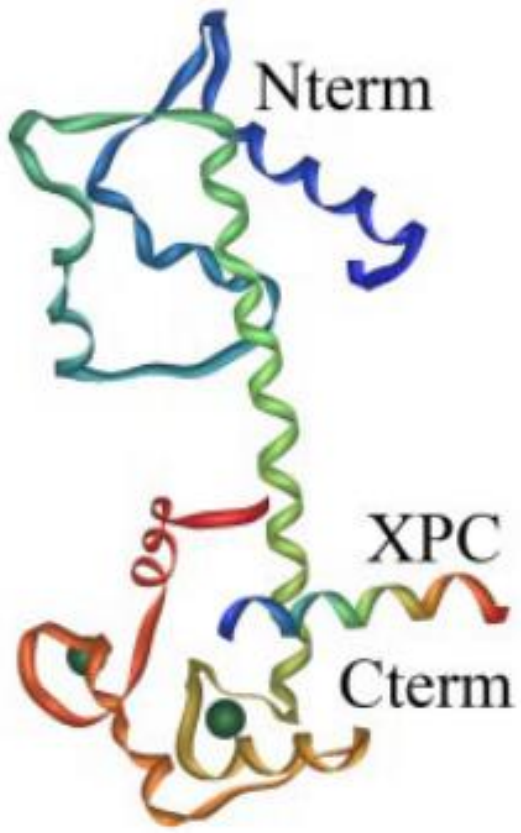


Figure 10. Structure of *Hs*cen2 bound to XPC peptide. An ordered helical linker separates N-terminal and C-terminal domains, and XPC peptide is bound to the C-terminal domain. *Adapted with permission from American Society for Biochemistry and Molecular Biology, Thompson, 2006.* [15]

Kar1

Kar1, another centrin target, plays an important role in the intranuclear and extranuclear microtubules, helping Cdc31, the centrin homolog in yeast, to initiate spindle pole body duplication. Mutagenesis and gene replacement studies in 1987 demonstrated that Kar1 is essential for cell division. These *in vitro* studies demonstrated defects in spindle body duplication and chromosome disjunction when the *Kar1* gene is mutated.[50] Studies in the yeast model found that Kar1, Cdc31, and Sfi1 form a complex to initiate spindle body duplication. Also, binding assays demonstrate the interaction between Kar1 and Sfi1. On a model developed by Seybold et al. 2015 suggest that Cdc31 and Kar1 provide cross-links, while Sfi1 stabilizes the bridge to ensure spindle pole body separation during cell division.[49] A structural model was presented by Meyn et al. 2006 [50] validating the interaction between Cdc31 and a segment of Kar1 (Figure 11).

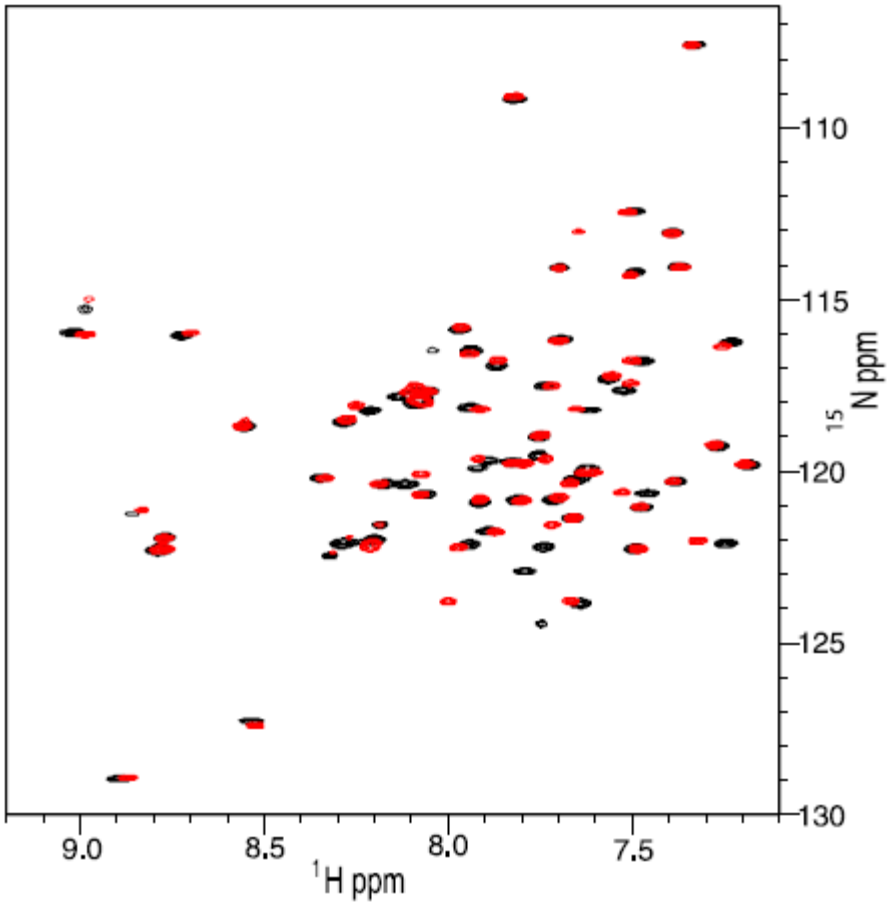


Figure 11. Structural characterization of the complex of Cdc31 and Kar1 peptide. Superimposition of HSQC NMR spectra acquired at 328 K for Cdc31 (black spectrum) and the complex (red spectrum). *Adapted with permission from Elsevier, Meyn, 2006*) [50]

Krr1, a novel centrin target

Krr1 is a 381 amino acids protein required for ribosome biogenesis within the nucleus (Figure 12), a pre-requisite step for the subsequent synthesis of proteins within the cell.[24] In 1996, Gromadka et al. [51] suggested the name *KRR1* for the gene that encodes this protein, based on its highly basic sequence. Krr1 is required for the proper processing of pre-rRNA, for the synthesis of 18S (Svedberg unit S, 10^{-13}s), and for the assembly of the 40S subunit.[52] The ribosome is a supramolecular complex composed of RNA and protein subunits. These subunits have been characterized by sedimentation analysis, and are therefore referred to by their sedimentation coefficient. In standard conditions, moderate expression of *Hs*Krr1 was found in a group of tissues (Figure 13A) according to the Human Protein Atlas project.[53] Furthermore, abnormalities in ribosome biogenesis cause specific clinical syndromes and have recently been associated with tumorigenesis. Melanoma, skin, liver, renal, testicular, colorectal, and breast cancers showed a strong Krr1 signal in the nucleus of the cancer cells (Figure 13B).[54] *In vivo* studies in 2000 showed that the depletion of Krr1 causes a significant reduction of 40S ribosomal subunit, suggesting the dysregulation of rRNA synthesis.[55] Studies in *Saccharomyces cerevisiae* (*Sc*) present the important role of Krr1 in ribosome formation, specifically in the coordination and folding of the pre-rRNA complex. Zheng et al. 2014 published a co-crystal structure (Figure 14) [60] of Krr1 and Faf1, both of which are members of the small ribosomal subunit complex. Mutations on *Sc*Krr1 have a detrimental effect on the complex formation causing cell lethality. The structure of the *Sc*Krr1-*Sc*Faf1 complex reveals that Krr1 contains two homology domains that play different roles in complex

formation. The first homology domain acts as assembly factor, and the second contains an RNA binding surface.

*The structure of Hs*Krr1 *has not been determined to date*, although this centrin target does contain a KH domain (Figure 1) [6] which has been identified as a nucleic acid recognition motif.

```
          10         20         30         40         50         60
     MASPSLERPE KGAGKSEFRN QKPKPENQDE SELLTVPDGW KEPAFSKEDN
PRGLLEESSF

          70         80         90        100        110        120
     ATLFPKYREA YLKECWPLVQ KALNEHHVNA TLDLIEGSMT VCTTKKTFDP
YIIIRARDLI

         130        140        150        160        170        180
     KLLARSVSFE QAVRILQDDV ACDIIKIGSL VRNKERFVKR RQRLIGPKGS
TLKALELLTN

         190        200        210        220        230        240
     CYIMVQGNTV SAIGPFSGLK EVRKVVLDTM KNIHPIYNIK SLMIKRELAK
DSELRSQSWE

         250        260        270        280        290        300
     RFLPQFKHKN VNKRKEPKKK TVKKEYTPFP PPQPESQIDK ELASGEYFLK
ANQKKRQKME

         310        320        330        340        350        360
     AIKAKQAEAI SKRQEERNKA FIPPKEKPIV KPKEASTETK IDVASIKEKV
KKAKNKKLGA

          370        380
     LTAEEIALKM EADEKKKKKK K
```

Figure 12. The sequence of *Hs*Krr1 protein. Underlined is the sequence corresponding to the KH domain, and highlighted in yellow is the specific sequence that interacts with oligonucleotides, the GXXG loop described by Valverde et al. 2008.[6] (Swiss Protein Database Accession number: Q13601)

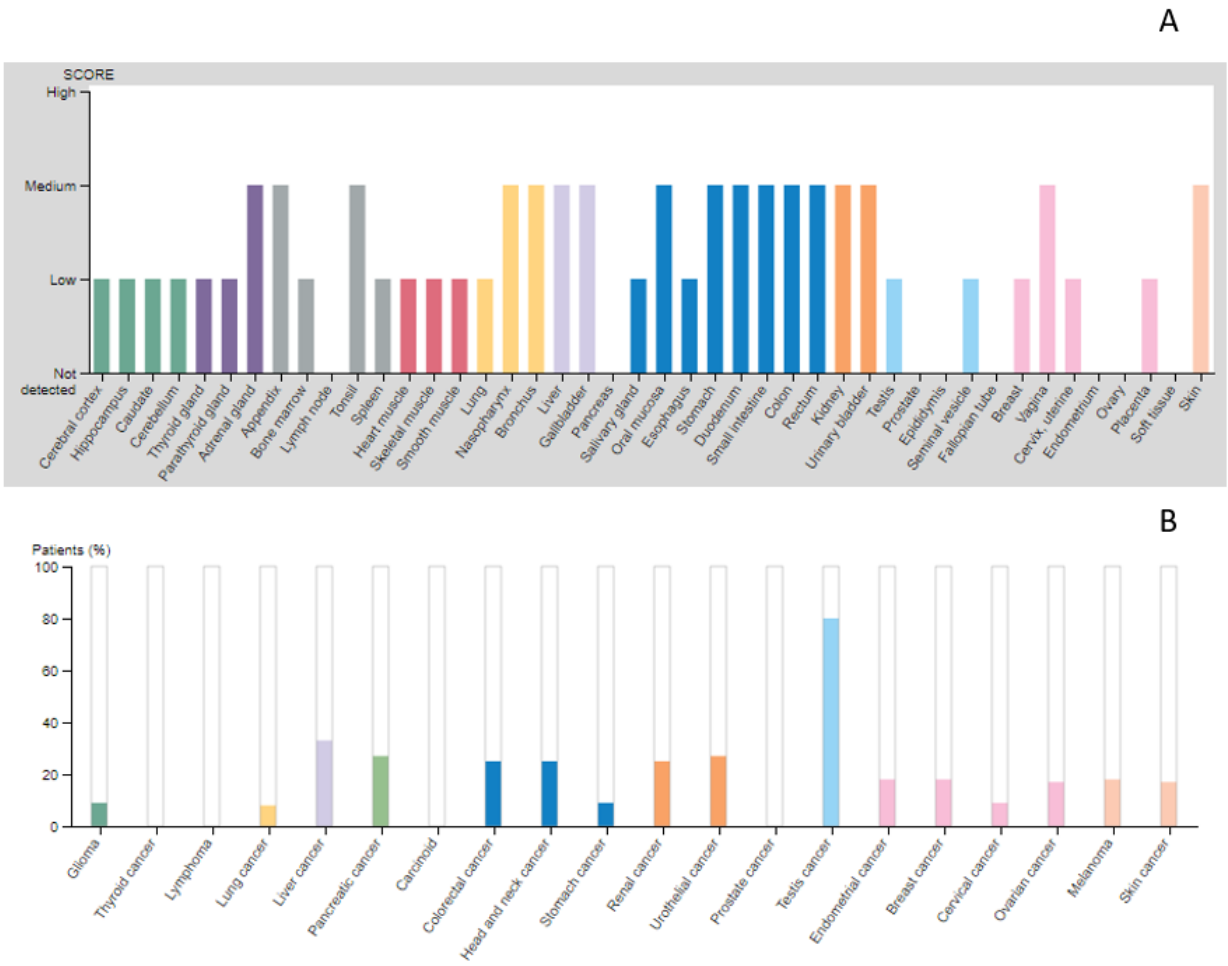


Figure 13. *Hs*Krr1 expression levels in different human tissues. (A) Protein expression data for 44 human tissues. Color coding is based on tissue groups. In normal conditions, a moderate level of expression is shown for 17 types of tissue, according to the Human Protein Atlas project. (B) *Hs*Krr1 expression was detected in melanoma, skin, liver, renal, testicular, colorectal and breast cancers, showing strong nuclear/nucleolar staining.*Adapted with permission from the American Association for the Advancement of Science, Uhlen, 2015 and 2017*. [53,54]

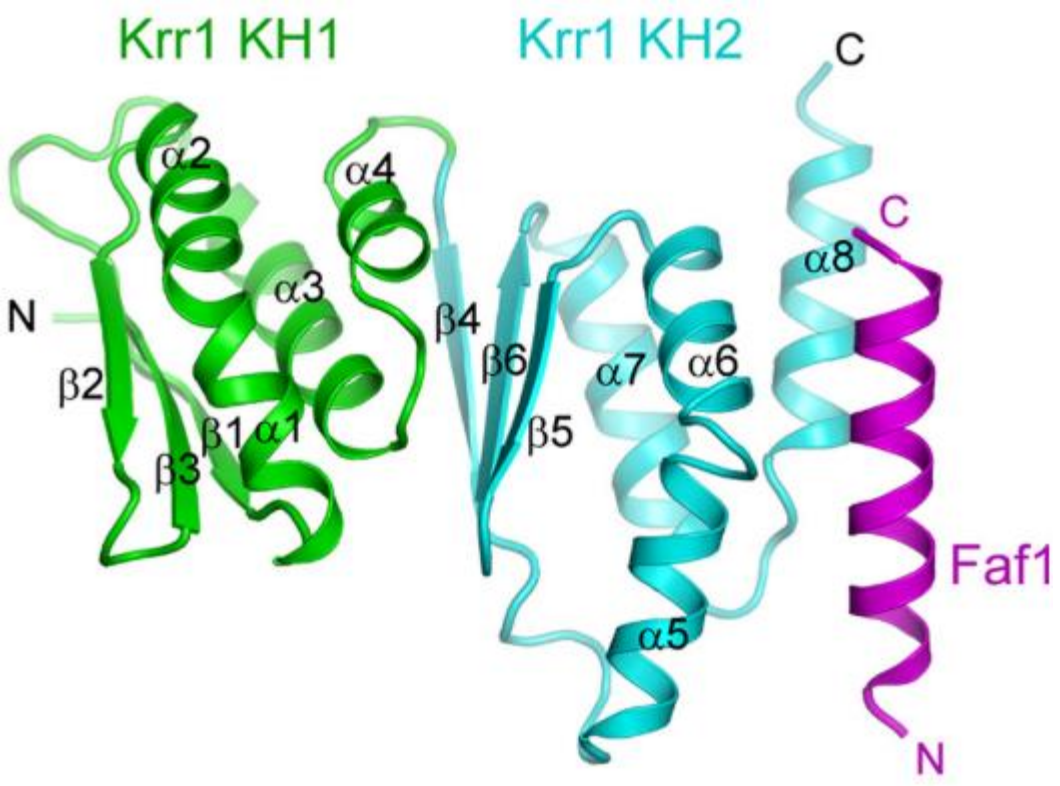


Figure 14. Structure of the *Sc*Krr1-*Sc*Faf1 complex. The first KH is shown in green and the second KH domain in cyan. Faf1 is presented in magenta. The Faf1 peptide interacts with Krr1 by hydrogens bonds. *Adapted with permission from American Society for Biochemistry and Molecular Biology, Zheng, 2014.* [60]

K homology domain

The KH domain was first identified in small nuclear ribonucleoprotein particles. The KH domain is a conserved domain present in ribosomal proteins and proteins involved in transcription regulation, splicing, and translational control, i.e., regulatory functions within the nucleus.[15–18] This domain is composed of approximately 52 amino acids, and crystallographic data confirms the existence of two types of secondary structures: *type 1* are typically in proteins found in eukaryotic organisms and consist of two or three antiparallel β-strands exposed on one surface and two or three α-helices on the other surface; *type 2* are typically found in proteins in prokaryotic organisms and consist of three β-strands—two in parallel and the last one antiparallel—exposed on one surface and two or three α-helices on the other surface (Figure 1). The KH domain has been found to contribute to the recognition of specific RNA or ssDNA segments.[15,19–21]

The KH domain GXXG Loop play an important role in how the proteins work, including Krr1. The interaction between the nucleic acids and the protein takes place at the helical surface of the KH domain in the GXXG loop (Figure 1).[20,22,23] In general, several weak interactions are observed within the KH domain-nucleic acid complex such as hydrogen bonding and electrostatic and hydrophobic interactions between the bases and amino acid side chains.[15] Hollingworth et al. 2012 designed KH domain double mutation specifically in the GXXGlp to evaluate the nucleic acid recognition and binding capacity.[20] (Figure 15). These mutations did not affect the structural conformation of the domain but affected the affinity for nucleic acids and the mode of recognition.

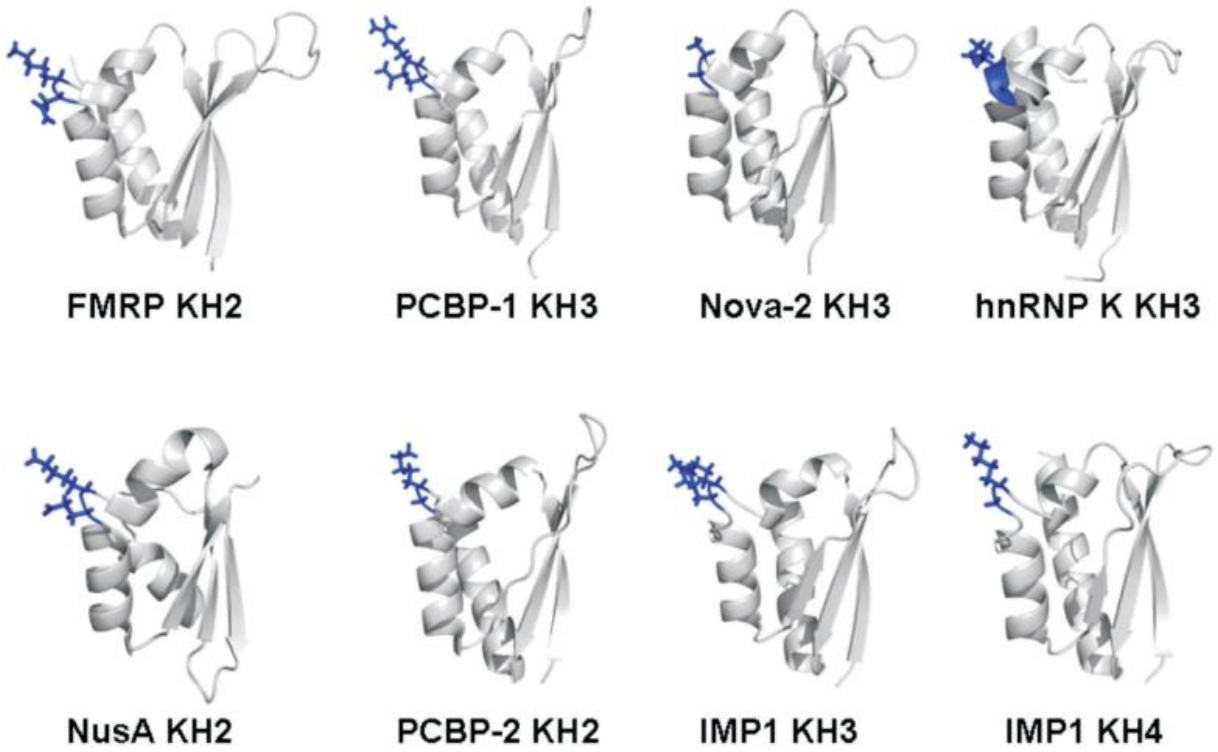


Figure 15. The GXXG loop in KH domain proteins. A ribbon cartoon representation of the structure of eight KH domains from different proteins, in white, highlights the amino acids of the GXXG loop, displayed in blue. *Adapted with permission from Oxford University Press, Hollingworth, 2012.* [57]

CHAPTER III
MATERIALS AND METHODS

Protein Expression and Purification

The recombinant proteins and domains *(Hs*cen2, *Hs*Krr1, *Hs*cen2 C-terminal, and *Hs*cen2 N-terminal) (Table 1) were overexpressed in *E. coli.* The cDNA of *Hs*Krr1, *Hs*Krr1 KH domain, *Hs*cen2 C-terminal, and *Hs*cen2 N-terminal were obtained by Blue Heron Biotech LLC and were subcloned in a pEX-1 expression vector. The cDNA of *Hs*cen2 was generously supplied by W. Chazin (Vanderbilt University, Nashville, TN). The plasmids were transformed into BL21 DE3 *E. coli* cells. The recombinants were then sequenced to verify the reading frame by Roswell Park Cancer Institute (Buffalo, NY) and Genewiz (Boston, MA). The cells were grown in Terrific Broth medium and incubated at 37 °C with 250 rpm agitation. Recombinant protein synthesis was induced at the onset of the log phase with isopropyl-β-D-thiogalactopyranoside (0.5 mM) for 4 h. The cells were harvested by centrifugation at 8,000 rpm. After purification (see below), the concentration of the proteins was measured using a UV spectrophotometer. The extinction coefficient (ε) of *Hs*cen2 and *Hs*cen2 C-terminal is 1,490 $M^{-1}cm^{-1}$ at 280 nm; the ε of *Hs*Krr1 is 26,930 $M^{-1}cm^{-1}$ at 280 nm. *Hs*cen2 N-terminal concentration was determined by drying a protein sample and a buffer sample with the same excipient conditions and calculating the weight difference between samples because the protein sequence does not contain Trp, Tyr, or Cys and thus is not visible by UV spectrophotometry. Direct UV absorbance measurements were taken only after dialysis of the desired protein against the appropriate buffer. All

proteins samples were preserved at -80°C and thawed before conducting biophysical experiments.

Table 1. Sequence details of *Hs*cen2, *Hs*Krr1, *Hs*cen2 C-terminal, and *Hs*cen2 N-terminal

RECOMBINANT PROTEIN	SEQUENCE
***HS*CEN2**	HHHHHHGAMASNFKKANMASSSQRKRMSPKPELTEEQKQEIREAFDLFDADGTGTIDVKELKVAMRALGFEPKKEEIKKMISEIDKEGTGKMNFGDFLTVMTQKMSEKDTKEEILKAFKLFDDDETGKISFKNLKRVAKELGENLTDEELQEMIDEADRDGDGEVSEQEFLRIMKKTSLY
***HS*CEN2 N-TERMINAL**	MASNFKKANMASSSQRKRMSPKPELTEEQKQEIREAFDLFDADGTGTIDVKELKVAMRALGFEPKKEEIKKMISEIDKEGTGKMNFGDFLTVMTQKMSEK
***HS*CEN2 C-TERMINAL**	MEKDTKEEILKAFKLFDDDETGKISFKNLKRVAKELGENLTDEELQEMIDEADRDGDGEVSEQEFLRIMKKTSLY
***HS*KRR1**	MASPSLERPEKGAGKSEFRNQKPKPENQDESELLTVPDGWKEPAFSKEDNPRGLLEESSFATLFPKYREAYLKECWPLVQKALNEHHVNATLDLIEGSMTVCTTKKTFDPYIIIRARDLIKLLARSVSFEQAVRILQDDVACDIIKIGSLVRNKERFVKRRQRLIGPKGSTLKALELLTNCYIMVQGNTVSAIGPFSGLKEVRKVVLDTMKNIHPIYNIKSLMIKRELAKDSELRSQSWERFLPQFKHKNVNKRKEPKKKTVKKEYTPFPPPQPESQIDKELASGEYFLKANQKKRQKMEAIKAKQAEAISKRQEERNKAFIPPKEKPIVKPKEASTETKIDVASIKEKVKKAKNKKLGALTAEEIALKMEADEK

***H*scen2**

The harvested pellet of His-tagged *H*scen2 was homogenized to lysate with 25 mM Tris (pH 8.0), 300 mM NaCl, 20 mM imidazole, 1 mM phenylmethanesufonyl fluoride (PMSF), 1% Nonidet P-40 (NP-40), and a cocktail of protease inhibitors EDTA-free using 5 non-consecutive minutes of sonication at an amplitude of 50, with 5 seconds on, followed by 10 seconds off on ice. The remaining solution was ultracentrifuged at 20,000 rpm for 20 min. The supernatant was filtered and loaded onto a previously equilibrated nickel column. The sample in the nickel column was washed three times with 20 mM Tris (pH 8.0), 100 mM NaCl, and 20 mM imidazole and then the recombinant protein was eluted with 20 mM Tris (pH 8.0), 100 mM NaCl, and 500 mM imidazole. The protein sample was dialyzed against the desired buffer to perform the biophysical studies.

*Hs*cen2 C-terminal domain

The purification protocol has been established in our laboratory for the full length centrin isoforms.[24–27] Variations in the protocol were conducted to improve the purification of the C-terminal domain. The harvested pellet of *Hs*cen2 C-terminal was homogenized to lysate with 50 mM Tris (pH 7.4), 500 mM NaCl, 1% Tween 20, and a cocktail of protease inhibitors EDTA-free using a microfluidizer processor system (Microfluidics, Newton, MA). The remaining solution was ultracentrifuged at 20,000 rpm for 20 min. The supernatant was filtered and loaded onto a previously equilibrated Phenyl Sepharose CL4B affinity column followed by anion exchange chromatography. The protein sample was dialyzed against the desired buffer to perform the biophysical studies.

*Hs*cen2 N-terminal domain

The harvested pellet of *Hs*cen2 N-terminal was homogenized to lysate with 50 mM Tris (pH 7.4), 500 mM NaCl, 1% Tween 20, and a cocktail of protease inhibitors EDTA-free using a microfluidizer processor system (Microfluidics, Newton, MA). The remaining solution was ultracentrifuged at 20,000 rpm for 20 min. To the total volume of the supernatant, 4 mM of $CaCl_2$ and $MgCl_2$ were added and the sample was heated at 45 °C for 30 min. After the heating step, the sample was centrifuged at 8,000 rpm for 15 min. The supernatant was filtered and loaded onto a previously equilibrated weak anion exchange chromatography column. The protein sample was dialyzed against the desired buffer to perform the biophysical studies.

***Hs*Krr1**

The harvested pellet of *Hs*Krr1 was homogenized to lysate with 20 mM Tris (pH 7.4), 50 mM NaCl, 1% Tween 20, and a cocktail of protease inhibitors EDTA-free using a microfluidizer processor system (Microfluidics, Newton, MA). The remaining solution was ultracentrifuged at 20,000 rpm for 20 min. The supernatant was filtered and loaded onto a previously equilibrated cation exchange chromatography column followed by size exclusion chromatography using Sephacryl resin and a 20 mM HEPES, 150 mM NaCl (pH 7.4) buffer.

Peptide Preparation

*Hs*Krr1 KH domain GXXG loop

The *Hs*Krr1 KH domain GXXG loop (GXXGlp, Ac-KRRQRLIGPKGSTLKALELLTNCW-NH2) was purchased from Biosynthesis, Inc. Part of the sample was subjected to TFA removal by repeated lyophilization with 0.1 N HCl, followed by exhaustive dialysis against the appropriate buffer as per our established method.[25,28–30] The peptide concentration was determined using a UV/VIS spectrophotometer. The ε of GXXGlp is 5,500 $M^{-1}cm^{-1}$ at 280 nm.

ssDNA Synthesis and Purification

Two single stranded oligonucleotides (5'ACCCCA'3, 6mer, MW=1,800g/mol; 5'TTCCCCTCCCCA'3, 13mer, MW=3,800g/mol) were synthesized using an AKTA Oligopilot Plus system. The oligonucleotide mixture came off as a bound ssDNA chain on solid support. The ssDNA was cleaved from the solid support using acetonitrile 40% (w/v) and heated for 8 h at 50 °C in a water bath. The purification of the oligonucleotides was performed using anion exchange chromatography followed by a desalting G-10 Sephadex column. The oligonucleotides were quantified spectrophotometrically at 260 nm.

FT-IR Spectroscopy

A comparative analysis of GXXGlp under various conditions was performed using FT-IR spectroscopy. Also, studies of ssDNA 13mer and GXXGlp-ssDNA13mer complex were performed using FT-IR spectroscopy. The samples were (a) 33.5 mg/mL GXXGlp in the presence of TFA (i.e., commercial peptide product without the removal of the solid phase synthesis by product), and in the absence of TFA, (b) high-concentration 50.1 mg/mL GXXGlp, (c) low-concentration 16.7 mg/mL GXXGlp, and (d) GXXGlp/ssDNA13mer complex (16mg/mL:21.6mg/mL). The buffer for each sample consisted of 50 mM HEPES, 150 mM NaCl, and 4 mM $CaCl_2$, at pH 7.4. All the samples were fully H → D exchanged as previously described.[25–28,31–33] The final volume of the sample, 60 µL, was placed on CaF_2 custom milled cells (Spectral Systems, LLC from Hopewell, NY). A reference cell was prepared with only the buffer (50 mM HEPES, 150 mM NaCl, and 4 mM $CaCl_2$, pH 7.4) and fully H → D exchanged. Both cells were set in a custom dual-chamber cell holder. The FT-IR spectrophotomer used was a Jasco Model 6200 equipped with a MCT detector and sample shuttle interface. Thermal control was achieved and monitored with a thermocouple in close contact with the sample cell. The data acquisition was performed at 5 °C intervals. Only after the desired temperature was reached and thermal equilibrium (15 min) was achieved did acquisition begin. The data collection was 512 scans, apodized with a triangular function and Fourier transformed to provide a resolution of 4 cm^{-1} with data encoded every 2 cm^{-1}.[24–28,31–33]

2D IR Correlation Spectroscopy and Spectral Analysis

2D IR correlation spectroscopy is a technique developed by Dr. Isao Noda[34–37] that uses the FT-IR series of sequential spectra as a function of a perturbation, in this case temperature (5−90 °C), to obtain a difference spectral data set by subtraction of the initial spectrum from all subsequent spectra (see Appendix A for details).

A cross correlation function was applied to the spectral data sets, which results in two separate, symmetrical and asymmetrical 2D plots. The first plot is referred to as the synchronous plot. It contains positive peaks on the diagonal, known as the auto peaks, and summarizes the changes observed in the spectral data set. The relationship established in this synchronous plot relates the spectral intensity changes that are in-phase with one another. The second 2D plot is known as the asynchronous plot and relates the out-of-phase intensity changes, enhances the resolution of the spectral region of interest, and can easily be distinguished from the previous plot because it lacks peaks on the diagonal.

Both plots contain off-diagonal peaks, which are referred to as cross peaks. These peaks correlate the spectral changes observed. Spectral intensity changes observed are due to the incremental thermal perturbation applied to the peptide sample. Therefore, the information in both plots allows for the determination of the sequential order of molecular events that occur during the perturbation, following Noda's rules (see Appendix B).[34–37] The spectral data was not manipulated except for baseline correction.

Circular Dichroism

Far-UV CD spectra of the protein, peptide, and oligonucleotide samples (20.7 μM TFA free GXXGlp, 2.7 μM ssDNA 13mer, GXXGlp/ssDNA13mer complex at equimolar concentration, and 2.3μM *Hs*Krr1) in 2.5 mM HEPES, 7.5 mM NaCl, 0.2 mM $MgCl_2$, and 0.2 mM $CaCl_2$, at pH 7.4 were subjected to a 10-scan acquisition within the spectral range of 250−193 nm, at a scan rate of 20 nm/min at 20°C using a Jasco spectropolarimeter model J-810 (Tokyo, Japan) equipped with a temperature controller. In addition, temperature dependence spectra of GXXGlp, GXXGlp/ssDNA13mer complex, and *Hs*Krr1 were also collected from 5 to 95°C at a rate of 2°C/min, while monitoring 222 nm. The CD absorbance data were normalized to mean residue molar ellipticity and smoothed with an adaptive smoothing method.[38]

Isothermal Titration Calorimetry

Titration studies of *Hs*Krr1 and *Hs*cen2 C-terminal were performed using Nano ITC (TA Instruments, New Castle, DE). The buffer for the experiments was 20 mM HEPES, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$ at pH 7.4 and 20 °C. Both proteins were exhaustively dialyzed against the buffer conditions mentioned before. A series of consecutive 5µL titrations of *Hs*Krr1 at 10 µM *were performed to Hs*cen2 C-terminal 100µM.

Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) ^{15}N-labeled *H*scen2 samples were obtained by expressing the protein using a minimal medium culture (M9) containing $^{15}NH_4Cl$ as the source of nitrogen in the medium. The titration experiments were performed using the unlabelled GXXGlp. All the samples at the 0.3-0.9 mM concentration were obtained after exhaustive dialysis against 20 mM Tris (pH 6.5), 100 mM NaCl, 5 mM $CaCl_2$, and 4% (v/v) D_2O. The unlabelled GXXGlp was added in the following molarity ratios: 1:1 and 1:2. Furthermore, the titration experiments were performed with *Cr*cen (obtained from Dr. Chazin laboratory) and the unlabelled peptide with the following molarity ratios: 1:1.2 and 1:2. Two-dimensional heteronuclear single quantum coherence (HSQC) spectroscopy was performed on Bruker Avance AV-III 800 spectrometers (Massachusetts, US) at 308 K. The NMR data was collected and processed with Bruker Topspin 3.2 software from Massachusetts, US.

Crystallization Tray

*Hs*Krr1 crystals were grown using the sitting drop vapor diffusion method at 23 °C by mixing 2 µL of 4 mg/mL protein in 20 mM HEPES, 150 mM NaCl at pH 7.4 and 2 µL of the reservoir solution containing 0.100 M triammonium citrate pH 8.0 and 15% (w/v) PEG 3350. Amorphous solid appeared after 2 days.

Small Angle X-ray Scattering and Modelling

Small angle X-ray scattering (SAXS) experiments were performed with *Hs*Krr1 (6.5 mg/mL) in 20 mM HEPES, 150 mM NaCl at pH 7.4 at room temperature. The measurements were taken using the Beamline G1, located at the Cornell High Energy Synchrotron Source (CHESS) at Cornell University. Scattering experiments on the protein sample were performed using a microfabricated sample cell. Background intensities were obtained by performing scattering experiments on the exact buffers of the protein sample using the same cell. The data obtained by 10 sec exposure were analyzed using RAW software from New York, US.

CHAPTER IV
RESULTS AND DISCUSSION

Bacterial Protein Expression and Purification

*Hs*cen2

The overexpression of His-tagged *Hs*cen2 using Terrific Broth medium resulted in a high-yield cell culture, obtaining a pellet of 3.7 g/L. As shown in Figure 16, the protein was overexpressed after induction with IPTG, showing a prominent band around 22 kDa. To purify the protein, the pellet was lysed with the lysis buffer and underwent an ultracentrifugation step that removed significant amount of cell debris. The supernatant was filtered with a 0.45 micron polyethersulfone (PES) membrane to clarify the sample. After the isolation step, the sample was loaded into a nickel affinity column. A 4-20% Bis-Tris SDS-PAGE gel was run of the elution fractions, confirming the purity of the sample removing the host cell proteins, showing the presence of a pure protein band near 22 kDa (Figure 16).

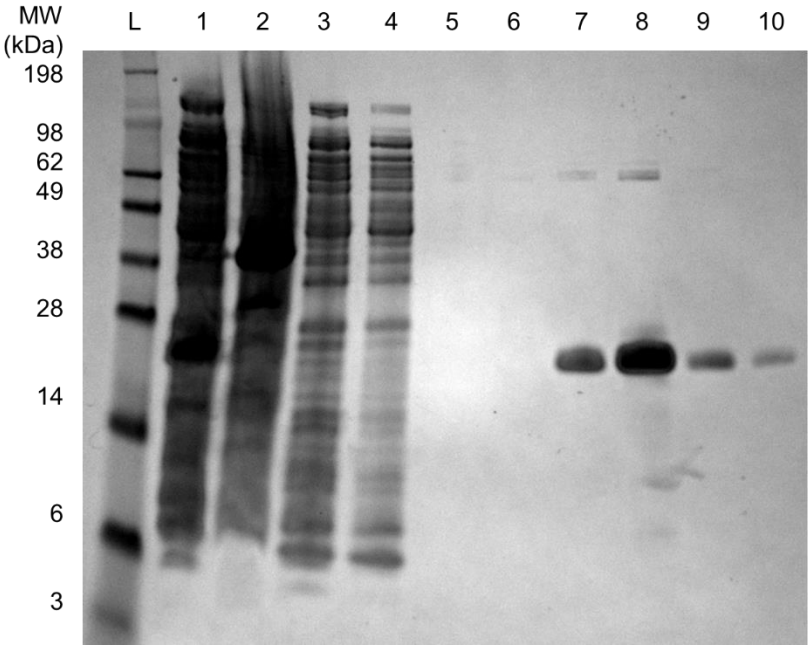


Figure 16. 4-20% Bis-Tris gradient SDS-PAGE for the purification of *Hs*cen2. L: molecular weight ladder, 1: centrifugation step supernatant, 2-6: wash steps in nickel column, 7-10: elution fractions from nickel column. Samples in fractions 7-10 were pooled and filtered with a PES membrane with a 30 kDa cut off to eliminate the high molecular weight impurity.

*Hs*cen2 C-terminal

Isolated supernatant of *Hs*cen2 C-terminal was filtered with a 0.45 micron polyethersulfone (PES) membrane filter to clarify the sample. For purification the protein sample was loaded onto a Phenyl Sepharose CL-4B affinity column and, as a polishing step, an anion exchange column. To analyze the purity of the sample, SDS-PAGE was used (Figure 17). The sample was collected and buffer exchanged into 20 mM HEPES, 150 mM HEPES, 4 mM $CaCl_2$, and 4 mM $MgCl_2$ at pH 7.4 for further biophysical studies.

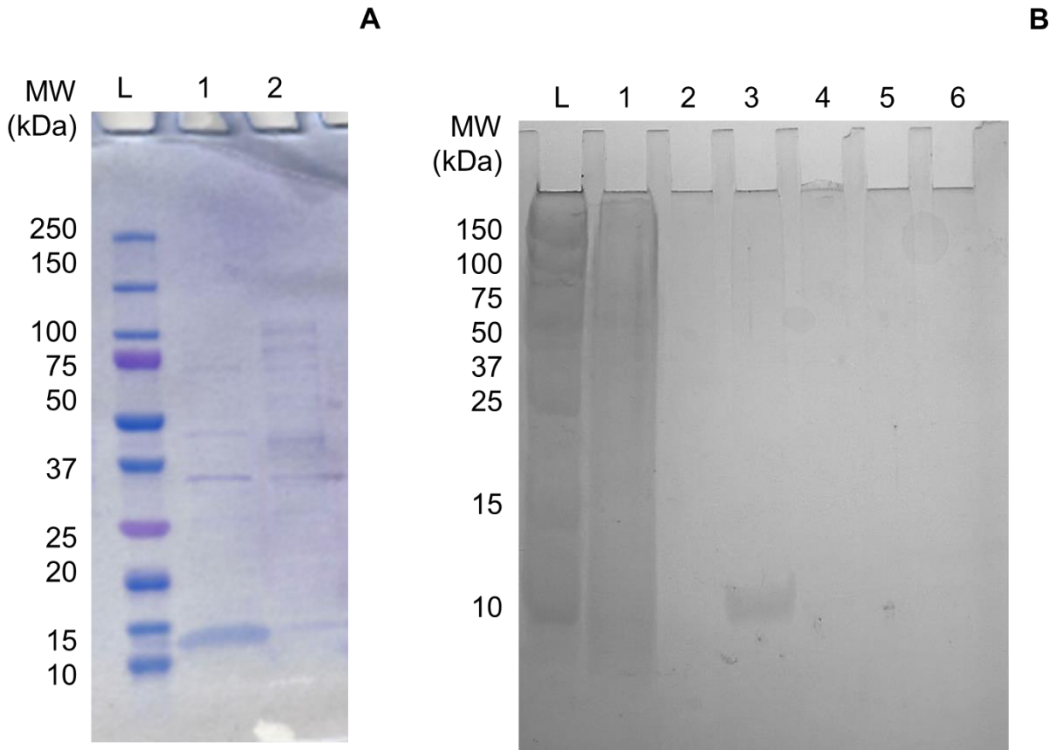


Figure 17. Expression and purification of *Hs*cen2 C-terminal. (A) 4-20% Bis-Tris gradient SDS-PAGE for the expression of *Hs*cen2 C-terminal. L: molecular weight ladder, 1: post-induction cell lysis sample, 2: pre-induction cell lysis sample. (B) 15% SDS-PAGE for the purification of *Hs*cen2 C-terminal. L: molecular weight ladder, 1: unbound fraction on anion exchange chromatography, 2: wash step with buffer in the absence of NaCl, 3: elution fraction from the anion exchange column with 20% NaCl, 4: elution fraction from the anion exchange column with 40% NaCl, and 5: elution fraction from the anion exchange column with 60% NaCl. The sample in 20% NaCl was collected with the desired protein.

*H*scen2 N-terminal

The lysate solution of *H*scen2 N-terminal was ultracentrifuged at 20,000 rpm for 20 min. To the total volume of the supernatant, 60 mL, 4 mM of $CaCl_2$ and 4 mM $MgCl_2$ were added and the sample was heated at 45 °C for 30 min in a water bath. After the heating step most of the *E. coli* host proteins became denatured and the solution turned cloudy. The sample was centrifuged at 8,000 rpm for 15 min and the supernatant was filtered with PES membrane centrifuge concentrators. In the first filtration step the sample was filtered with a membrane with a cut off size of 50 kDa. In this first step most of the high molecular weight protein remained in the retentate, clarifying the sample more. A second filtration step was performed with a 3 kDa membrane pore size to concentrate the sample and load it onto a previously equilibrated weak anion exchange chromatography column. The target protein elutes at 20% NaCl. To analyse the purity of the sample, SDS-PAGE was used (Figure 18). The sample was collected and buffer exchanged into 20 mM HEPES, 150 mM NaCl, 4 mM $CaCl_2$, and 4 mM $MgCl_2$ at pH 7.4 for further biophysical studies.

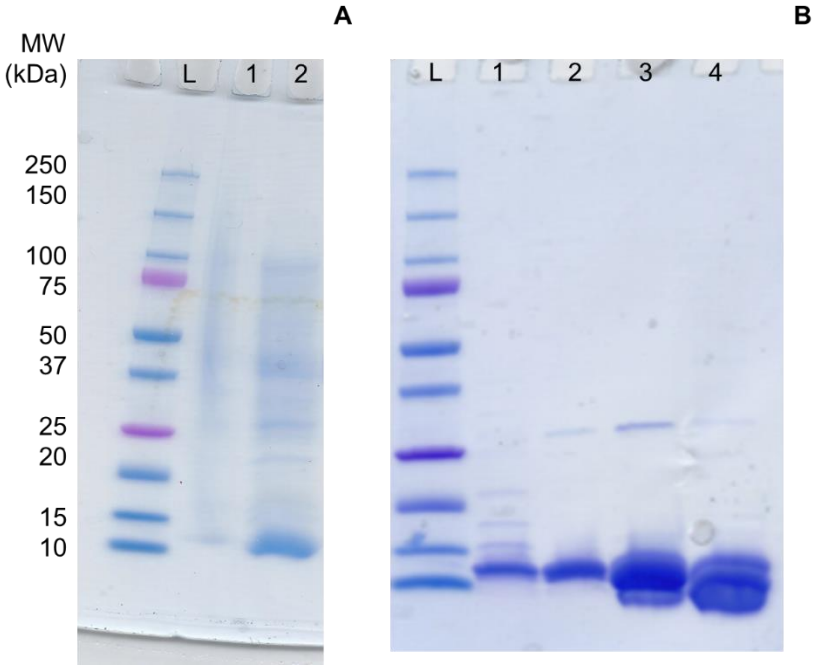


Figure 18. Expression and purification of *Hs*cen2 N-terminal. (A) 4-20% Bis-Tris gradient SDS-PAGE for the expression test of *Hs*cen2 N-terminal. L: molecular weight ladder, 1: pre-induction cell lysis sample, 2: post-induction cell lysis sample. (B) 4-20% Bis-Tris gradient SDS-PAGE for the purification of *Hs*cen2 N-terminal. L: molecular weight ladder, 1: unbound fraction from anion exchange chromatography, 2-4: elution fractions from the anion exchange column with 20% NaCl. Samples in 20% NaCl were pooled and filtered with a PES membrane filter with a cut-off of 30k Da to remove the high molecular weight impurity.

***Hs*Krr1**

The lysate solution of *Hs*Krr1 was ultracentrifuged at 20,000 rpm for 20 min. The supernatant was filtered with PES membrane centrifuge concentrators to eliminate the remaining cell debris and was loaded onto a previously equilibrated strong cation exchange chromatography column. This type of ion chromatography was selected since the theoretical isoelectric point of the protein is 9.67. This means that at the current pH of the actual buffer conditions, the protein is positively charged and feels affinity to the cation exchange resin. The target protein elutes when the column contains 600 mM NaCl. As a polish step, size exclusion chromatography was selected using 20 mM HEPES, 150 mM NaCl at pH 7.4 to eliminate the lower molecular weight contaminants and exchange the buffer of the sample. To analyze the purity of the sample, SDS-PAGE was used (see Figure 19). The first half of the peak was collected for further biophysical studies. Partial amino acid sequencing of HsKrr1 was performed at the Tufts University Core Facility (TUCF), Boston, MA. The partial amino acid sequence result for *Hs*Krr1 shows the expected amino acid sequence (Appendix C). The loss of the first methionine residue is observed at the N-terminus.

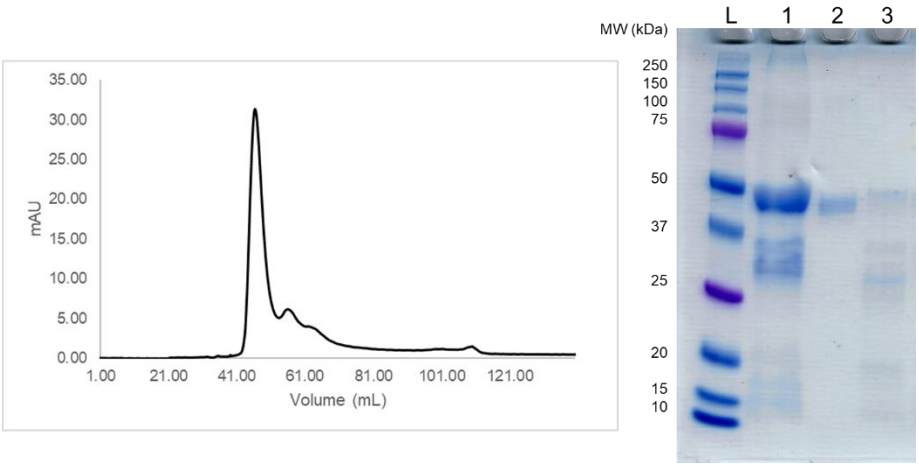
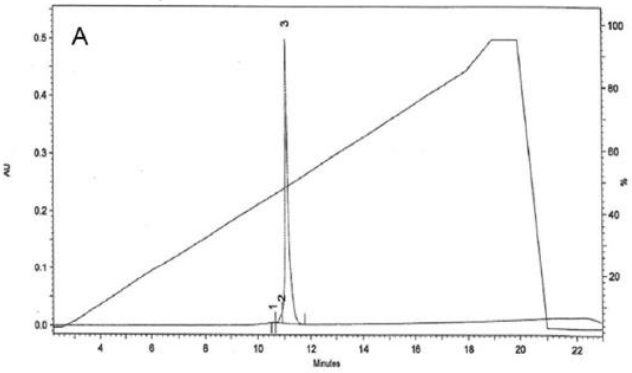


Figure 19. Purification of *Hs*Krr1. In the size exclusion chromatogram, the black line corresponds to the UV-Vis spectrum at 280 nm. The large molecular weight protein in the sample eluted first, *Hs*Krr1 (peak 1, SDS-PAGE lane 2). Peaks 2 and 3 represent low molecular weight impurities (SDS-PAGE lane 3). In SDS-PAGE, lane 1 was the input sample after the cation exchange chromatography.

Analysis of GXXG Loop Peptide

The purity and the molecular weight of the GXXGlp peptide were determined by reverse phase high performance liquid chromatography (RP-HPLC) and mass spectrometry (MS) analysis, respectively (Figure 20). The purity of GXXGlp obtained by the RP-HPLC analysis was 97.5% and the calculated molecular weight for GXXGlp was 4041.35 Da, while the experimental result obtained by the MS report was 4042.76. Accordingly, these results show that the GXXGlp sample had high purity.



Pk #	Retention Time	Area	Area Percent
1	10.550	1861	0.045
2	10.883	101332	2.427
3	11.083	4072575	97.529
Totals		4175768	100.000

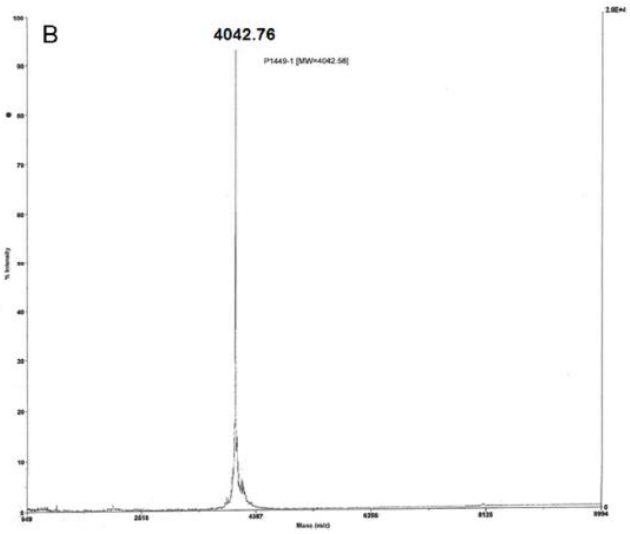


Figure 20. Peptide purification. The GXXG loop peptide was purchased from Bio-Synthesis, Inc. (A) Purity level analysis with RP-HPLC. (B) Mass spectrometry analysis with Voyager-DE RP. *Adapted with permission from American Chemical Society, Rodriguez Nassif 2017.* [59]

Analysis of ssDNA, a GXXG Loop Target

After the synthesis of the 6mer and 13mer oligonucleotides (5'ACCCCA'3 and 5'TTCCCCTCCCCA'3) in the AKTA Oligopilot Plus system, the ssDNA was cleaved from the solid support using acetonitrile 40% (w/v) and heated for 8 h at 50 °C in a water bath. The purification of the oligonucleotides was performed using anion exchange chromatography followed by a desalting G-10 Sephadex column. The oligonucleotides were quantified spectrophotometrically at 260 nm. The 6mer eluted at 25% NaCl while the 13mer eluted at 50% NaCl in the gradient purification method performed (Figure 21). To remove the NaCl from the sample, a MiniTrap G-10 gravity column from GE Health Care was used for buffer exchange against water. Figure 22 presents the HPLC analysis performed to validate the purity of the oligonucleotides, obtaining more than 95% purity in both samples.

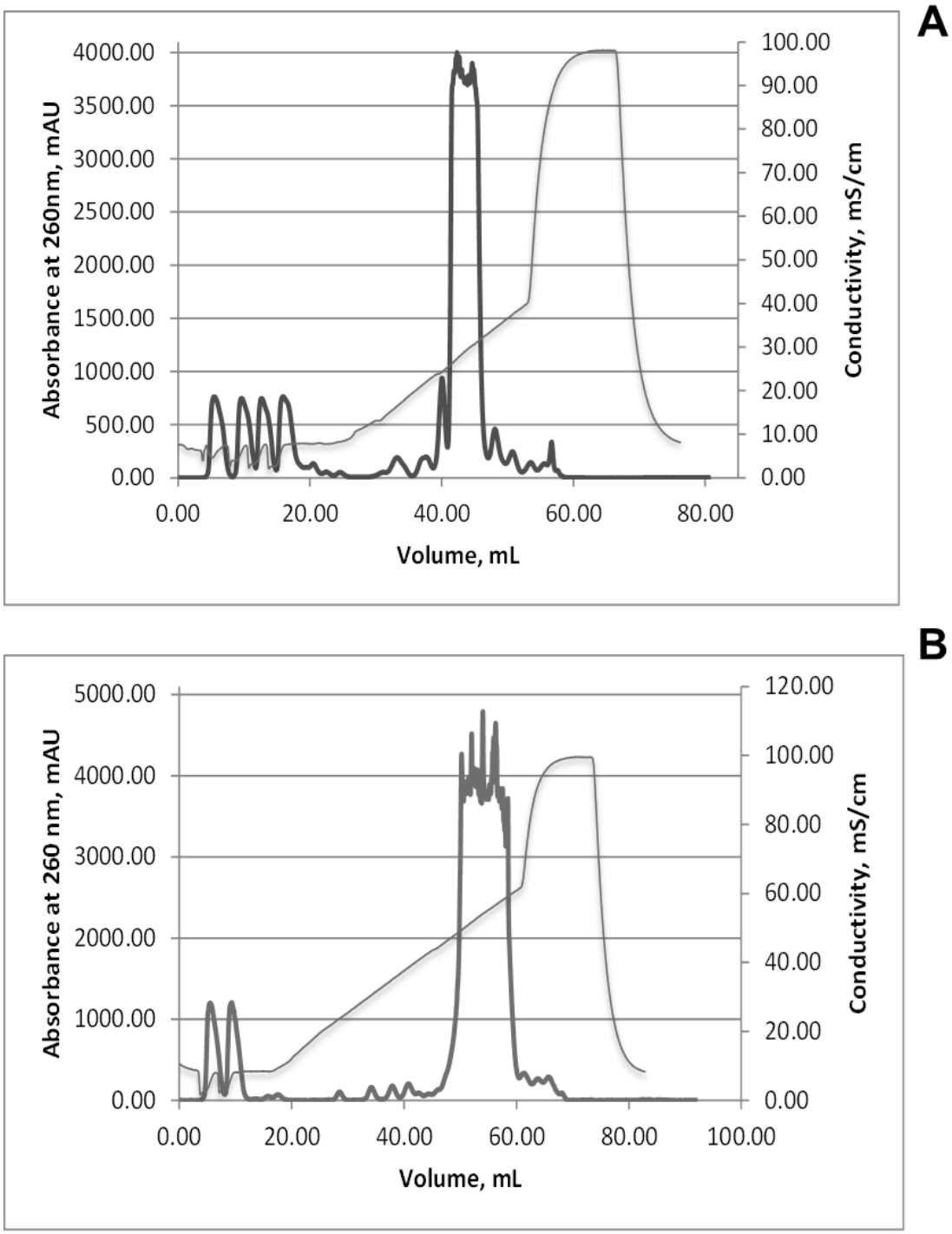
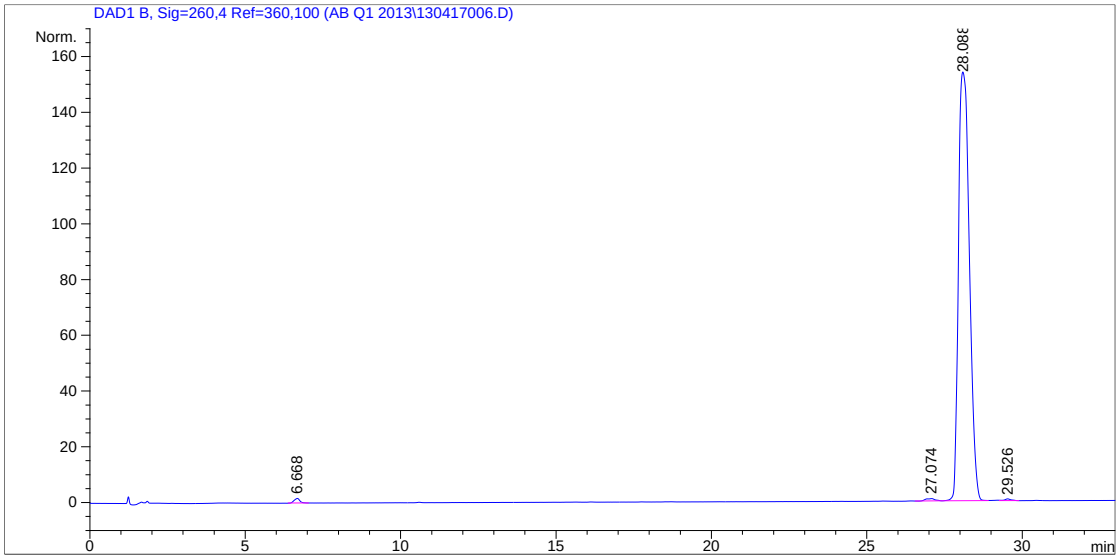


Figure 21. Anion exchange chromatography for oligonucleotide purification. (A) ssDNA 6mer purification. (B) ssDNA 13mer purification. Fractions where the main peak eluted were pooled for the 6mer and the 13mer separately.

A



B

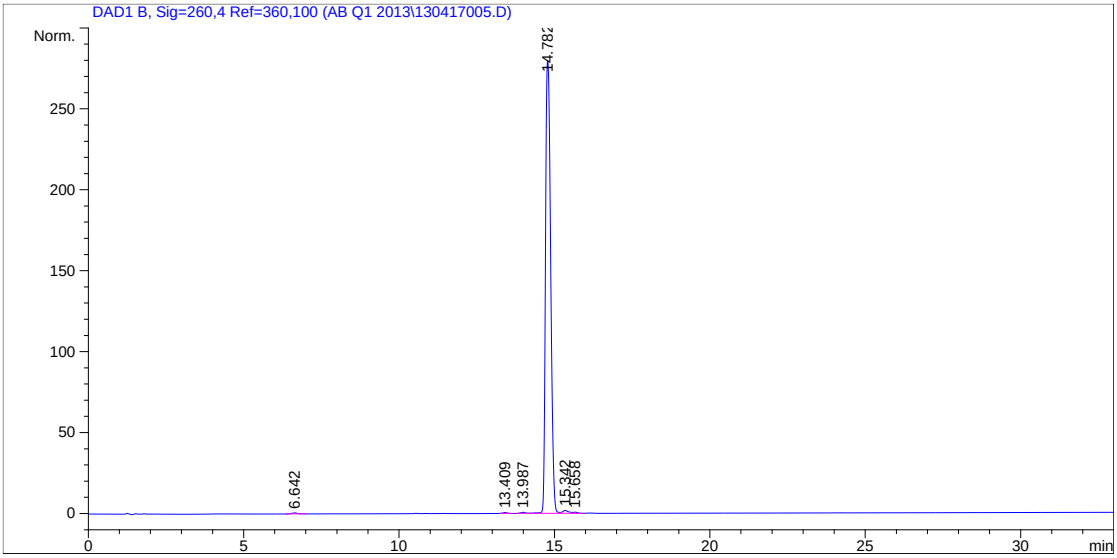


Figure 22. HPLC analysis to validate the purity of the oligonucleotides. (A) ssDNA 6mer chromatogram, and (B) ssDNA 13mer chromatogram.

Analysis of Secondary Structure and Stability Studies of *Hs*Krr1 and its GXXG Loop

Far-UV CD was employed to verify the secondary structure composition of the GXXGlp in the absence of TFA (Figure 23A) and *Hs*Krr1 (Figure 23B). The results suggest that the predominant secondary structure of both samples is α-helical. The mean and standard deviation are as follows: GXXGlp helical contribution was determined at 222 nm and 25.0 °C to be $[\theta]_{MR} = -18.6 \times 10^3 \pm 8 \times 10^3$ degrees cm^2/dmol, while the random coil component was determined at 201 nm and at 25.0 °C to be $[\theta]_{MR} = -6.3 \times 10^4 \pm 1 \times 10^3$ degrees cm^2/dmol for triplicate spectra. In addition, the mean and standard deviation of GXXGlp helical contribution was determined at 222 nm and 5.0 °C to be $[\theta]_{MR} = -15.6 \times 10^3 \pm 2 \times 10^2$ degrees cm^2/dmol for triplicate spectra.

The helical component of GXXGlp and *Hs*Krr1 were monitored using CD spectroscopy. Thermal dependence plots are shown in Figures 24 and 25. When the temperature increased, the GXXGlp increased its α-helical content. Studies with the GXXGlp/ssDNA13mer complex suggest that the stability of the peptide increased even more. That increment suggests the interaction between both components and how this interaction stabilizes the peptide (see Figure 24). Furthermore, the thermal dependence study of *Hs*Krr1 presents a sigmoidal curve, suggesting that cooperativity defines the thermal unfolding process at the molecular level (see Figure 25).

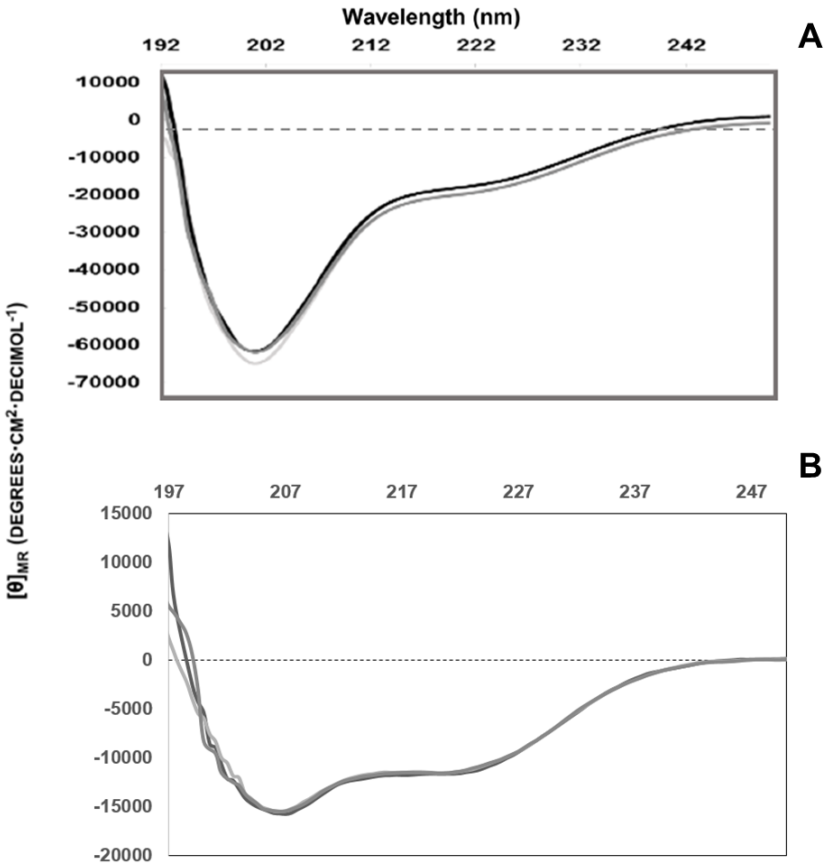


Figure 23. CD analysis of GXXGlp and *Hs*Krr1 in triplicate. (A) Far-UV CD spectra at 25.0 °C in the spectral region of 193−250 nm for 20.7 μM GXXGlp in 2.5 mM HEPES, 7.5 mM NaCl, and 0.2 mM $CaCl_2$ at pH 7.4. Adapted from Rodriguez Nassif 2017.[22] (B) Far-UV CD spectra at 5.0 °C in the spectral region of 197−250 nm for 2.3 μM *Hs*Krr1 in 6.7 mM HEPES, 50 mM NaCl at pH 7.4. Helical and random coil contributions were observed for the peptide and helical contribution for the full-length protein.

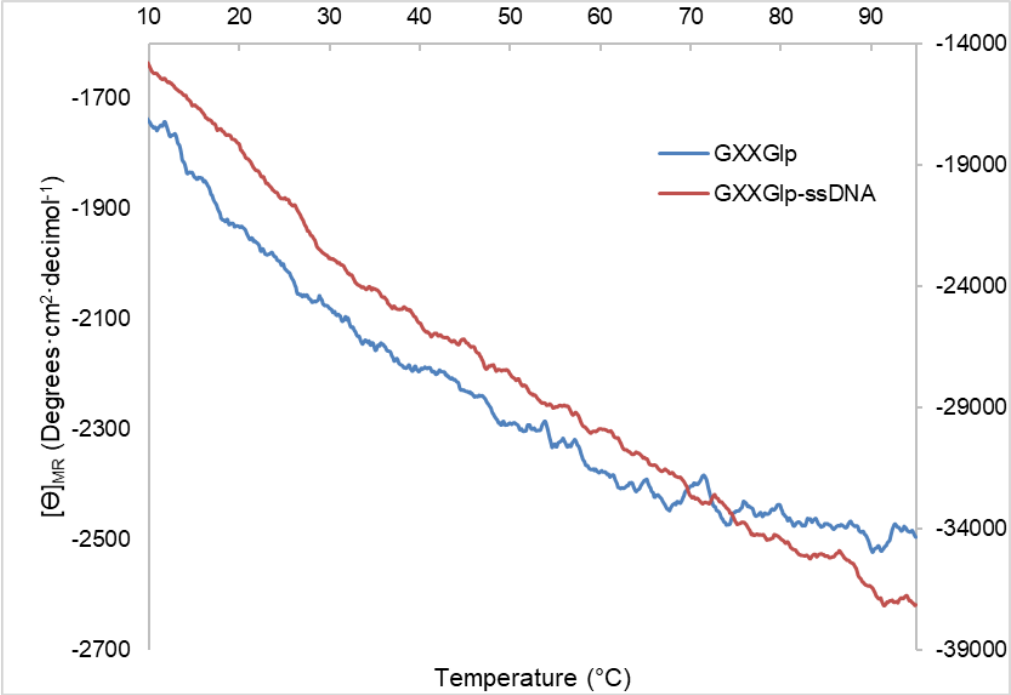


Figure 24. Thermal dependence plots of GXXGlp and GXXGlp/ssDNA13mer complex.

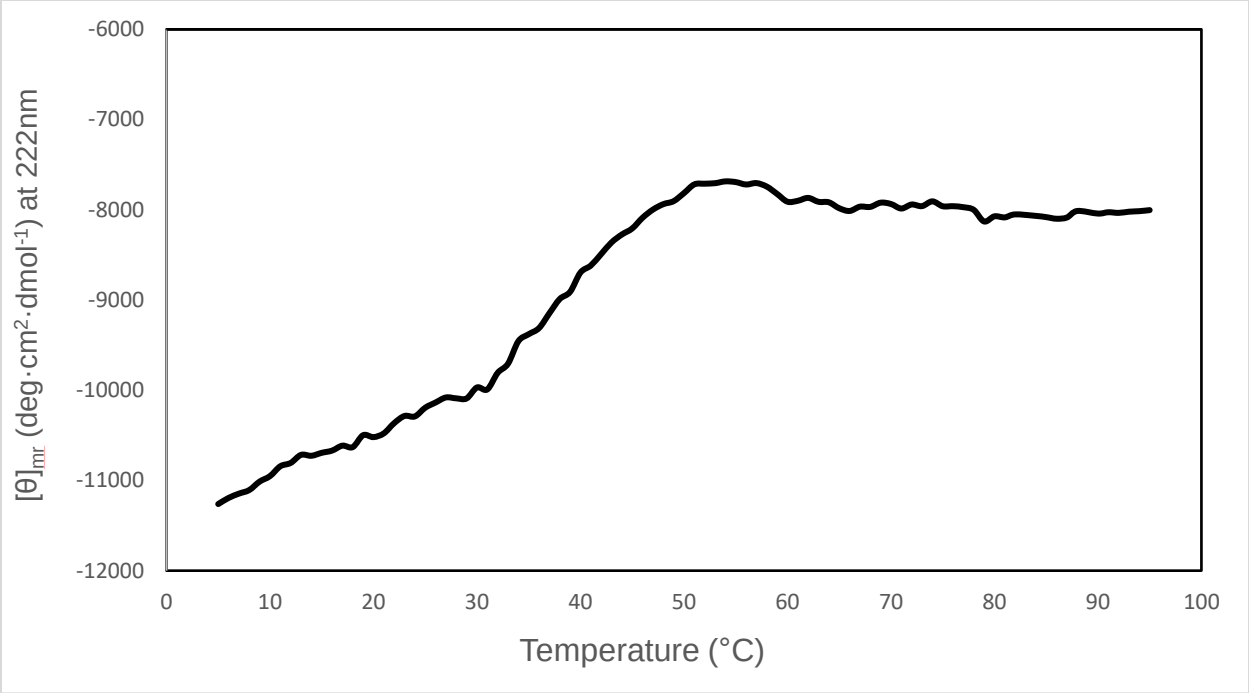


Figure 25. Thermal dependence plot of *Hs*Krr1. CD spectra were obtained using $[\theta]_{mr}$ at 222 nm in the temperature range of 5-95 °C.

Preliminary Studies of the Interaction between *Hs*Krr1 and *Hs*cen2 C-terminal

To explore the interaction between Hscen2 C-terminal and HsKrr1, preliminary analyses were performed by ITC to obtain the thermodynamic parameters that govern the interaction. As shown in Figure 26, the results of the ITC at 20°C presents negative peaks suggesting an exothermic interaction between the complex. The experimental data were best fitted using the single site binding model. As shown in Table 2, the number of biding sites of *Hs*Krr1 was 0.347. This fractional value (0.347) is obtained from the data analysis using the concentrations calculated by a nanodrop UV/VIS equipment and the Beer Lambert Law. In this preliminary study, because *Hs*centrin2 C-terminal have a low molecular extinction coefficient, the concentration is inaccurate and the number of binding site suggests that the concentration has been estimated. Otherwise, *Hs*Krr1 may contain more than one binding site for Hscen2 C-terminal or that Hscen2 N-terminal also interacts with HsKrr1 to stabilize the complex. A 12.82 kcal/mol -TΔS value suggests an unfavorable entropy contribution, compensated partially by the enthalpy parameter (-23.9 kcal/mol, see Table 2). The unfavorable entropy contribution could be related to the interaction being studied, i.e., with the C-terminal of *Hs*cen2 and not the full-length protein. Furthermore, phosphorylation of serines and threonines in the KH domain of HsKrr1 and in the Hscen2 C-terminal may cause the initiation of conformational changes of the protein, causing the activation of its function.

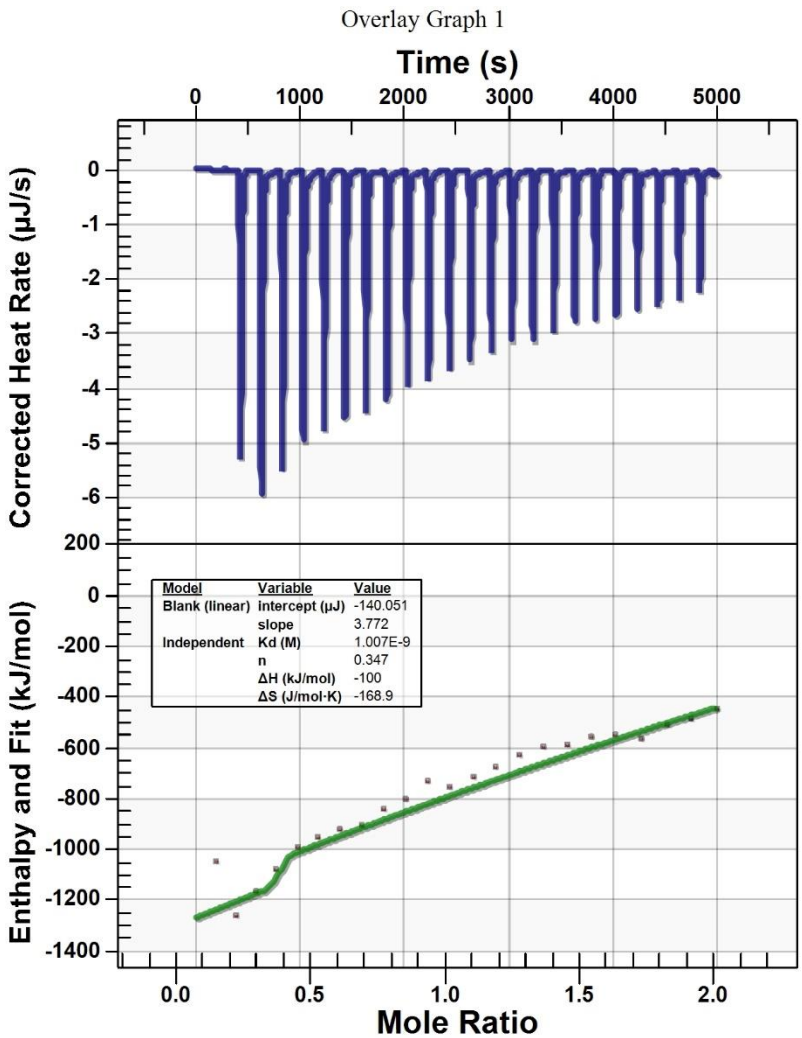


Figure 26. ITC isotherm of the interaction between *Hs*Krr1 and *Hs*cen2 C-terminal at 20 °C. The upper panel shows the raw data and the lower panel shows the integrated enthalpy fitted to a single site binding model.

Table 2. Thermodynamic data of the interaction between *Hs*Krr1 and *Hs*cen2 C-terminal at 20 °C.

Temperature (°C)	K_d (10^9 M)	K_a (10^8 M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
20	1.01	9.93	-10.08	-23.9	12.82

Nuclear Magnetic Resonance Spectroscopy Studies of the *Hs*cen2/*Cr*cen and *Hs*Krr1 GXXG Loop Complex

Preliminary analyses of the interaction between *Hs*cen2 and *Hs*Krr1 were conducted by NMR titration experiments with *Hs*cen2/GXXGlp and *Cr*cen/GXXGlp (Figures 27 and 28). Figure 27 compares the HSQC spectra of *Hs*cen2 and the complex formed with GXXGlp. Each signal represents a protein backbone N-H group. The *Hs*cen2 spectra shown a typical spectrum of an aggregated protein, suggesting that the His tag the protein contains promotes self-association. Titration experiments with the GXXGlp changed the spectrum, adding and shifting signals. These signals changes, most of them in the C-terminal domain, suggest a complex formation between Hscentrin2 and GXXGlp. The results obtained were compared with Yang et al. (2006)[13] by superimposition to determine the amino acids that play an important role in the interaction. The signals in the C-terminal domain correspond to the new signal peaks in the spectrum when the GXXGlp is present (S158, V157, D139, F123, and D116). D139, V157, and S158 amino acids belong to the fourth EF hand domain bound to calcium in *Hs*cen2. Furthermore, ^{15}N *Cr*cen was used to validate the interaction because *Cr*cen showed a high degree of homology to *Hs*cen2. In Figure 28 the HSQC spectra of *Cr*cen and the complex formed with GXXGlp are shown. Adding the peptide to the protein solution changed the spectrum considerably and reduced the number of observable signal peaks (81 peaks) in comparison with the original protein spectrum (96 peaks). These findings validate the formation of the complex.

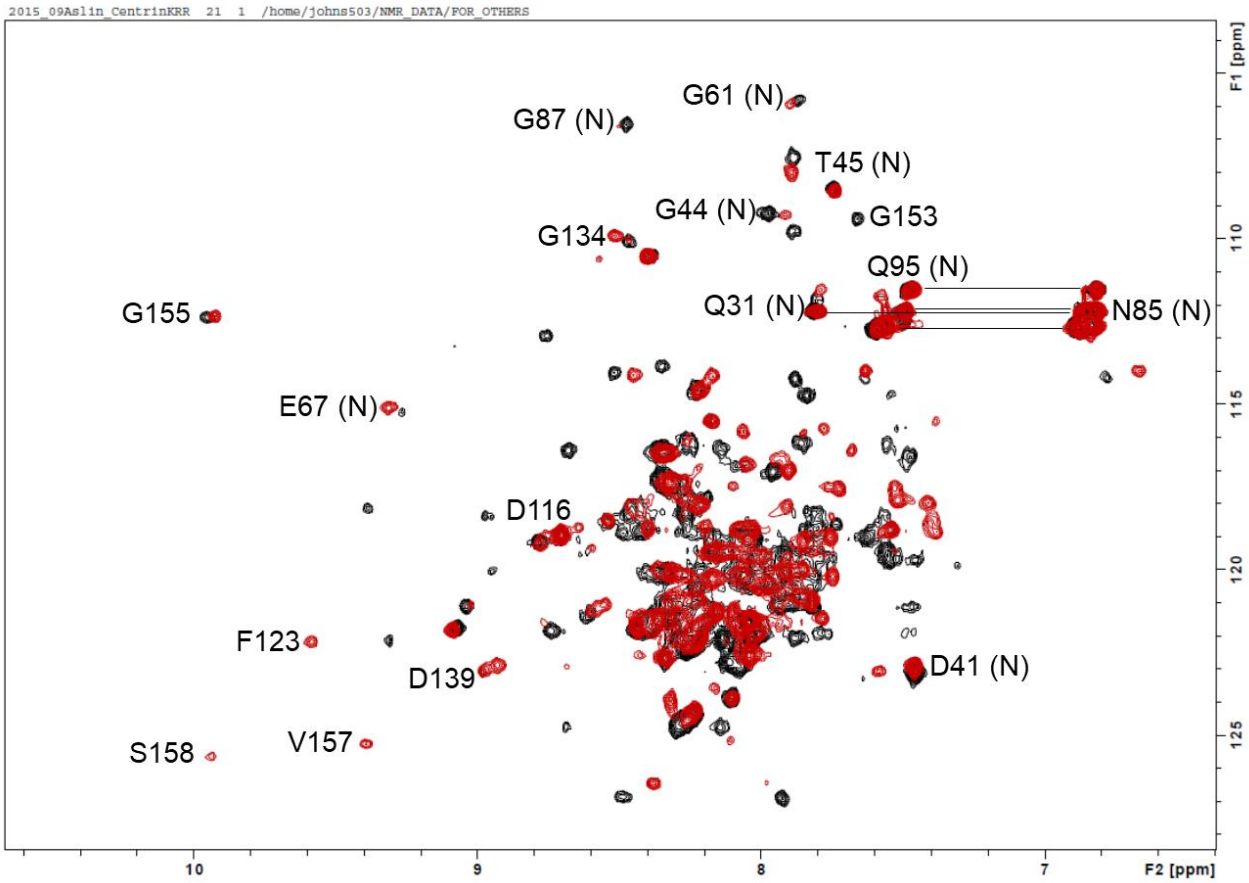


Figure 27. HSQC spectra of uniformly ^{15}N-labeled *H*scen2. Black: *H*scen2 in the absence of GXXGlp peptide. Red: *H*scen2 in complex with 2 molar equivalents of the peptide. The protein is 0.3 mM in Tris buffer at pH 6.7, and the spectra were recorded at 308 K. The assignments indicated by the one letter code for the amino acids and the number in the sequence. The paired peaks connected by a line correspond to the side chain amino acid groups in Asparagine and Glutamine amino acids. The most shifted resonances assigned to the C-terminal domain according to Yang et al. (2006).

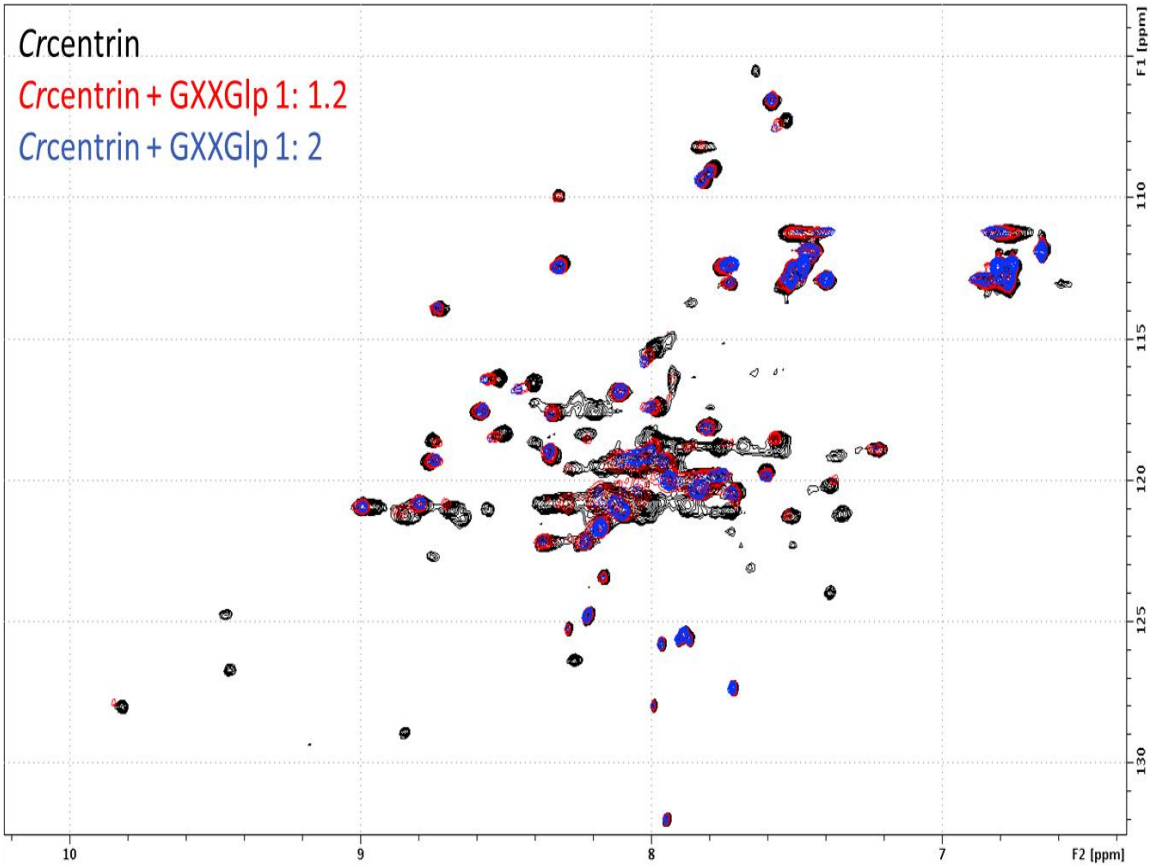


Figure 28. HSQC spectra of uniformly ^{15}N-labeled *Cr*cen. Black: *Cr*cen in the absence of GXXGlp peptide. Red and blue: the complex of *Cr*cen with 1.2 and 2 molar equivalents of the peptide, respectively. The protein is at 0.3 mM in Tris buffer at pH 6.7, and the spectra were recorded at 308 K.

FT-IR Spectroscopy Analysis of the GXXG Loop and its Complex with ssDNA

The molecular behavior of fully H→D exchanged GXXGlp in the presence of TFA, in the absence of TFA at various peptide concentrations, and in complex with ssDNA 13mer using FT-IR spectroscopy within the spectral region of 1750−1500 cm^{-1} (Figure 29) was explored. The spectral region of interest comprised the amide I′ band (1700−1600 cm^{-1}) and shoulders due to side chain modes (1600−1500 cm^{-1}). The complete H→D exchange simplifies the amide I′ band contour, allowing for the helical and loop vibrational modes associated with carbonyl stretching modes ($\upsilon_{(C=O)}$) to be observed. These carbonyls stretching vibrations are highly sensitive to conformational changes. Similarly, the side chain modes were studied as shoulders within the contour. The side chain modes are limited to the spectral region of 1600−1500 cm^{-1} and comprise arginine guanidinium asymmetric and symmetric stretching vibrational modes ($\upsilon_{(N\text{-}D)}$), a single glutamate carboxylate stretching vibrational mode ($\upsilon_{(COO^-)}$), and the nucleotide base pair stretching modes for the ssDNA 13mer. The glutamate residue at position 18 of the 24-residue peptide, located near the C-terminal end within the second helical motif of the peptide sequence, serves as a probe for this region of the peptide.

In general, for the peptide spectral overlay, the amide I′ band maximum shifted toward higher wavenumbers as the temperature increased, and a concomitant decrease in intensity was observed, suggesting a transition toward unfolding (Figure 29). Moreover, in the complex spectral overlay, the amide I’ band maximum shifted

toward lower wavenumbers as the temperature increased and an associated increase in intensity was observed, suggesting an increment in the peptide stability.

The intensity of the side chain band of GXXGlp also decreased with increasing temperature in the absence of TFA at both high and low peptide concentration and in the presence of ssDNA 13mer (Figure 29A, B, D), in contrast to the TFA-containing sample (Figure 29C), which may be indicative of the strong interaction between the TFA and the peptide. In the absence of TFA and at high peptide concentration, aggregation of the peptide was observed as a shoulder at 1620.7 cm^{-1} (Figure 29B). In this peptide sample, the intensity of the aggregation peak continued to increase and shift to higher wavenumbers until the temperature reached 50 °C, after which the intensity of the aggregation peak (shoulder at 1620.7 cm^{-1}) decreased and shifted to lower wavenumbers. Subsequent spectral changes followed the unfolding process of the peptide, suggesting that at temperatures above T_m, the backbone dynamics governed the process.

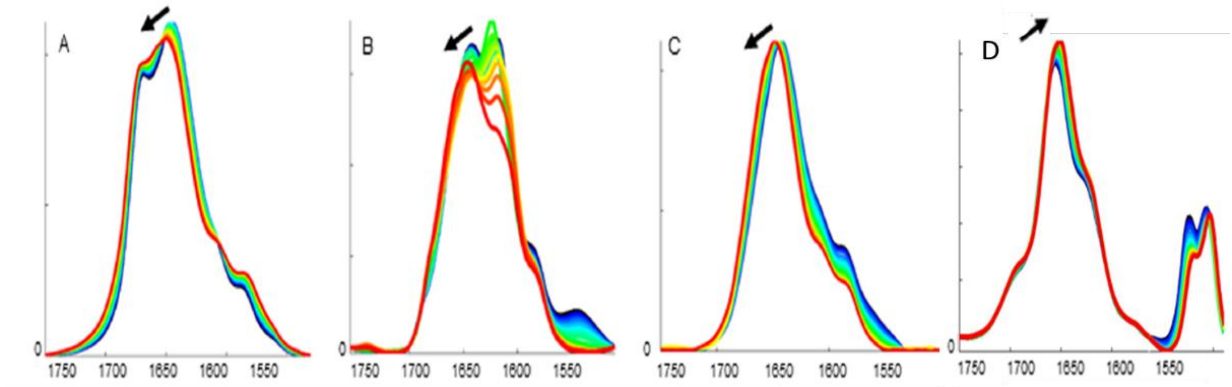


Figure 29. FT-IR spectral overlay of GXXGlp. The temperatures range is indicated in color for 5 °C (dark blue), 10−30 °C (blue), 35−65 °C (green), 70 °C (yellow), 75−85 °C (orange), and 90 °C (red). The spectra show the amide I′ and side chain bands in the spectral region of 1750−1500 cm^{-1} for (A) 33.5 mg/mL GXXGlp in the presence of TFA, (B) 50.1 mg/mL GXXGlp in the absence of TFA, (C) 16.7 mg/mL GXXGlp in the absence of TFA, and (D) 16 mg/mL GXXGlp and 21.6 mg/mL ssDNA 13mer complex. *Adapted with permission from American Chemical Society, Rodriguez Nassif 2017.* [59]

2D IR Correlation Spectroscopy in the Determination of Aggregation and Stability of the GXXG Loop

2D IR correlation spectroscopy was performed to improve our understanding of the relationship between the GXXGlp backbone dynamics and side chain interactions involved in the aggregation process and determine the extent of aggregation. The band assignment values are the mean values for each assignment (Table 3). The TFA- containing GXXGlp sample spectrum has TFA absorption[22,25,40,41] as a sharp peak at 1675.3 cm^{-1} (Figure 30A). The backbone vibrational modes for GXXGlp have two different loop types (loop A and loop B) that are referring to the single GXXGlp found in the peptide (see Figure 1) with varying degrees of flexibility because of hydrogen bonding found in a mixed population of the peptide in solution. Loop A (1684.3 cm^{-1}) will be considered the flexible loop present in lower abundance within the sample, and loop B (1661.5 cm^{-1}) will be considered the rigid loop present at higher abundance. We also assigned the helical component (1642.2 cm^{-1}) and the β-strand (1624.1 cm^{-1}) component located at the C-terminal end of the peptide. Aggregation was observed only for the highly concentrated TFA-free sample (Figure 30B) as a shoulder at 1620.7 cm^{-1}. The side chain modes included the following: the positively charged arginine residue (R_{166}, R_{167}, and R_{169}) guanidinium stretching modes $\upsilon_{a(N-D)}$ and $\upsilon_{s(N-D)}$) at 1605.7 and 1585.5 cm^{-1}, respectivelyy. Also, a single glutamate residue (E_{182}) assigned to a stretching vibrational mode $\upsilon_{(COO-)}$ at 1558.7 cm^{-1}.

Analysis of the synchronous (Figure 30D−F) and asynchronous plots (Figure 30G−I) involves the evaluation of peaks observed in these contour plots. In general,

the synchronous plot contains peaks on the diagonal known as auto peaks. These peaks are always positive and define the magnitude of the overall spectral intensity changes observed. For both the synchronous and asynchronous plots, the off-diagonal peaks are referred to as cross peaks. Cross peaks can be either positive or negative. The sign of the cross peaks is used to determine the order of molecular events, as per Noda's rules. Covariance between the 2D IR correlation plots can be observed for GXXGlp (Figure 30D−I), suggesting distinct peptide behavior due to thermal stress in each sample in the presence and absence of TFA and at varying peptide concentration.

Each sample was analyzed using the asynchronous and synchronous plots to generate the order of events for GXXGlp in the presence and absence of TFA at varying peptide concentrations within the spectral region of 1750−1500 cm^{-1} in the temperature range of 5−90 °C. Hence, the initial events of perturbation within the peptide occur at low temperatures and therefore are the least stable regions of the peptide. Similarly, the last event of perturbation occurs at high temperatures and therefore corresponds to the most stable regions of the peptide. The results of the interpretation are shown in Figure 31. For the GXXGlp in the presence of TFA (Figure 31A), the helical component (1642.2 cm^{-1}) and the β-strand (1624.1 cm−1) located in the C-terminal end were perturbed initially, followed by the flexible loop A and the rigid loop B (1684.3 and 1661.5 cm^{-1}, respectively); the arginine (1605.7 and 1585.5 cm^{-1}), which was perturbed by TFA via ion-pair interaction; and then, the glutamate (1558.7 cm^{-1}). The final event was the dissociation of the ion pairing with the TFA (1675.3

cm^{-1}). Finally, the intramolecular salt-bridge interactions were perturbed (Glu^{-}, 1558.7 cm^{-1}; Arg, 1585.5 cm^{-1}).

In the absence of TFA and at high GXXGlp concentration (Figure 31B) the initial perturbation occurred within the flexible loop A (1684.3 cm^{-1}, the least stable motif) followed by the β-strand (1624.1 cm^{-1}) located in the C-terminal end; then the aggregated species (1620.7 cm^{-1}) was observed to increase involving the α-helical motif (1642.2 cm^{-1}), and as the temperature increased the intermolecular salt-bridge interactions involving the single glutamate (E_{182}) residue (1558.7 cm^{-1}) and the arginine (1585.5, 1605.7 cm^{-1}) located in the N-terminal helix were broken, disrupting the self-association. Finally, the most stable motif was the rigid loop B (1661.5 cm^{-1}). In this sample, there were two separate molecular events occurring as the temperature increased: the first was the increase in aggregation or self-association, which upon closer inspection revealed molecular details about the self-association involving the salt-bridge interaction (see model in Figure 32). The second event was the cross peaks evident in the interaction that support the self-association mechanism.

In the absence of TFA at low GXXGlp concentration (Figure 31C), which represents the optimal formulation conditions, both loop types representing the ssDNA binding region of the peptide were perturbed initially (loop A at 1684.3 cm^{-1} and loop B at 1661.5 cm^{-1}). As the temperature was increased, the helical (1642.2 cm^{-1}) and β-strand (1624.1 cm^{-1}) motifs were perturbed, followed by the disruption of the intramolecular salt-bridge interaction involving the single glutamate (E_{182}) residue (1558.7 cm^{-1}) and the arginines (1585.5, 1605.7 cm^{-1}) located in the N-terminal end

of the α-helical motif. Consequently, the intramolecular salt-bridge interaction is key to the structural integrity of the peptide.

Table 3. Summary of Peak Assignments for the GXXGlp in D_2O in the Spectral Region of 1750−1500 cm^{-1} *Adapted with permission from American Chemical Society, Rodriguez Nassif et al. (2017).*[59]

Peak Assignment	GXXGlp Peak Positions			
	Mean (cm^{-1})	presence of TFA (cm^{-1})	absence of TFA high concentration (cm^{-1})	absence of TFA low concentration (cm^{-1})
TFA	1675.3	1675.3	---a	---a
Backbone modes				
$loop_A$	1684.3	1693.7	1679.6	1679.6
$loop_B$	1661.5	1658.7	1661.1	1664.8
α-helix	1642.2	1642.3	1640.5	1643.7
β-strand	1624.1	1623.3	1622.3	1626.7
Aggregation	1620.7	---b	1620.7	---b
Side chain modes				
Arginine guanidinum ν_a(N-D)	1605.7	1606.5	1600.9	1609.7
ν_s(N-D)	1585.5	1584.5	1587.4	1584.7
Glutamate $\nu_s(COO^-)$	1558.7	1564.5	1546.4	1565.3

(a) N/A because the sample does not contain TFA.

(b)The sample is aggregate free.

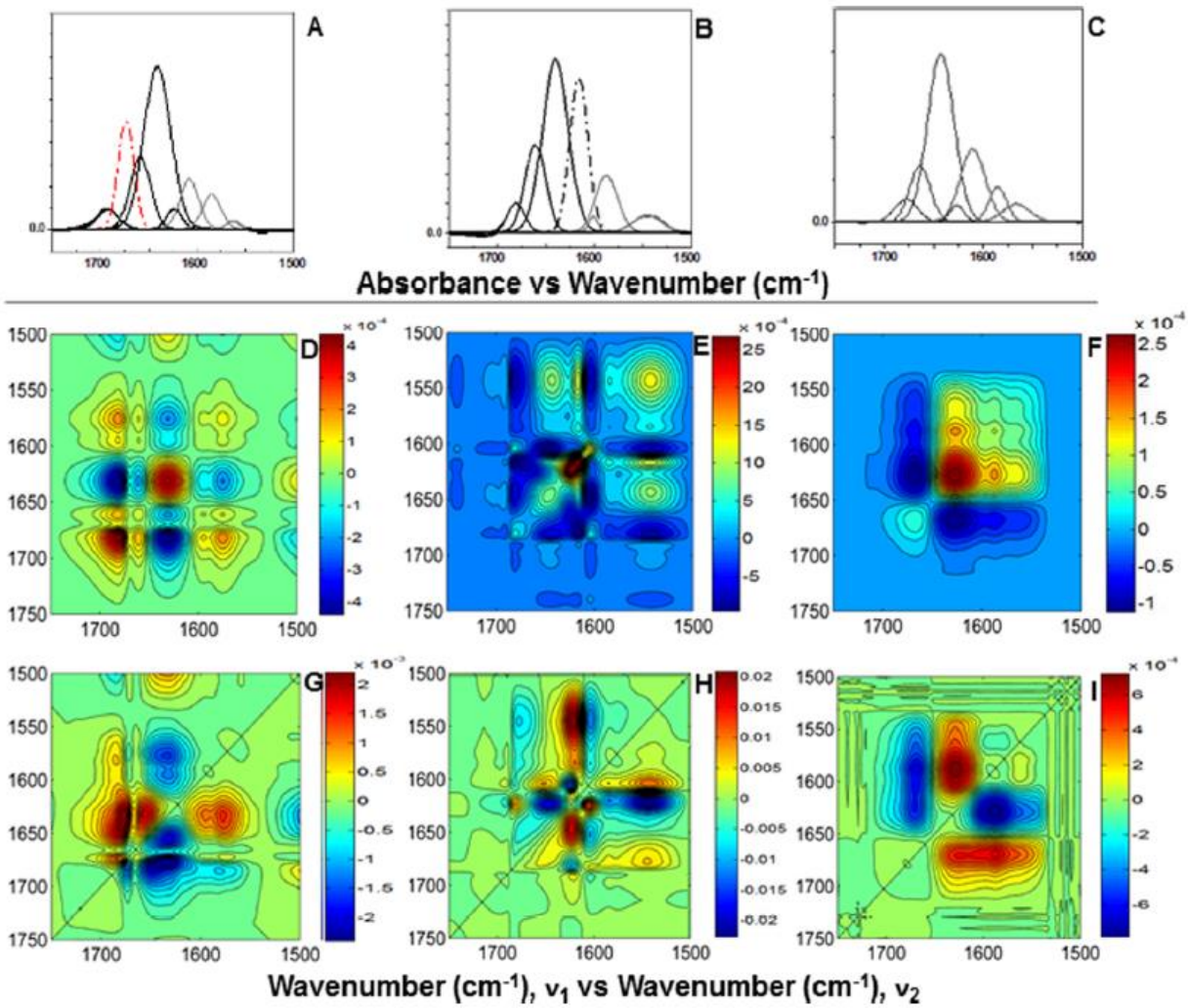


Figure 30. 2D IR correlation spectroscopy during thermal perturbation of the GXXGlp. (A−C). The thermally induced spectral changes in GXXGlp in the presence of TFA at 33.5 mg/mL (A) and in the absence of TFA at 50.1 (B) and 16.7 mg/mL (C) are shown as sub-bands associated with the backbone vibrational modes (black lines), side chain modes (grey lines), TFA (red dash/dot line) and aggregation (black dash/dot line). The curve fitting analysis of each condition is shown in Appendix D. (D−F) Synchronous and (G−I) asynchronous 2D plots. *Adapted with permission from American Chemical Society, Rodriguez Nassif et al. (2017).* [59]

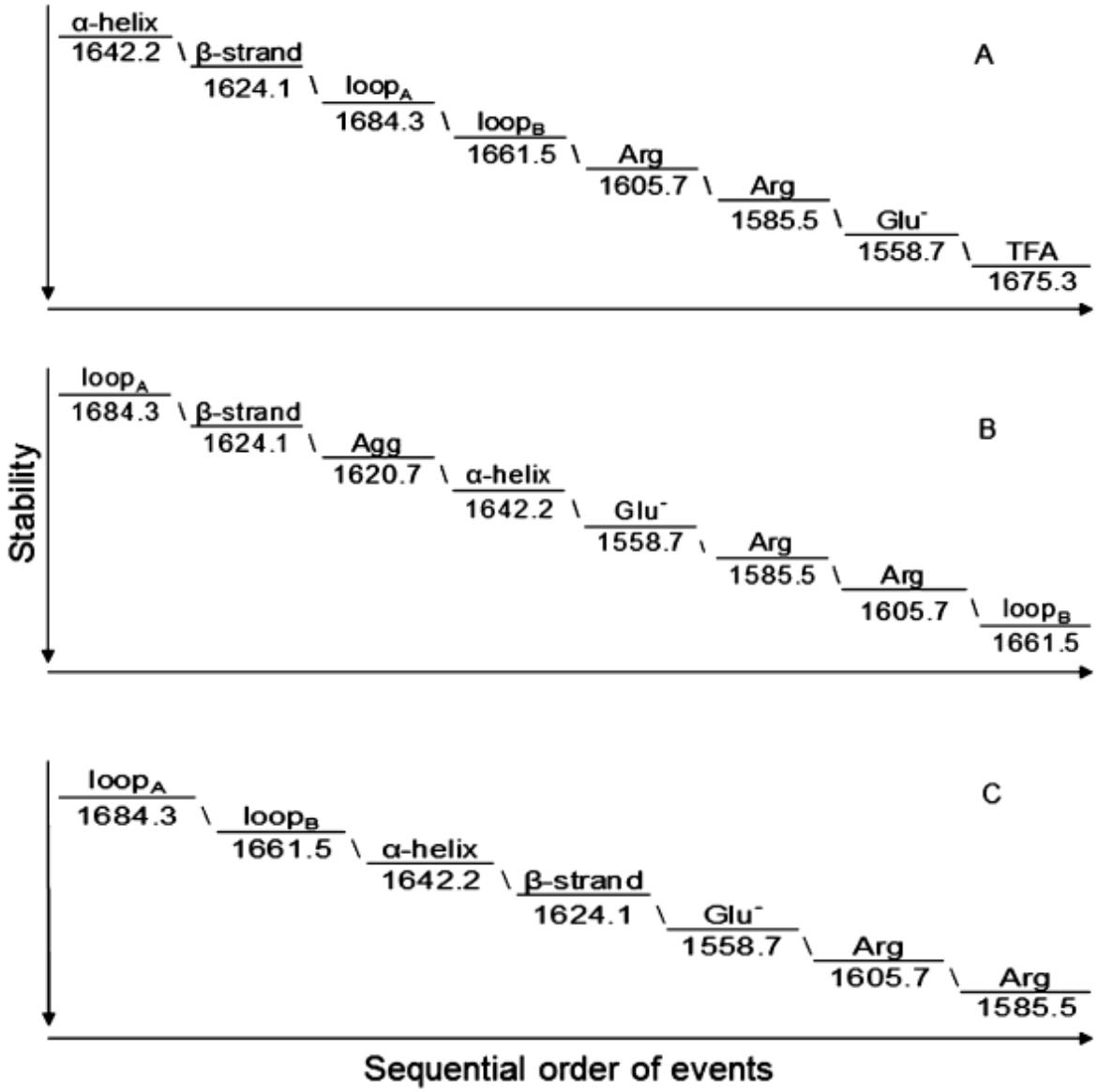


Figure 31. Sequential order of molecular events for all three GXXGlp samples. Schematic representation summarizing the molecular events during the thermal perturbation of (A) 33.5 mg/mL GXXGlp in the presence of TFA, (B) 50.1 mg/mL GXXGlp in the absence of TFA, and (C) 16.7 mg/mL GXXGlp in the absence of TFA. Differences in stability are attributed to the presence of TFA and the aggregate species at high peptide concentration versus the optimized conditions in the absence of TFA at lower concentration of the peptide. *Adapted with permission from American Chemical Society, Rodriguez Nassif et al. (2017).*[59]

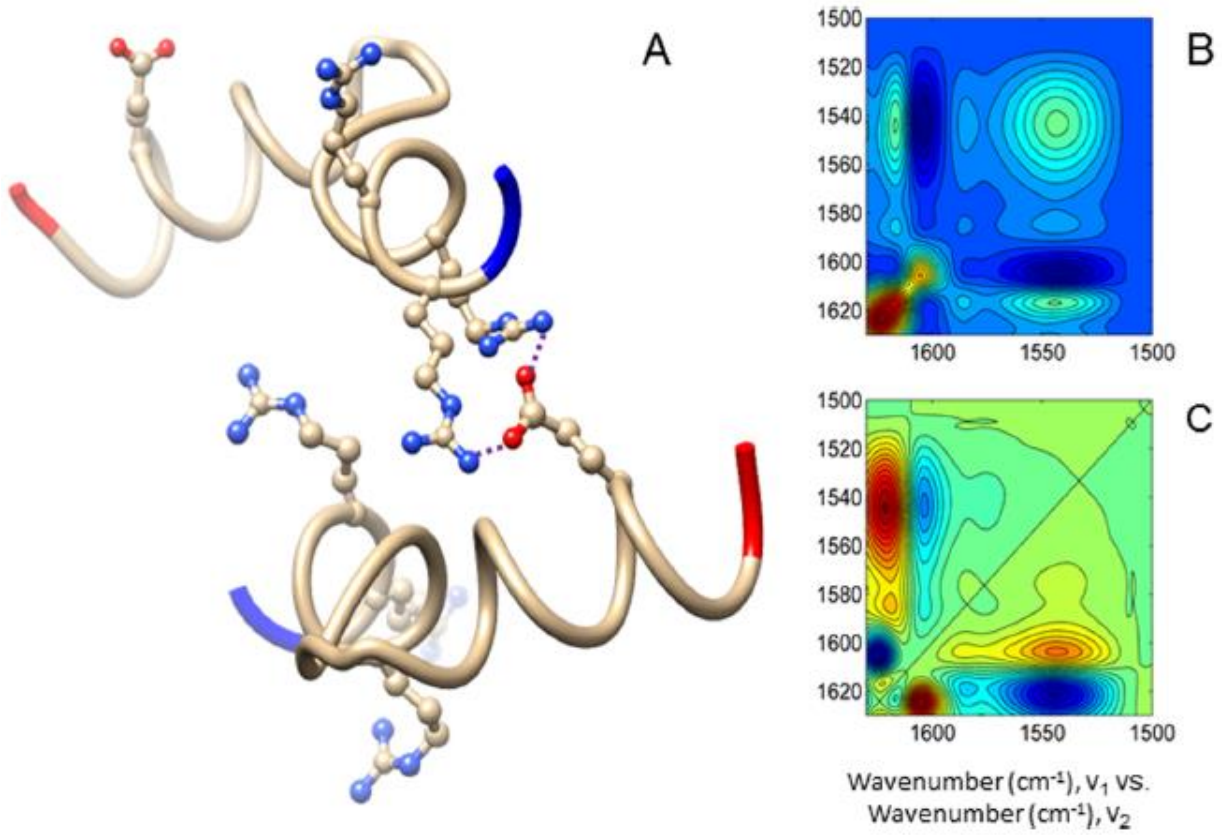


Figure 32. Model of the intermolecular salt-bridge interaction that mediates the self-association of the GXXGlp. (A) Representative cartoon model depicting the GXXGlp based on the *Sc*Krr1 structure available in the Protein Data Bank (PDB ID 4QMF). The N-terminal and C-terminal ends of the peptide are indicated in blue and red, respectively. The ball-and-stick side chains in blue and red correspond respectively to the arginines (R165, 166, 168) located at the C-terminal end and the internal glutamate (E182), showing the intermolecular salt-bridge interaction. (B−C) 2D IR synchronous and asynchronous plots within the spectral range of 1630−1500 cm^{-1} and in the temperature range of 5−90 °C providing the molecular evidence of the salt-bridge interaction that is responsible for the aggregation (self-association) of the peptide. *Adapted with permission from American Chemical Society, Rodriguez Nassif et al. (2017).*[59]

Small Angle X-ray Scattering, Modelling, and Preliminary Crystallization Tray of *Hs*Krr1

To obtain structural information on *Hs*Krr1 in solution, a SAXS experiment was performed. The mass of the protein particles in solution determined by SAXS using the Guinier test and the theoretical mass of *Hs*Krr1 based on its sequence concordat: 46 kDa and 42.7 kDa, respectively. To calculate the molecular weight of *Hs*Krr1, the first 10 measurements were eliminated as a partial aggregation pattern was observed in the sample, suggesting that at 6.5 mg/mL, in these buffer conditions, the protein is unstable. Furthermore, a Kratky plot was carried out to check the globularity and flexibility of *Hs*Krr1 (Figure 33B). In this case, the result suggested that the sample contains folded protein, because the plot exhibited only one "bell-shaped" peak at low q value. The plot does not converge to the q axis at high q values, suggesting that the protein contains a degree of flexibility. Moreover, an *ab initio* model of averaged molecular envelopes from DAMMIN was created (Figure 33A).

To determine the structure of *Hs*Krr1, a crystallization tray was set up. Figure 34 shows the amorphous solid was obtained after two days, by the sitting drop vapor diffusion method at room temperature.

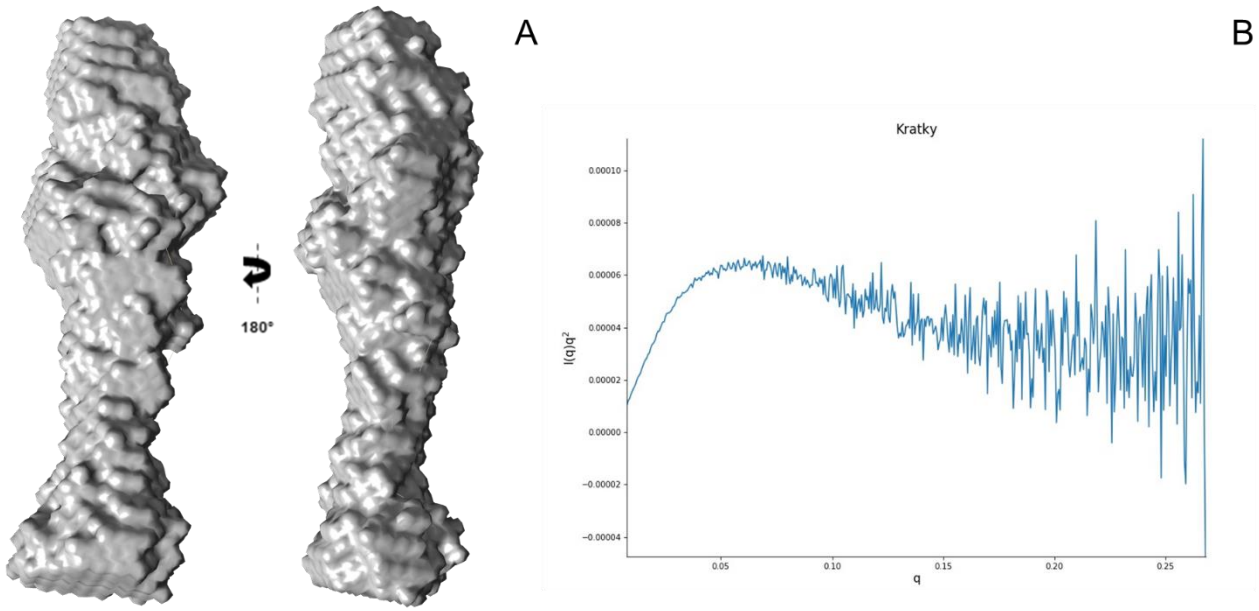


Figure 33. *Hs*Krr1 small angle X-ray scattering model. (A) Ab initio model of *Hs*Krr1 calculated with DAMMIN and averaged with DAMAVER. (B) *Hs*Krr1 Kratky plot at 25 °C in buffer containing 20 mM HEPES and 150 mM NaCl, pH 7.4.

Figure 34. Microcrystal of *Hs*Krr1. The precipitant solution used was 50% PEG 3350 in 1.5 M triammonium citrate, pH 8. The crystal is growing at the interface of phase separation.

CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The expression and purification of *Hs*cen2, *Hs*cen2 C-terminal, *Hs*cen2 N-terminal, and *Hs*Krr1 has been described in this dissertation. The purification protocol described was performed based on the biochemical characteristics of the recombinant proteins. A successful modification of Dr. Pastrana-Rios' laboratory centrin purification protocols gave an excellent yield of protein for *Hs*cen2 C-terminal and *Hs*cen2 N-terminal. CD and SAXS, were presented as useful techniques to partially characterize, for the first time, a novel centrin target, *Hs*Krr1. To understand in more detail the conformation and relative stability of the protein, the characterization of its GXXGlp was performed by analyzing the structural motifs within the peptide and validating the optimal conditions for the peptide formulation using FT-IR spectroscopy and 2D correlation spectroscopy. The secondary structure of the GXXGlp in the absence of TFA was characterized by CD, and was also validated by FT-IR and 2D IR correlation spectroscopies. In addition, FT-IR and 2D IR correlation spectroscopies were used to establish the differences in conformational stability of the GXXGlp in the presence and absence of TFA. The results presented herein have demonstrated the wealth of information the combination of these molecular biophysical techniques can provide in the following order: 2D IR correlation > FT-IR > CD. More importantly, the determination of the mechanism and extent of aggregation/self-association and the optimal formulation conditions for the GXXGlp, allowing novel peptide-target

interactions to be characterized. Moreover, CD and FT-IR analysis confirm the interaction of GXXGlp and ssDNA 13mer. Thermal dependence studies of the GXXGlp/ssDNA13mer complex suggest that the stability of the peptide increases in the presence of the ssDNA. That increment suggests an interaction between both components. Furthermore, the interaction between *Hs*cen2 and GXXGlp was characterized by NMR and preliminary studies using ITC. NMR titration experiments with the GXXGlp showed changes in the *Hs*cen2 spectrum, including appearance and shifting signals. These changes suggest a complex formation between *Hs*cen2 and GXXGlp.

FT-IR and 2D correlation spectroscopy analysis should be performed to characterize the relative stability of *Hs*Krr1 and its ternary complex (*Hs*Krr1/ssDNA/*Hs*cen2). Further investigation of the interaction of *Hs*Krr1 and *Hs*cen2 C-terminal and N-terminal will be useful to determine the thermodynamic parameters that govern the complex. Finally, structural information defined by NMR will be very helpful to determine the binding site of this novel complex to understand more about the biological processes inside the nucleus.

REFERENCES

(1) Ciuffreda, L., Sanza, C. Di, and Milella, U. C. I. and M. (2010) The mTOR Pathway: A New Target in Cancer Therapy. *Curr. Cancer Drug Targets*. *10* (5), 484-495.

(2) Drygin, D., Rice, W. G., and Grummt, I. (2010) The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol. 50*, 131–156.

(3) Freed, E. F., Bleichert, F., Dutca, L. M., and Baserga, S. J. (2010) When ribosomes go bad: diseases of ribosome biogenesis. *Mol. Biosyst. 6*, 481–93.

(4) Narla, A., and Ebert, B. L. (2010) Ribosomopathies: Human disorders of ribosome dysfunction. *Blood 115*, 3196–3205.

(5) van Riggelen, J., Yetil, A., and Felsher, D. W. (2010) MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer 10*, 301–309.

(6) Valverde, R., Edwards, L., and Regan, L. (2008) Structure and function of KH domains. *FEBS J. 275*, 2712–2726.

(7) Yang, B., and Rodgers, M. T. (2014) Base-pairing energies of proton-bound heterodimers of cytosine and modified cytosines: Implications for the stability of DNA i-motif conformations. *J. Am. Chem. Soc. 136*, 282–290.

(8) Jin, P., and Warren, S. T. (2000) Understanding the molecular basis of fragile X syndrome. *Hum. Mol. Genet. 9*, 901–908.

(9) Sabaratnam, M. (2006) Fragile-X syndrome. *Psychiatry 5*, 325–330.

(10) Hartwig, A., Kasper, P., Madle, S., Speit, G., Staedtler, F., and Sengstag, C. (2001) The potential use of mutation spectra in cancer related genes in genetic toxicology: a statement of a GUM working group. *Mutat. Res. Mol. Mech. Mutagen. 473*, 263–267.

(11) Renaud, E., Miccoli, L., Zacal, N., Biard, D. S., Craescu, C. T., Rainbow, A. J., and Angulo, J. F. (2011) Differential contribution of XPC, RAD23A, RAD23B and CENTRIN 2 to the UV-response in human cells. *DNA Repair (Amst). 10*, 835–847.

(12) Charbonnier, J., Renaud, E., Miron, S., Hélène, M., Du, L., Blouquit, Y.,

Duchambon, P., Christova, P., Shosheva, A., Rose, T., Angulo, J. F., Craescu, C. T., and Paris-sud, U. (2007) Structural , Thermodynamic , and Cellular Characterization of Human Centrin 2 Interaction with Xeroderma Pigmentosum Group C Protein *J. Mol. Biol. 373*, 1032–1046.

(13) Yang, A., Miron, S., Mouawad, L., Duchambon, P., Blouquit, Y., and Craescu, C. T. (2006) Flexibility and Plasticity of Human Centrin 2 Binding to the Xeroderma Pigmentosum Group C Protein (XPC) from Nuclear Excision Repair. *Biochemistry. 45* (11), 3653–3663.

(14) Durussel, I., Blouquit, Y., Middendorp, S., Craescu, C. T., and Y, J. A. C. (2000) Cation- and peptide-binding properties of human centrin 2. *FEBS*. *472*, 208–212.

(15) 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.

(16) Díaz Casas, A., Chazin, W. J., and Pastrana-Ríos, B. (2017) Prp40 Homolog A Is a Novel Centrin Target. *Biophys. J. 112*, 2529–2539.

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

(18) Pastrana-Ríos, B., Reyes, M., De Orbeta, J., Meza, V., Narváez, D., Gómez, A. M., Rodríguez Nassif, A., Almodovar, R., Díaz Casas, A., Robles, J., Ortiz, A. M., Irizarry, L., Campbell, M., and Colón, M. (2013) Relative stability of human centrins and its relationship to calcium binding. *Biochemistry 52*, 1236–1248.

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

(20) Ortiz, 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.

(21) Pastrana-Rios, 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.

(22) Zhao, Y., Yan, J., Feng, Y., Liang, A., and Yang, B. (2011) Analysis of the role of Mg2+ on conformational change and target recognition by Ciliate Euplotes octocarinatus centrin. *J. Photochem. Photobiol. B Biol. 105*, 60–68.

(23) Hildebrandt, F., Benzing, T., and Katsanis, N. (2011) Ciliopathies. *N. Engl. J. Med.* (Schwartz, R. S., Ed.) *364*, 1533–1543.

(24) Miron, S., Durand, D., Chilom, C., Pérez, J., and Craescu, C. T. (2011) Binding of Calcium, Magnesium, and Target Peptides to Cdc31, the Centrin of Yeast Saccharomyces cerevisiae. *Biochemistry 50*, 6409–6422.

(25) Vonderfecht, T., Stemm-Wolf, A. J., Hendershott, M., Giddings, T. H. J., Meehl, J. B., and Winey, M. (2011) The two domains of centrin have distinct basal body functions in Tetrahymena. *Mol. Biol. Cell 22*, 2221–2234.

(26) Grabarek, Z. (2006) Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. *J. Mol. Biol. 359*, 509–525.

(27) Chazin, W. J. (2011) Relating Form and Function of EF-Hand Calcium Binding Proteins. *Acc. Chem. Res. 44*, 171–179.

(28) Lancaster, M. A., and Gleeson, J. G. (2009) The primary cilium as a cellular signaling center: lessons from disease. *Curr. Opin. Genet. Dev. 19*, 220–229.

(29) Salisbury, J. L., Suino, K. M., Busby, R., and Springett, M. (2002) Centrin-2 Is Required for Centriole Duplication in Mammalian Cells. *Curr. Biol. 12*, 1287–1292.

(30) Middendorp, S., Küntziger, T., Abraham, Y., Holmes, S., Bordes, N., Paintrand, M., Paoletti, A., and Bornens, M. (2000) A role for centrin 3 in centrosome reproduction. *J. Cell Biol. 148*, 405–415.

(31) Dantas, T. J., Wang, Y., Lalor, P., Dockery, P., and Morrison, C. G. (2011) Defective nucleotide excision repair with normal centrosome structures and functions in the absence of all vertebrate centrins. *J. Cell Biol. 193*, 307–318.

(32) Errabolu, R., Sanders, M. a, and Salisbury, J. L. (1994) Cloning of a cDNA encoding human centrin, an EF-hand protein of centrosomes and mitotic spindle poles. *J. Cell Sci. 107*, 9–16.

(33) Laoukili, J., Perret, E., Middendorp, S., Houcine, O., Guennou, C., Marano, F., Bornens, M., and Tournier, F. (2000) Differential expression and cellular distribution of centrin isoforms during human ciliated cell differentiation in vitro. *J. Cell Sci. 113*, 1355–1364.

(34) Salisbury, J. L., and Salisbury, J. L. (2002) Centrin-2 is required for centriole duplication in mammalian cells. *Curr. Biol. 12*, 1287–1292.

(35) Johnson, C. N., Damo, S. M., and Chazin, W. J. (2014) EF-Hand Calcium-Binding Proteins, in *eLS* John Wiley & Sons, Ltd.

(36) 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.

(37) 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-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem. 281*, 356–364.

(38) 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*, 960–7.

(39) 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.

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

(41) Fairman-Williams, M. E., Guenther, U. P., and Jankowsky, E. (2010) SF1 and SF2 helicases: Family matters. *Curr. Opin. Struct. Biol. 20*, 313–324.

(42) Sokoloski, J. E., Kozlov, A. G., Galletto, R., and Lohman, T. M. (2016) Chemo-mechanical pushing of proteins along single-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A. 113*, 6194–9.

(43) Martinez-Sanz, J., Yang, A., Blouquit, Y., Duchambon, P., Assairi, L., and Craescu, C. T. (2006) Binding of human centrin 2 to the centrosomal protein hSfi1.

FEBS J. 273, 4504–4515.

(44) Rady, I., Siddiqui, I. A., Rady, M., and Mukhtar, H. (2017) Melittin, a major peptide component of bee venom, and its conjugates in cancer therapy. *Cancer Lett. 402*, 16–31.

(45) Heinen, T. E., and Gorini da Veiga, A. B. (2011) Arthropod venoms and cancer. *Toxicon 57*, 497–511.

(46) 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.

(47) Nishi, R., Sakai, W., Tone, D., Hanaoka, F., and Sugasawa, K. (2013) Structure-function analysis of the EF-hand protein centrin-2 for its intracellular localization and nucleotide excision repair. *Nucleic Acids Res. 41*, 6917–6929.

(48) Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P., and Craescu, C. T. (2003) Xeroderma Pigmentosum Group C Protein Possesses a High Affinity Binding Site to Human Centrin 2 and Calmodulin *J. Biol. Chem. 278*, 40252–40261.

(49) Seybold, C., Elserafy, M., Rüthnick, D., Ozboyaci, M., Neuner, A., Flottmann, B., Heilemann, M., Wade, R. C., and Schiebel, E. (2015) Kar1 binding to Sfi1 C-terminal regions anchors the SPB bridge to the nuclear envelope. *J. Cell Biol. 209*, 843–861.

(50) 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.

(51) Gromadka, R., Kaniak, A., Slonimski, P. P., and Rytka, J. (1996) A novel cross-phylum family of proteins comprises a KRR1 (YCL059c) gene which is essential for viability of Saccharomyces cerevisiae cells. *Gene 171*, 27–32.

(52) Sturm, M., Cheng, J., Baßler, J., Beckmann, R., and Hurt, E. (2017) Interdependent action of KH domain proteins Krr1 and Dim2 drive the 40S platform assembly. *Nat. Commun. 8*.

(53) Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K.,

Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Ponten, F. (2015) Tissue-based map of the human proteome. *Science. 347*, 1260419–1260419.

(54) Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S., Nilsson, P., Mattsson, J., Schwenk, J. M., Brunnström, H., Glimelius, B., Sjöblom, T., Edqvist, P.-H., Djureinovic, D., Micke, P., Lindskog, C., Mardinoglu, A., and Ponten, F. (2017) A pathology atlas of the human cancer transcriptome. *Science. 357*.

(55) Gromadka, R., and Rytka, J. (2000) The KRR1 gene encodes a protein required for 18S rRNA synthesis and 40S ribosomal subunit assembly in Saccharomyces cerevisiae. *Acta Biochim. Pol. 47*, 993–1005.

(56) Peled-zehavi, H., Berglund, J. A., and Rosbash, M. (2001) Recognition of RNA Branch Point Sequences by the KH Domain of Splicing Factor 1 (Mammalian Branch Point Binding Protein) in a Splicing Factor Complex *21*, 5232–5241.

(57) Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., and Ramos, A. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res. 40*, 6873–6886.

(58) Nicastro, G., Taylor, I. A., and Ramos, A. (2015) KH-RNA interactions: Back in the groove. *Curr. Opin. Struct. Biol. 30*, 63–70.

(59) Rodríguez Nassif, A., De La Arada, I., Arrondo, J. L., and Pastrana-Rios, B. (2017) 2D IR Correlation Spectroscopy in the Determination of Aggregation and Stability of KH Domain GXXG Loop Peptide in the Presence and Absence of Trifluoroacetate. *Anal. Chem. 89*, 5765–5775.

(60) Zheng, S., Lan, P., Liu, X., and Ye, K. (2014) Interaction between ribosome assembly factors Krr1 and Faf1 is essential for formation of small ribosomal subunit in yeast. *J. Biol. Chem. 289*, 22692–22703.

(61) Pastrana-Rios, B. (2001) Mechanism of unfolding of a model helical peptide. *Biochemistry 40*, 9074–9081.

(62) Pastrana-Rios, 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*, 041711-1 041711-11.

(63) Roux, S., Zekri, E., Rousseau, B., Paternostre, M., Cintrat, J.-C., and Fay, N. (2008) Elimination and exchange of trifluoroacetate counter-ion from cationic peptides: a critical evaluation of different approaches. *J. Pept. Sci. 14*, 354–359.

(64) Iloro, I., and Pastrana-Rios, B. (2006) Simulation of FT-IR spectra and 2D-COS analysis for the H/D exchange of two related ligands. *J. Mol. Struct. 799*, 153–157.

(65) Gaussier, H., Morency, H., Lavoie, M. C., and Subirade, M. (2002) Replacement of trifluoroacetic acid with HCl in the hydrophobic purification steps of pediocin PA-1: A structural effect. *Appl. Environ. Microbiol. 68*, 4803–4808.

(66) Barth, A. (2007) Infrared spectroscopy of proteins. *Biochim. Biophys. Acta - Bioenerg. 1767*, 1073–1101.

(67) Noda, I., and Ozaki, Y. (2004) Two-dimensional correlation spectroscopy: applications in vibrational and optical spectroscopy.

(68) Noda, I. (2004) Advances in two-dimensional correlation spectroscopy. *Vib. Spectrosc. 36*, 143–165.

(69) Noda, I. (2016) Techniques useful in two-dimensional correlation and codistribution spectroscopy (2DCOS and 2DCDS) analyses. *J. Mol. Struct. 1124*, 29–41.

(70) Noda, I. (2014) Frontiers of two-dimensional correlation spectroscopy. Part 2. Perturbation methods, fields of applications, and types of analytical probes. *J. Mol. Struct. 1069*, 23–49.

(71) Berova, N., Nakanishi, K., and Woody, R. (Eds.). (2000) Circular Dichroism: Principles and Applications 2nd. Edition. Wiley Online Library.

(72) Yang, A., Miron, S., Mouawad, L., Duchambon, P., Blouquit, Y., and Craescu, C. T. (2006) Flexibility and plasticity of human centrin 2 binding to the xeroderma pigmentosum group C protein (XPC) from nuclear excision repair. *Biochemistry 45*, 3653–3663.

(73) Valenti, L. E., Paci, M. B., Pauli, C. P. De, and Giacomelli, C. E. (2011) Infrared study of trifluoroacetic acid unpurified synthetic peptides in aqueous solution: Trifluoroacetic acid removal and band assignment. *Anal. Biochem. 410*, 118–123.

(74) Ito, F. (2014) Stable isomers for trifluoroacetic acid (TFA) pentahydrates obtained from density functional calculations. *Vib. Spectrosc. 71*, 57–61.

APPENDIX A

2D Correlation Spectroscopy Equations

Difference spectra used in 2D IR correlation spectroscopy are defined as:

$$\tilde{A}(\nu_j, t_k) = \begin{cases} A\,(\nu_j, t_k) - \bar{A}(\upsilon_j) & for\ 1 \leq k \leq m \\ 0 & otherwise \end{cases} \quad (1)$$

Where, $\bar{A}(\upsilon_j)$ is the initial spectrum of the data set to generate the covariance spectra. Synchronous 2D correlation intensities of the covariance spectral data are defined by:

$$\Phi(\nu_1, \nu_2) = \tilde{A}(\nu_1, t_j) \cdot \tilde{A}(\nu_2, t_j) \quad (2)$$

The resulting correlation intensity $\Phi(\nu_1, \nu_2)$ as a function of two independent wavenumber axes, ν_1 and ν_2, is the synchronous plot. Asynchronous 2D correlation intensities of the covariance spectral data are defined by:

$$\Psi(\nu_1, \nu_2) = \tilde{A}(\nu_1, t_j) \cdot N_{ij}\tilde{A}(\nu_2, t_i) \quad (3)$$

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

$$N_{ij} = \begin{cases} 0 & for\ i = j \\ \frac{1}{\pi(j-i)} & otherwise \end{cases} \quad (4)$$

APPENDIX B

Noda's Rules

(I) Asynchronous cross peak v_2: If positive, then v_2 is perturbed prior to v_1 ($v_2 \rightarrow v_1$).

(II) Asynchronous cross peak v_2: If negative, then v_2 is perturbed after v_1 ($v_2 \leftarrow v_1$).

(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) If the corresponding synchronous cross peak is negative and the asynchronous cross peak is positive, then the order is reversed.

(V) The order of events can be established for each peak observed on the v_2 axis.

APPENDIX C

Table. Biochemical characterization of *Hs*Krr1. Partial amino acid sequencing analysis of the N-terminal end of the recombinant *Hs*Krr1 validating the identity.

Protein	Accession Number	Amino Acids									
***Hs*Krr1 (WT)**	**Q13601**	**1**			**5**					**10**	
		A	**S**	**P**	**S**	**L**	**E**	**R**	**P**	**E**	**K**
		A	**S**	**P**	**S**	**L**	**E**	**R**	**P**	**E**	**K**

APPENDIX D

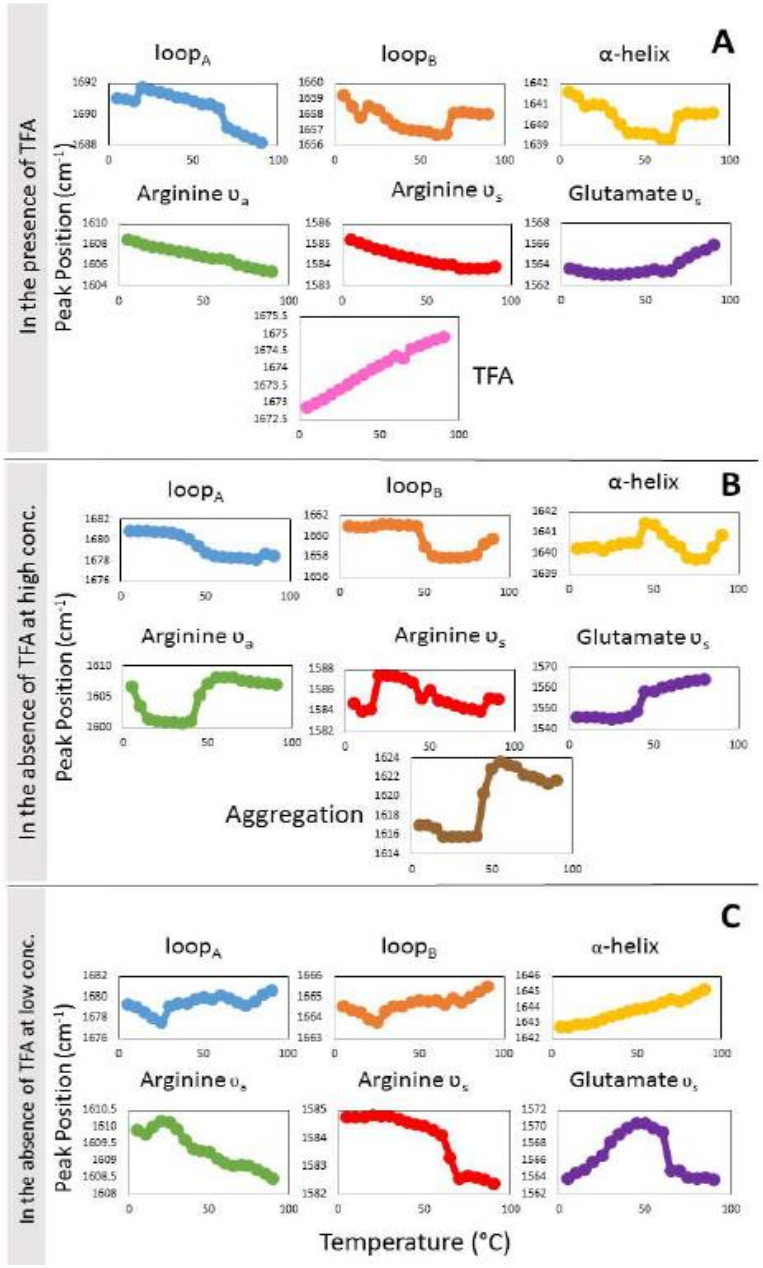


Figure. Thermal dependence plots from the curve fitting analysis (Figure 30) as a function of increasing temperature (5-90 °C) are shown for each spectral component corresponding to (A) GXXGlp in the presence of TFA, (B) GXXGlp in the absence of TFA at high concentration, and (C) GXXGlp in the absence of TFA at low concentration. *Adapted with permission from* American Chemical Society, Rodriguez Nassif 2017.[59]

APPENDIX E

Biographical Sketch

ASLIN M RODRIGUEZ NASSIF

Department of Chemistry
University of Puerto Rico, Mayagüez, Puerto Rico 00681-9019
aslin.rodriguez@upr.edu

A. Professional Preparation:

University of Puerto Rico Cayey	Chemistry	B.S. 2009

B. Professional Appointments:

2007-2009	Research Assistant, Chemistry Department, UPR, Cayey, PR
2010-2018	Ph. D. Candidate, Chemistry Department, UPR, Mayaguez, PR
2015	Scientist, Protein Dynamics Solutions LLC, Mayaguez, PR
2017	COOP Student Quality Control Dept., Lilly del Caribe, Carolina, PR

Professional Honors and Awards:

2010-2018	Alfred P. Sloan Foundation Minority Graduate Scholarship Programs, New York, NY
2011-2015	N**IH-RISE TO BEST**, University of Puerto Rico, Mayaguez, PR
2016-2017	Yale Ciencia Academy Award, Ciencia Puerto Rico, PR
2016-2018	BioXFEL UPR Graduate Student Fellowship, BioXFEL, Buffalo, NY

C. Publications

1. Sosa, L. D.V., Alfaro, E., Santiago, J., Narváez, D., Rosado, M.C., **Rodríguez, A.**, Gómez, A.M., Schreiter, E., and Pastrana-Rios, B. The Structure, Molecular Dynamics, and Energetics of Centrin-Melittin Complex (**2011**) *(Protein, Structure, Function, and Bioinformatics) DOI:10.1002/prot.23142.*

2. Pastrana-Ríos, B.; Reyes, M.; De Orbeta, J.; Meza, V.; Narváez, D.; Gómez, A.M.; **Rodríguez Nassif, A.**; Almodovar, R.; Díaz Casas, A.; Robles, J.; Ortiz, A.M.; Irizarry,

L.; Campell, M.; Colón, M. Relative Stability of Human Centrins and Its Relationship to Calcium Binding. (**2013**) *Biochemistry* 52:1236-1248.

3. **Rodríguez Nassif A**, de la Arada I, Arrondo JL, Pastrana-Rios B. "2D IR Correlation Spectroscopy in the determination of aggregation and stability of KH domain GXXG loop peptide in the presence and absence of trifluoroacetate." (**2017**) *Analytical Chemistry* 89 (11):5765-75.

D. Synergistic Activities

- **Student advisor at the National Society of Leadership and Success Mayaguez Chapter**
  - *August - December 2016*
    - *Mayaguez, Puerto Rico*

Internships

- **Vanderbilt University, Dr. Walter Chazin Laboratory**
  - *July-November 2015*
    - *Nashville, Tennessee*

- **Hauptman Woodward Institute, Dr. William Bauer Laboratory**
  - *January-July 2018*
    - *Buffalo, New York*

Posters and Oral Presentations

- **Characterization of Krr1-ssDNA Complex by Molecular Biophysical Studies**
  - *Poster Presentation*
    - *BioXFEL Annual Conference, Ponce, PR*
      - *January 13-15, 2015.*
- **Molecular Biophysical Characterization of Krr1p-ssDNA Complex**
  - *Poster Presentation (Awardee)*
    - *PepTalk 2015, San Diego, CA*
      - *January 19-23, 2015*
- **Biophysical Characterization and Molecular Behavior of Krr1p-ssDNA Complex**
  - *Poster Presentation*
    - *38th Senior Technical Meeting ACS Puerto Rico Local Section, San Juan, PR*
      - *November 7, 2014*
- **Molecular Biophysical Characterization of Krr1p-ssDNA Complex**
  - *Poster Presentation (Awardee)*
    - *The American Society of Cell Biology 2014 Annual Meeting, Philadelphia, PA*
      - *December 6-10, 2014*

- **Cloning a Novel Centrin Target Peptide**
  - *Poster Presentation, SIGMA XI XVI POSTER DAY. Mayagüez, PR*
    - *April 12, 2011*
  - *Poster Presentation, 2011 Lilly Academy Technical Forum. Carolina, PR*
    - *April 15, 2011*
  - *Poster Presentation, IUPAC 2011. San Juan, PR*
    - *Jul 31-August 5, 2011*
  - *Poster Presentation, ASCB 2011. Denver, CO*
    - *Dec 3-5, 2011 (MAC Award)*
  - *Oral Presentation, Ely Lilly 2011. Carolina, PR*
    - *Dec 8, 2011*
  - *Poster Presentation, 2012 Lilly Academy Technical Forum. Carolina, PR*
    - *Apr 2012*
  - *Poster Presentation, Biophysical Society Annual Meeting. Philadelphia, USA*
    - *Feb 2012*
- **Methodology to assess the inhibition of α-amylase by plant extracts with anti-diabetic potential**
  - *Poster Presentation, Biominds Research Symposium. Cayey, PR. May 2008*
  - *Poster Presentation, 19thAnnual Student Research Symposium. Cayey, PR. August 29, 2008. Honorific mention*
  - *Oral Presentation, Proposals Student Presentation. Cayey, PR. Sep 27, 2008.*
  - *Poster Presentation, Annual Biomedical Research Conference for Minority Students (ABRCMS). Orlando, FL. Nov 5-9, 2008*
  - *Poster Presentation, Biominds Poster Annual Day. Mayaguez, PR. Mar 21, 2009*
  - *Poster Presentation, Society of Toxicology Annual Meeting. Baltimore, MD. Mar 15-19, 2009*

Attended Workshop

- *Proteomics Ponce School of Medicine and Health Sciences. Ponce, PR. June 1, 2012*
- *Protein Production and Purification University of Puerto Rico at Mayaguez, Mayaguez, PR. May 11th, 2016.*