The Effects of Centrin Mutations on the Affinity, Stability, and Structure of the Centrin-Sfi1p$_{21}$ Complex

by

Adriana Oliveras-Cabrera

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

MASTER OF SCIENCE
in
CHEMISTRY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2017

Approved by:

_____	_____
Aikomarí Guzmán-Martínez, PhD Member, Graduate Committee	Date
_____	_____
Jorge Ríos-Steiner, PhD Member, Graduate Committee	Date
_____	_____
Belinda Pastrana-Ríos, PhD President, Graduate Committee	Date
_____	_____
Carlos Velazquez Figueroa, PhD Representative of Graduate Studies	Date
_____	_____
Enrique Meléndez, PhD Chairperson of the Department	Date

Abstract

Centrin is a small, calcium binding protein part of the EF-hand superfamily. As one of the 347 eukaryotic signature proteins (ESPs), centrin has no bacterial or archaea homologue. Centrin is required for centriole duplication and assembly of cilia and flagella. In human cells, centrin is found as three isoforms, termed centrin 1-3, respectively. Centrin 1 is found in the base of the flagella in the sperm, centrin 2 and centrin 3 are ubiquitously expressed in all somatic cells, but centrin 2 has also been found to regulate DNA excision repair within the nucleus and export mRNA and other proteins from the nucleus. As an ESP, centrin has a wide range of biological targets with whom it interacts, one of them being Sfi1. Sfi1 is a vital centriolar protein required for its duplication. Variations in centrin's sequence can affect its structure and function, resulting in weaker interactions with its biological targets. Circular dichroism (CD) and isothermal titration calorimetry (ITC) are spectroscopic and thermodynamic techniques, respectively, that were used to characterize non-conserved mutations in centrin 1 and centrin 2 and their effect on centrin's secondary structure and its interaction with a fragment of human Sfi1 (*Hs*Sfi1p$_{21}$). From CD analyses it was observed that the non-conserved variations contribute to the helical content and increase thermal stability. Thermodynamic binding parameters obtained by ITC indicated that the non-conserved variation affected the interaction interface between the studied variant and *Hs*Sfi1p$_{21}$.

Resumen

Centrin es una pequeña proteína que enlazante de calcio y es parte de la superfamilia de la mano EF. Como una de las 347 proteínas esenciales para la vida (ESP, por sus siglas en inglés), centrin no tiene un homólogo en bacterias o arqueas. Centrin es requerida para la duplicación del centriolo y ensamblaje de cilios y flagela. En células humanas, centrin se encuentra como tres isoformas, nombradas centrin 1-3, respectivamente. Centrin 1 se encuentra en la base de la flagela en los espermatozoides, centrin 2 y centrin 3 son expresados de manera ubicua en las células somáticas, pero centrin 2 también se ha encontrado que regula la reparación del DNA dentro del núcleo y exporta mRNA y otras proteínas del núcleo. Como una ESP, centrin tiene un amplio rango de blancos biológicos con los cuales interacciona, uno de ellos siendo Sfi1. Sfi1 es una proteína centriolar vital requerida para la duplicación de éste. Variaciones en la secuencia de centrin pueden afectar su estructura y función, resultando en interacciones más débiles con sus blancos biológicos. Dicroísmo circular (CD, por sus siglas en inglés) y titulación calorimétrica isotermal (ITC, por sus siglas en inglés) son técnicas espectroscópicas y termodinámicas, respectivamente, que se utilizaron para caracterizar variaciones no-conservadas en centrin 1 y centrin 2; su efecto en sobre la estructura secundaria de centrin y su interacción con un fragmento de Sfi1 humano (*Hs*Sfi1p$_{21}$). A partir de los análisis de CD se observó que las variaciones no conservadas contribuyen al contenido helicoidal y aumentan la estabilidad termal. Los parámetros termodinámicos de enlazamiento obtenidos por ITC indicaron que la variación no-conservada afectó la interacción entre la variante estudiada y *Hs*Sfi1p$_{21}$.

Acknowledgement

I want to thank the Chemistry Department at the University of Puerto Rico at Mayaguez for providing financial support. A special thanks to the Chair of my committee, Dr. Belinda Pastrana-Ríos for accepting me as a graduate student and allowing me to be part of her research laboratory team. I want to thank her for her unfaltering support, motivation and respect throughout the process of completing this thesis.

Thanks to Dr. Enrique Melendez's laboratory team for their collaboration with the circular dichroism experiments and to Dr. Michael Berne from Tufts University for his collaboration in the amino acid sequencing of the proposed variants of centrin.

I would also like to thank my committee: Dr. Aiko Guzmán and Dr. Jorge Ríos-Steiner, for their recommendations and support.

To Dr. Pastrana-Ríos' graduate students-Aslín Rodríguez, Adalberto Díaz, Bianca Alamo, Tatiana Garcés, Yesenia Acevedo and Michael Meléndez- thanks for your help and friendship throughout these years. And to the lab's undergraduates Sherly Nieves and Gabriela Casanova for their tireless support throughout the experiments.

Table of Contents

List of Tables

List of Figures

I. Justification

Understanding protein-protein interactions (PPI's) is the basis for the development of novel drug therapies. As a eukaryotic signature protein (ESP), centrin is highly conserved and in higher eukaryotes it is present as various isoforms [1]. These characteristics make centrin a possible drug target for diseases in which centrin is overexpressed [4] [5] [6] [7]. Centrin is a small calcium-binding protein of approximately 172 residues and a molecular weight of 20 kDa. Localized in the nucleus, centrioles, basal bodies, primary cilium, cilia, flagella and the connecting cilium; centrin has various biological targets with whom it interacts, such as Sfi1, XPC and Prp40 Homolog A, among others. Centrioles are the cell's organelles responsible for spindle formation and DNA separation during the cell cycle. Centrioles can also differentiate to form basal bodies, giving rise to cilia and flagella, organelles critical for sensory perception and cell locomotion [7] [8]. Correlated to a vast number of human diseases, centriole instability can be the result of defects in centriole duplication, structure or function or defects in the formation or maintenance of cilia and flagella [9].

Centrin's role in centriole and basal body duplication and structure has been thoroughly studied ever since its discovery in the contractile fibers of the basal body in green algae [10]. In *Chlamydomonas,* centrin has an essential role in microtubule severing and flagellar excision [11], and its deficiency causes defects in basal body duplication and growth in both *Chlamydomonas reinhardtii* [12] and *Paramecium tetraurelia* [13]. As previously mentioned, in higher eukaryotes centrin has up to four isoforms, three of which are found in humans. In human somatic cells, centrin isoform 2 (*Hs*cen2) is localized within the centriole and basal body where it interacts with Sfi1. In

2002, Salisbury established that centrin 2 is necessary for centriole duplication in mammalian cells [14].

Centrin's targets have an idiosyncratic consensus sequence of variable length designated as a centrin binding site (CBS) to which centrin binds in a 1:1 ratio [15]. The CBSs are characterized by a highly conserved hydrophobic triad with a tryptophan residue in the first position and two leucines in the fourth and eighth positions ($W_1L_4L_8$). Amongst centrin's targets we chose the centriolar protein Sfi1 [16] for the focus of this research. *Homo sapiens* Sfi1 (*Hs*Sfi1) consists of 1242 residues and comprises approximately 23 tandem CBSs. Based on the 1:1 ratio between the CBS and centrin, we selected the 21st CBS (*Hs*Sfi1p$_{21}$) to study the interaction between the peptide and full-length human centrin 1. The peptide fragment contains the consensus sequence $AX_7LLX_3F/LX_2WK/R$, where X represents any amino acid. It is also important to remark that the hydrophobic triad is reversed: $L_1L_4W_8$, rather than $W_1L_4L_8$.

Variations in centrin's sequence can alter its secondary structure, its thermal stability and its binding affinity towards its targets and as result it can decrease the centrin-target complex's stability. For this research project, we will carry out the recombinant expression, purification and biophysical characterization of three *Homo sapiens* (*Hs*) centrin 1 and centrin 2 variants using circular dichroism (CD) and one *Hs*cen1 variant in complex with *Hs*Sfi1p$_{21}$ using isothermal titration calorimetry (ITC). CD is a spectroscopic technique that provides information on the protein's secondary structure, and by applying a thermal perturbation we can determine the structure's stability. ITC is a powerful thermodynamic technique that provides the complex's binding parameters. The results obtained in this research are unprecedented because neither these centrin variants nor

the *Hs*cen1(E24K)-*Hs*Sfi1p$_{21}$ complex have been studied. The results will further complement and help in the understanding of the protein-protein interactions between full-length human centrin 1 and centrin 2 variants and *Hs*Sfi1p$_{21}$.

Objectives

The objectives of this research are:

- to characterize the secondary structure of *Hs*cen1(E24K), *Hs*cen2(E32A) and *Hs*cen2(E24K/E105K) using circular dichroism (CD),
- perform thermal dependence studies of said variants using CD,
- to compare the variants' secondary structure with wild-type human centrin
- to perform isothermal titration calorimetry studies to determine the thermodynamic binding parameters of *Hs*cen1(E24K) with the peptide *Hs*Sfi1p$_{21}$
- to compare *Hs*cen1(WT)-*Hs*Sfi1p$_{21}$ binding parameters with *Hs*cen1(E24K)-*Hs*Sfi1p$_{21}$

Hypothesis

Through this research project we aim to corroborate and validate the following hypothesis. The non-conserved variations in *Hs*cen1 and *Hs*cen2 will not affect to a significant extent the protein's secondary structure in comparison to wild-type *Hs*cen1 and *Hs*cen2. The variant centrin-containing complex will have a lower binding affinity when compared to its wild-type counterpart.

II. Literature Review

A. The Centrosome, Centrioles and Basal Bodies

Eukaryotic cells depend on microtubule organizing centers (MTOCs) to carry out pivotal functions such as cell division, locomotion, cell signaling and sensory perception [17]. MTOCs can be found as centrosomes or spindle pole bodies (SPBs). Animal cells contain both centrosomes and SPBs, whereas yeast contain SPBs which function as a centrosome equivalent. Centrosomes are located near the nucleus, do not have a membrane and they are comprised of approximately 250 conserved proteins throughout the animal kingdom [17]. The centrosome's primary function is to regulate the cell cycle through the duplication of centrioles **(Figure 1)** a role of outmost importance for proper chromosome alignment and separation during cell division [7]. This process involves centrosome duplication which occurs ***only once per cell cycle***. Hence, abnormal centriole duplication is detrimental to the cell, impairing the proper transfer of genetic material with observed unipolar centrioles, hypertrophy of the centrosomes which often leads to chromosome instability [7] [18]. Therefore, the clinical relevance of the organelle and its association with cancer.

Centrosomes, basal bodies, primary cilium and cilia have very similar protein components, but they differ in their location and function. Structurally, basal bodies are essentially modified centrioles [19]. They lack a surrounding membrane and consist of three fundamental structural domains: (1) a pair of orthogonally arranged centrioles which are embedded in (2) an amorphous matrix known as the pericentriolar matrix (PCM) and (3) gamma-tubulin complexes. Centrioles are the only structural component of the centrosome with two orthogonally arranged cylindrical-shaped structures consisting of

nine sets of triplet microtubules [20] [21] [9]. The triplet microtubules are composed of 13 -tubulin and β-tubulin-containing proto-filaments, called the A-tubule, to which two successive 10 proto-filament microtubules, called B- and C-tubules, are assembled. The microtubules triplet arrangement often become doublets toward the distal end **(Figure 2)** [22] [20], a region called the transition zone. In most organisms, the transition zone shows a stellate organization with an array of thin fibers linking alternate microtubules to one another. These fibers consist of a centrin-rich protein complex (**Figure 3** and **Figure 4**). The centrin-rich stellate fibers act as calcium-sensitive contractile connectors between the microtubules of the transition zone, mediating changes in the overall structure of the distal region of the centriole, especially during cilia or flagellar loss and re-growth [23]. The PCM serves as a platform in which protein complexes regulate microtubule assembly. Major components of the PCM are Cep 135, ninein, pericentrin and other coiled-coil proteins. The entirety of the PCM components and their interactions are yet to be characterized and wholly understood, with recent advancements being made using cryoelectron tomography [24], [25]. In order to nucleate the microtubules, ϒ-tubulin complexes are anchored to the centrosome by pericentrin and kendrin, which in turn are anchored by protein kinase A [23].

In animal cells, centrioles have been observed to interconvert to basal bodies during the G1 phase of the cell cycle where they move to the cell's surface and become anchored through a basal foot **(Figure 5)** [26]. Basal bodies are found in most organisms in the animal and protist kingdoms and in algae [27]. Basal bodies give rise to flagella and cilia, which are critical structures for cell locomotion and sensory perception [7], [8].

As previously mentioned, centrosome duplication is semiconservative occurring only once per cell cycle. Each pair of centriole consists of a mature (mother) centriole and an immature (daughter) pro-centriole, linked through their proximal regions by specialized structures that form near the distal end of the mother centriole. There are two possible pathways for new centriole formation: (1) the semiconservative centriole duplication, where a daughter centriole buds from the proximal end of the mother centriole and the (2) *de novo* pathway, where centrioles form without contribution of pre-existing centrioles. In most cells, pro-centrioles are formed from pre-existing centrioles. This process has been thoroughly studied in the basal body formation of model unicellular organisms such as *Paramecium tetraurelia* and *Chlamydomonas reinhardtii* and has been confirmed that the fundamental features are common to higher eukaryotes [28]. In semiconservative centriole duplication, the first stage of centriole duplication is the separation of the centriole pair, followed by the development of a cartwheel structure with ninefold symmetry embedded in electron-dense material outside the proximal end of the mother centriole. The cartwheel structure serves as a template for the nine “singlet” microtubules to bud, then converting to “doublet” microtubules by the addition of proto-filaments which share a portion of the pre-existing microtubule wall, then transform into “triplets” as they elongate to form the ninefold blades of a pro-centriole growth of the pro-centriole continues within the distal end until finally a daughter centriole is achieved. In the *de novo* pathway centrioles form in the cytoplasm, adjacent to amorphous and electron-dense granular masses of the PCM. **(Figure 1)** [27] [29] [30]

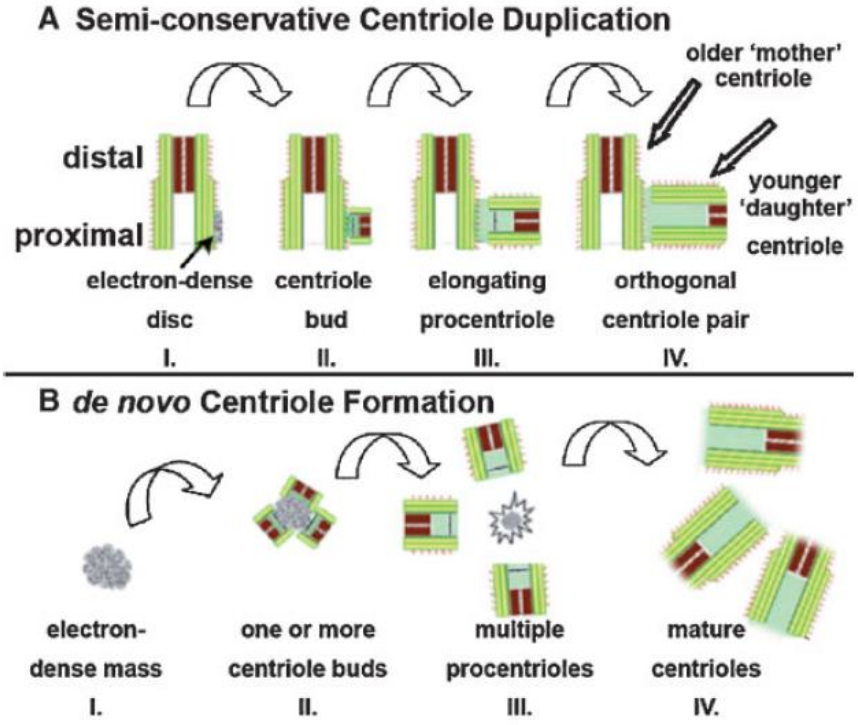


Figure 1. Centriole duplication. (a) semi-conservative duplication and (b) de novo formation. (Adapted from Salisbury 2007 *[23]*)

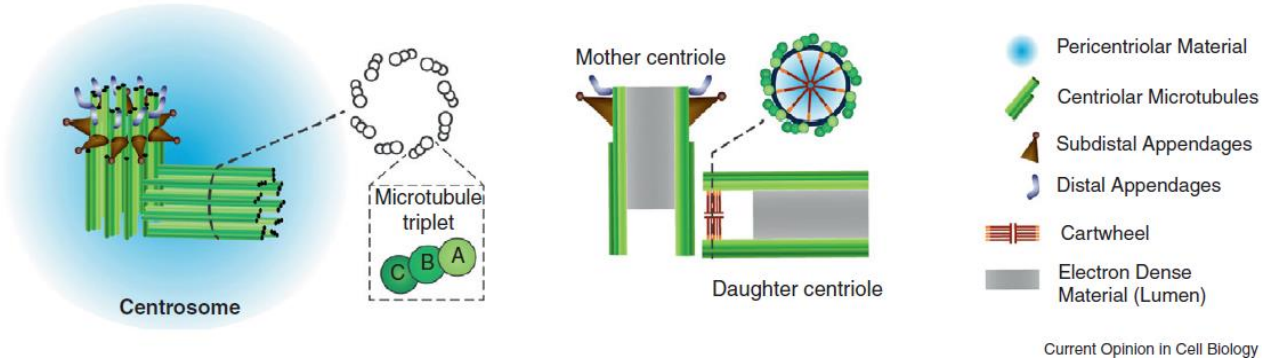


Figure 2. Animal centriole pair. The centrosome consists of two microtubule-based structures, the centrioles, embedded in the PCM. Each pair of centriole consists of a mature (mother) centriole and an immature (daughter) pro-centriole, linked through their proximal regions by linker proteins. (Adapted from Bettencourt-Dias 2012 *[20]*)

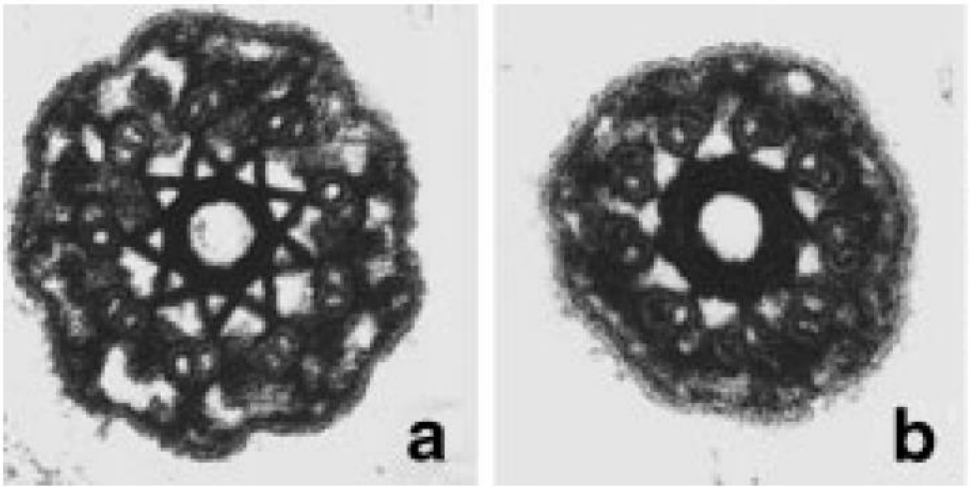


Figure 3. Electron micrographs of the transition zone in the distal region of *Chlamydomonas reinhardtii* centriole. (a) extended (b) contracted centrin-rich fibers. (Adapted from Salisbury 2007 *[23]*)

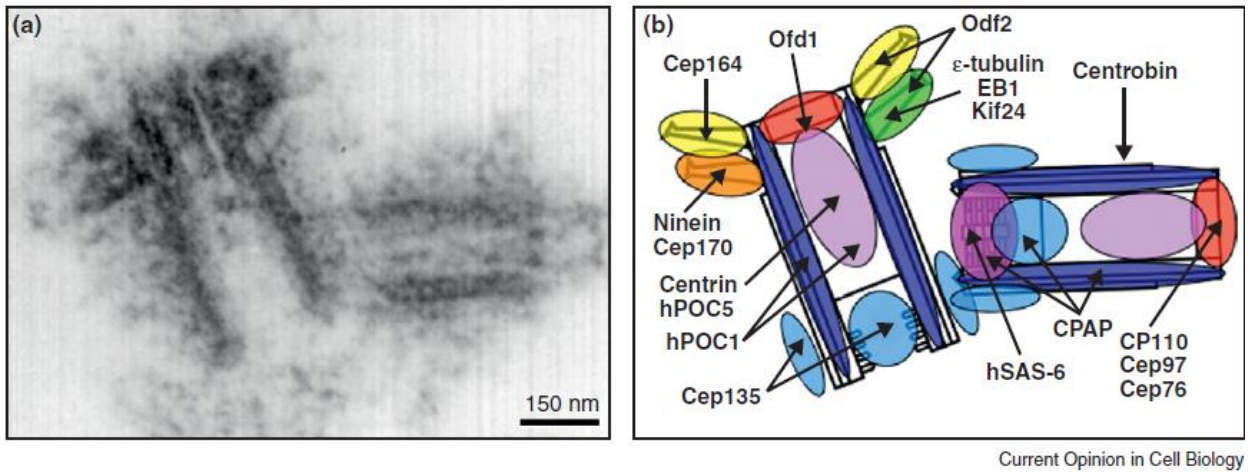


Figure 4. Centriole pair. (a) Electron micrograph of a longitudinal section of a human centrosome. (b) The localization of several centriolar proteins. The transition zone is represented by the purple ellipsoid. (Adapted from Bettencourt-Dias 2012 *[20]*)

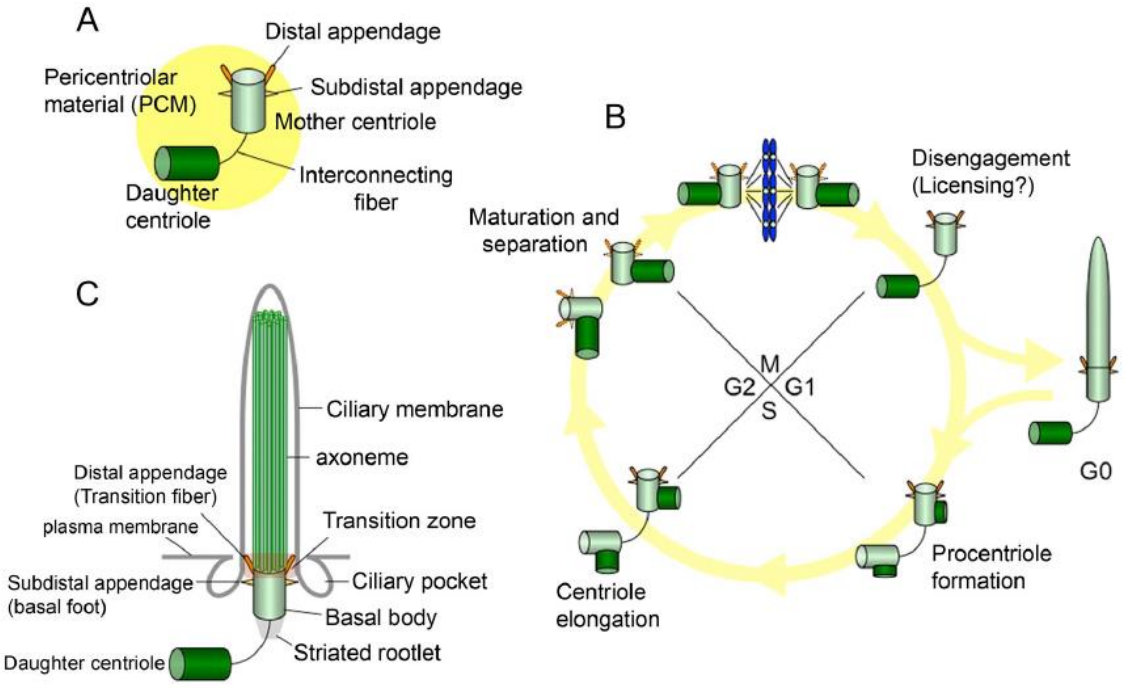


Figure 5. Centrosome, the cell cycle and the cilia. (a) mother and daughter centrioles in the PCM (b) semiconservative centriole duplication in the cell cycle and (c) basal body structure. The basal body is localized near the plasma membrane where it nucleates to a primary cilium. The mother centriole converts to the basal body, presenting structures such as the transition fibers/distal appendage, basal foot/sub-distal appendage and striated rootlet. The transition fibers tether the basal body to the plasma membrane in the transition zone. (Adapted from Kobayashi et al. 2011 *[26]*)

B. Centrin

Centrin was first identified in the unicellular green alga *Tetraselmis striata* by Dr. Jeffrey L. Salisbury in 1984. Salisbury found that *T. striata's* contractile organelles were mainly composed of a 20 kilo Dalton (kDa) protein sensitive to calcium, this protein was later termed centrin [10]. Centrins are highly conserved, small calcium-binding proteins (CaBP) members of the EF-hand superfamily with a molecular weight of approximately 20 kDa [31]. In 2002, in a feat to understand the origin of the eukaryotic cell or more specifically the nucleus, Hartman and Fedorov identified 347 proteins in eukaryotic cells that have no significant homology to proteins in Archaea and Bacteria. Hartman and Fedorov designated these proteins as eukaryotic signature proteins (ESPs) [1]. Within these 347, proteins 108 of them were associated to signaling systems. Among the ESPs associated to signaling systems are centrin and calmodulin (CaM), the latter being a ubiquitous CaBP critical for cellular signal transduction pathways. Whilst CaM is a ubiquitous protein, centrin is usually localized to MTOCs, yet has also been found in the nucleus and the nuclear pore [32]. Even though CaM and centrin function in distinct signaling pathways they share approximately more than 50% sequence identity, and both are comprised of four EF-hand motifs. EF-hand proteins have two globular domains linked by a tethered helix. Each globular domain is composed of two helix-loop-helix motifs, each motif capable of binding calcium with different affinities [33] **(Figure 6)**. Chazin's group determined the calcium dissociation constants for *Chlamydomonas reinhardtii* centrin (*Cr*cen) to be 1.2 μM for sites I and II near the N-terminal domain and 160 μM for sites III and IV near the C-terminal domain. More so, in 2014 Dr. Belinda Pastrana-Ríos validated the conformational changes upon calcium binding in the terminal domain fragments and

in full-length *Cr*cen [31]. Previously, Pastrana-Ríos' group had also characterized the relative stability of full-length human centrins through thermodynamic and spectroscopic techniques [33].

In lower eukaryotes, centrin-based fibers have essential roles. In the unicellular green algae *Chlamydomonas reinhardtii*, these centrin-based fibers in the basal body of the flagella play a fundamental role in microtubule severing at the time of flagellar excision during sexual reproduction [11]. Similarly, in *Saccharomyces cerevisiae*, a unicellular yeast and model organism, the centrin homologue Cdc31 plays a crucial role in SPB duplication. In 2006 Kilmartin's group determined the first high-resolution structure of Cdc31-Sfi1 complex, one of its primary targets [34]. In *S. cerevisiae* the nucleation of the spindle and cytoplasmic microtubules is performed by the SPB. On one side of each SPB there is a structure called the half-bridge, which is essential for SPB duplication. The half-bridge is formed of fibers rich in centrin, Sfi1 and other centrin targets.

Up to four centrin genes have been identified in higher eukaryotes **(Figure 7)**. In the rodents *Mus musculus* and *Rattus norvegicus* the four centrin genes were identified (*MmCetn*1-4 and *RnCetn*1-4, respectively.) However, in *Homo sapiens (Hs*) only three genes have been identified (*HsCETN*1-3) and a predicted fourth centrin pseudogene is found on chromosome 4, but the potential gene transcript encodes a very short 98 amino acid long peptide. It is uncertain whether this transcript will exist as a functional polypeptide [35]. In humans, the centrin isoforms are abbreviated as *Hs*cen1 to *Hs*cen3, respectively. *Hs*cen1 is localized in the base of the flagella in the sperm [36], where centrin deficiency has been correlated to human male infertility [37] [38]. *Hs*cen2 is perhaps the most versatile of centrin isoforms, ubiquitously expressed in all somatic cells,

has also been found to regulate DNA excision repair within the nucleus [2] and export mRNA and other proteins from the nucleus [32]. Like *Hs*cen2, *Hs*cen3 is ubiquitously expressed in all somatic cells but linked to the spindle pole bodies and chromatin separation.

As an eukaryote signature protein, centrins are highly conserved and variations in centrin's sequence are detrimental to cellular function such as improper centriole duplication and impaired sensory perception [38] [14]. Variations in centrin's secondary structure can reduce the binding affinity for its various biological targets resulting in weaker and less stable protein-protein interactions (PPIs) [3] [39] [40] **(Figure 8)**. Among these targets are Sfi1 and Prp40, whose interactions with centrin were recently validated and characterized by Pastrana-Ríos and Díaz-Casas [40] [41].

Improper centrosome duplication and defective assembly of cilia can result in various diseases. Improper centrosome duplication has long been correlated to breast cancer among other types of cancer [7] [44] [18]. There are two types of cilia: motile and non-motile or primary cilia. Primary cilia are the cell's main sensory organelle and are found throughout most organs, predominantly in the kidneys, pancreas, eyes, neurons and thyroid gland among many others [42] [43]. Defective cilia have been determined to be the causative agent of a wide class of diseases and syndromes grouped as ciliopathies [45]. Ciliopathies comprise conditions such as such as polycystic kidney disease, some forms of retinal degeneration, Bardet-Biedl syndrome, nephronophthisis and obesity among others [45].

As a key protein in centrosome duplication and ciliary function, centrin is a high-potential target for small-molecule drug therapy development. Recently, centrin was

identified as a novel target for a live attenuated vaccine against *Leishmania donovani*, a protozoan parasite. Classified by the Center of Disease Control (CDC) and Prevention as a Neglected Tropical Disease, leishmaniasis currently affects about 4 to 12 million people and no licensed human vaccine is available. Studies with live attenuated *Leishmania* vaccines with centrin deleted *L. donovani* parasites showed protective immunity in animal models [5]. Centrin has also been identified as a target for monastrol, an inhibitor of the mitotic spindle with a high potential as a potential anti-cancer drug [4].

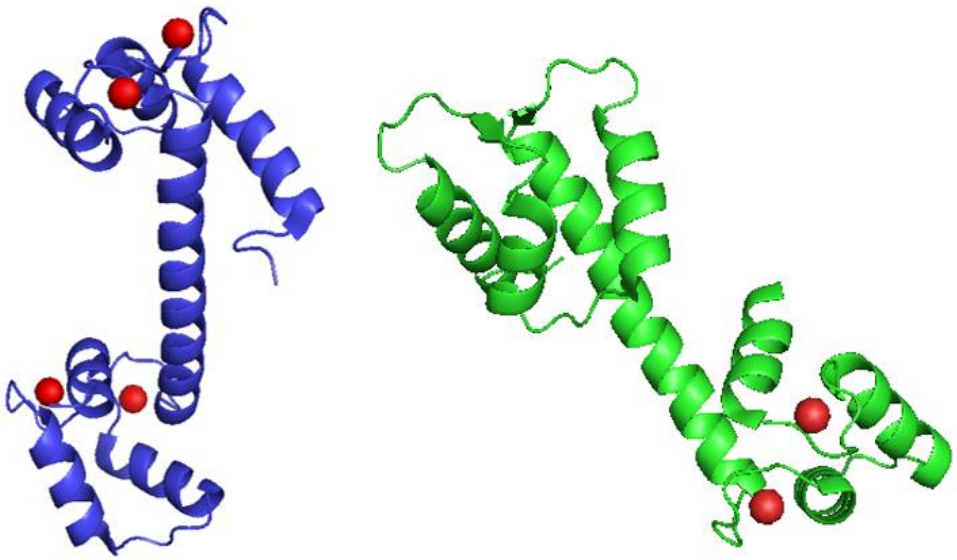


Figure 6. Ribbon model for archetypal proteins of the EF-hand superfamily. *Gallus domesticus* calmodulin (*Gd*CaM, blue ribbon) and Human centrin 2 (*Hs*cen2, green ribbon). In CaM two EF-hands at each terminal bind a calcium ion. In centrin only the two EF-hands at the C-terminal are binding calcium. The calcium ions are represented as red spheres bound by the helix-loop-helix motifs. Structures were generated from the Protein Data Bank PDB ID: **1UP5** and **2GGM**, respectively and using PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

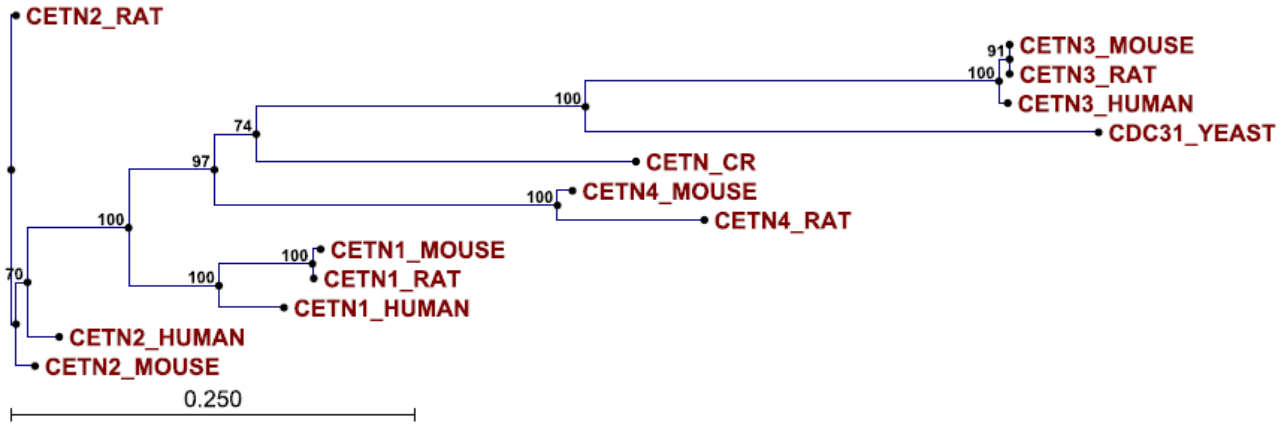


Figure 7. Phylogenetic relationship of centrin throughout model organisms. The relationship includes *Hs*cen 1, 2 and 4 with *Cr*cen, while *Hs*cen 3 and Cdc31 are more closely related, which can imply two divergent centrin subfamilies.

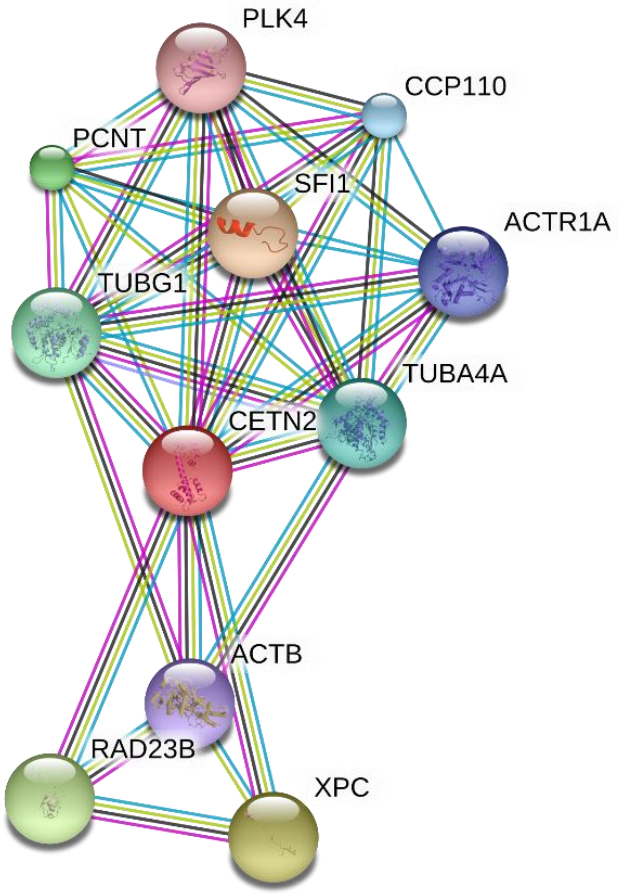


Figure 8. Hscen2 interactome was generated using the search tool for the retrieval of interacting genes/proteins (STRING) v. 10.5 from the Swiss Institute of Bioinformatics.

C. Sfi1: a centrin target

Sfi1 was first identified by Dr. Patrick Van Dijck's group in 1999 in *S. cerevisiae*, they determined that it is required for progression from the G_2 to M transition of the cell cycle [46]. *S. cerevisiae* Sfi1 (*Sc*Sfi1) is a protein of 946 amino acids and a molecular weight of 113 kDa. In *Saccharomyces* species Sfi1 tends to be a variable protein, with only a 19.4% sequence identity between *S. cerevisiae* and *S. pombe*, however, it tends to be more conserved in higher eukaryotes, with a 64.2% sequence identity between *H. sapiens* and *Mus musculus*. Regardless of its sequence variability, Sfi1 has a series of internal repeats containing a consensus sequence important for its interaction with centrin. In 2003, Kilmartin determined through pull-down experiments that Cdc31 (centrin's yeast homolog) binds to Sfi1 directly in a 1:1 ratio [16].

As mentioned earlier, Kilmartin's group published the structure of the Cdc31-Sfi1 complex at low and high Ca^{2+} concentrations. *Sc*Sfi1 has approximately 20 tandem centrin binding sites of variable length separated by gaps of 23 to 35 amino acids [47] whereas *Hs*Sfi1 has 23 sites with a length of 33 amino acids and gaps of 10 amino acids. The gaps serve as interaction sites between neighboring centrins [48]. The centrin binding sites have the consensus sequence $AX_7LLX_3F/LX_2WK/R$ where the X represents variable amino acids. The consensus sequence tends to be variable, except for the following hydrophobic amino acids: the tryptophan, the phenylalanine/leucine, the leucine and the alanine in the positions 17, 14, 8 and 1, respectively. Of the four residues, the tryptophan is the most conserved one. The residues W_{17}, F/L_{14} and L_8 consist of the hydrophobic triad present in the centrin binding sites found along centrin's targets **(Figure 9).** Important features to note from the structure is (a) that the complex's conformation did not vary

much at high and low Ca^{2+} concentrations **(Figure 10)** and (b) the structure shows that centrin's N-terminal domain binds to the conserved alanine in Sfi1's N-terminal half and centrin's C-terminal domain binds to the C-terminal half of the Sfi1 fragment, the specific location of which is not known because it was not visible in the electron density map. The terminal domain interactions are particularly important in SPB duplication. Extending from the mother SPB is the half-bridge, an area consisting of Kar1, Msp3, Sfi1 and centrin **(Figure 11a)**. At the moment of SPB duplication the half-bridge elongates through the head-to-tail binding of centrin/Sfi1 fibers extending from the C-terminal domain of Sfi1 in the mother SPB **(Figure 11b)**. The complete half-bridge provides an anchor site at the N-terminal domain of the centrin/Sfi1 fiber where SPB components start to assemble. After the SPB components are added the two SPBs separate by cleavage of the bridge, leaving a half bridge with each SPB **(Figure 11c)**. [23] [15]

Craescu's group analyzed the interaction, in terms of secondary structure, of a *Hs*Sfi1 fragment (from residues 475-494) upon binding the C-terminal domain of *Hs*cen2 by circular dichroism. They found that the Sfi1 fragment by itself had a disordered conformation but upon binding centrin the conformation shifted onto an α-helical one **(Figure 12)** [48]. More relevantly, Pastrana-Ríos' group also used CD to carry-out a comparative analysis of the three human centrin isoforms in their full-length **(Figure 13).**

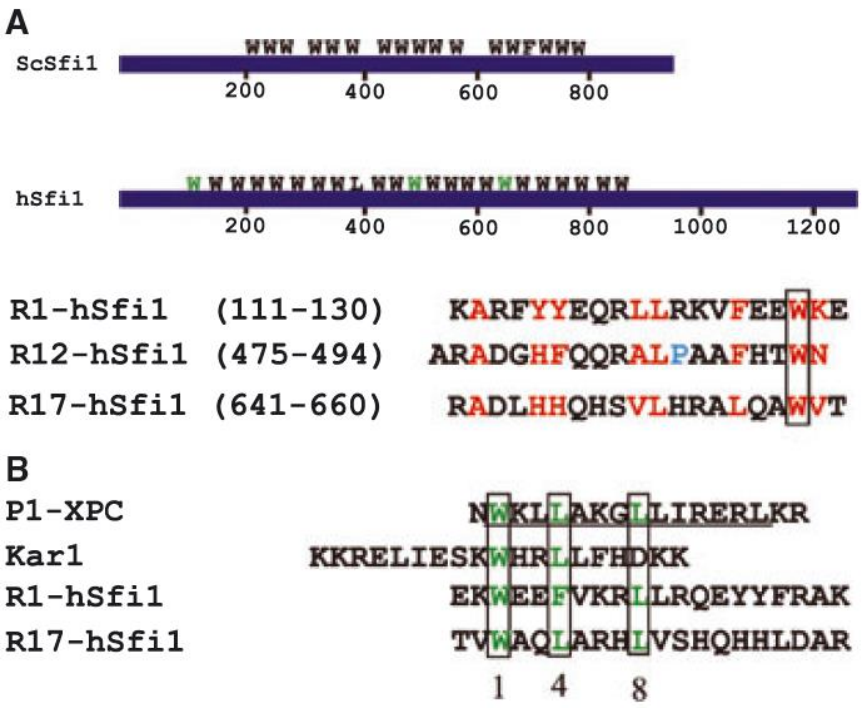


Figure 9. Consensus sequence in Sfi1. (A: upper panel) Schematic representation of the consensus sequence in full-length *Sc*Sfi1 and *Hs*Sfi1. (A: middle panel) sequence alignment of three centrin binding sites, positions #1, #12 and #16, respectively. (B) Sequence alignment of the centrin binding sites in three of centrin's targets. Note: Sfi1's sequence is reversed for purpose of the alignment. (Adapted from Martínez-Sanz et al 2006 *[48]*)

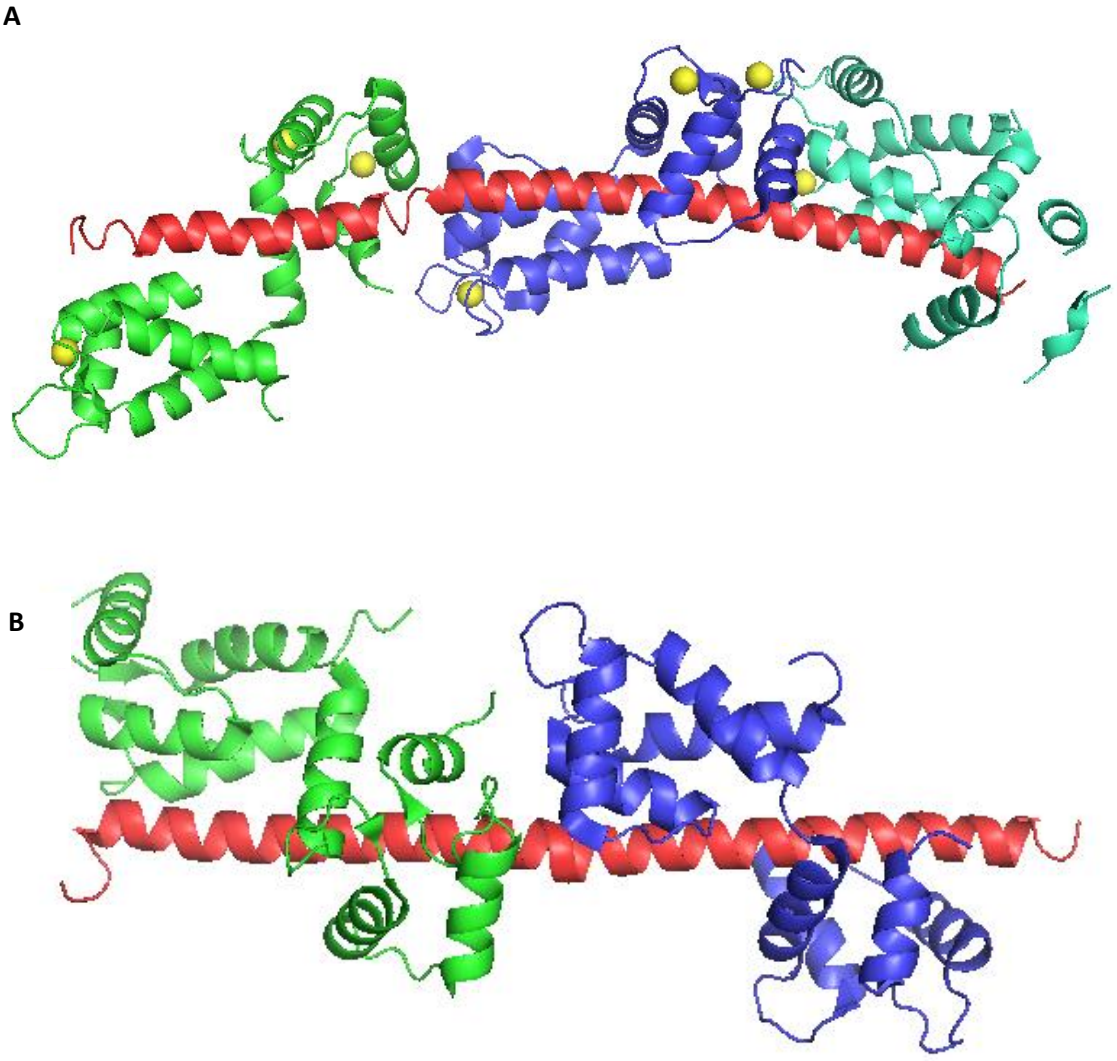


Figure 10. Ribbon models for the Cdc31-Sfi complex in the high and low Ca^{2+} concentrations. (a) Ribbon model for the Cdc31-Sfi1 complex containing three Sfi1 repeats and three centrins at high Ca^{2+} concentration. Sfi1 is the red helix interacting with three calcium-bound centrins (green, blue and cyan helices) at three centrin binding sites. The calcium ions are represented by the yellow spheres (b) Ribbon model for the Cdc31-Sfi1 complex containing two Sfi1 repeats (red helix) and two centrins (green and blue helices) at low Ca^{2+} concentration. The structures were generated from the Protein Data Bank entries PDB ID: **2DOQ** and **2GV5**, respectively and using PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

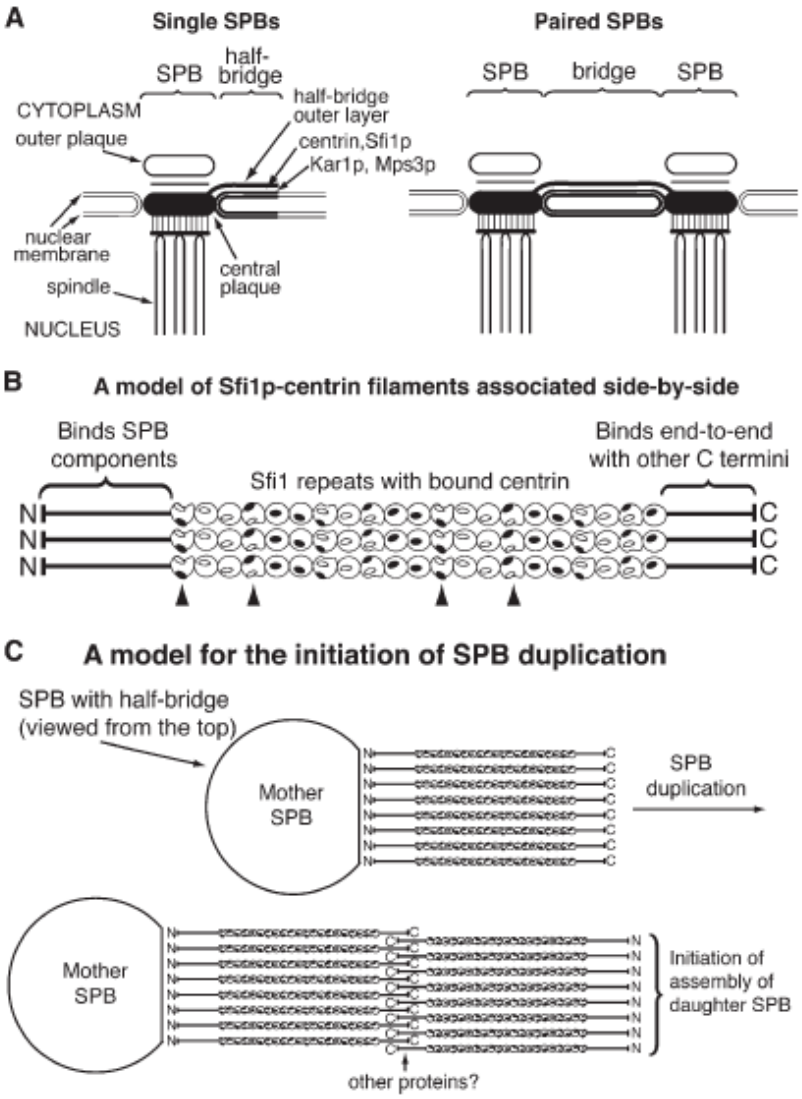


Figure 11. Single and paired SBPs and a model for the initiation of SPB duplication. (a) Single and paired SPBs showing the location of the half-bridge and its components. (b) A model for Sfi1-centrin fibers. Centrin have a 65 twist along Sfi1 and are shown as semi-spheres with filled and unfilled patches, where filled patches interact with unfilled patches. (c) Model for the initial step in SPB duplication. (Adapted from Li et al. 2006 *[15]*)

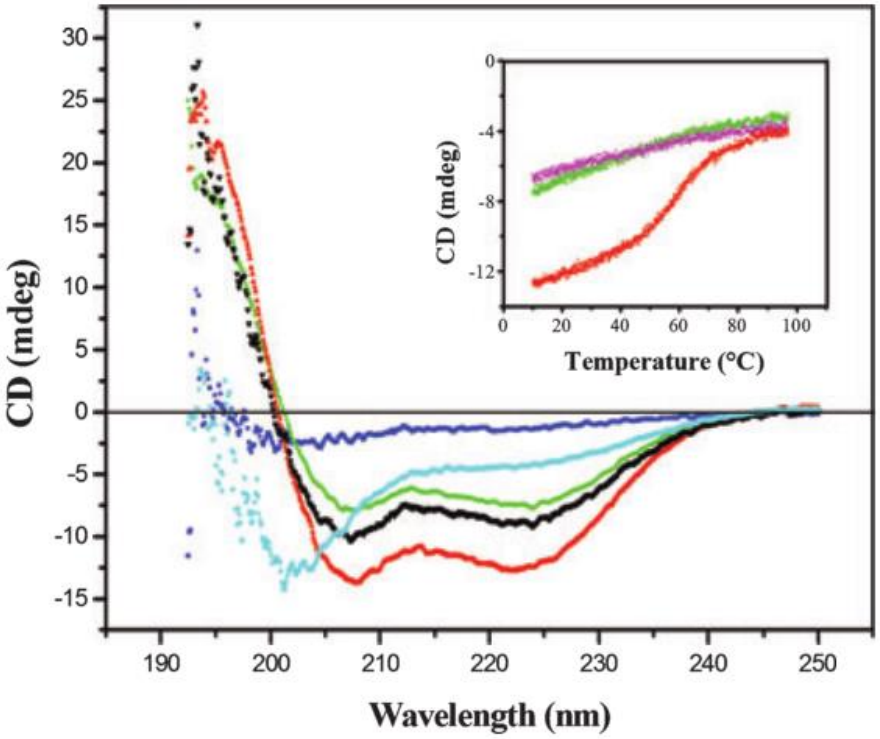


Figure 12. CD spectra of *Hs*Sfip$_{17}$ at 10 μM (blue) and 50 μM (cyan); C-*Hs*cen2 (green); C-*Hs*cen2 in complex with *Hs*Sfi1p$_{17}$ (red) at 10 μM. The spectral sum of the isolated components at the same concentration is shown in black. The inset shows the thermal denaturation curves, monitored by the ellipticity at 222 nm for the C-*Hs*cen2-*Hs*Sfi1p$_{17}$ complex in the presence and absence of Ca^{+2} (red and green, respectively) and C-*Hs*cen2 bound to Ca^{2+} (magenta). Adapted from Martínez-Sanz et al. 2006 *[48]*

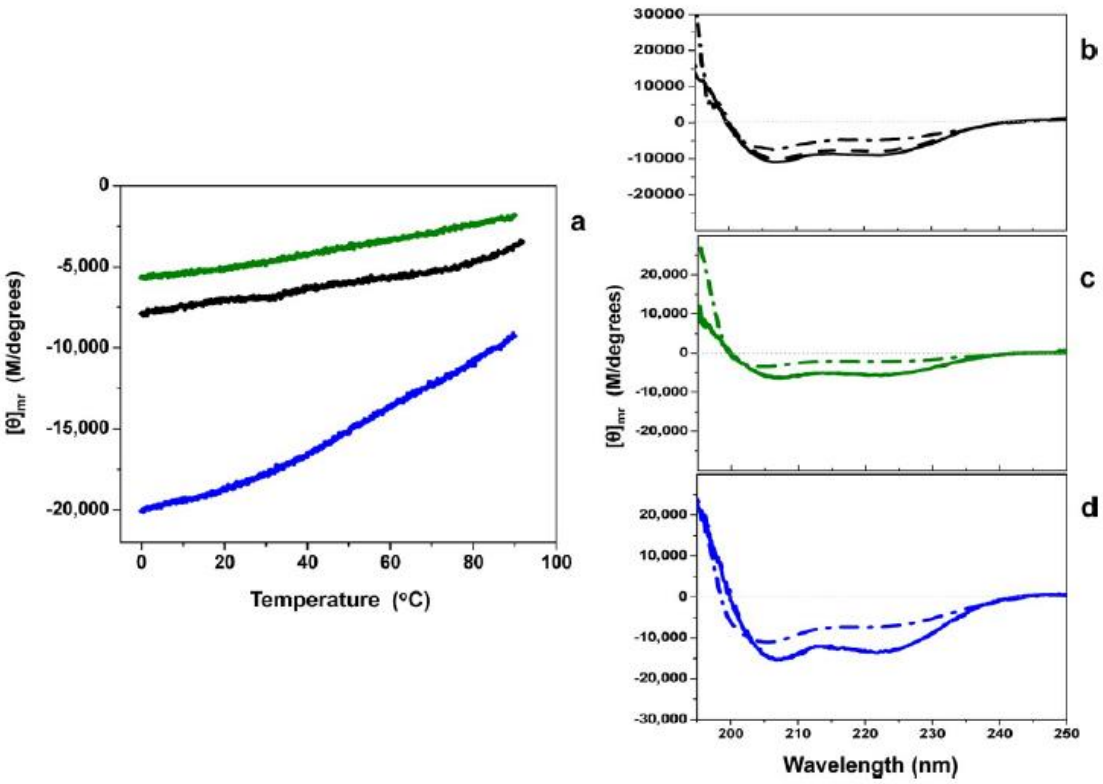


Figure 13. Relative thermal stability of human centrin by CD. (a) Thermal dependence of the helical regions among centrins *Hs*cen1 (black), *Hs*cen2 (green) and *Hs*cen3 (blue). (b-d) Overlay spectra of far-UV CD at 5 (—), 25(— — —), and 85°C (— · —) within the spectral region of 195-250 nm for (b) *Hs*cen1, (c) *Hs*cen2, and (d) *Hs*cen3. Data adapted from Pastrana-Ríos, 2013 *[33]*.

D. Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique that measures the difference between right and left-circularly polarized light. When either of them are absorbed at different magnitudes the resulting radiation would be said to possess elliptical polarization. A CD signal is observed when a chromophore is (a) optically active (chiral) because of its structure, (b) it is covalently linked to a chiral center in the molecule or (c) it is placed in a regular folded environment by the 3-dimensional structure adapted by the molecule. When the left-circularly polarized light is absorbed at a greater extent than the right-circularly polarized light the resulting spectrum is positive, therefore when the right-circularly polarized light is absorbed at a greater extent that the left-circularly polarized light the resulting spectrum is negative **(Figure 14)**. The resulting CD spectrum is obtained in terms of ellipticity (Φ) in degrees as function of wavenumber (λ), but because the CD signal from biological samples is intrinsically very small, ellipticity values are reported in terms of milli degrees (mdeg). [49] CD data is normalized by scaling to molar concentration the whole molecule. For proteins and peptides, the number of amino acids, the molecular mass, concentration and path-length are considered. For proteins, the mean residue molar ellipticity is given by:

$$[\Phi]_{MRW,\lambda} = \frac{MRW * \Phi_{\lambda}}{10 * c * l} \quad (1)$$

where MRW =M/(N-1), where M is the molecular mass of the protein (in Da), and N is the number of amino acids in the chain. Φ_{λ} is the observed ellipticity (mdeg) at wavelength λ, *c* is the concentration (g/mL) and *l* is the path-length (cm).

A CD instrument (known as a spectropolarimeter) consists of a radiation source, typically a xenon lamp, to generate plane polarized radiation. The radiation source passes through a series of prisms, mirrors and slits where it is focused by a lens and passed through a filter to the modulator. The circularly polarized components are passed through the shutter to the sample compartment before detection by the photomultiplier [49] **(Figure 15)**.

CD is a useful technique for the analysis of biological samples due to the fact that the vast majority of biological molecules are chiral. For instance, 19 out of the 20 common amino acids that make up most proteins are chiral, as well as the structure of higher molecules such as RNA and DNA. In proteins, the peptide bond is optically active, absorbing in the Far UV from 240 to 180 nm with different ellipticity values based on the conformation of the protein's secondary structure. Most proteins are levo-rotatory, therefore their secondary structures present negative CD spectra [50]. **Figure 16** shows typical spectra for proteins with an α-helical, β-sheet and disordered structure. α-helical proteins display characteristic spectral features between 203 and 240 nm including two minima at approximately 209 and 222 nm. Proteins with a β-sheet secondary structure show a single minimum between 210 and 225 nm. When proteins have both α-helical and β-sheet components the CD spectra can be considered as mixture, in which the stronger or more dominant structural component dominates the spectrum. Finally, for the random coil a single minimum at or near 197-202 nm is typically observed. [51]

Apart from being a useful technique to determine a protein's secondary structure, CD can also be used to study its thermal stability and conformational changes caused by the binding of ligands. Most proteins are sensitive to thermal perturbations and their

structures suffer conformational changes upon unfolding as a result of the perturbation. Consequently, the resulting spectra of the unfolded protein will differ from the one of the folded protein. For that reason, to study a protein's thermal stability, the CD signal measured at a fixed wavelength as function of temperature gives a measure of thermal stability. A protein's structural changes are essential when studying the mechanisms of action and regulation of biological activity. Proteins have a considerable array of ligands they interact with which in turn can alter the protein's conformation. These ligands range from ions to peptides, DNA, RNA or simply other proteins. Therefore, CD spectra can be used to determine the conformational changes that occur when a protein binds a ligand.

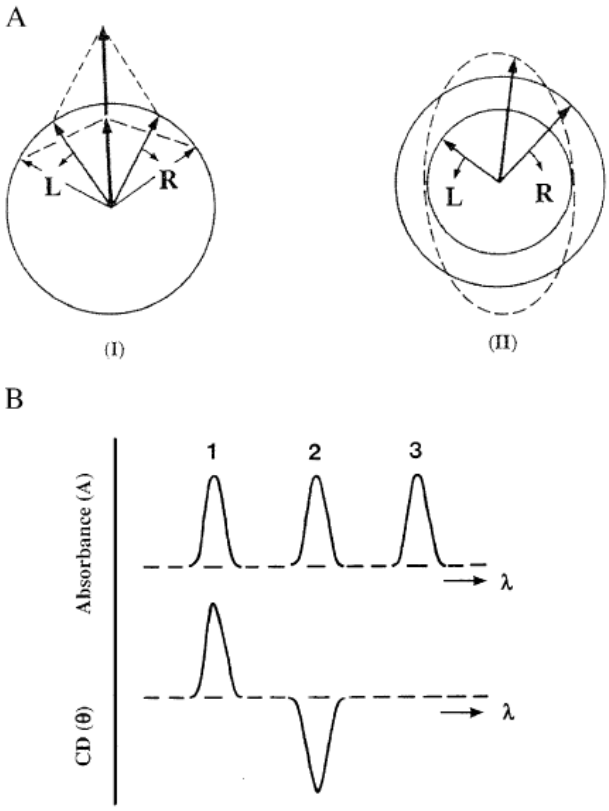


Figure 14. Origin of the CD effect. (a) Left and right circularly polarized light. (I) The left (L) and right (R) components have the same magnitude and when combined they generate plane polarized radiation. (II) The two components are of different magnitudes and when combined the generated resultant is elliptically polarized. (b) Band 1 has a positive CD spectrum because L absorbed more than R. Band 2 has a negative CD spectrum because R is absorbed more than L. Band 3 is an achiral component. (Adapted from Price et al. 2005 *[49]*)

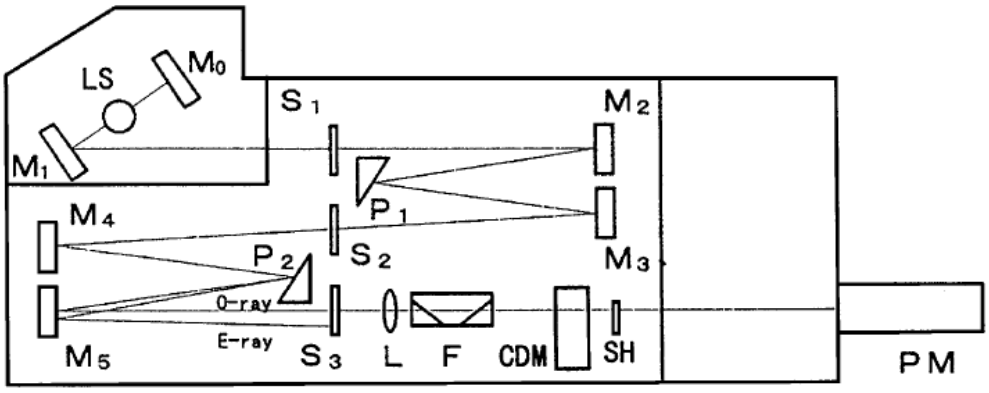


Figure 15. Block diagram of a spectropolarimeter (Jasco J-810 model). Circularly polarized light is generated by the passage of a radiation or light source (LS) through two prisms (P_1 and P_2) and a series of mirrors (M_0 to M_5) and slits (S_1 to S_3). The ordinary ray (O) is focused by a lens (L) and passed through a filter (F) to the modulator (CDM). The circularly polarized light is the passed though the shutter (SH) to the sample compartment, before detection by the photomultiplier (PM). (E represents extra-ordinary ray) (Adapted from Price et al. 2005 *[49]*)

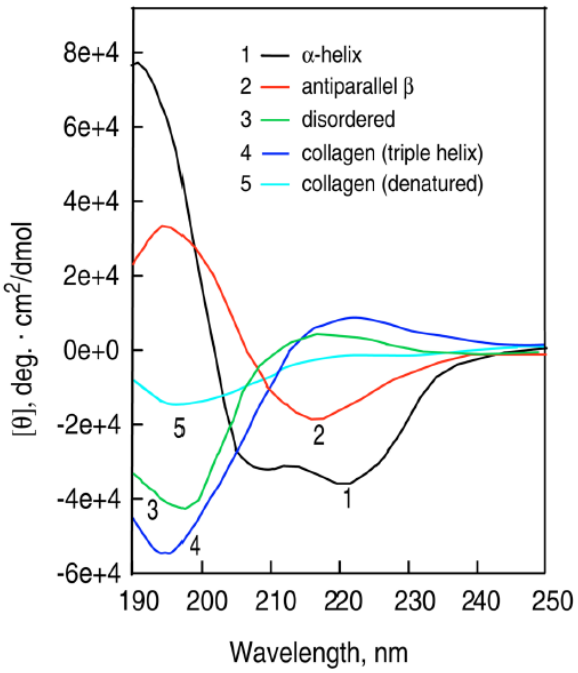


Figure 16. CD spectra of secondary structures. The black line corresponds to α-helix, the red line corresponds to a β-sheet, the green line corresponds to a disordered structure, the blue line corresponds to a triple helix and the cyan line corresponds to a denatured protein sample. (Adapted from Greenfield 2006 *[51]*)

E. Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a thermodynamic technique used to determine the binding parameters between a protein and its target. In biological terms the target could be a substrate, metal ion, peptide, protein, coenzyme, DNA or RNA or any kind of molecule thought to non-covalently interact at a specific site of a protein [52]. ITC works without the need of probes or target modification by directly measuring the heat released or absorbed as the two molecules interact at a constant temperature. By measuring the heat transfer upon binding, the accurate determination of stoichiometry (n), change in enthalpy (ΔH_B), change in Gibbs free energy (ΔG_B), change in entropy (ΔS_B) and binding affinity constant (K_a) can be carried out. At a constant temperature, the change in Gibbs free energy is given by:

$$\Delta G_B = -RT \ln(K_a) = \Delta H_B - T\Delta S_B \quad (2)$$

where R is the gas constant. The enthalpy change is the experimentally measured variable and it is a global property of the system. The enthalpy change represents the *total* heat released or absorbed in the system by addition of the ligand, it accounts for contributions arising from non-specific effects. In the case of protein-protein interactions, it includes contributions from the formation of hydrogen bonds, salt-bridges and hydrophobic interactions and the desolvation of polar groups [52]. Favorable enthalpy indicates that the interactions between the ligand and protein are strong enough to compensate for the interactions with surrounding water molecules. Therefore, unfavorable enthalpy indicates that polar groups are desolvated without being able to form weak interactions with the target. Binding entropy includes contributions from changes in solvation and conformational entropy. Solvation entropy is favorable and

results from the release of water molecules as the targets interact, hence, favorable solvation entropy is the predominant force associated with the binding of hydrophobic groups. The conformational entropy change is almost always unfavorable, because as the binding occurs it involves the loss of conformational degrees of freedom of both molecules. The k_a value indicates the affinity of the interaction and the change in Gibbs free energy provides information about the stability of the complex. [53]

A schematic representation of a typical ITC instrument is shown in **Figure 17**. The instrument consists of two identical cells surrounded by an adiabatic jacket. The receptor molecule is placed in one of the cells (the sample cell) to which the ligand is added via an automatic titrator, meanwhile the reference cell contains the buffer or water. Sensitive thermophile/thermocouple circuits detect the temperature differences between the two cells and the jacket as heat is absorbed or released as the ligand is titrated. Heaters on both cells and the jacket are activated when necessary to maintain identical temperatures between reference and sample cells.

The variable measured in an ITC experiment is the time-dependent input of power needed to maintain the desired experimental temperature in the sample cell equal to the temperature in the reference cell. During the titration of the ligand, if heat is released (exothermic), the temperature in the sample cell increases and the feedback power will decrease to maintain equal temperature between the cells. Thus, for endothermic reactions the opposite will occur, the feedback power will increase to the sample cell to maintain the desired temperature. After each titration, the signal (feedback power) returns to zero as the reaction reaches equilibrium. The molar ratio gradually increases as the ligand is titrated and the receptor molecule becomes more saturated and less binding

occurs. Finally, when all the binding sites in the receptor molecule are completely bound there is excess ligand in the sample cell and the reaction becomes saturated. Hence, the measured heat is proportional to the amount of binding between the molecules, therefore it is of upmost importance to accurately determine the initial concentration of the ligand and receptor molecule. Experimental data is plotted as raw power data, which represents each injection as a function of the feedback power and time. The power data is then integrated to obtain an isotherm from which the thermodynamic parameters are obtained (**Figure 18**) [54].

In 2012, Diaz-Casas studied the interaction of full-length human centrin and *Hs*Sfi1p$_{21}$ by ITC. **Table 1** summarizes part of the obtained thermodynamic data between human centrin isoforms and *Hs*Sfi1p$_{21}$ at 30°C. The complex formation with all three isoforms presented favorable enthalpy.

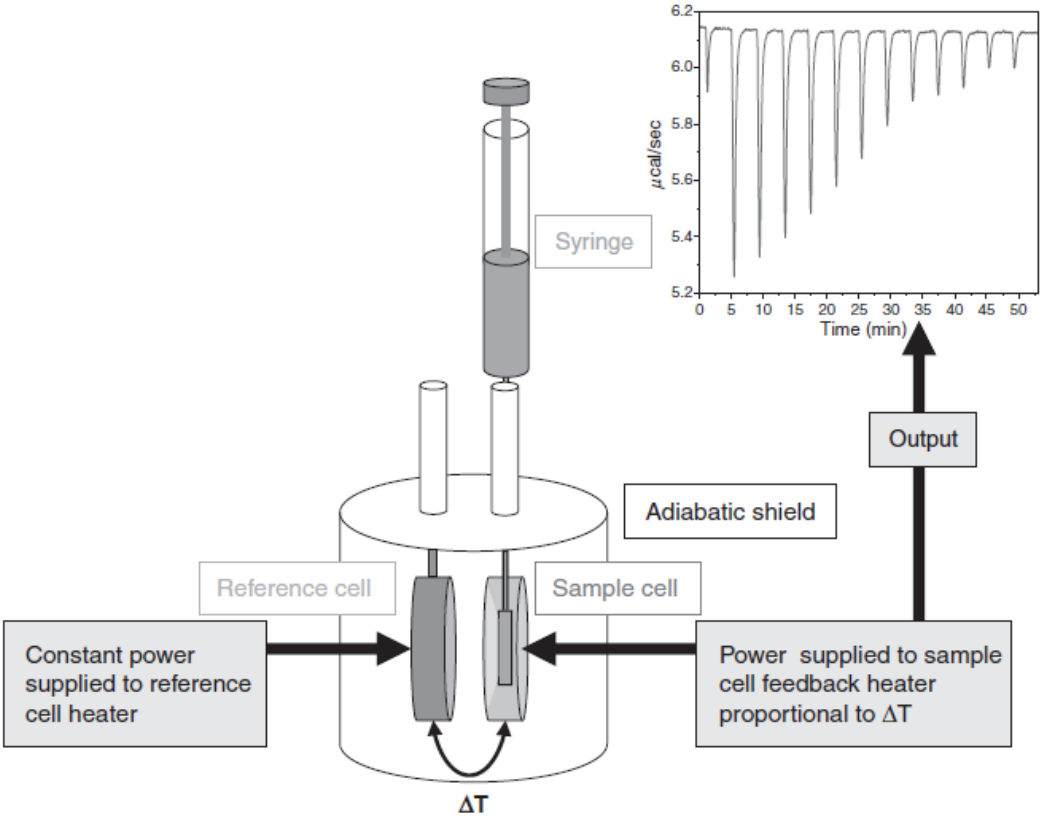


Figure 17. Representation of an ITC instrument. The reference and sample cell are contained within an adiabatic shield. As the ligand is titrated onto the sample cell, power is supplied to maintain the temperature equal to the reference cell temperature. The feedback power is measured and plotted as a function of time to obtain the raw power data. Adapted from Freyer and Lewis 2008 *[52]*.

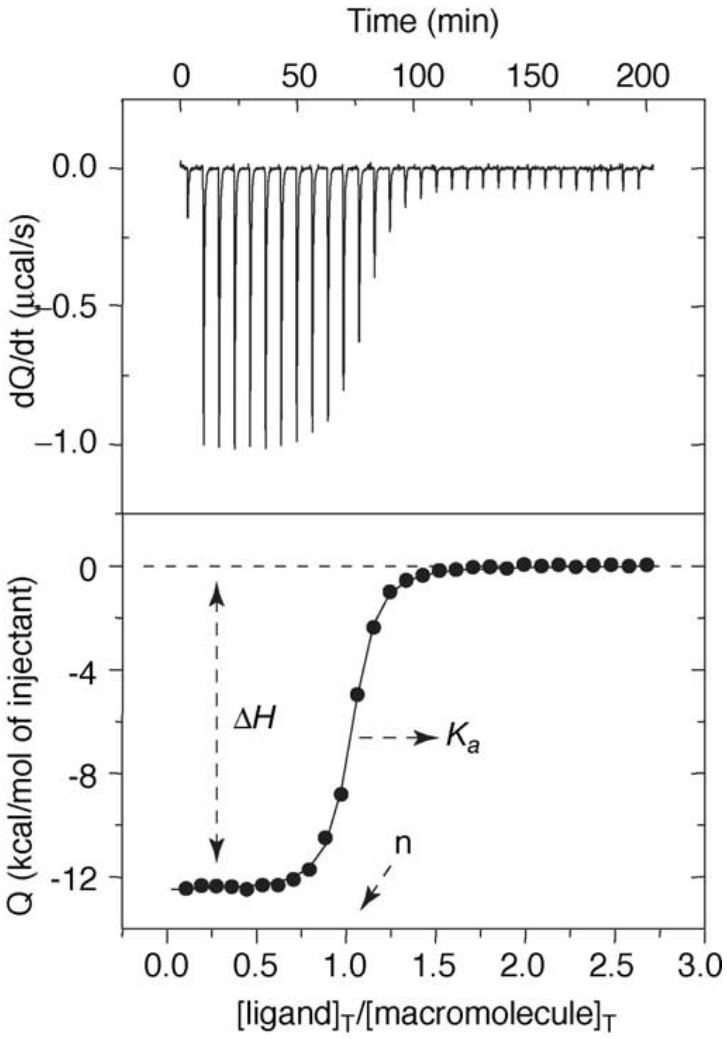


Figure 18. Sample ITC raw data and integrated power data for an exothermic interaction. The upper panel represents the raw data. Each peak corresponds to the heat liberated by the titration of an aliquot of the ligand into the sample cell containing the receptor molecule. As the binding sites become more saturated the peak height become less prominent. The lower panel represents the integrated data. Data was normalized and integrated using the Origin software provided by the ITC manufacturer (in this case MicroCal) using a non-linear regression and a one-site binding model. Adapted from Freire 2004 *[53]*.

Table 1. Summary of thermodynamic parameters of the interaction between wild type human centrin isoforms and *Hs*Sfi1p$_{21}$ at 30°C. Table adapted from Diaz-Casas 2012 *[41]*.

Protein	Peptide	K_a (10^7 M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
***Hs*cen1(WT)**	*Hs*Sfi1p$_{21}$	1.6	-10.0	-26.7	16.7
***Hs*cen2(WT)**	*Hs*Sfi1p$_{21}$	0.9	-9.6	-18.2	8.5
***Hs*cen3(WT)**	*Hs*Sfi1p$_{21}$	5.1	-10.7	-15.8	5.1

Materials and Methods

A. Protein and Peptide Preparation

Centrin variants samples were expressed, isolated and purified following our laboratory's established protocol (**Figures 19-20**) [33]. The variant's identity was verified by sequencing analyses performed by Dr. Michael Bern from Tufts University. These results are presented to corroborate the expression and purification of the variants (**Figures 21-23**). Bio-Synthesis (Lewisville, Texas USA) was subcontracted for the custom synthesis of the *Hs*Sfi1p$_{21}$ peptide (A_1QQRLQLERAVQHHHRQLLLWGLARWKTHHLLQ_{33}). The purity and identity analyses carried out by Bio-Synthesis confirmed that the peptide's purity was >95%.

Calibration curves were carried out using a Jasco (Tokyo, Japan) UV/Vis Spectrophotometer Model V-560 and a 1 cm quartz cuvette. The absorption coefficient (ε) at 280 nm for all the centrin 1 and centrin 2 variants is 1,490 M^{-1} cm^{-1} and 5,550 $M^{-1}cm^{-1}$ for *Hs*Sfi1p$_{21}$.

Figure 24 and **Figure 25** represent the UV scan and calibration curve for the centrin variants, respectively. Likewise, **Figure 26** and **Figure 27** represent the UV scan and calibration curve for *Hs*Sfi1p$_{21}$. UV scan data was baseline corrected from 300-320 nm.

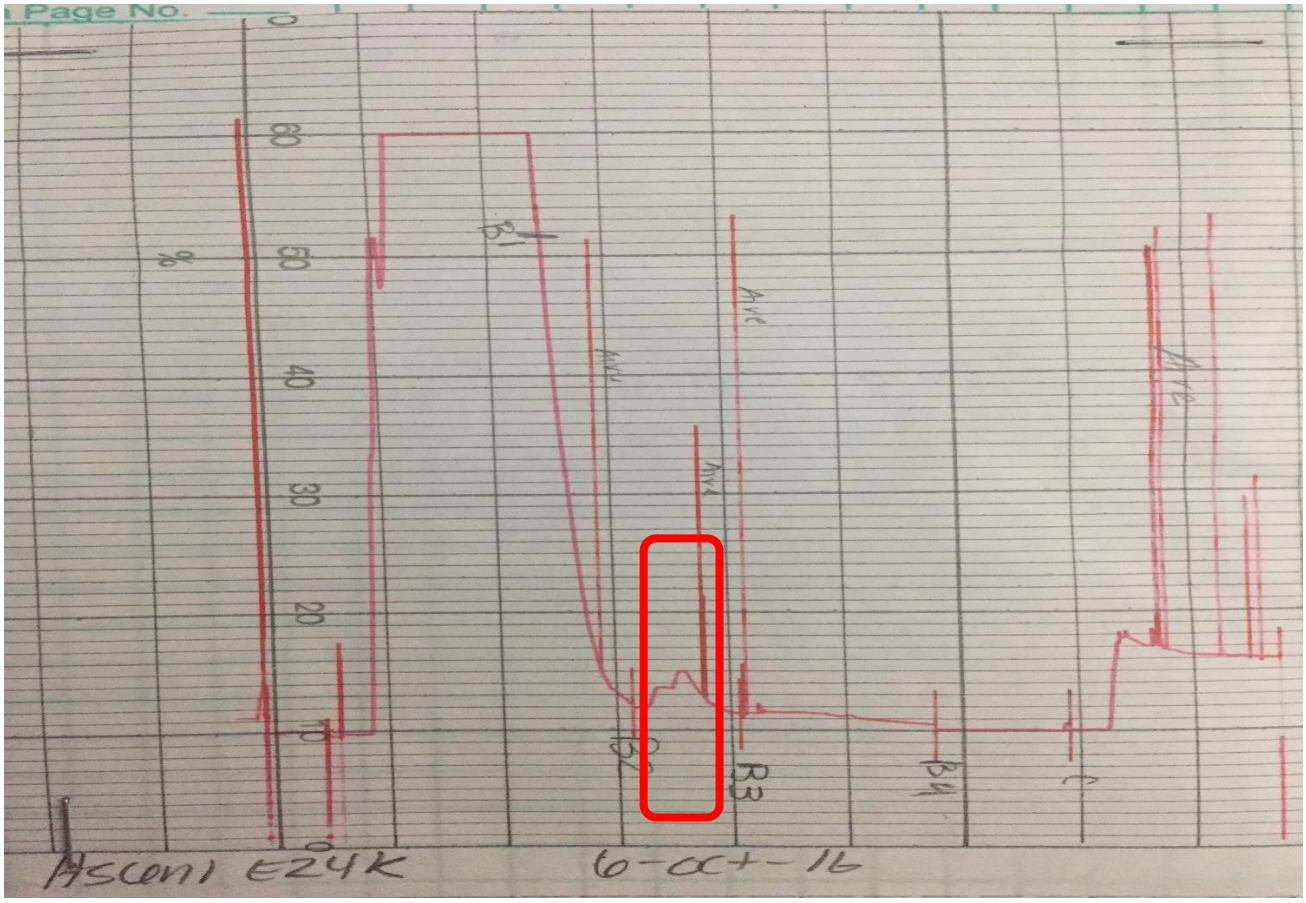


Figure 19. Affinity chromatogram for *Hs*cen1(E24K). Three significant peaks were observed. The first peak corresponds to the elution of *E. coli*'s native proteins in Buffer A, the second peak is outlined in the red box and corresponds to centrin's elution in Buffer B, and the third peak corresponds to proteins eluted in Buffer C. To guarantee clean separation the three buffers were added 4x the column volume.

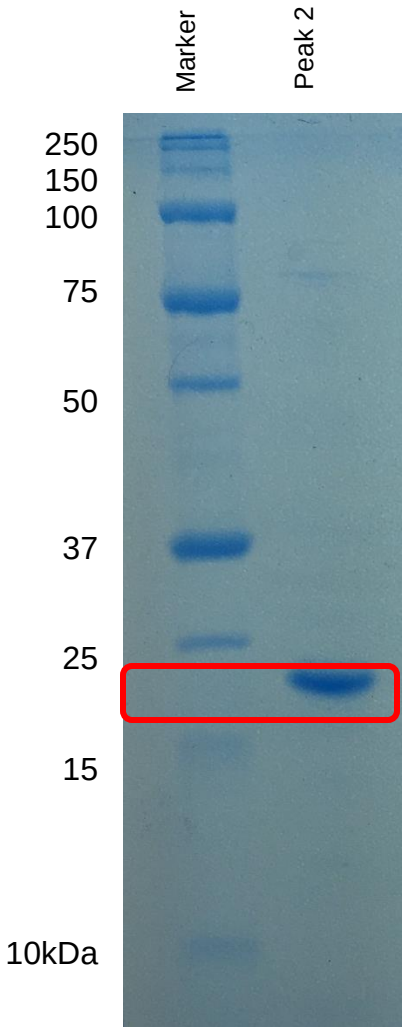
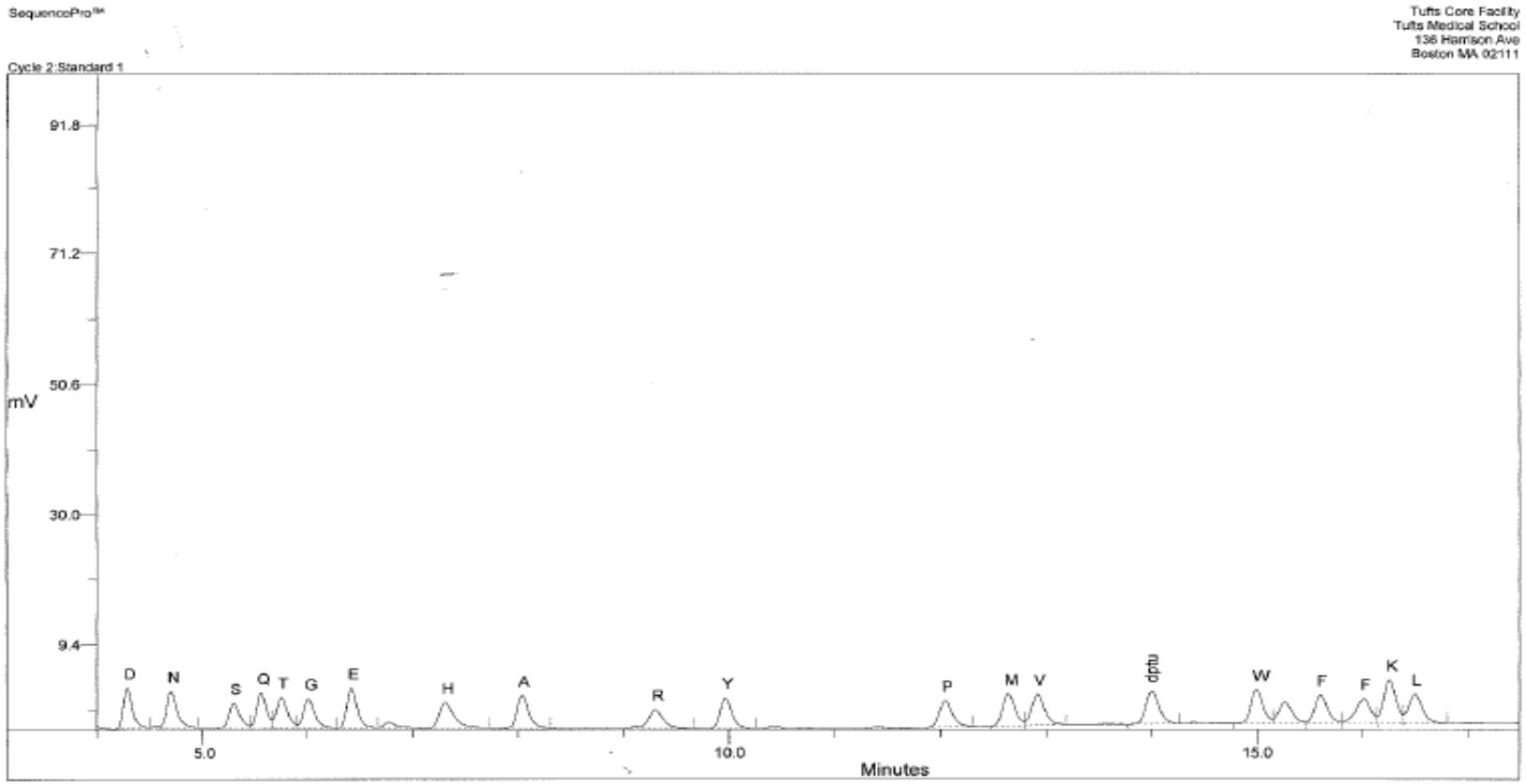


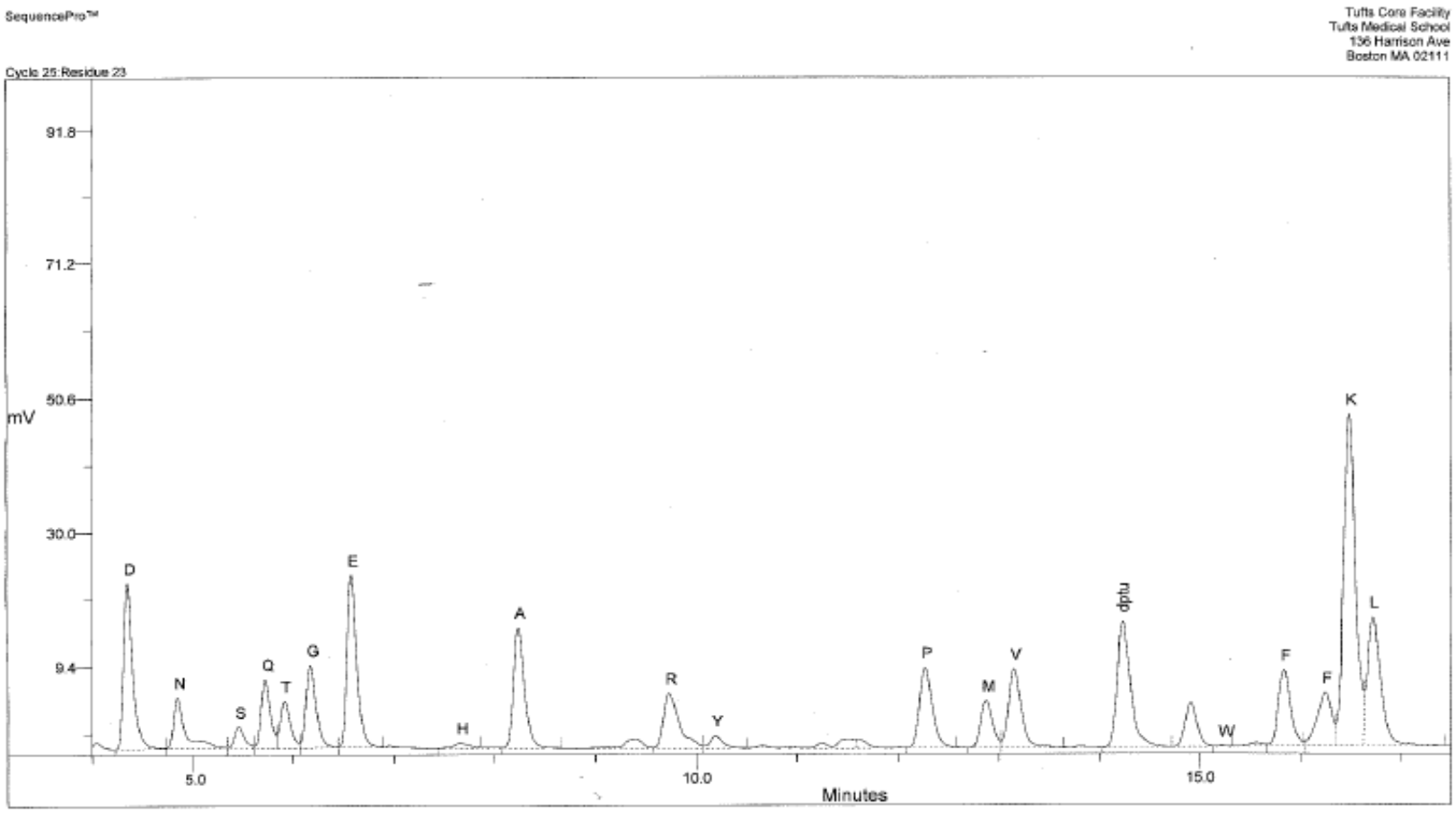
Figure 20. SDS-PAGE at 15% separation of the second peak obtained in the affinity chromatogram. On the first lane is the molecular marker and in the second lane is a 10 µL aliquot of the second peak of the affinity column. Outlined in the red box is centrin's band at approximately 20kDa. The gel was stained with Coomasie Blue G50.



PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
D	4.29	4.29	6.362	10.000
N	4.70	4.70	5.832	10.000
S	5.30	5.30	3.768	10.000
Q	5.56	5.56	5.430	10.000
T	5.75	5.75	4.761	10.000
G	6.01	6.01	4.524	10.000
E	6.42	6.42	6.229	10.000
H	7.31	7.31	3.925	10.000
A	8.04	8.04	5.073	10.000
R	9.30	9.30	2.878	10.000

PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
Y	9.97	9.97	4.787	10.000
P	12.05	12.05	4.128	10.000
M	12.64	12.64	5.046	10.000
V	12.92	12.92	4.879	10.000
dptu	14.00	14.00	5.185	10.000
W	15.00	15.00	5.338	10.000
F	15.60	15.60	4.500	10.000
F	16.02	15.60	3.868	10.000
K	16.26	16.26	6.845	10.000
L	16.50	16.50	4.585	10.000

Figure 21. Sequencing calibration curve. From Tufts Core Facility.



PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
D	4.34	4.29	25.279	39.735
N	4.84	4.70	7.806	13.861
S	5.45	5.30	3.199	8.489
Q	5.72	5.56	10.520	19.374
T	5.91	5.75	7.133	14.983
G	6.17	6.01	12.672	28.010
E	6.58	6.42	26.429	42.430
H	7.66	7.31	0.746	1.902
A	8.23	8.04	16.401	36.274
R	9.73	9.30	8.381	29.118

PEAK ID	R.TIME (mins)	C.TIME (mins)	HEIGHT (mV)	PMOL HT
Y	10.19	9.97	1.879	3.925
P	12.27	12.05	12.382	28.997
M	12.87	12.64	7.202	14.273
V	13.15	12.92	12.155	24.912
dptu	14.23	14.00	19.284	37.194
W	15.21	15.00	0.104	0.196
F	15.84	15.60	11.755	26.120
F	16.25	15.60	8.208	18.236
K	16.46	16.26	51.018	74.533
L	16.73	16.50	19.758	43.092

Figure 22. Sequencing analysis of *Hs*cen1(E24K). From Tufts Core Facility.

PROTEIN	**1**									**10**	**11**									**20**	**21**			**24**
Hs cen1(WT)	M	A	S	G	F	K	K	P	S	A	A	S	T	G	Q	K	R	K	V	A	P	K	P	E
Hs cen1(E24K)	-	A	S	G	F	K	K	P	S	A	A	S	T	G	Q	K	R	K	V	A	P	K	P	**K**

Figure 23. Sequencing results for *Hs*cen1(E24K). Top row *Hs*cen1(WT) (accession number: Q12798). Results were obtained from the Tufts Core Facility.

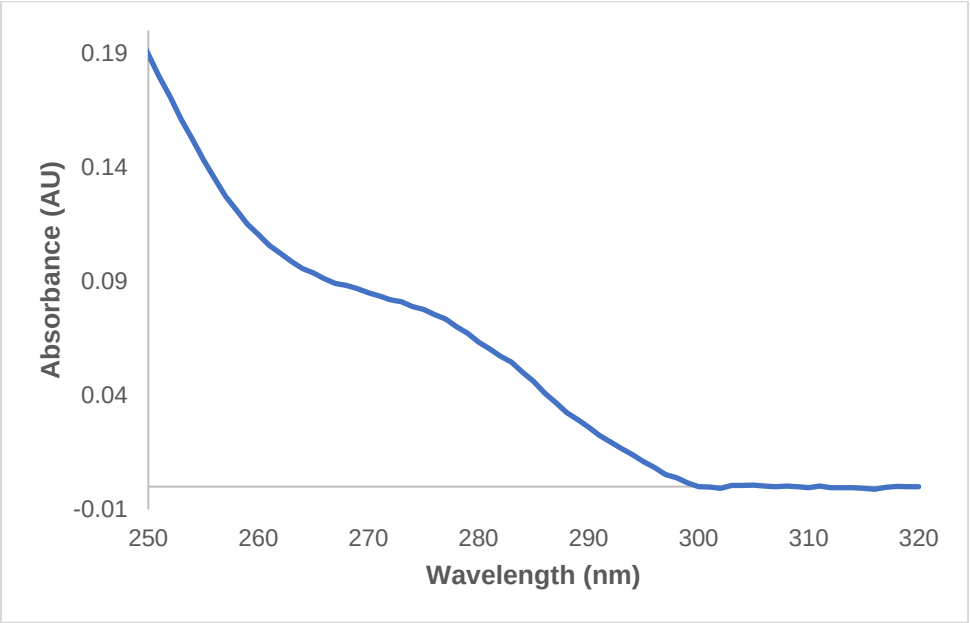


Figure 24. UV scan of the *Hs*cen1(E24K) variant. The other variants presented a very similar spectrum. Data was baseline corrected from 300-320 nm

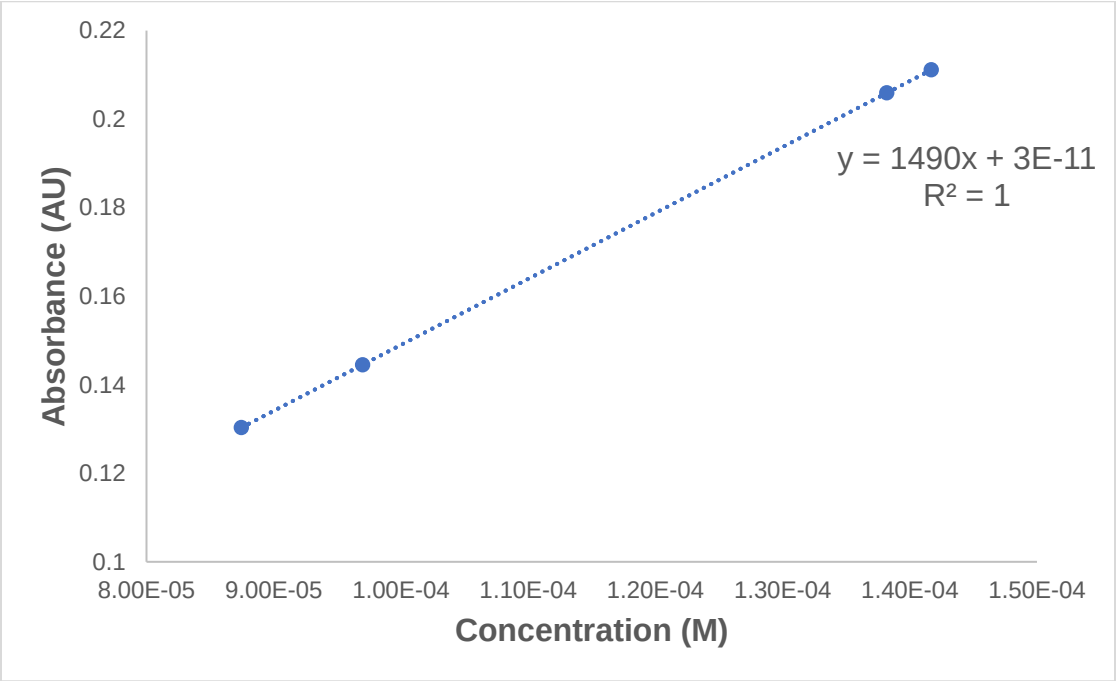


Figure 25. Calibration curve for the *Hs*cen1 and *Hs*cen2 variants using a 1 cm path-length quartz cuvette.

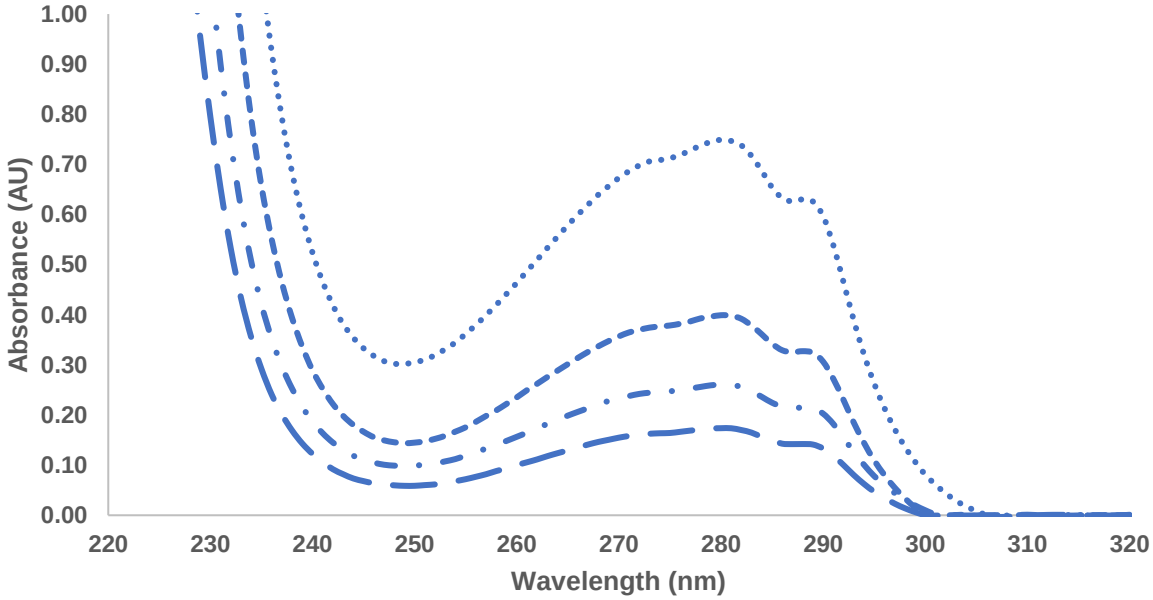


Figure 26. UV scan of *Hs*Sfi1p$_{21}$ from 320-220 nm. Data was baseline corrected from 300-320 nm.

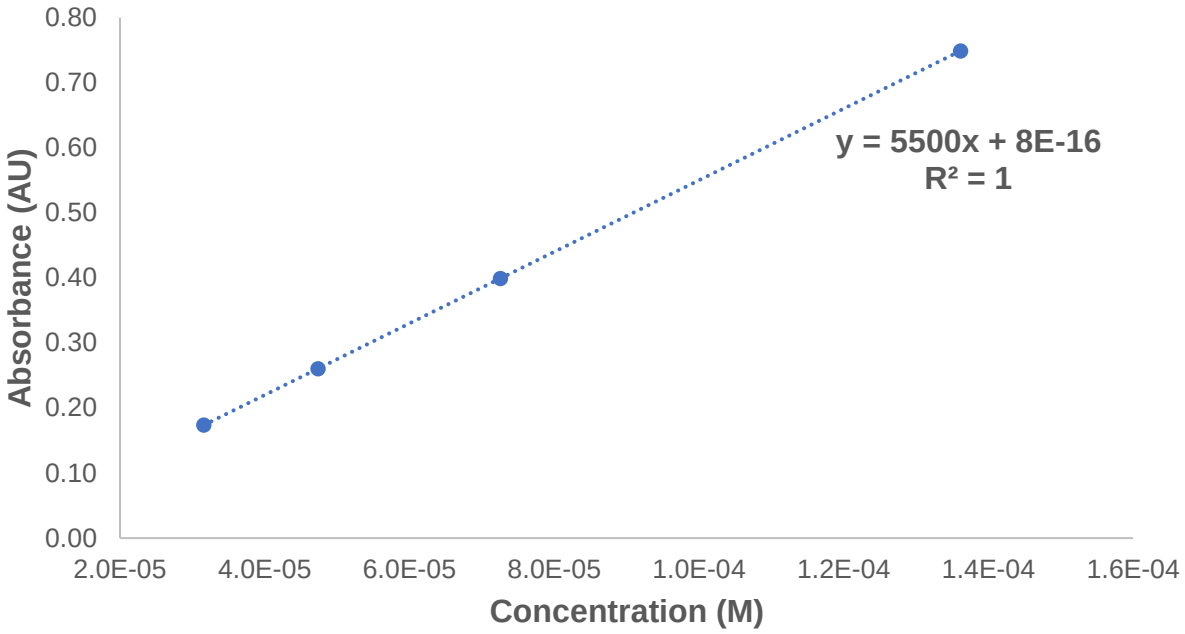


Figure 27. Calibration curve for *Hs*Sfi1p$_{21}$ using a 1 cm path-length quartz cuvette.

B. Circular Dichroism

CD analyses were carried out using a Jasco (Tokyo, Japan) spectropolarimeter model J-810 equipped with a Peltier temperature controller accessory. CD experiments were carried out to monitor secondary structure and its stability as a function of temperature. To determine secondary structure, ten scans within the spectral range of 250 to 190 nm were collected at a scan rate of 20 nm/min at 5°C, 25°C, 85°C and 95°C. To determine thermal stability, absorbance was recorded at 222 nm from 5°C to 95°C at a rate of 1°C/min. After the analysis, CD mdeg data was baseline corrected and converted to mean residue molar ellipticity ($[\Phi]_{MR}$) to allow proper comparison between the centrin 1 and centrin 2 variants.

The samples were dialyzed using a Spectra/Por (California, USA) Dialysis Membrane Biotech CE Tubing with a 500 Da molecular weight cutoff in the following buffer conditions: 8 mM HEPES, 2 mM $CaCl_2$, 2 mM $MgCl_2$ and 50 mM NaCl at a pH of 7.4. Protein concentration was 2.5 µM and it was measured using a Jasco (Tokyo, Japan) UV/Vis Spectrophotometer Model V-560.

C. Isothermal Titration Calorimetry

Titrations were carried out using a VP-ITC microcalorimeter (MicroCal LLC, Northampton, MA) at an experimental temperature of 25°C. The *Hs*cen1(E24K) variant and the *Hs*Sfi1p$_{21}$ peptide were dialyzed in the following buffer conditions: 50 mM HEPES, 150 mM NaCl, 4 mM $CaCl_2$ and 4 mM $MgCl_2$ at pH 7.4, using a Spectra/Por (California, USA) Dialysis Membrane Biotech CE Tubing with a 500 Da molecular weight cutoff. Previous the experiments samples were degassed at 35°C for 15 minutes. *Hs*cen1(E24K) concentration was 64 µM and *Hs*Sfi1p$_{21}$ concentration was 8 µM, and were measured using a Jasco (Tokyo, Japan) UV/Vis Spectrophotometer Model V-560.

Results and Discussions

A. *Hs*Sfi1 CBS sequence analysis and implications

Homo sapiens Sfi1 (*Hs*Sfi1, **Figure 28**) is a protein of 1242 amino acids and a molecular weight of 148 kDa. *Hs*Sfi1 has approximately 23 tandem CBS that contain the consensus sequence $AX_7LLX_3F/LX_2WK/R$, where X represents any amino acid. After generating a sequence alignment of the 23 CBSs it is evident that the consensus sequence is not always conserved, with non-conserved substitutions occurring in some of the key hydrophobic amino acids of the centrin binding site **(Figure 29)**. The plot on **Figure 30** compares the substitution frequency and the extent of non-conserved substitutions in the consensus sequence. The tryptophan residue in the 17th position (W_{17}) is 100% conserved, with only one substitution being of conserved nature. The phenylalanine/leucine residue in the 14th position (F/L) has a high rate of substitution but in a conserved manner, yielding 100% of conservation. The leucine in the 10th position (L_{10}) has a high rate of substitutions, many of which are not conserved, yielding only 25% conservation. The leucine in the 9th position (L_9) also has a high rate of substitution, but many of which are conserved, yielding a higher rate of conservation at 72.73%. Finally, the alanine in the 1st position (A_1) only has three substitutions, of which two are not conserved, yielding a 33% of conservation.

Helical wheel plots were generated to analyze the distribution of the binding site hydrophobic triad. Plots for the 19th, 20th, 21st, 22nd and 23rd CBS are shown in **Figure 31**. Integrating the analysis from the sequence alignment and the generated helical wheels, it is evident that the 21st CBS comprises the most conserved consensus sequence, with the hydrophobic residues distributed uniformly through the helix, making it a stable and

readily soluble fragment. This made the *Hs*Sfi1p$_{21}$ peptide an ideal candidate for the study. The 19th CBS has a non-conserved substitution in one of the residues of the consensus sequence; L_{10} is substituted by a cysteine. CBS 20 has a non-conserved substitution in one of the residues of the consensus sequence, where the L_9 is substituted by a cysteine residue. CBS 22 has two non-conserved substitutions, the hydrophobic residues A_1 and L_{10} are both substituted by serine residues, increasing its polarity and possibly decreasing its binding affinity for centrin. In addition, of the 18 residues in CBS 23, four are tryptophan and ten other residues are also hydrophobic. This high frequency of hydrophobic residues makes this CBS very hydrophobic, which can present solubility and instability problems. Finally, CBS 23 has two substitutions in the consensus sequence, one of which is not conserved, L_9 is substituted by a glutamine residue.

A hydrophobicity plot was generated to further understand the potential for solubility of the CBS present in *Hs*Sfi1. From **Figure 32** it is evident that the CBSs contain hydrophobic segments. Upon closer inspection of *Hs*Sfi1p$_{21}$ it is clear that the peptide fragment (from residues 768 to 801) has a hydrophobic segment that corresponds to the hydrophobic triad of the consensus sequence (**Figure 33**).

MKNLLTEKCISSHNFHQKVIKQRMEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIRRELPSTS
HLVQYRGTHTCTRQGRLRELRIRCVARKFLYLWIRMTFGRVFPSKARFYYEQRLLRKVFEEWKEE
WWVFQHEWKLCVRADCHYRYYLYNLMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAW
KSWLIYVVVRRTKLQMQTTALEFRQRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQ
AWSQWREQLLYVQKEKQKVVSAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHV
TVLQIYFCDWQQAWERRESLYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEE
HHRHSQLYFCFRALKDNVTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLH
AAWDHYRIALLCKCIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQEN
VLSARATRFHRETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAAR
HQEQEWQTVACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRE
CLALRGAERQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRR
WKENTMARVDEAKKTFQASTHYRRTICSKVLVQWREAVSVQMYYRQQEDCAIWEAQKVLDRGC
LRTWFQRWWDCSRRSAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLL
AQRLSRTCFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQW
ALQAYQGQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGG
KPQPLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTAHSA
RKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPGALSSAPGP
KQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVPDPHLLLPGDFSATR
AGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRRWLELNREEPGPEDQEVEQQ
VQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC

Figure 28. *Hs*Sfi1 sequence, accession number A8K8P3. The *Hs*Sfip$_{21}$ fragment sequence is highlighted and underlined.

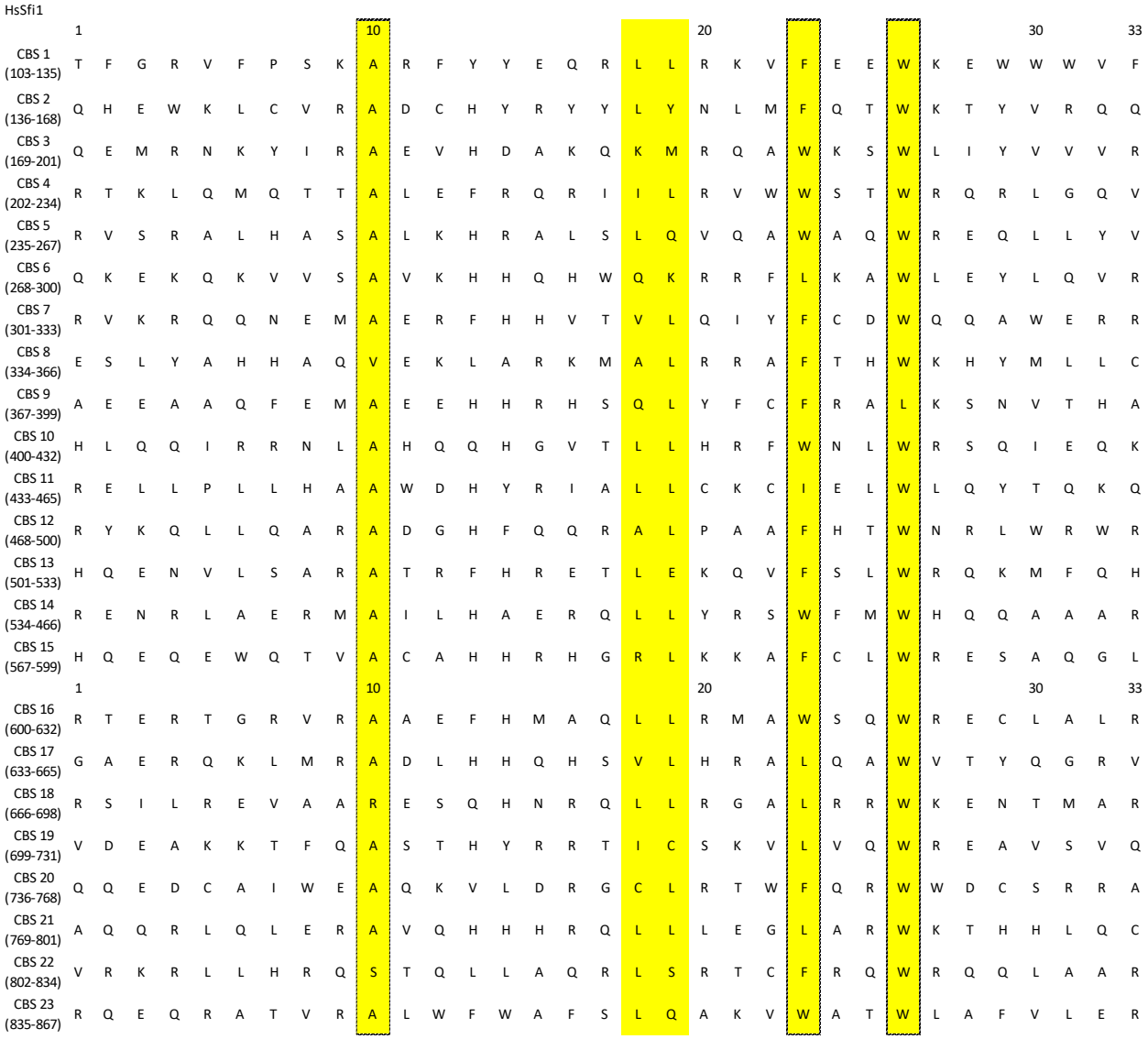


Figure 29. 23 tandem centrin binding sites (CBS) in full-length *Hs*Sfi1 protein. The consensus sequence AX_7LLX_3F/LX_2W is highlighted in yellow and the hydrophobic triad is outlined.

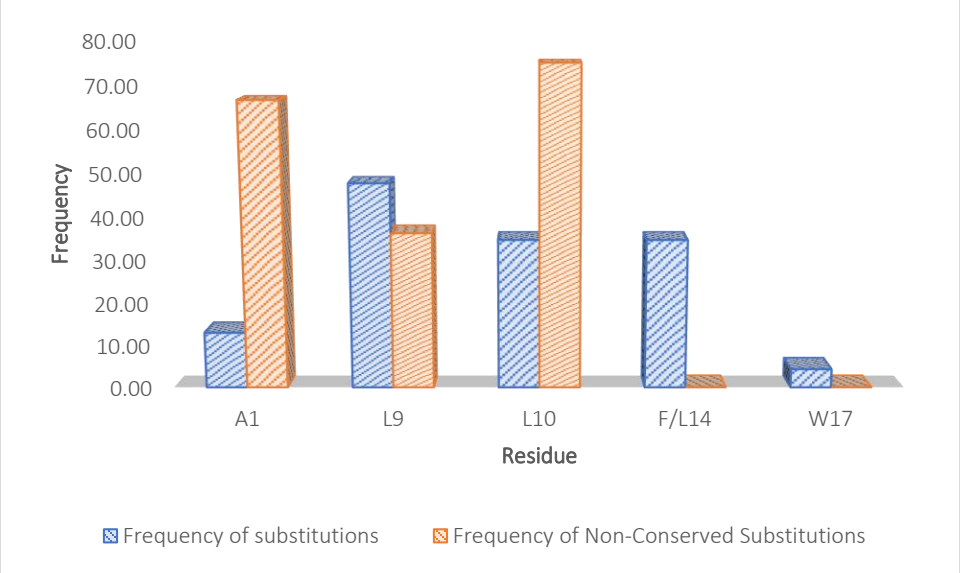


Figure 30. Frequency distribution of conserved and non-conserved substitutions of centrin binding site (CBS) hydrophobic residues in *Hs*Sfi1's consensus sequence.

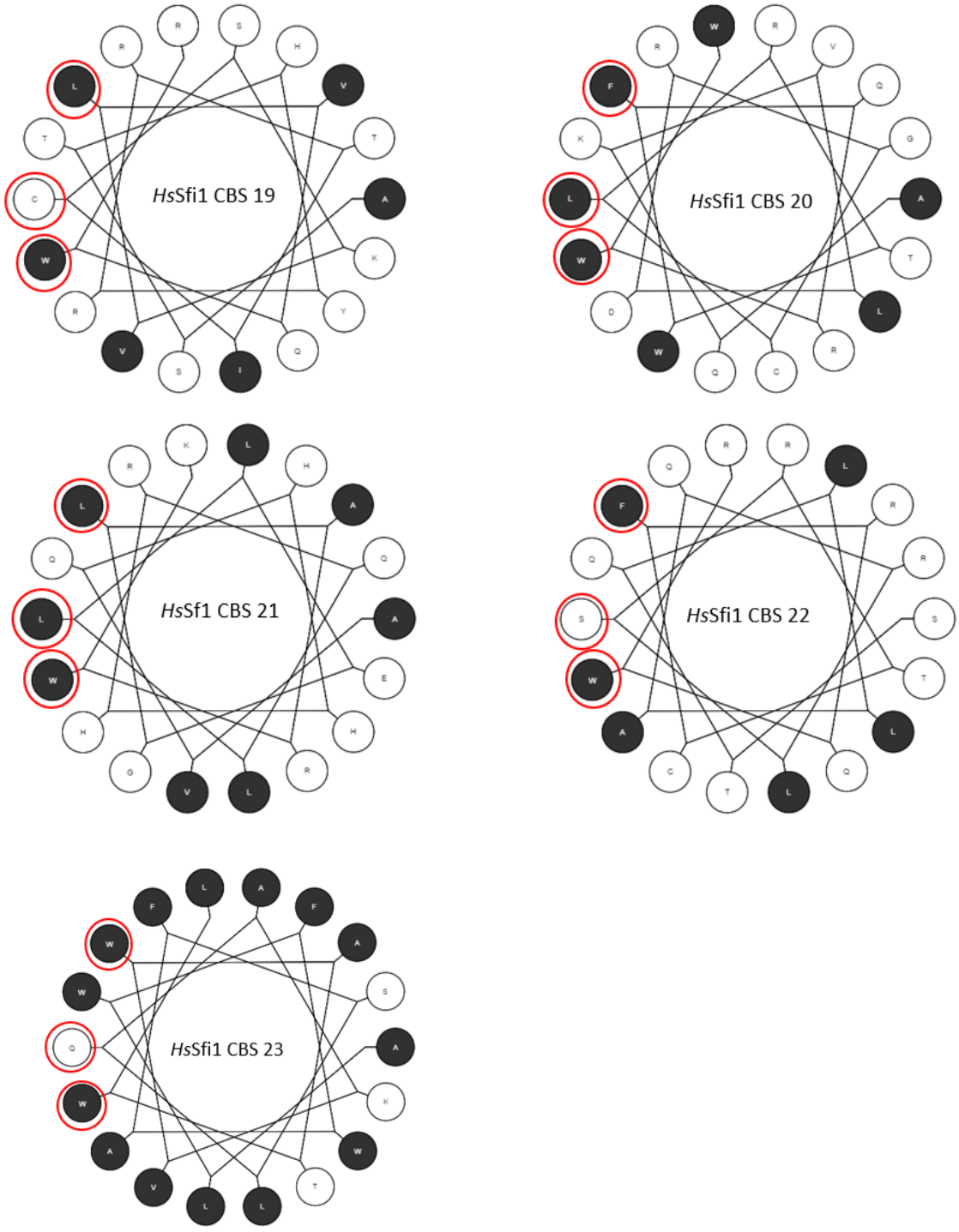


Figure 31. Helical plots of *Hs*Sfi1's 19th to 23rd centrin binding sites. Hydrophobic residues have a black background and the hydrophobic triad contains a red concentric circle to show its distribution in the helix. Plots were generated using a software provided by Kael Fischer.

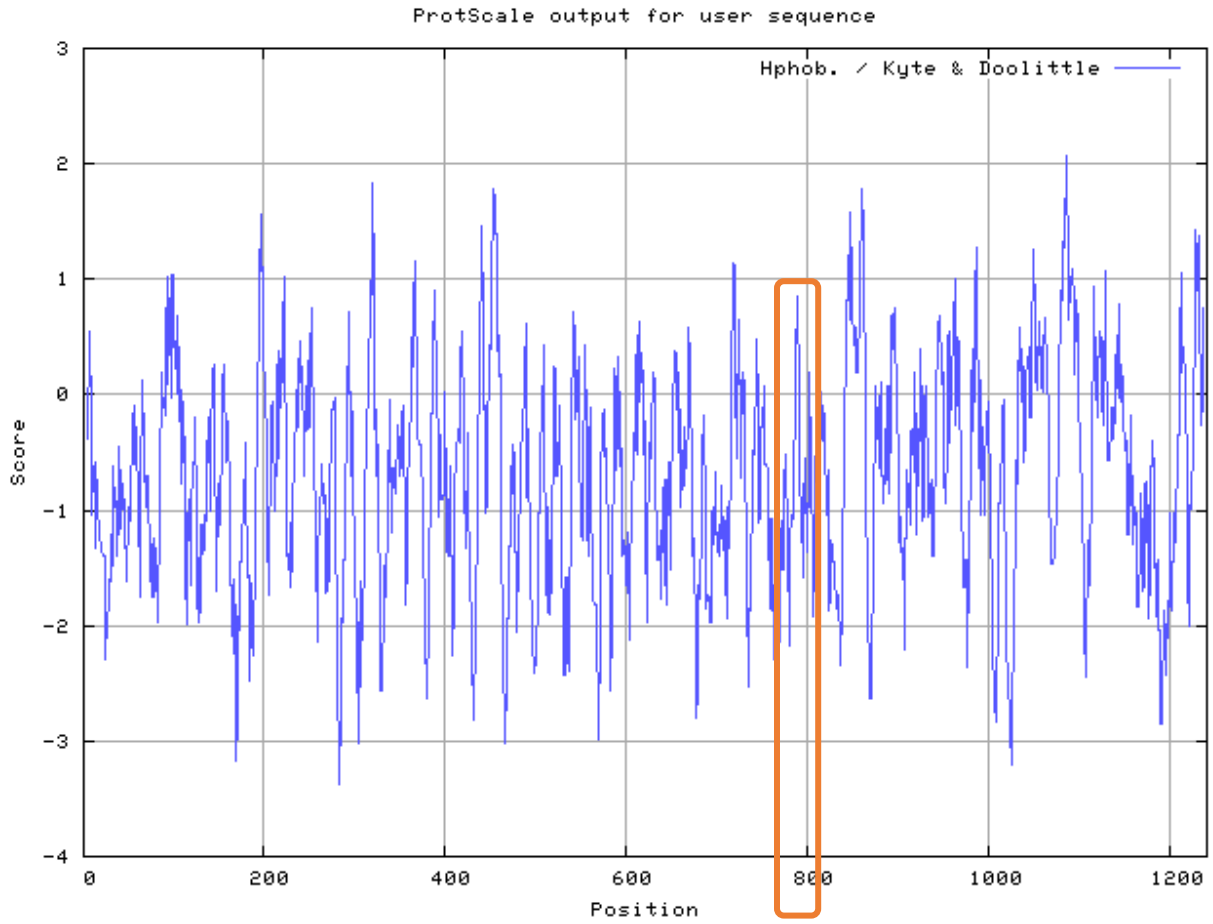


Figure 32. Hydrophobicity plot for *Hs*Sfi1. *Hs*Sfi1p$_{21}$ is outlined. The plot was generated using the ProtScale tool from web.expasy.org.

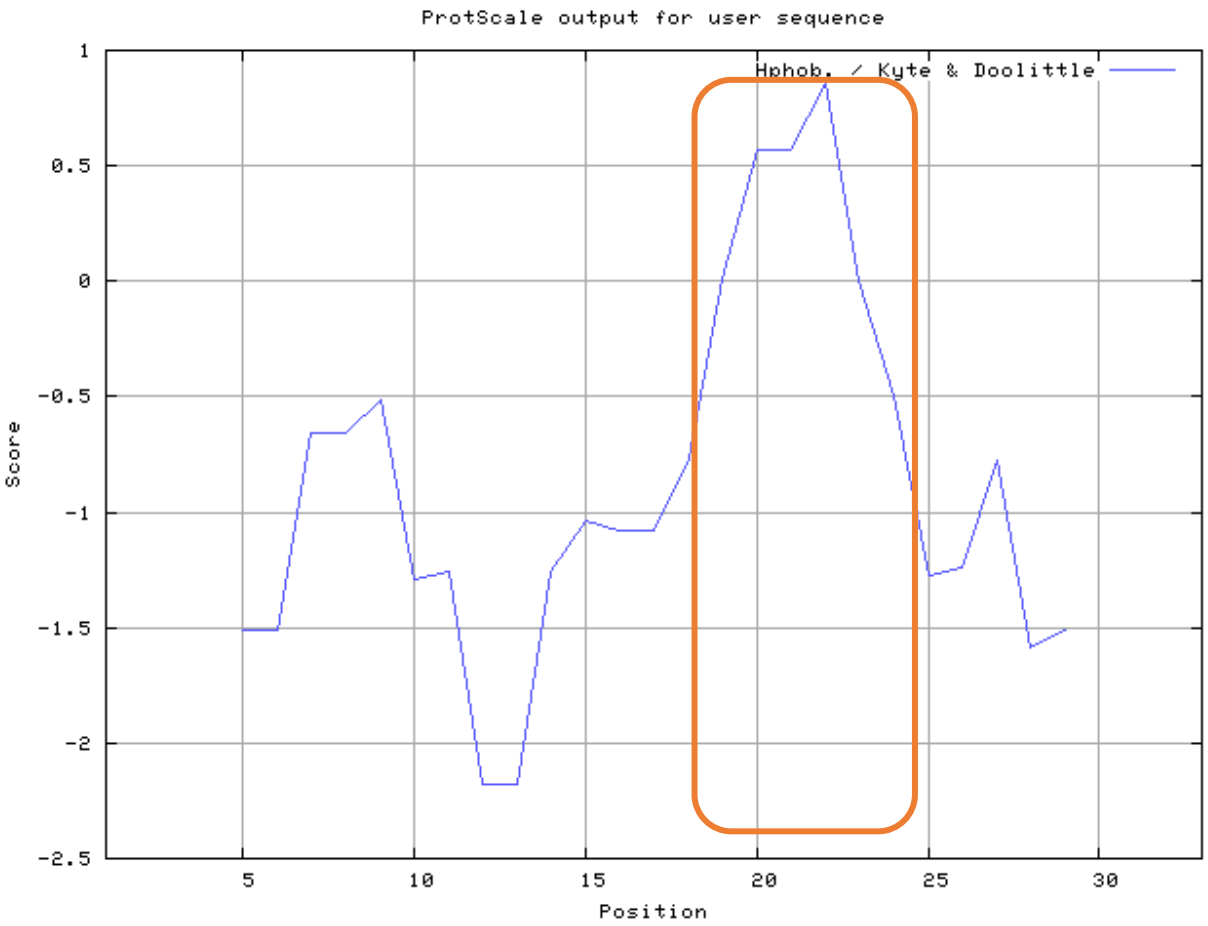


Figure 33. Hydrophobicity plot for the 33-residue partial CBS of *Hs*Sfi1p$_{21}$ from A$_{769}$-C$_{801}$. The hydrophobic triad required for centrin binding is outlined. The plot was generated using the ProtScale tool from web.expasy.org.

B. Characterization of centrin variants

Far-UV scans and helical content thermal dependence analyses of the centrin variants were performed to determine to what extent the variations affect the secondary structure in comparison to the *Hs*cen1 and *Hs*cen2 wild-type (WT) structures. Relative thermal stability was measured in terms of crossover point and minimum mean residue molar ellipticity values by scanning the Far UV region and monitoring helical content at 222 nm in a thermal dependence plot. **Figure 34** represents the relative thermal dependence of the secondary structure of the variants *Hs*cen1(E24K), *Hs*cen2(E32A), and *Hs*cen2(E24K/E105K) at 5°C, 25°C, 85°C and 95°C in the spectral region of 190-250 nm. **Figure 35** represents helical content in terms of mean residue molar ellipticity at 222 nm as a function of temperature from 5°C to 95°C.

From the Far UV scan, it is noticeable the protein's helical content is affected by the variations and changes in temperature. At 5°C, 25°C and 85°C the greatest helical content was observed in *Hs*cen1(E24K), presenting a helical content percentage of 52.40%, 47.03% and 25.13%, respectively. At 95°C the greatest helical content was observed in *Hs*cen2(E24K/E105K) presented by 17.3% of helical content.

Comparing *Hs*cen1(E24K) to *Hs*cen1(WT), the variant presented a 20% increase in helical content. The *Hs*cen2 variants also presented an increase in helical content, with the *Hs*cen2(E24K/E105K) variant presenting a 20% increase and the *Hs*cen2(E32A) variant presenting only a 3% increase in helical content when compared to their wild-type counterpart; these increments imply that the non-conserved variations contribute to higher helical content and stability (**Figure 13** and **Figure 34**) This data is also summarized in **Table *2***.

In the thermal dependence analysis, the variants *Hs*cen1(E24K) and *Hs*cen2(E24K/E105K) presented a sharp decrease in helical contribution near 80°C. This decrease is indicative that the proteins are undergoing a conformational change and adopting a more disordered structure. This relationship is consistent with previous DSC results of wild-type human centrin 1 and centrin 2 [33]. Considering that wild-type centrin 1 and centrin 2 have a T_m of 89.5°C and 81.9°C, respectively, it was expected that the proteins still have some helical contributions and do not present a completely disordered secondary structure. The decrease in helical content near 80°C was not observed in the *Hs*cen2(E32A) variant, suggesting that this protein did not undergo a change in structure. Comparing *Hs*cen1(E24K)'s thermal dependence analysis to its wild-type counterpart, both exhibited the same behavior making this variant the most suitable for the ITC analysis in complex with *Hs*Sfi1p$_{21}$.

A comparative analysis of the crossover points and helical mean residue molar ellipticity values determined at 5°C, 25°C, 85°C and 95°C and the thermal dependence analysis suggest that like the wild-type centrins, these variants undergo a series of conformational changes that do not involve a typical helix-to-random coil transition and instead undergo a different type of transition.

Table 2. Summary of the relative thermal stability of wild-type centrin 1 and centrin 2 and their variants. *Wild-type (WT) data adapted from Pastrana-Ríos et. al 2013 *[33]*.

Protein	Temperature (°C)	Crossover point (nm)	minimum $[\theta]_{MR}$ (nm/M/degree)	Helical content (%)	$[\theta]_{MR}$ at 222nm (M/degree)
H*scen1 (WT)	5	199.7	206.8/-10,975.9	33.26	-9,036.1
	25	200.0	207.5/-9,154.8	27.74	-7,237.8
	85	199.8	206.9/-7,621.3	23.09	-4,830.4
H*scen2 (WT)	5	199.9	205.3/-6,019.3	18.24	-5,700.8
	25	200.0	206.4/-6,185.9	18.75	-5,598.5
	85	199.6	204.7/-3,441.1	10.43	-2,175.6
***H*scen1(E24K)**	5	202.6	208.4/-17,293.6	52.40	-10,054.1
	25	204.0	209.5/-15,520.9	47.03	-9,461.1
	85	208.9	214.5/-8,293.7	25.13	-5,603.3
	95	206.7	212.0/-4,745.3	14.38	-4,448.7
***H*scen2(E32A)**	5	199.9	208.8/-7,033.2	21.31	-8,705.3
	25	199.8	207.7/-6,822.3	20.67	-8,561.1
	85	204.1	205.7/-6,804.1	20.62	-6,756.1
	95	205.6	208.9/-5,555.4	16.83	-6,636.4
***H*scen2 (E24K/E105K)**	5	202.3	207.8/-12,954.1	39.25	-11,266.9
	25	203.3	208.5/-12,058.1	36.54	-10,790.3
	85	209.0	214.8/-6,295.1	19.08	-7,091.3
	95	210.7	217.6/-5,715.2	17.32	-6,100.7

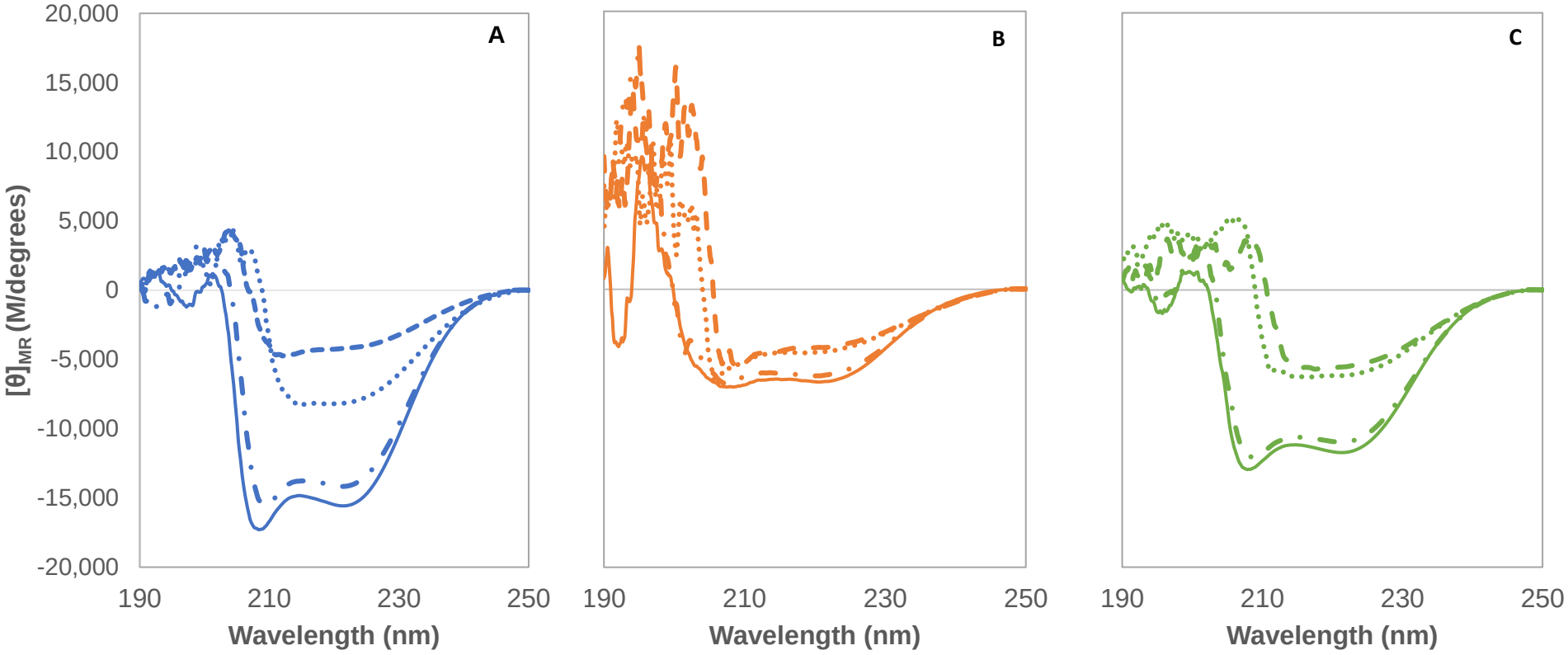


Figure 34. Overlay of Far-UV CD spectra at 5°C (—), 25°C (— · —), 85°C (·) and 95°C (— — —) of the centrin variants (A) *Hs*cen1(E24K), (B) *Hs*cen2(E32A) and (C) *Hs*cen2(E24K/E105K) within the spectral region of 190-250 nm.

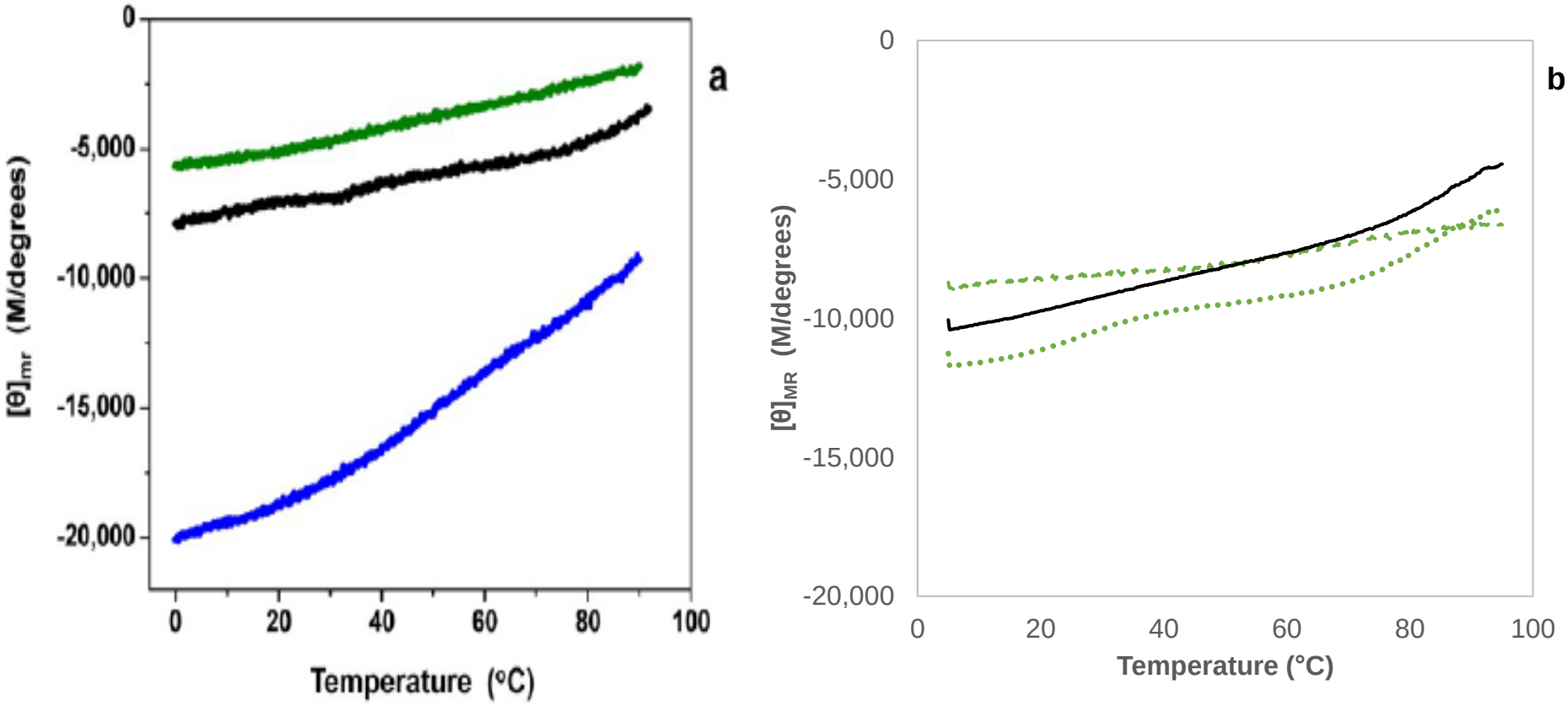


Figure 35. Thermal dependence of the helical content among centrin isoforms and variants. (a) *Hs*cen1 (black), *Hs*cen2 (green), *Hs*cen3 (blue)*. (b) *Hs*cen1(E24K) (black), *Hs*cen2(E32A) (green, - -) and *Hs*cen2(E24K/E105K) (green, ⋯) monitored at 222 nm from 5°C to 95°C. *Adapted from Pastrana 2013

C. *Hs*cen1(E24K)-*Hs*Sfi1p$_{21}$ complex formation

The analysis of the interaction between *Hs*cen1(E24K) and *Hs*Sfip$_{21}$ was carried out using ITC at an experimental temperature of 25°C. **Figure 36** represents the raw power data and integrated enthalpy data fitted with a one site binding model. The interaction process was exothermic ($\Delta H = -3.10 \times 10^3$ cal/mol), the stoichiometry of the interaction was n= 0.5 and the affinity constant (k_a) for the complex was $4.93 \times 10^5 M^{-1}$ or a dissociation constant (k_d) of 0.41 µM. **Table 3** summarized the obtained binding parameters.

Although a negative change in Gibbs free energy was observed suggesting exergonic complex formation, the *Hs*cen1(E24K)-*Hs*Sfi1p$_{21}$ complex involves half of the binding interface as suggested from the stoichiometry value of 0.5. Also, the change in enthalpy was 1/8 the magnitude of that observed for *Hs*cen1(WT)-*Hs*Sfi1p$_{21}$ complex by Díaz-Casas in 2012 [41], indicative that a lower number of weak interactions are being formed in the variant containing complex. Finally, the -TΔS is negative suggestive of no conformational change during complex formation which also validates the low stoichiometry value of less than 1 (**Figure 37**). This can be attributed to the need for the glutamate in this position which is involved in a crucial salt-bridge interaction between wild type centrin and *Hs*Sfi1p$_{21}$.

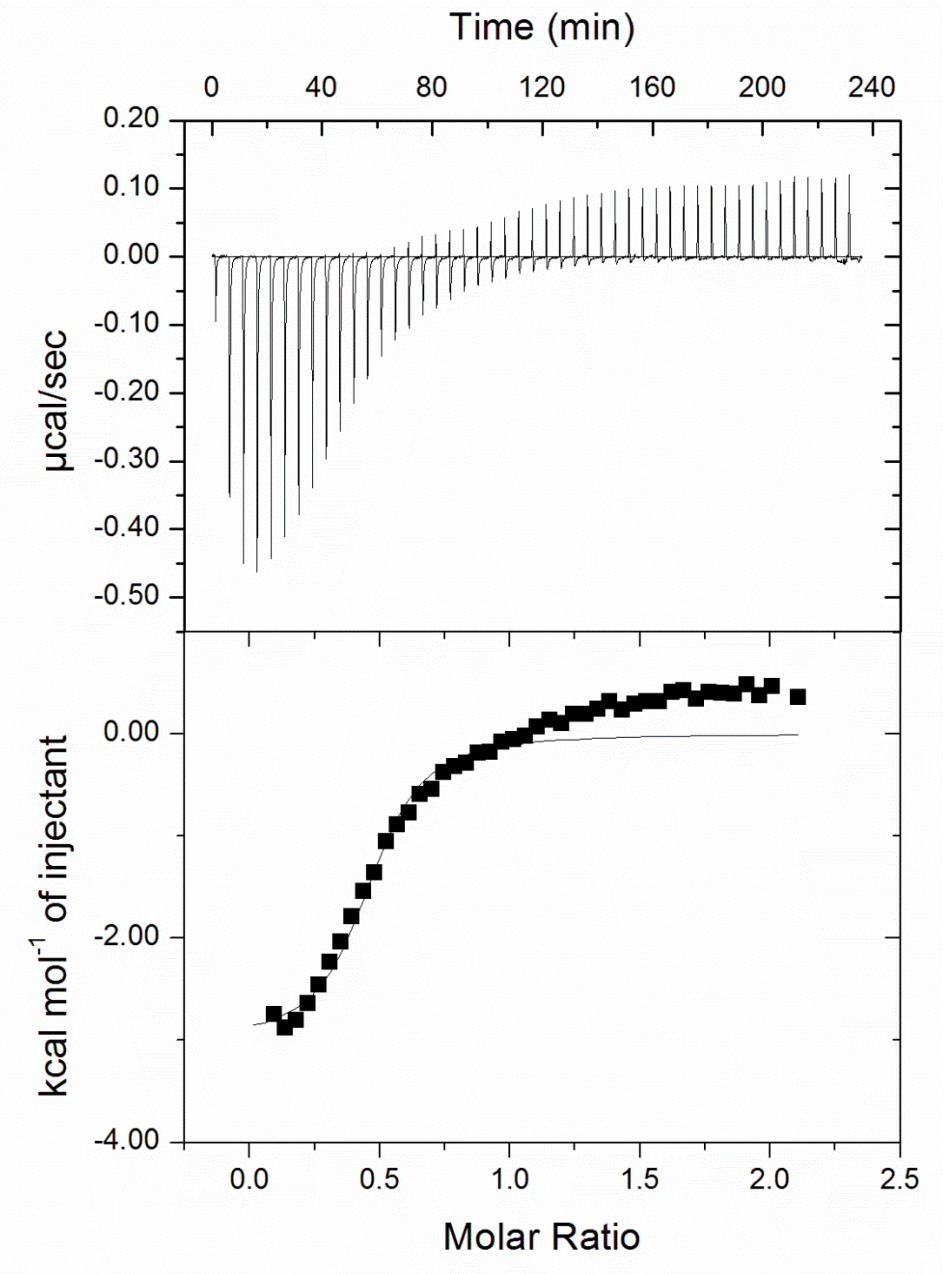


Figure 36. ITC isotherm of the interaction between *Hs*cen1(E24K) and *Hs*Sfi1p$_{21}$ at 25°C. The upper panel shows the raw power data and the lower panel shows the integrated enthalpy data fitted to a one-site binding model.

Table 3. ITC parameters for *Hs*cen1(E24K) and *Hs*cen1(WT)* interacting with *Hs*Sfi1p$_{21}$ at 25°C. Molarity ratio for *Hs*cen1(WT)/*Hs*Sfi1p$_{21}$ was 10 μM and for *Hs*cen1(E24K)/*Hs*Sfi1p$_{21}$ was 8 μM. **Hs*cen1(WT) adapted from Díaz-Casas 2012.

Protein	Peptide	n	K_a (M^{-1})	K_d (M)	ΔH (cal/mol)	ΔG (cal/mol)	-TΔS (cal/mol)
***Hs*cen1(E24K)**	*Hs*Sfi1p$_{21}$	0.5	$4.9x10^5$ ($\pm 2x10^5$)	$2.03x10^{-6}$	$-3.0x10^3(\pm 0.2)$	-7.8	-4.7
Hs*cen1(WT)	*Hs*Sfi1p$_{21}$	1.0	$1.5x10^7$ ($\pm 1x10^6$)	$6.7x10^{-8}$	$-24.2x10^3$ (± 0.1)	-9.8	14.5

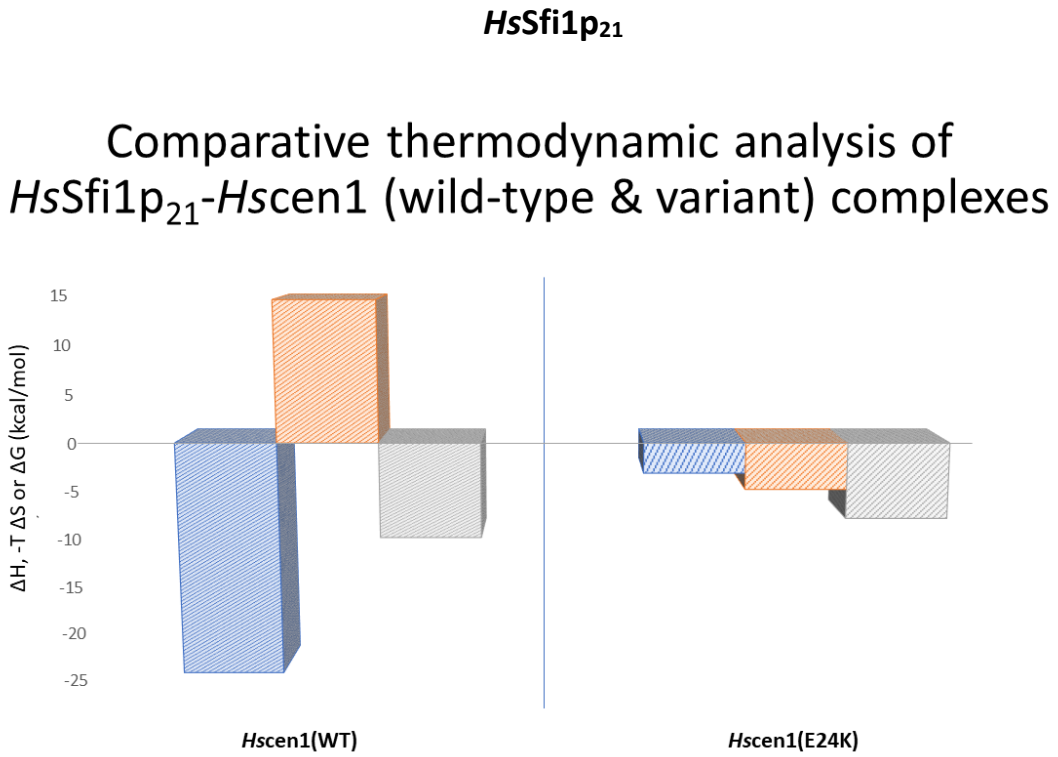


Figure 37. A comparative analysis of *Hs*Sfi1p$_{21}$interactions with *Hs*cen1(WT) and *Hs*cen1(E24K). Comparison of the enthalpic (ΔH, blue bar), entropic (-TΔS, orange bar) and Gibbs free energy (ΔG, grey bar) contributions between *Hs*cen1(WT) and *Hs*cen1(E24K) at 25°C. *Hs*cen1(WT) data adapted from Diaz-Casas 2012 *[41]*.

D. Discussion and implications of results

The recombinant expression and purification of *Hs* centrin variants was performed along with the biochemical and biophysical characterization of these proteins. Included as evidence of their characterization are UV, HPLC and Circular Dichroism analyses. The interaction of the protein with its target ligand was also ascertained via isothermal titration calorimetry. The results yielded important information as to the stability and affinity of the resulting *Hs*cen1(E24K)-*Hs*Sfi1p$_{21}$ complex. These results were compared with the wild-type counterpart as referenced from previous work. CD results showed that the mutations have a significant effect on helical content and their relative thermal stability. The *Hs*cen2 variants showed a significant decrease in ellipticity values, therefore they contribute to higher helical content. It was observed that variants in which the substitution involved a polar residue such as, *Hs*cen1(E24K) and *Hs*cen2(E24K/E105K), presented greater helical content than the variant in which the polar residue was substituted for a non-polar residue (*Hs*cen2E32A). This effect can be attributed to charged residues that contribute to non-covalent interactions, such as the formation of hydrogen bonds and salt-bridge, that stabilize the structure. This analysis was not limited to empirical characterization, but also significant progress was made using bioinformatics tools that are also summarized herein and in the appendix.

Complex formation analysis. Although the variants presented higher thermal stability, ITC results from the interaction of *Hs*cen1(E24K) with *Hs*Sfi1p$_{21}$ showed a substantial decrease in the thermodynamic binding parameters, suggesting that the glutamate is essential towards complex formation due to salt-bridge interaction. As a

result, only half the CBS interacts with *Hs*cen1(E24K), which is observed by the stoichiometry value of 0.5.

Conclusions

The biophysical characterization of the centrin variants proved valuable to the understanding to the effect of the relationship of sequence-structure-stability. Dedicated molecular biophysical approach such as ITC are essential towards the study of protein-protein interactions.

Future Work

Differential Scanning Calorimetry (DSC) experiments can also be carried out to determine the variant's T_m. Also, High resolution structure determination of the complex whether multi-dimensional NMR experiments that provide the solution structure and dynamics of the complex or X-ray diffraction to obtain the crystal structure of the complex.

Appendix

QUIM 8995: Bioinformatics: Proteins & Protein-Protein Interactions (PPI's)

Objectives

In my thesis research project, I will be using different biophysical methods to analyze how non-conserved mutations in human centrin 1 and 2 affect the interaction with one of its biological targets, Sfi1. Homo sapiens Sfi1 (*Hs*Sfi1) has 23 tandem centrin binding sites, which are not conserved. As part of this course requirement, the objective of this project is to use different bioinformatics databases to analyze Sfi1 orthologs, specifically its 23 binding sites. Bioinformatics tools will be used to establish evolutionary relationships, percent identity, sequence alignments and other comparative analyses to further study these centrin binding sites (CBS) in Sfi1, complementing the biophysical experimental data for a more thorough analysis and validation.

Part I: Sequence Alignments and Dendrogram

In the subsequent pages are the individual sequences of the target protein (Sfi1) in Homo sapiens (*Hs*) and Saccharomyces cerevisiae (*Sc*). The following analyses are also presented: sequence alignments using both CLC Genomics Workbench and ClustalW, a comparative table and dendrogram generated using CLC Genomics Workbench. A sequence analysis of the CBS is also presented, with the conserved sequence AX_7LLX_3F/LX_2W highlighted in yellow (the X representing any given residue). Said conserved sequence is thought to essential for the interaction and defines the selectivity for centrin. A frequency distribution analysis of substitutions in the CBS is also presented.

*Hs*Sfi1

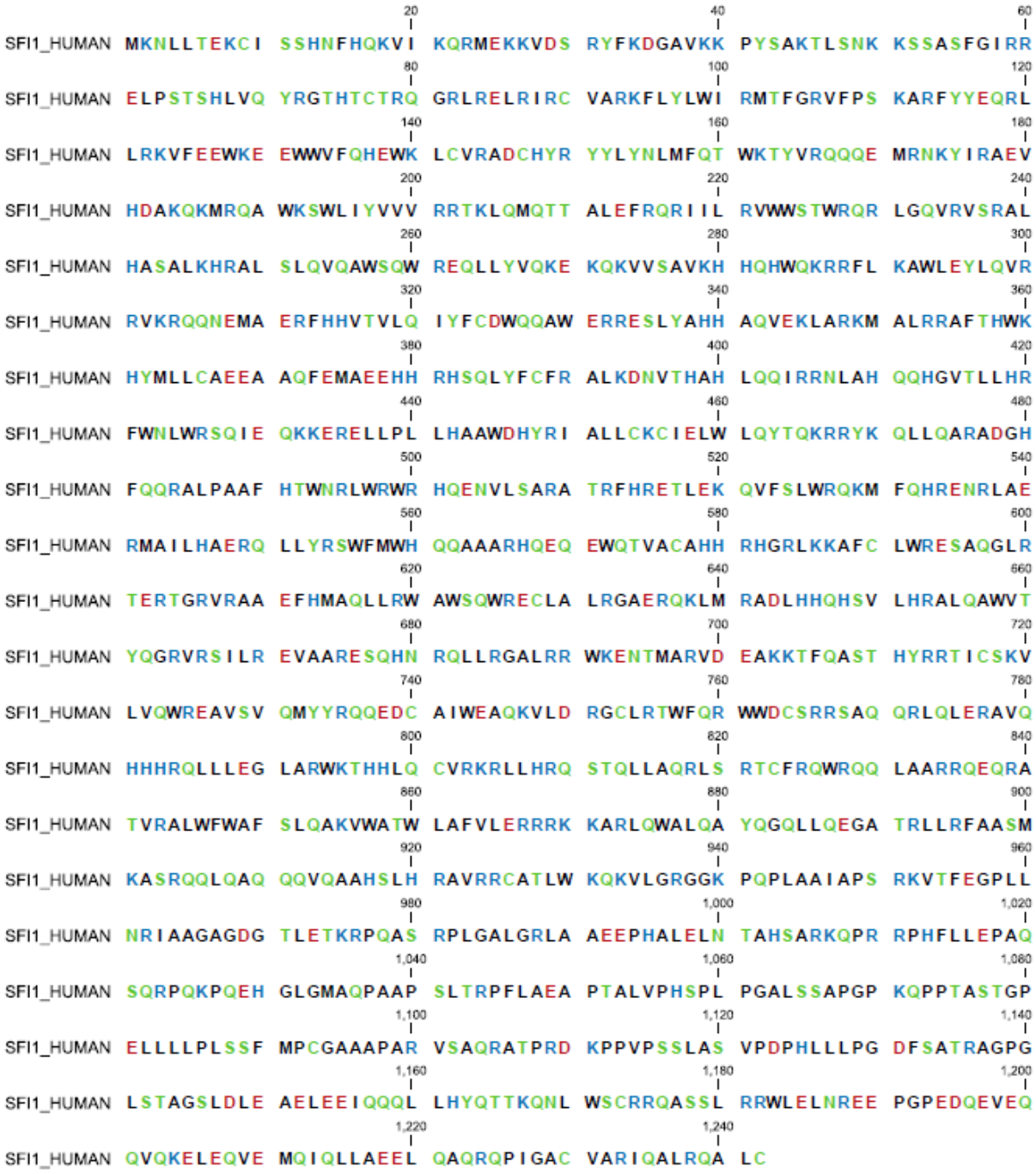


Figure A1. *Hs*Sfi1 sequence compromised of 1242 amino acids. Generated using CLC Workbench. Residues are color coded according to polarity: negatively charged residues are red, positively charged residues are blue, hydrophobic residues are black and polar residues are green.

Frequency of substitutions in the *Hs*Sfi1 consensus sequence

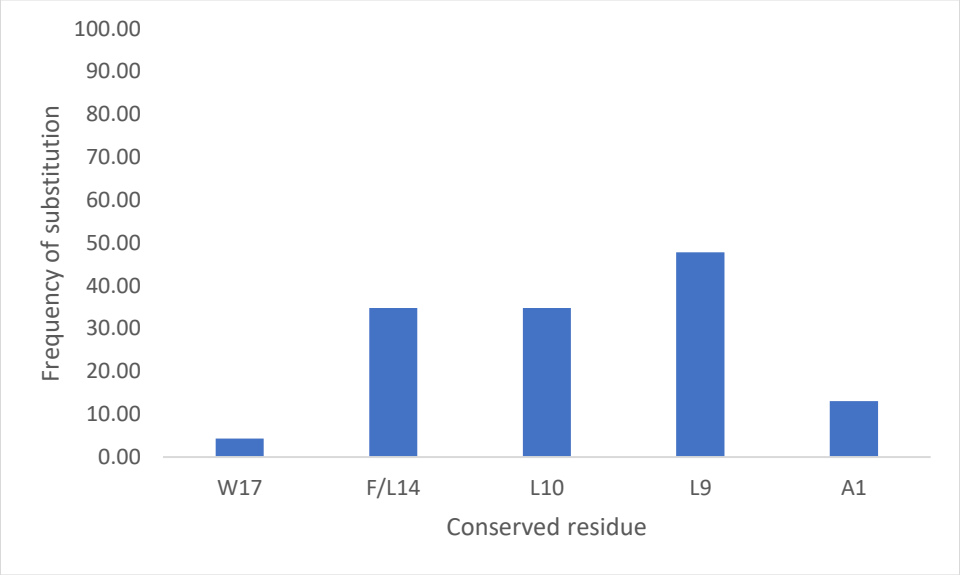


Figure A2. Frequency distribution of substitutions in *Hs*Sfi1.

*Sc*Sfi1

Figure A3. *Sc*Sfi1 sequence containing 946 amino acids and compromised of 17 tandem CBS. Generated using CLC Workbench. Residues are color coded according to polarity: negatively charged residues are red, positively charged residues are blue, hydrophobic residues are black and polar residues are green.

*Sc*Sfi1 CBS alignment

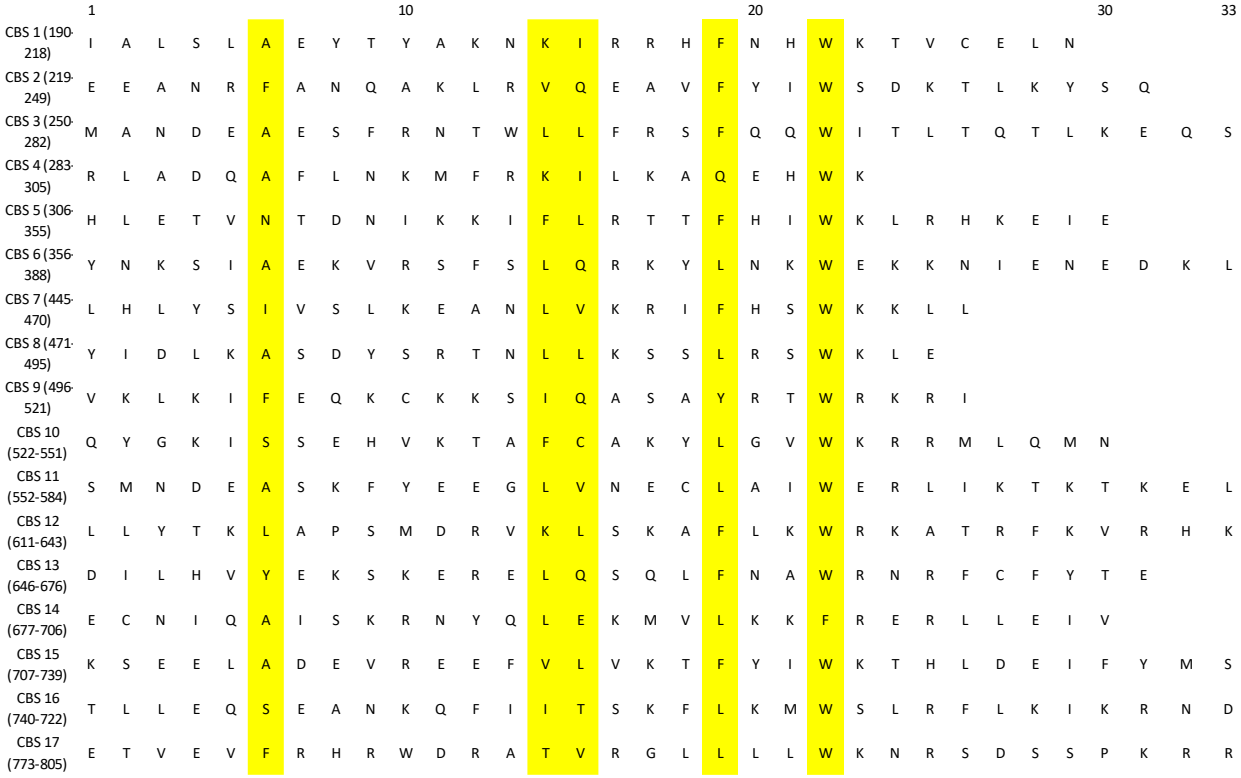


Figure A4. Sequence alignment of the 17 tandem CBS in *Sc*Sfi1. The conserved residues of the consensus sequence AX_7LLX_3F/LX_2W are highlighted in yellow. Length variability is observed.

Frequency of substitutions in the *Sc*Sfi1 consensus sequence

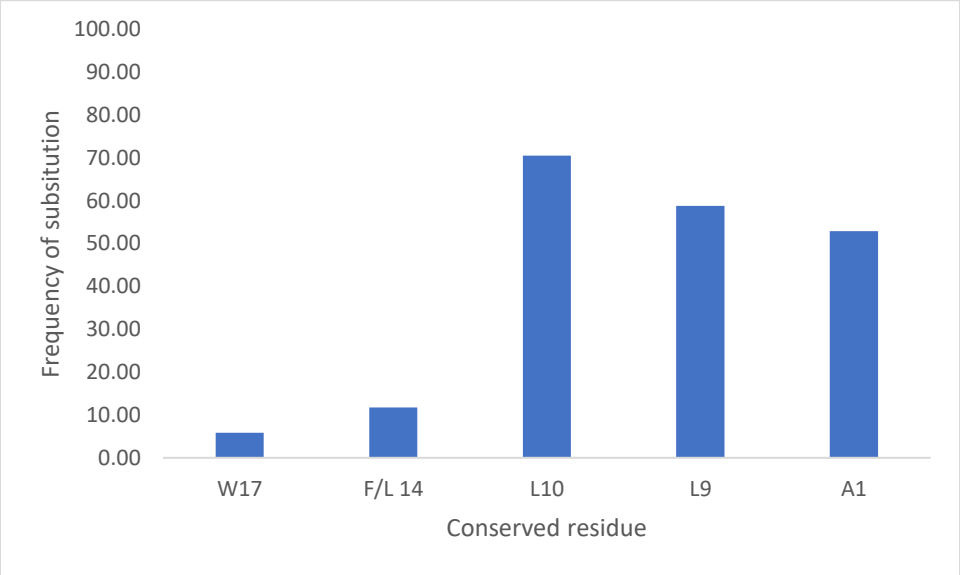


Figure A5. Frequency distribution of substitutions in *Sc*Sfi1

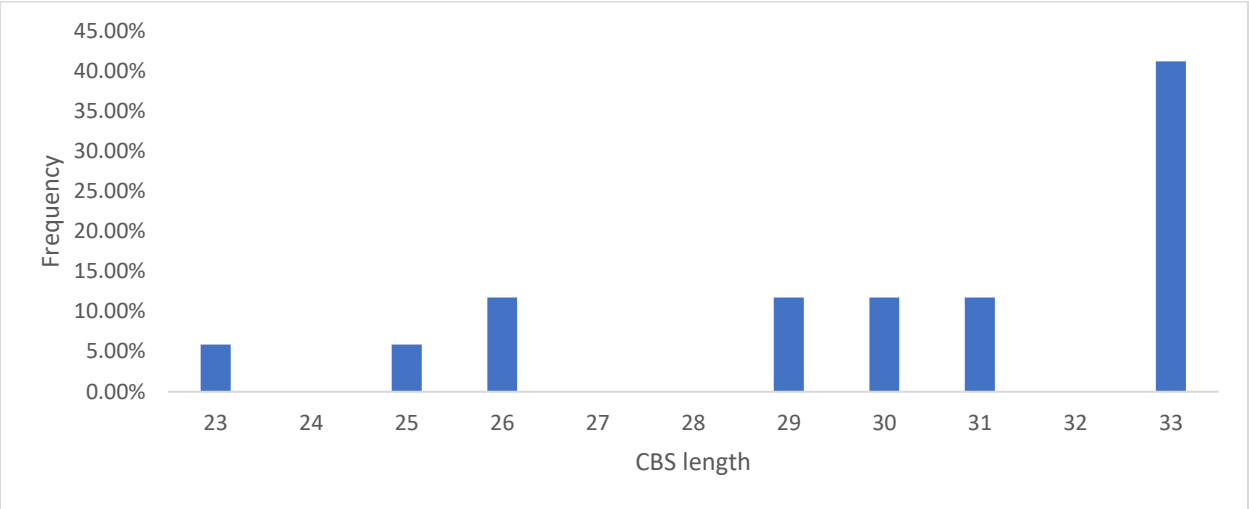
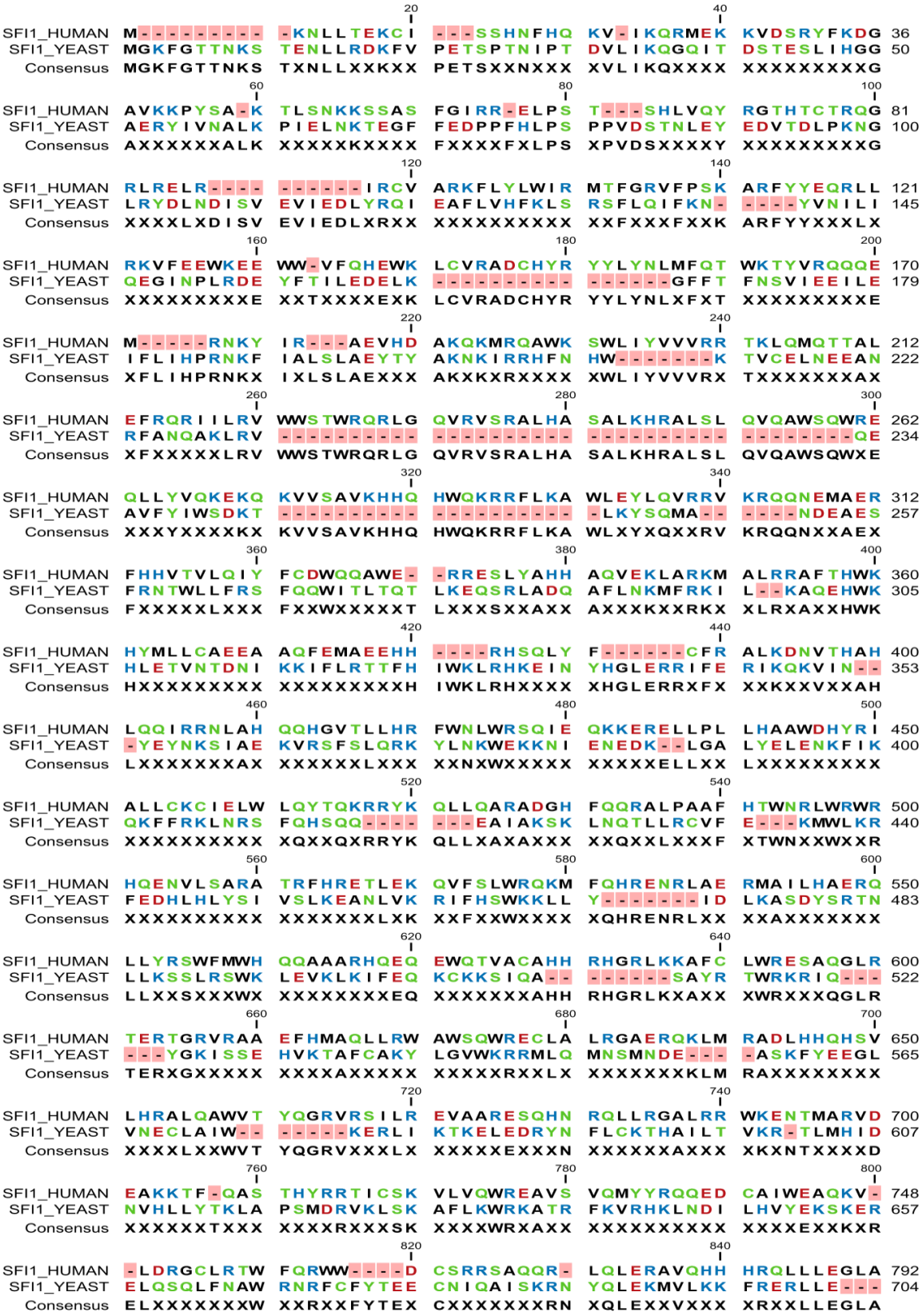


Figure A6. Frequency of *Sc*Sfi1 CBS lengths.

CLC Genomics Workbench Sequence Alignment



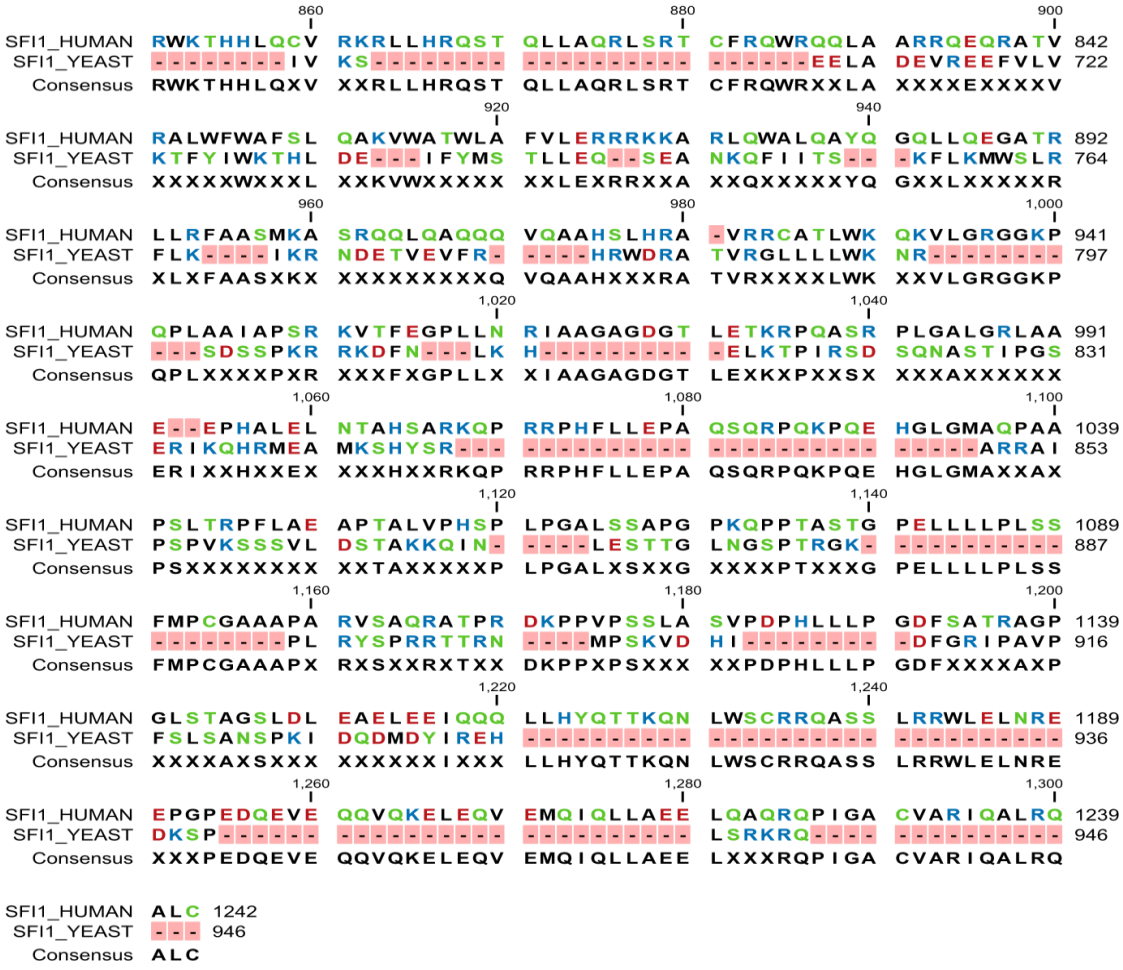


Figure A7. Sequence alignment of *Hs*Sfi1 and *Sc*Sfi1, generated using CLC Workbench. Residues are color coded according to polarity: negatively charged residues are red, positively charged residues are blue, hydrophobic residues are black and polar residues are green. Gaps are represented by (-) and highlighted in pink.

ClustalW Sequence Alignment

```
sp|A8K8P3|SFI1_HUMAN      ----------MKNLLTEKCISSH---NFHQK-VIKQR----------------------  23
sp|Q12369|SFI1_YEAST      MGKFGTTNKSTENLLRDKFVPETSPTNIPTDVLIKQGQITDSTESLIHGGAERYIVNALK  60
                                    :*** :* : .    *:  . :***

sp|A8K8P3|SFI1_HUMAN      -MEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIRRELPSTSHLVQYRGTHTCTRQGR  82
sp|Q12369|SFI1_YEAST      PIELNKTEGFFEDPPFHLPSP------------------PVDSTNLEYEDVTDLPKNGL  101
                           :* :  . :*:*  .: *                   *  *  ::*. .    ::*

sp|A8K8P3|SFI1_HUMAN      LR---ELRIRCVAR---KFLYLWIRMTFGRVFPSKARFYYEQRLLRKVFEEWKEEWWVFQ  136
sp|Q12369|SFI1_YEAST      RYDLNDISVEVIEDLYRQIEAFLVHFKLSRSFLQIFKNY-VNILIQEGINPLRDEYFTIL  160
                               :: :. :     ::  : :::.:.* * .  : *  : *::: ::  ::*::.:

sp|A8K8P3|SFI1_HUMAN      HEWKLCVRADCHYRYYLYNLMFQT-WKTYVRQQQEMRNKYIR---AEVHDAKQKMRQAWK  192
sp|Q12369|SFI1_YEAST      --------EDELKGFFTFNSVIEEILEIFL---IHPRNKFIALSLAEYTYAKNKIRRHFN  209
                                  *    :: :* :::   : ::    . ***:*    **   **:*:*: ::

sp|A8K8P3|SFI1_HUMAN      SWLIYVVVRRTKLQMQTTALEFRQRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSL  252
sp|Q12369|SFI1_YEAST      HWKTVCELNEEANR-------FANQ------------------------------AKLRV  232
                           *     :..   :       * ::                                 * :

sp|A8K8P3|SFI1_HUMAN      QVQAWSQWREQLLY-VQKEKQKVVSAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAE  311
sp|Q12369|SFI1_YEAST      QEAVFYIWSDKTLKYSQMAND---EAESFRNTWLLFRSFQQWITLTQTLK---EQSRLAD  286
                          *  .:  * :: *   *  ::   .* ..:: *   * :: *:   *. :   :*..:*:

sp|A8K8P3|SFI1_HUMAN      RFHHVTVLQIYF---CDWQQAWERRESLYAHHAQVEKLARKMALRRAFTHWKHYMLLCAE  368
sp|Q12369|SFI1_YEAST      Q----AFLNKMFRKILKAQEHWKHLETVNT------DNIKKIFLRTTFHIWKLRHKEINY  336
                          :    :.*:  *    . *: *:: *:: :      .  :*: ** :*  **

sp|A8K8P3|SFI1_HUMAN      EAAQFEMAEEHHRHSQLYFCFRALKDNVTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQ  428
sp|Q12369|SFI1_YEAST      -------------HGLERRIFERIKQKVINY---EYNKSIAEKVRSFSLQRKYLNKWEKK  380
                                    *.     *. :*::* .    : .:.:*.: :..:* ::: * *..:

sp|A8K8P3|SFI1_HUMAN      IEQKKER-E-------------LLPLLHAAWDH-----------YRIALLCKCIELWLQY  463
sp|Q12369|SFI1_YEAST      NIENEDKLGALYELENKFIKQKFFRKLNRSFQHSQQEAIAKSKLNQTLLRCVFEKMWLKR  440
                            :::::               ::  *. :::*            :  * *   ::**:

sp|A8K8P3|SFI1_HUMAN      TQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHRETLEKQVF  523
sp|Q12369|SFI1_YEAST      FEDHLHLYSIV----SLKEANLVKRIFHSWKKLLYID-------LKASDYSRTNLLKSSL  489
                           :.: :   :     .  : . :   **:*:*            :*: : * .* *. :

sp|A8K8P3|SFI1_HUMAN      SLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTVACAHHRHG  583
sp|Q12369|SFI1_YEAST      RSWKLEVKLKIFEQK-----CKKSIQASAYRTWRKRIQ----------YGKIS-----SE  529
                            *: ::  :  ::        :: :   **:*    *          : .::

sp|A8K8P3|SFI1_HUMAN      RLKKAFC-----LWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAERQK  638
sp|Q12369|SFI1_YEAST      HVKTAFCAKYLGVWKRRMLQMNSMN---DEASKFYEEGLVNECLAIWKERLIKTKE----  582
                          ::*.***     :*:.    :.: .    .*::*:   *:. . : *:* *

sp|A8K8P3|SFI1_HUMAN      LMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENTMAR  698
sp|Q12369|SFI1_YEAST      -----LEDRYN-----------FLCKTHAILTVK---------------------RTLMH  605
                               *..::.           :  :.::**                       .*: :

sp|A8K8P3|SFI1_HUMAN      VDEAKKTF-QASTHYRRTICSKVLVQWREAVSVQMYYRQQEDCAIWEAQK--VLDRGCLR  755
sp|Q12369|SFI1_YEAST      IDNVHLLYTKLAPSMDRVKLSKAFLKWRKATRFKVRHKLNDILHVYEKSKERELQSQLFN  665
                          :*:.:  : : :    *.  **.:::**:*. .:: :: ::   ::* .*   *:   :.

sp|A8K8P3|SFI1_HUMAN      TWFQRWW----DCSRRS-AQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQ  810
sp|Q12369|SFI1_YEAST      AWRNRFCFYTEECNIQAISKRNYQLEKMVLKKFRE-------------------RLLEIV  706
                          :* :*:     :*. :: :::. ***: * ::.*:                   ***.

sp|A8K8P3|SFI1_HUMAN      STQLLAQRLSRTCFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRK  870
sp|Q12369|SFI1_YEAST      KS----------------EELADEVREEFVLVKTFYIWKTHLDEIFYMST---LLEQSEA  747
                          .:               ::** . :*: . *:::::*   *:  .: :    :**: .
```

Figure A8. Sequence alignment of *Hs*Sfi1 and *Sc*Sfi1, generated using ClustalW. Residues are color coded according to their chemical properties: hydrophobic residues are colored red, polar residues are colored green, negatively charged residues are colored blue and positively charged residues are colored lilac. Gaps are represented by (-), consensus is represented with the following signs: (*) represents maximum conservation, (:) represents conservation between groups with very similar characteristics and (.) represents conservation between groups with slightly similar characteristics.

Comparative Table

1.1 Sequence information

Information	SFI1_HUMAN	SFI1_YEAST
Sequence type	Protein	Protein
Length	1242 aa	946 aa
Organism	Homo sapiens	Saccharomyces cerevisiae
Name	SFI1_HUMAN	SFI1_YEAST
Description	RecName: Full=Protein SFI1 homolog; Short=hSFI1.	RecName: Full=Protein SFI1; AltName: Full=Suppressor of fermentation induced loss of stress resistance protein 1
Modification Date	20-JAN-2016	01-NOV-1996
Weight	147.662 kDa	112.977 kDa
Isoelectric point	10.96	10.01
Aliphatic index	78.494	84.419

1.2 Half-life

Information	SFI1_HUMAN	SFI1_YEAST
N-terminal aa	Methionine	Methionine
Half-life mammals	30 hours	30 hours
Half-life yeast	>20 hours	>20 hours
Half-life E.Coli	>10 hours	>10 hours

1.3 Extinction coefficient

Conditions	SFI1_HUMAN	SFI1_YEAST
Non-reduced cysteines Extinction	323,060	149,550

coefficient at 280nm		
Non-reduced cysteines Absorption at 280nm 0.1% (=1 g/l)	2.188	1.324
Reduced cysteines Extinction coefficient at 280nm	321,620	149,070
Conditions	SFI1_HUMAN	SFI1_YEAST
Reduced cysteines Absorption at 280nm 0.1% (=1 g/l)	2.178	1.319

1.4 Counts of atoms

Atoms	SFI1_HUMAN	SFI1_YEAST
hydrogen (H)	10,383	8,076
carbon (C)	6,587	5,113
nitrogen (N)	2,071	1,428
oxygen (O)	1,725	1,416
sulphur (S)	46	24

1.5 Frequencies of atoms

Atoms	SFI1_HUMAN	SFI1_YEAST
hydrogen (H)	0.499	0.503
carbon (C)	0.317	0.318

nitrogen (N)	0.100	0.089
oxygen (O)	0.083	0.088
sulphur (S)	0.002	0.001

1.6 Counts of hydrophobic and hydrophilic residues

Type	SFI1_HUMAN	SFI1_YEAST
Hydrophobic (A,F,G,I,L,M,P,V,W)	558	382
Hydrophilic (C,N,Q,S,T,Y)	310	247
Other	374	317

Type	SFI1_HUMAN	SFI1_YEAST
Hydrophobic (A,F,G,I,L,M,P,V,W)	0.449	0.404
Hydrophil (C,N,Q,S,T,Y)	0.25	0.261
Other	0.301	0.335

1.7 Frequencies of hydrophobic and hydrophilic residues

1.8 Counts of charged residues

Charge type	SFI1_HUMAN	SFI1_YEAST
Negatively Charged (D & E)	98	115
Positively Charged (R & K)	215	174
Other	929	657

1.9 Frequencies of charged residues

Charge type	SFI1_HUMAN	SFI1_YEAST
Negatively Charged (D & E)	0.079	0.122
Positively Charged (R & K)	0.173	0.184
Other	0.748	0.695

1.10 Counts of amino acids

Amino acid	SFI1_HUMAN	SFI1_YEAST
Alanine (A)	134	43
Cysteine (C)	24	8
Aspartic Acid (D)	19	39
Glutamic Acid (E)	79	76
Phenylalanine (F)	39	54
Glycine (G)	39	20
Histidine (H)	61	28
Isoleucine (I)	24	63
Lysine (K)	67	102
Leucine (L)	147	101
Methionine (M)	22	16
Asparagine (N)	18	53
Proline (P)	43	26
Glutamine (Q)	128	36
Arginine (R)	148	72
Serine (S)	61	69

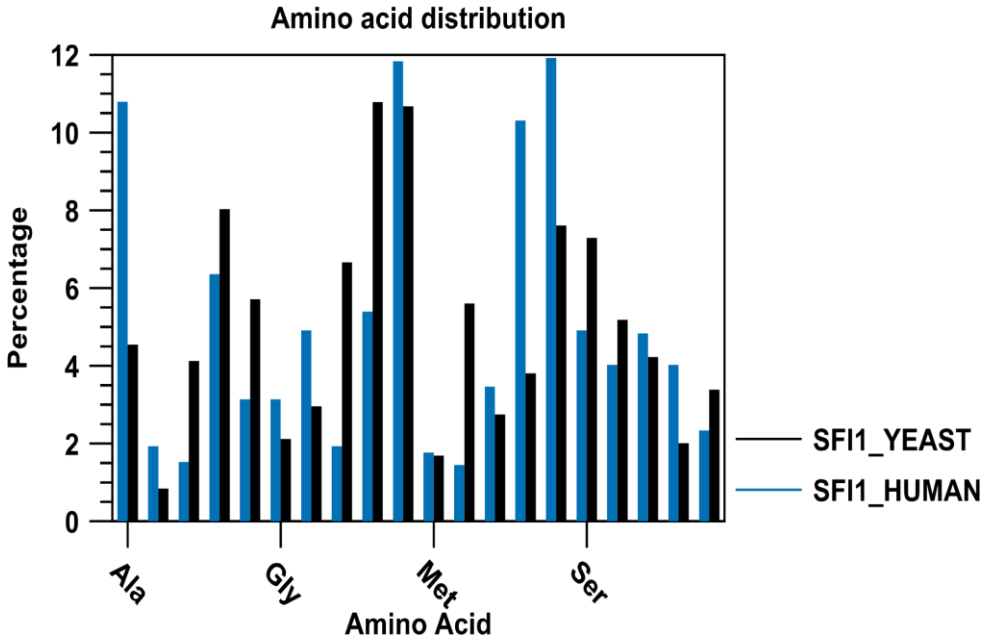
Threonine (T)	50	49
Amino acid	SFI1_HUMAN	SFI1_YEAST
Valine (V)	60	40
Tryptophan (W)	50	19
Tyrosine (Y)	29	32

1.11 Frequencies of amino acids

Amino acid	SFI1_HUMAN	SFI1_YEAST
Alanine (A)	0.108	0.045
Cysteine (C)	0.019	0.008
Aspartic Acid (D)	0.015	0.041
Glutamic Acid (E)	0.064	0.080
Phenylalanine (F)	0.031	0.057
Glycine (G)	0.031	0.021
Histidine (H)	0.049	0.030
Isoleucine (I)	0.019	0.067
Lysine (K)	0.054	0.108
Leucine (L)	0.118	0.107
Methionine (M)	0.018	0.017
Asparagine (N)	0.014	0.056
Proline (P)	0.035	0.027
Glutamine (Q)	0.103	0.038
Arginine (R)	0.119	0.076

Serine (S)	0.049	0.073
Threonine (T)	0.040	0.052
Valine (V)	0.048	0.042
Tryptophan (W)	0.040	0.020
Tyrosine (Y)	0.023	0.034

1.12 Histogram of amino acid frequencies



1.13 Counts of annotations

Feature type	SFI1_HUMAN	SFI1_YEAST
Alpha-helix	0	5
Chain	0	1
Gene	1	0
Modified site	0	2

Protein	1	0
Region	29	0
Source	1	0
Turn	0	1
source	0	1

Figure A9. Comparative table for *Hs* and *Sc* Sfi1. Generated using CLC Workbench.

Dendrogram

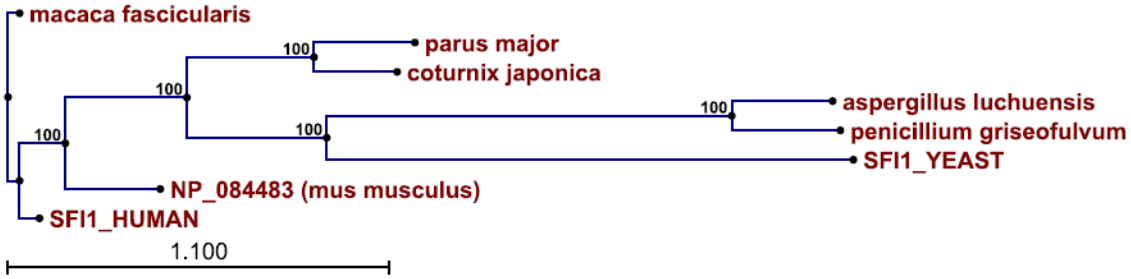


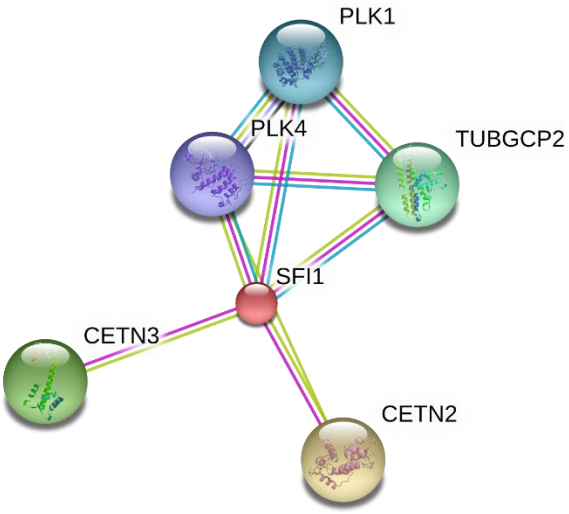
Figure A10. Dendrogram comparing Sfi1 homologues from different organisms. Generated using CLC Workbench.

Summary

From the dendrogram we can visualize the evolutionary distribution of Sfi1. In higher eukaryotes Sfi1 is localized in the centrioles, while in lower eukaryotes it is localized in the half bridge. We can see that while higher eukaryotes branch out from the same node, they are not clustered together, implying some degree of non-conservation within higher species. More primitive organisms, such as fungi are clustered together, with unicellular and multicellular fungi branching out from the same node, and the multicellular species are clustered together.

Part II: Interactome

The input for the following interactome was *Hs*Sfi1. The interactome was generated using the Search Tool for the Retrieval of Interacting Genes/Proteins, better known as the STRING database within the Swiss Institute of Bioinformatics (SIB) website.



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
SFI1	Sfi1 homolog, spindle assembly associated (yeast) (1242 aa) *(Homo sapiens)*									

Predicted Functional Partners:

CETN2	centrin, EF-hand protein, 2; Plays a fundamental role in microtubule-organizing center structur [...] (172 aa)					•		•		0.978
CETN3	centrin, EF-hand protein, 3; Plays a fundamental role in microtubule-organizing center structur [...] (167 aa)					•		•		0.960
TUBGCP2	tubulin, gamma complex associated protein 2; Gamma-tubulin complex is necessary for microtubule [...] (902 aa)						•			0.934
PLK1	polo-like kinase 1; Serine/threonine-protein kinase that performs several important functions t [...] (603 aa)						•			0.924
PLK4	polo-like kinase 4; Serine/threonine-protein kinase that plays a central role in centriole dupl [...] (970 aa)						•			0.924

Figure A11. *Hs*Sfi1 interactome. Generated using STRING database. Interactions are represented by color-coded lines. Known interactions are either represented by pink and cyan lines. Pink lines represent interactions that have been determined experimentally while cyan lines represent interactions from curated databases. Predicted interactions are colored green, red or dark blue. Green lines represent gene neighborhood, red lines represent gene fusions and dark blue lines represent gene co-occurrence. Other interactions lines are represented by lime green, black and light blue. Lime green lines represent textmining, black lines represent co-expression and light blue lines represent protein homology.

Summary

From the interactome we can see that *Hs*Sfi1 interacts with various proteins. Known interactions supported by experimental data and textmining are with the following proteins: centrin 2, centrin 3, gamma-tubulin complex, polo-like kinase 1 and 4. Centrin 2 and 3 have fundamental roles in microtubule-organizing structure and function, while the gamma- tubulin complex is necessary for microtubule nucleation at the centrosome. Protein homology is also present between HsSfi1 and gamma-tubulin complex and polo-like kinase 1 and 4.

Part III: ProtParam

We obtained physical and chemical parameters of the full-length peptide, *Hs*Sfi1, using the ProtParam tool, which is part of the SIB. The parameters obtained include molecular weight, theoretical pI, extinction coefficient and amino acid composition.

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RYFKDGAVKK PYSAKTLSNK KSSASFGIRR

        70         80         90        100        110        120
ELPSTSHLVQ YRGTHTCTRQ GRLRELRIRC VARKFLYLWI RMTFGRVFPS KARFYYEQRL

       130        140        150        160        170        180
LRKVFEEWKE EWWVFQHEWK LCVRADCHYR YYLYNLMFQT WKTYVRQQQE MRNKYIRAEV

       190        200        210        220        230        240
HDAKQKMRQA WKSWLIYVVV RRTKLQMQTT ALEFRQRIIL RVWWSTWRQR LGQVRVSRAL

       250        260        270        280        290        300
HASALKHRAL SLQVQAWSQW REQLLYVQKE KQKVVSAVKH HQHWQKRRFL KAWLEYLQVR

       310        320        330        340        350        360
RVKRQQNEMA ERFHHVTVLQ IYFCDWQQAW ERRESLYAHH AQVEKLARKM ALRRAFTHWK

       370        380        390        400        410        420
HYMLLCAEEA AQFEMAEEHH RHSQLYFCFR ALKDNVTHAH LQQIRRNLAH QQHGVTLLHR

       430        440        450        460        470        480
FWNLWRSQIE QKKERELLPL LHAAWDHYRI ALLCKCIELW LQYTQKRRYK QLLQARADGH

       490        500        510        520        530        540
FQQRALPAAF HTWNRLWRWR HQENVLSARA TRFHRETLEK QVFSLWRQKM FQHRENRLAE

       550        560        570        580        590        600
RMAILHAERQ LLYRSWFMWH QQAAARHQEQ EWQTVACAHH RHGRLKKAFC LWRESAQGLR

       610        620        630        640        650        660
TERTGRVRAA EFHMAQLLRW AWSQWRECLA LRGAERQKLM RADLHHQHSV LHRALQAWVT

       670        680        690        700        710        720
YQGRVRSILR EVAARESQHN RQLLRGALRR WKENTMARVD EAKKTFQAST HYRRTICSKV

       730        740        750        760        770        780
LVQWREAVSV QMYYRQQEDC AIWEAQKVLD RGCLRTWFQR WWDCSRRSAQ QRLQLERAVQ

       790        800        810        820        830        840
HHHRQLLLEG LARWKTHHLQ CVRKRLLHRQ STQLLAQRLS RTCFRQWRQQ LAARRQEQRA

       850        860        870        880        890        900
TVRALWFWAF SLQAKVWATW LAFVLERRRK KARLQWALQA YQGQLLQEGA TRLLRFAASM

       910        920        930        940        950        960
KASRQQLQAQ QQVQAAHSLH RAVRRCATLW KQKVLGRGGK PQPLAAIAPS RKVTFEGPLL

       970        980        990       1000       1010       1020
NRIAAGAGDG TLETKRPQAS RPLGALGRLA AEEPHALELN TAHSARKQPR RPHFLLEPAQ

      1030       1040       1050       1060       1070       1080
```

SQRPQKPQEH GLGMAQPAAP SLTRPFLAEA PTALVPHSPL PGALSSAPGP KQPPTASTGP

1090 1100 1110 1120 1130 1140
ELLLLPLSSF MPCGAAAPAR VSAQRATPRD KPPVPSSLAS VPDPHLLLPG DFSATRAGPG

1150 1160 1170 1180 1190 1200
LSTAGSLDLE AELEEIQQQL LHYQTTKQNL WSCRRQASSL RRWLELNREE PGPEDQEVEQ

1210 1220 1230 1240
QVQKELEQVE MQIQLLAEEL QAQRQPIGAC VARIQALRQA LC

Number of amino acids: 1242

Molecular weight: 147663.5

Theoretical pI: 10.82

Amino acid composition:

Ala (A)	134	10.8%
Arg (R)	148	11.9%
Asn (N)	18	1.4%
Asp (D)	19	1.5%
Cys (C)	24	1.9%
Gln (Q)	128	10.3%
Glu (E)	79	6.4%
Gly (G)	39	3.1%
His (H)	61	4.9%
Ile (I)	24	1.9%
Leu (L)	147	11.8%
Lys (K)	67	5.4%
Met (M)	22	1.8%
Phe (F)	39	3.1%
Pro (P)	43	3.5%
Ser (S)	61	4.9%
Thr (T)	50	4.0%
Trp (W)	50	4.0%
Tyr (Y)	29	2.3%
Val (V)	60	4.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 98
Total number of positively charged residues (Arg + Lys): 215

Atomic composition:

Carbon	C	6587
Hydrogen	H	10383
Nitrogen	N	2071
Oxygen	O	1725
Sulfur	S	46

Formula: $C_{6587}H_{10383}N_{2071}O_{1725}S_{46}$
Total number of atoms: 20812

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 319710
Abs 0.1% (=1 g/l) 2.165, assuming all pairs of Cys residues form cystines

Ext. coefficient 318210
Abs 0.1% (=1 g/l) 2.155, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 58.25
This classifies the protein as unstable.

Aliphatic index: 78.49

Grand average of hydropathicity (GRAVY): -0.690

Summary

From the data obtained through ProtParam we can say that *Hs*Sfi1 is an unstable protein since its instability index is 58.25 (40 or less being the value for stable proteins). The Grand Average of Hydropathicity (GRAVY) value is -0.690, indicating that it is a hydrophilic protein.

Part IV: PyMOL and Py-ETV

We used PyMOL molecular visualization software to generate ribbon and ribbon and stick diagrams for the *Saccharomyces cerevisae* Sfi1 - Cdc31p (centrin homolog) complex with bound Ca^{2+}. Both diagrams contain three Sfi1 repeats and three centrins. The three Sfi1 repeats are colored purple while the first centrin is colored from blue to light blue, the second from green to light green and the third from yellow to orange, bound Ca^{2+} atoms are represented as red spheres. Using the Py-ETV plugin we generated molecular representations at 10%, 30% and 90% of identity for the chain A of the Cdc31/Sfi1 complex. The red segments represent conserved residues, while pink segments are not conserved. Chain A represents one of the bound centrins.

PyMOL

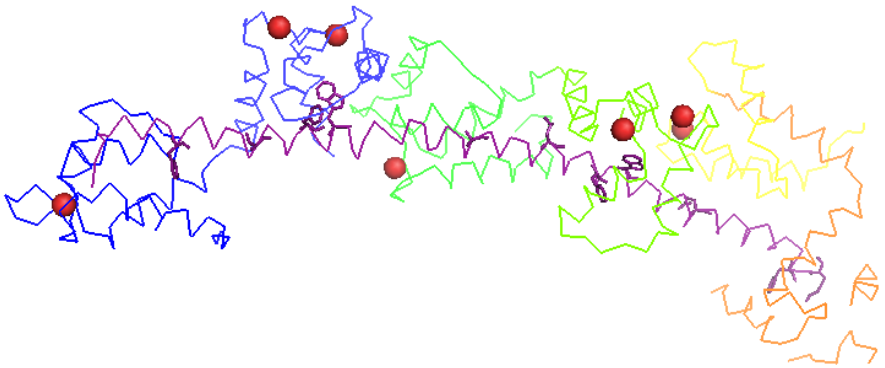


Figure A12. Carbon alpha trace representation for the Sfi1p-Cdc31p complex with Ca^{2+}. Generated using PyMOL, from Schrodinger. The three Sfi1 repeats are colored purple while the first centrin is colored from blue to light blue, the second from green to light green and the third from yellow to orange, bound Ca^{2+} atoms are represented as red spheres.

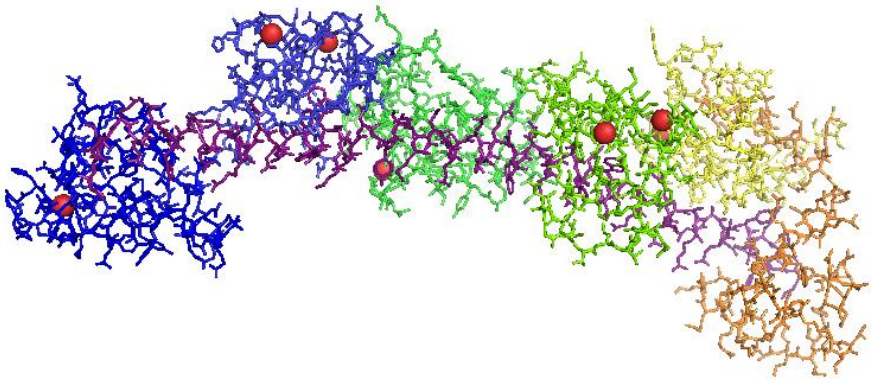


Figure A13. All atom representation for the Sfi1p-Cdc31p complex with Ca^{2+}. Generated using PyMOL, from Schrodinger. The three Sfi1 repeats are colored purple while the first centrin is colored from blue to light blue, the second from green to light green and the third from yellow to orange, bound Ca^{2+} atoms are represented as red spheres.

PyETV

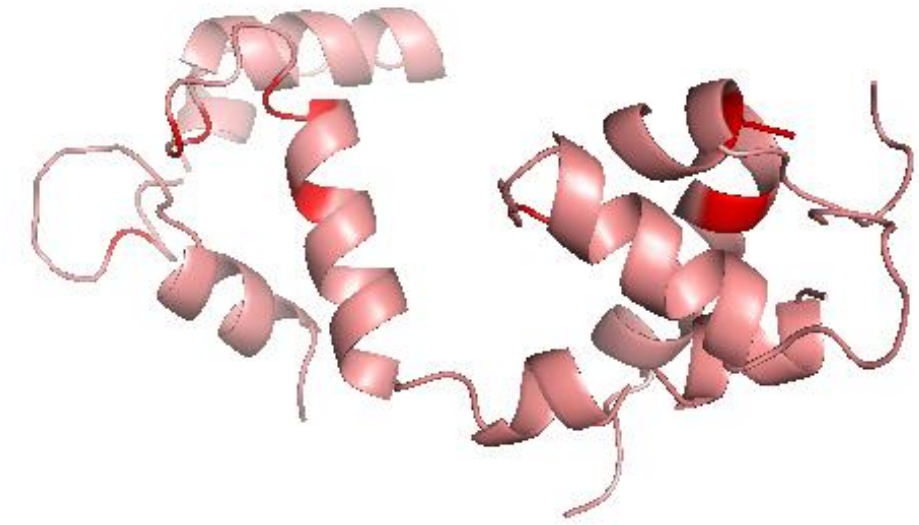


Figure A14. Evolutionary trace for chain A at 10% of identity, image A. Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues while pink segments are not conserved.

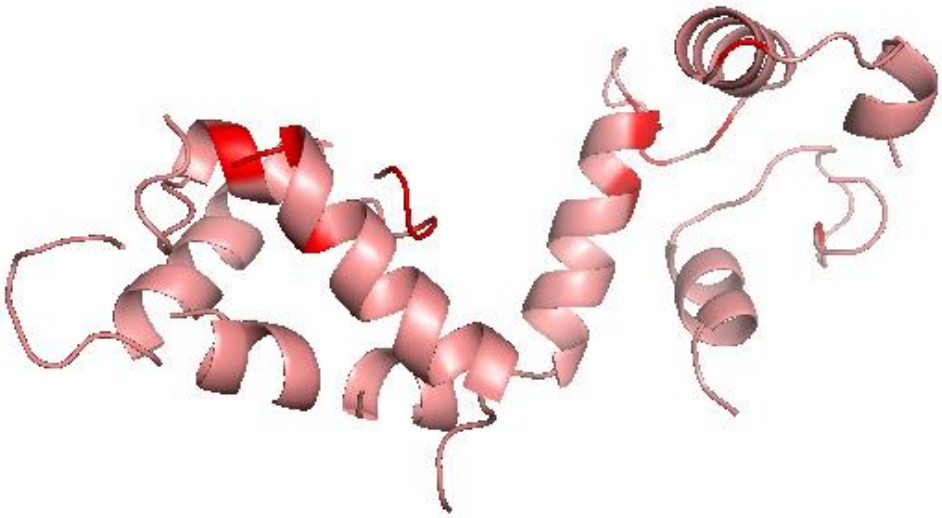


Figure A15. Evolutionary trace for chain A at 10% of identity, image B (rotated 180 on *x* axis). Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues, while pink segments are not conserved.

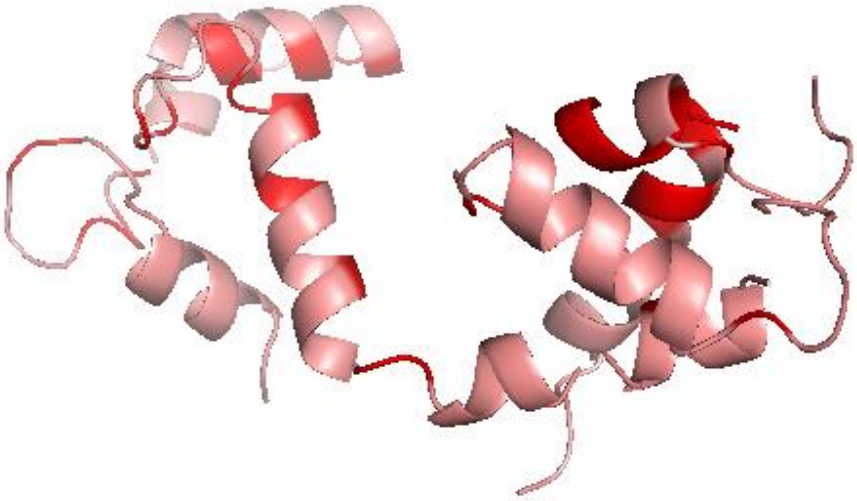


Figure A16. Evolutionary trace for chain A at 30% of identity, image A. Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues while pink segments are not conserved.

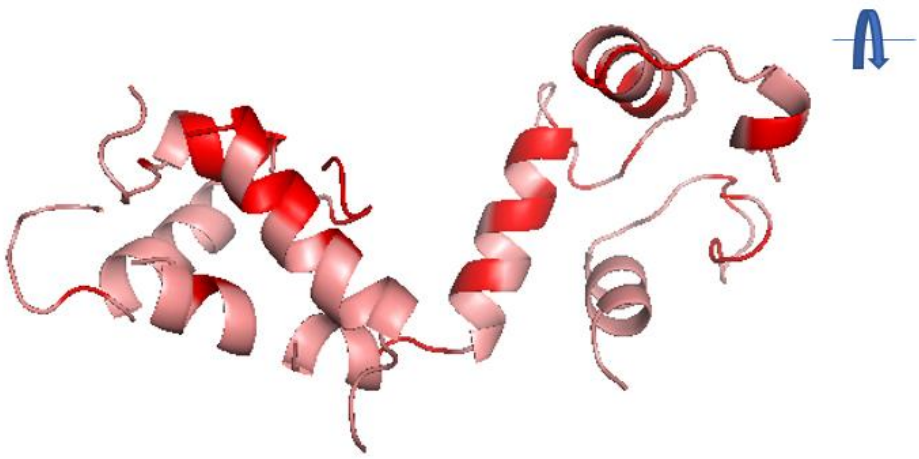


Figure A17. Evolutionary trace for chain A at 30% of identity, image B (rotated 180 on *x* axis). Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues while pink segments are not conserved.

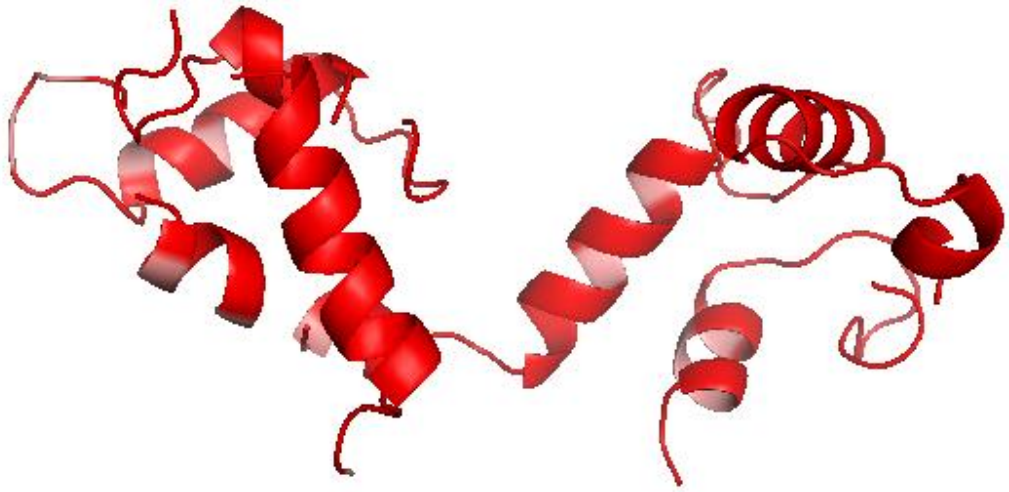


Figure A18. Evolutionary trace for chain A at 90% of identity, image A. Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues while pink segments are not conserved.

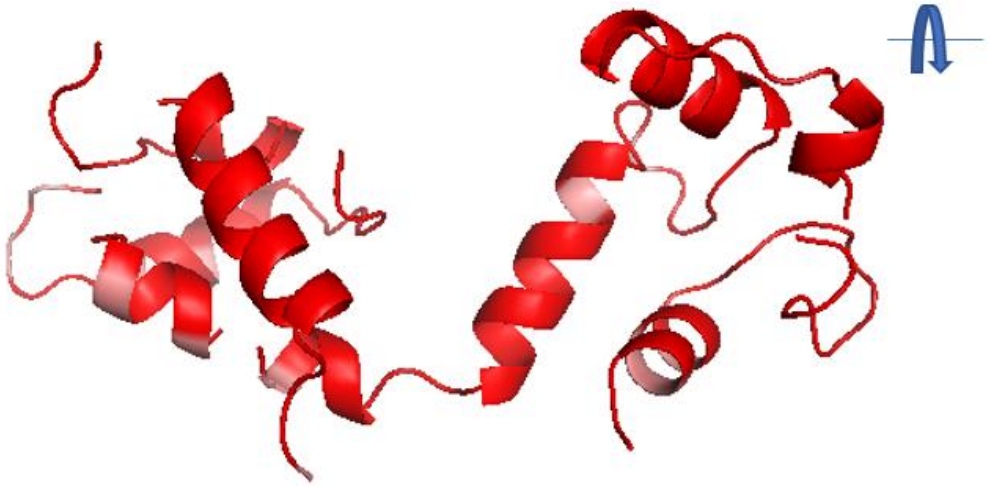


Figure A19. Evolutionary trace for chain A at 90% of identity, image B (rotated 180 on *x* axis). Generated using PyETV for PyMOL, from Schrodinger. The red segments represent conserved residues while pink segments are not conserved.

Summary

The Cdc31/Sfi1 complex was determined by Kilmartin in 2006. To generate the Cdc31/Sfi1 complex crystals they used a fragment of Sfi1, from N218-H306, which contained three CBS. They found that each CBS bound one centrin. They determined from the crystal structures that Sfi1 adopts an α-helical conformation that interacts with centrin in an extended conformation bound to each CBS. They also found that while Ca^{2+} may help to stabilize the structure; its absence did not affect conformation.

At 10% identity, seven of the 13 conserved residues at the interface with Sfi1 are not conserved. The non-conserved residues are the following: L118, A28, F154, E24, V121, E127 and L125. At 30% identity only four of the 13 residues in the 25% of the interface with Sfi1 are not conserved. The residues are the following: L118, F154 and E127. At 90% identity all residues at the interface with Sfi1 are conserved.

References

[1] H. Hartman and A. Fedorov, "The origin of the eukaryotic cell: A genomic investigation," *Proceedings of the National Academy of Sciences,* vol. 99, no. 3, pp. 1420-1425, 2002.

[2] R. Nishi, Y. Okuda, E. Watanabe, T. Mori, S. Iwai and F. Hanaoka, "Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein," *Molecular and Cellular Biology,* vol. 25, no. 13, pp. 5664-5674, 2005.

[3] J. V. Kilmartin, "Sfi1p has conserved centrin-binding sites and an essential function in budding yeast spindle pole body duplication," *Journal of Cell Biology,* vol. 162, no. 7, pp. 1211-1221, 2003.

[4] L. Duan, T.-Q. Wang, W. Bian, W. Liu, Y. Sun and B.-S. Yang, "Centrin: Another target of monastrol, an inhibitor of mitotic spindle," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 137, pp. 1086-1091, 2015.

[5] S. Gannavaram, J. Torcivia, L. Gasparyan, N. Ismail, V. Simonyan and H. Nakhasi, "Whole genome sequencing of live attenuated Leishmania donovani parasites reveals novel biomarkers of attenuation and enables product characterization," *Nature Scientific Reports,* vol. 7, no. 4718, pp. 1-10, 2017.

[6] A. Shumilov, M.-H. Tsai, Y. T. Schlosser, A. Kratz, K. Bernhardt, S. Fink, T. Mizani, X. Lin, A. Jauch, J. Mautner, A. Kopp-Schneider, R. Feederle, I. Hoffmann and H.-J. Delecluse, "Epstein–Barr virus particles induce centrosome amplification and chromosomal instability," *Nature Communications,* vol. 8, 2017.

[7] E. Nigg, "Origins and Consequences of Centrosome Aberrations in Human Cancers," *International Journal of Cancer,* vol. 119, pp. 2717-2723, 2006.

[8] T. Stearns and E. Firat-Karalar, "The Centriole Duplication Cycle," *Philosophical Transactions of the Royal Society of London,* vol. 369, pp. 1-10, 2014.

[9] E. Nigg and J. W. Raff, "Centrioles, Centrosomes, and Cilia in Health and Disease," *Cell,* vol. 113, pp. 663-678, 2009.

[10] J. L. Salisbury, "Striated Flagellar Roots: Isolation and Partial Characterization of a Calcium-modulated Contractile Organelle," *Journal of Cell Biology,* vol. 99, pp. 962-970, 1984.

[11] J. L. Salisbury and M. A. Sanders, "Centin Plays an Essential Role in Microtubule Severing during Flagellar Excision in Chlamydomonas reinhardtii," *Journal of Cell Biology,* vol. 124, no. 5, pp. 795-805, 1994.

[12] B. Koblenz, J. Schoppmeier, A. Grunow and K.-F. Lechtreck, "Centrin deficiency in Chlamydomonas causes defects in basal body replication, segregation and maturation," *Journal of Cell Science,* vol. 116, no. 13, pp. 2635-2646, 2013.

[13] F. Ruiz, N. G. de Loubresse, C. Klotz, J. Beisson and F. Koll, "Centrin Deficiency in Paramecium Affects the Geometry of Basal-Body Duplication," *Current Biology,* vol. 15, pp. 2097-2106, 2005.

[14] J. L. Salisbury, K. M. Suino, R. Busby and M. Springett, "Centrin-2 Is Required for Centriole Duplication in Mammalian Cells," *Current Biology,* vol. 12, pp. 1287-1292, 2002.

[15] S. Li, A. Sandercock, P. Conduit, C. Robinson, R. Williams and J. Kilmartin, "Structural Role of Sfi1p-centrin Filaments in Budding Yeast Spindle Pole Body Duplication," *The Journal of Cell Biology,* vol. 173, no. 6, pp. 867-877, 2006.

[16] J. Kilmartin, "Sfi1p has conserved centrin-binding sites and an essential function in budding yeast spindle pole body duplication," *The Journal of Cell Biology,* vol. 162, no. 7, pp. 1211-1221, 2003.

[17] E. A. Nigg and T. Stearns, "The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries," *Nature Cell Biology,* vol. 13, no. 10, pp. 1154-1160, 2014.

[18] J. Salisbury, A. D'Assoro and W. Lingle, "Centrosome Aplification and the Origin of Chromosomal Instability in Breast Cancer," *Journal of Mammary Gland Biology and Neoplasia,* vol. 9, no. 3, pp. 275-283, 2004.

[19] T. Avidor-Reiss, A. Maer, E. Koundakjan, A. Polyanovsky, T. Keil, S. Subramaniam and C. Zuker, "Decoding Cilia Function: Defining Specialized Genes Requiered for Compartmentalized Cilia Biogenesis," *Cell,* vol. 117, no. 4, pp. 527-539, 2004.

[20] D. A. Brito, S. Montenegro-Gouveia and M. Bettencourt-Días, "Deconstructing the centriole: structure and number control," *Current Opinion in Cell Biology,* vol. 24, no. 4, pp. 4-13, 2012.

[21] J. Azimzadeh and M. Bornens, "Structure and Duplication of the Centrosome," *Journal of Cell Science,* vol. 120, no. 13, pp. 2139-2142, 2007.

[22] M. Winey and E. O'Toole, "Centriole Structure," *Philosophical Transactions of the Royal Society B,* vol. 369, 2014.

[23] J. L. Salisbury, "A Mechanistic view on the evolutionary origin for centrin-based control of centriole duplication," *Journal of Cellular Physiology,* vol. 213, no. 1, pp. 420-428, 2007.

[24] D. Zyss and F. Gergely, "Centrosome function in cancer: guilty of innocent?," *Trends in Cell Biology,* pp. 334-346, 2009.

[25] J. Woodruff, O. Wuuseke and A. Hyman, "Pericentriolar Material Structure and Dynamics," *Philosophical Transactions of the Royal Society of London,* vol. 369, 2014.

[26] T. Kobayashi and B. D. Dynlacht, "Regulating the transition from centrioleto basal body," *Journal of Cell Biology,* vol. 193, no. 3, pp. 435-444, 2011.

[27] M. Winey and C. Kilburn, "Basal Bodies," *Current Biology,* vol. 18, no. 2, pp. 56-57, 2008.

[28] S. Jaspersen and M. Winey, "The Budding Yeast Spindle Pole Body: Structrure, Duplication, and Function," *Annual Review Cell Developmental Biology,* vol. 20, pp. 1-28, 2004.

[29] P. Conduit, A. Wainman and J. Raff, "Centrosome Function and Assembly in Animal Cells," *Nature Reviews: Molecular Cell Biology,* vol. 16, pp. 611-624, 2015.

[30] A. Paoletti, M. Moudjou, M. Paintrand, J. Salisbury and M. Bornens, "Most of centrin in animal cells is not centrosome-associated and centrosomal centrin is confined to the distal lumen of centrioles," *Journal of Cell Science,* vol. 109, pp. 3089-3102, 1996.

[31] B. Pastrana-Ríos and W. Ocaña, "Calcium titration of Chlamydomonas reinhardtii centrin and its structural changes," *Journalof Molecular Structure,* vol. 1069, pp. 73-78, 2014.

[32] K. K. Resendes, B. A. Rasala and D. J. Forbes, "Centrin 2 localizes to the vertebrate nuclear pore and plays a role in mRNA and protein export," *Molecular and Cellular Biology,* vol. 28, no. 5, pp. 1755-1769, 2008.

[33] B. Pastrana-Ríos, M. Reyes, J. De Orbeta, V. Meza, D. Narváez, A. M. Gómez, A. Rodriguez-Nassif and A. Ortiz, "Relative Stability of Human Centrins and Its Relationship to Calcium," *Biochemistry,* vol. 52, pp. 1236-1248, 2013.

[34] I. R. Adams and J. V. Kilmartin, "Localization of Core Spindle Pole Body (SPB) Components during SPB Duplication in Saccharomyces cerevisiae," *Journal of Cell Biology,* vol. 14, no. 4, pp. 809-823, 1999.

[35] P. Trojan, N. Krauss, H.-W. Choe, A. Gießl, A. Pulvermuller and U. Wolfrum, "Centrins in retinal photoreceptor cells: Regulators in the connecting cilium," *Progress in Retinal and Eye Research,* vol. 27, no. 3, pp. 237-259, 2008.

[36] G. Manandhar, H. Schatten and P. Sutovsky, " Centrosome Reduction During Gametogenesis and Its Significance," *Biology of Reproduction,* vol. 72, pp. 2-13, 2005.

[37] I. Hinduja, N. Baliga and K. Zaveri, "Correlation of human sperm centrosomal proteins with infertility," *Journal of Human Reproductive Sciences,* vol. 3, no. 2, pp. 95-101, 2010.

[38] P. Avasthi, J. F. Scheel, G. Ying, J. Frederick, W. Baehr and U. Wolfrum, "Germline deletion of Cetn1 causes infertility in male mice," *Journal of Cell Science,* vol. 126, pp. 3204-3213, 2013.

[39] D. Grecu, V. P. Irudayaraj, J. Martínez-Sanz, J.-M. Mallet and L. Assairi, "A chirality change in XPC- and Sfi1-derived peptides affects their affinity for centrin," *Peptides,* vol. 78, pp. 77-86, 2016.

[40] A. Díaz-Casas, W. J. Chazin and B. Pastrana-Ríos, "Prp40 Homolog A Is a Novel Centrin Target," *Biophysical Journal,* vol. 112, no. 12, pp. 2529-2539, 2017.

[41] A. Díaz-Casas, "Thermodynamics Governing Binding of Homo sapiens Centrin – Homo sapiens Sfi1p21 Complexes," *(Master's thesis),* p. Retrieved from http://grad.uprm.edu/oeg/TesisDisertacionesDigitales/Quimica/#2012, 2012.

[42] J. H. Lee and J. G. Gleeson, "The role of primary cilia in neuronal function," *Journal of Neurobiological Disorders,* vol. 38, no. 2, pp. 167-172, 2010.

[43] J. A. Green and K. Mykytyn, "Neuronal Primary Cilia: An Underappreciated Signaling and Sensory Organelle in the Brain," *Nuropsychopharmacology Reviews,* vol. 39, pp. 244-245, 2014.

[44] E. A. Nigg, "Centrosome duplication: of rules and licenses," *Trends in Cell Biology,* vol. 17, no. 5, pp. 215-221, 2007.

[45] J. L. Badano, N. Mitsuma, P. L. Beales and N. Katsanis, "The Ciliopathies: An Emerging Class of Human Genetic Disorders," *Annual Review of Genomics and Human Genetics,* vol. 7, pp. 125-148, 2006.

[46] P. Van Dijck, P. Ma, J. Winderikx, D. Nauwelaers, F. Dumortier, A. De Doncker and J. M. Thevelein, "Deletion of SFI1, a Novel Suppressor of Partial Ras–cAMP Pathway Deficiency in the Yeast Saccharomyces cerevisiae, Causes G2 Arrest," *Yeast,* vol. 15, no. 11, pp. 1097-1109, 1999.

[47] J. L. Salisbury, "Centrosomes: Sfi1p and Centrin Unravel a Structural Riddle," *Current Biology,* vol. 14, pp. R27-R29, 2004.

[48] J. Martinez-Sans, A. Yang, Y. Blouquit, P. Duchambon, L. Assairi and C. Craescu, "Binding of Human Centrin 2 to the Centrosomal Protein HSfi1," *Federation of European Biochemical Societies,* vol. 273, pp. 4504-4515, 2006.

[49] N. C. Price, S. M. Kelly and T. J. Jess, "How to study proteins by circular dichroism," *Biochimica et Biophysica Acta,* vol. 1751, pp. 119-139, 2005.

[50] I. N. Serdyuk, N. R. Zaccai and J. Zaccai, Methods in Molecular Biophysics, Cambridge: Cambridge University Press, 2007.

[51] N. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure.," *Nature Protocols,* vol. 1, no. 6, pp. 2876-2890, 2006.

[52] M. W. Freyer and E. A. Lewis, "Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions," *Methods in Cell Biology,* vol. 84, pp. 79-113, 2008.

[53] E. Freire, "Isothermal Titration Calorimetry: Controlling Binding Forces in Lead Optimization," *Drug Discovery Today: Technologies,* vol. 1, no. 3, pp. 295-299, 2004.

[54] M. M. Pierce, C. Raman and B. T. Nall, "Isothermal Titration Calorimetry of Protein–Protein Interactions," *Methods,* vol. 19, no. 2, pp. 213-221, 1999.

[55] C. Pearson, T. Giddings and M. Winey, "Basal Body Components Exhibit Differential Protein Dynamics during Nascent Basal Body Assembly," *Molecular Biology of the Cell,* vol. 20, pp. 904-914, 2009.