EXPRESSION, ISOLATION, PURIFICATION, AND CHARACTERIZATION OF RECOMBINANT HUMAN SFi1p$_{1-2}$

by

Marie Cely Rosado Arroyo

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

MASTER OF SCIENCE

in

CHEMISTRY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2013

Approved by:

Robert Ríos, Ph.D. Member, Graduate Committee	Date
Jorge Ríos-Steiner, Ph.D. Member, Graduate Committee	Date
Belinda Pastrana- Ríos, Ph.D. President, Graduate Committee	Date
Silvestre Colón, M.S./M.A. Representative of Graduate Studies	Date
René Vieta, Ph.D. Chairperson of the Department	Date

ABSTRACT

Homo sapiens Sfi1 (*Hs* Sfi1) is a 1242 amino acid protein containing 23 tandem binding sites for a calcium binding protein known as centrin. *Hs* Sfi1p_{1-2} is an 8,592 Da peptide that corresponds to the first two centrin binding sites (CBS) within *Hs* Sfi1. Both proteins are constituents of contractile fibers in the centrosome and play essential roles in centriole duplication and separation. The objectives of this work were to: express, isolate, purify, and biochemically characterize recombinant *Hs* Sfi1p_{1-2}. A bacterial stock culture of *Escherichia coli* BL21-(DE3) RIL was transformed with pET 100 expression vector containing His tagged-*Hs* Sfi1p_{1-2}. His tagged-*Hs* Sfi1p_{1-2} recombinant peptide has a molecular weight of 12,119 Da. The bacterial stock was grown in a five liter bench scale fermentor up to log phase and induced with isopropyl-ß-thiogalactoside (IPTG). Following the successful expression of recombinant His-*Hs* Sfi1p_{1-2}, the supernatant was subjected to His-tag affinity and anion exchange chromatography and a band was observed near the expected molecular weight of ~12 kDa by 4-20% (Bis-Tris) gradient SDS-PAGE. However, based on the elution pattern and UV/Vis analysis it was suspected that the recombinant peptide stayed in the pellet. An alternative isolation process was performed by an extraction method with organic solvents using $CHCl_3$:CH_3OH (2:1, v/v) extraction followed by a $CHCl_3$:CH_3OH (1:1, v/v) volume ratios. A 4-20% (Bis-Tris) gradient SDS-PAGE revealed the presence of a protein with similar molecular weight in both the aqueous and the organic phases. Partial amino acid sequencing confirmed the

presence of *Hs* $Sfi1p_{1-2}$ to be in the aqueous phase of the extraction. Alternate purification step involved subjecting the protein sample to size exclusion chromatography.

RESUMEN

Homo sapiens Sfi1 (*Hs* Sfi1) es una proteína de 1242 amino ácidos que contiene 23 lugares de enlace para la proteína enlazante de calcio centrin. *Hs* Sfi1p$_{1-2}$ es un péptido con un peso molecular de 8,592 Da que corresponde a los primeros dos lugares de enlace de centrin, dentro de *Hs* Sfi1. Ambas proteínas forman parte de las fibras contráctiles en el centrosoma y poseen un rol esencial en la separación y duplicación del centriolo. Los objetivos de este trabajo lo fueron: expresar, aislar, purificar y caracterizar el péptido recombinante *Hs* Sfi1p$_{1-2}$. Un cultivo bacteriano de *Escherichia coli* BL21-(DE3) RIL fué transformado con el vector de expresión pET 100 que contenía la secuencia de His-*Hs* Sfi1p$_{1-2.}$ El péptido recombinante His-*Hs* Sfi1p$_{1-2}$ posee un peso molecular de 12,119 Da. Las células transformadas fueron crecidas en un fermentador de cinco litros hasta alcanzar la fase log e inducidas mediante la adición de isopropil-ß-tiogalactosidasa (IPTG). Luego de la exitosa expresión de His-*Hs* Sfi1p$_{1-2}$, el sobrenadante fue sometido a cromatografía de afinidad por el contenido en su secuencia de histidinas y cromatografía de intercambio aniónico, en donde una banda pura con el peso molecular esperado fue observada por la técnica de SDS-PAGE (por sus siglas en inglés). Luego de analizar los patrones de elución y la absorción en la región espectral ultravioleta, se comenzó a sospechar que *Hs* Sfi1p$_{1-2}$ estaba permaneciendo en el precipitado de la centrifugación. Un procedimiento alternativo de aislación se realizó mediante extracción por solventes orgánicos utilizando $CHCl_3$:CH_3OH (2:1, v/v) seguido de otra extracción $CHCl_3$:CH_3OH (1:1, v/v) razones por volumen. La técnica de gradiente 4-20% (Bis-Tris) SDS-PAGE reveló la presencia de

una proteína con peso molecular similar (~12 kDa) para ambas fases: la acuosa y orgánica y la técnica de secuenciación parcial de amino ácido reveló la presencia de *Hs* Sfi1p_{1-2} en la fase acuosa de la extracción. La estrategia de purificación alterna consistió en utilizar cromatografía por exclusión de tamaño.

DEDICATION

*To **my family,** for their support and motivation throughout my scientific career.*

*To **my friends,** for being there all the time and for always putting a smile on my face.*

*To **those,** who are suffering from or lost the battle against cancer.*

ACKNOWLEDGEMENTS

First, I want to thank the Chemistry Department of the University of Puerto Rico at Mayagüez for accepting me as a graduate student and providing financial support. I would like to express my deepest gratitude to my research advisor, Dr. Belinda Pastrana-Ríos, for the opportunity of developing this research work under her guidance and provided a wealth of helpful suggestions and advice. I would like to acknowledge Tufts University for their services in sequencing and mass spectral analysis and the National Institutes of Health (NIH) for providing the research funds and resources that allowed for the development of this research project.

I also want to thank the Bioprocess Development and Training Center (BDTC), for letting me use their facilities at the beginning of my research and especially to Zuleika Sanoguet for the training and guidance provided. Kind appreciation is expressed to the personnel at Abbott Biotechnology Ltd. who inspire and encourage me to pursue graduate studies in Biochemistry. My deepest gratitude to the personnel at PR05 Lilly del Caribe, for giving me an internship opportunity and provide their support during the final experiments of this work.

I am especially grateful to, Ana María and Daniel, for teaching and giving me direction in laboratory work and for their wise advice in difficult times. I admire and acknowledge the work of all the undergraduate students, who contribute for the excellence in Dr. Pastrana's laboratory, especially to José Robles, an example of humbleness, dedication and commitment to our scientific work. I thank my graduate

colleagues at the laboratory for helping me with their ideas and skills every time I needed. I want to express my deepest appreciation for two exceptional persons, Ruth Almodóvar and Adalberto Díaz, for influencing my research work and my personal development in such a positive way. I am grateful for their tremendous support and for sharing with me their continuous vision of success and excellence. So many thanks to all the people directly involved in this research work with me and my most sincere apologies to those I have not named.

Table of Contents

ABSTRACT II
RESUMEN... IV
ACKNOWLEDGEMENTS..VII
TABLE OF CONTENTS..IX
LIST OF FIGURES..XI

1 INTRODUCTION 1

2 OBJECTIVES 4

3 LITERATURE REVIEW 5

3.1 The Centrosome: Structural Features and Duplication 5
3.2 Functions of the Centrosome 11
3.3 Centrosome Role in Disease... 12
3.4 Centrin Protein: Localization and Structural Features 16
3.5 Sfi1 Protein: Localization and Structural Features 19
3.6 Studies of Sfi1-Centrin and Sfi1-Cdc31 Complexes 23

4 EXPERIMENTAL METHODS 29

4.1 Bacterial protein expression of *Hs* Sfi1p_{1-2} 29
4.2 Isolation of *Hs* Sfi1p_{1-2} 30
4.3 Purification by Affinity and Anion Exchange Chromatography.........32
4.4 Alternate Isolation by Extraction Method 35
4.5 Alternate Purification by Size Exclusion Chromatography 37

5 RESULTS DISCUSSION.. 38

5.1 Bacterial protein expression of *Hs* Sfi1p_{1-2}.............................. 38
5.2 Isolation of *Hs* Sfi1p_{1-2}.. 41
5.3 Purification by Affinity and Anion Exchange Chromatography.........44
5.4 Alternate Isolation by Extraction Method 53
5.5 Alternate Purification by Size Exclusion Chromatography........... 62

6 CONCLUSIONS AND FUTURE WORK 66

REFERENCES 68

APPENDIX A. **73**
BIOGRAPHICAL SKETCH. **144**

List of Figures

FIGURES **Page**

Figure 1. Histidine tagged *Homo sapiens* Sfi1p$_{1-2}$ amino acid sequence 3

Figure 2. The centrosome of a human cell..7

Figure 3. The spindle pole body (SPB) in yeasts....................................8

Figure 4. Duplication of the centrosome and the cell cycle....................... 10

Figure 5. The structure of the primary cilium and intraflagellar transport 15

Figure 6. Full length centrin (*Hs*cen2) comprised of four EF-hands 18

Figure 7. The consensus sequence of Sfi1 in different species 21

Figure 8. A model for Sfi1-centrin contraction..................................... 22

Figure 9. Sequence comparison of Sfi1 repeats in three organisms. 24

Figure 10. Duplication of the yeast SPB regulated by Sfi1-centrin complex. ... 25

Figure 11. Crystal structure of Sfi1p-Cdc31 complex (PDB ID 2DOQ). 27

Figure 12. Model for the initiation of SPB duplication 28

Figure 13. Bacterial expression, isolation and purification of His-*Hs* Sfi1p$_{1-2}$... 34

Figure 14. Alternate extraction method for isolation 36

Figure 15. Growth curve of *E. coli* transformed cells with pET-100 39

Figure 16. SDS-PAGE of the bacterial pellet lysate after high level expression 40

Figure 17. 4-20% (Bis-Tris) gradient SDS-PAGE of the isolation process 43

Figure 18. Cobalt Affinity Chromatogram 46

Figure 19. 4-20% (Bis-Tris) SDS-PAGE of affinity chromatography fractions 47

Figure 20. Anion exchange chromatogram and 4-20% (Bis-Tris) gradient SDS-PAGE for a step gradient of 1-1.0 M NaCl................................ 50

Figure 21. Characteristic UV/Vis spectra of Hs Sfi1p$_{1-2}$ after anion exchange chromatography 51

Figure 22. Anion exchange chromatogram and 4-20% (Bis-Tris) gradient SDS-PAGE for a step gradient of 0-0.70 M NaCl 52

Figure 23. Expasy ProtParam parameters comparison of His- *Hs* Sfi1p$_{1-2}$ and *Hs* Sfi1p$_{1-2}$ 55

Figure 24. Pellet 1 dilutions (D1-D3) to identify a protein at 12 kDa 56

Figure 25. SDS-PAGE of the Aqueous Phase (AP) and the Organic Phase (OP) for $CHCl_3$:CH_3OH (2:1, v/v) 57

Figure 26. SDS-PAGE of the Aqueous Phase (AP) and the Organic Phase for $CHCl_3$:CH_3OH (1:1, v/v).. 59

Figure 27. SDS-PAGE of consecutive extractions for the Aqueous Phase (A), the Organic Phase (B) and a comparison of relative mobility for both phases (C) 60

Figure 28. Partial amino acid sequencing results from the proteomics facility at Tufts University 61

Figure 29. Size exclusion chromatogram of the aqueous phase sample 63

Figure 30. SDS-PAGE of size exclusion pooled fractions indicating the presence of *Hs* Sfi1p$_{1-2}$ 64

Figure 31. Characteristic UV/Vis spectra of *Hs* Sfi1p$_{1-2}$ fraction after size exclusion chromatography 65

1 INTRODUCTION

To date, the scientific challenge is the understanding of vital biological processes and events related to diseases. Most of these diseases are specifically related to defects in protein-protein interactions (PPI's). Protein-protein interactions involve hydrophobic, ionic and hydrogen bonding interactions, among others, between specific side chains or backbone. These interactions are important for several reasons including: signal transduction, protein complex formation and protein conformational changes associated with the interaction. Thus, while the analysis of the structure and function of individual proteins is crucial for the understanding of their role in biological processes, it has a limited capability to explain the processes themselves. Information about these interactions improves our understanding of diseases and can provide the basis for new therapeutic approaches.

For the purpose of this study, our goal was to express, isolate, purify, and characterize Homo *sapiens (Hs)* Sfi1p_{1-2} peptide. *Hs* Sfi1p_{1-2} peptide corresponds to the first two centrin binding sites (CBS) of *Hs* Sfi1. *Hs* Sfi1 is comprised of 1242 amino acids which include 23 tandem CBS [19]. These 23 CBS are not well conserved and are expected to have variability in the affinity for *Hs* centrin. These binding sites are composed of 33 amino acids, and upon closer examination of the sequences, a pattern of hydrophobic residues within these sites can be established as $AX_7LLX_3F/LX_2WK/R$ (15). The analysis includes the use of the recombinant Histidine tagged (His tag) *Hs* Sfi1p_{1-2}. The His tag will potentially allow for the purification of His-*Hs* Sfi1p_{1-2}

through affinity chromatography. Figure 1, shows the sequence of His-*Hs* Sfi1p_{1-2} recombinant peptide. It has an additional sequence of 96 amino acids at the N-terminal end of the peptide fragment resulting in a molecular weight of 12,119 Da.

Kilmartin and co-workers have determined that *Hs c*en2 binds to *Hs* Sfi1 in yeast and HeLa cells, in events governing microtubule organization within these eukaryotic cells [19-20]. Eukaryotic cells have a very important organelle known in general as the microtubule organizing center (MTOC) which is important in cell cycle regulation and division [3]. Our long term goal will include further experiments to study the particular interaction with *Hs* Sfi1p_{1-2} (centrin binding sites 1 and 2) and wild type *Hs* variant *Hs* cen2 (E105K), using different biophysical techniques in order to elucidate the thermodynamics governing binding, complex stability, and the conformational changes involved in complex formation. This will aid in the future design of new treatments for diseases like cancer, which involves aberrant chromosome segregation and uncontrolled cell division, as a consequence of protein defects and aberrant protein interactions [21].

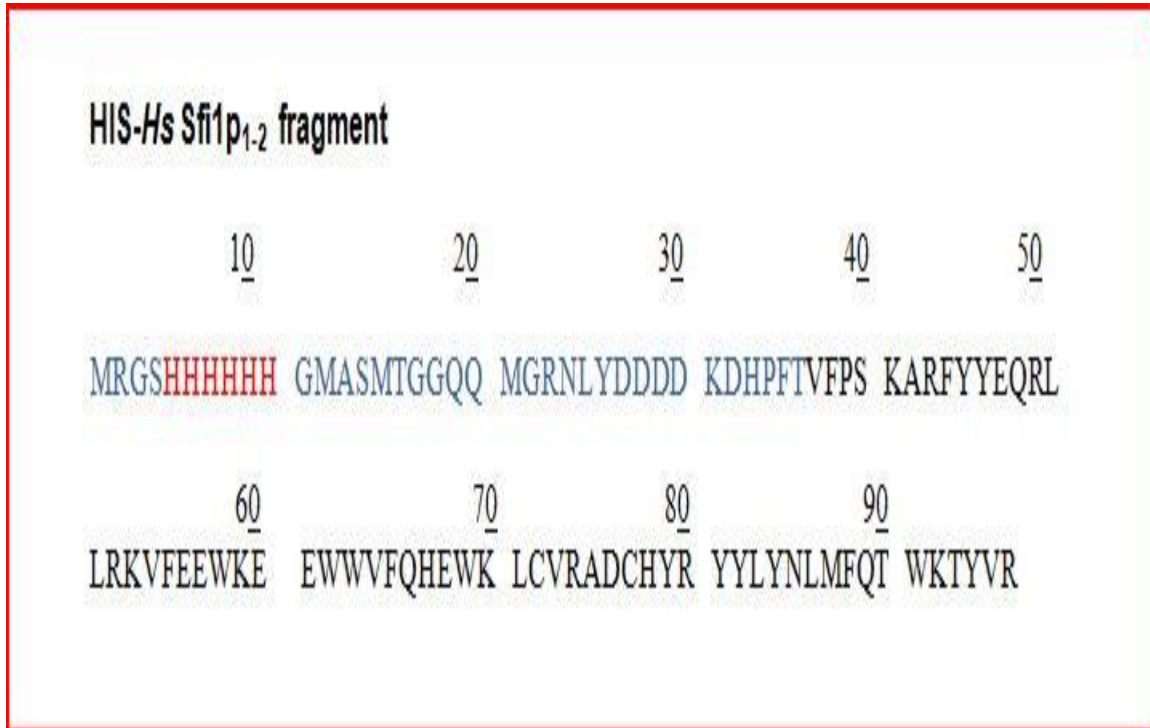


Figure 1. Histidine tagged *Homo sapiens* Sfi1p$_{1\text{-}2}$ amino acid sequence. The blue letters represents the sequence of amino acids, which includes the His tag (red letters). Black letters represent the sequence of interest, corresponding to the first two domains of *Hs* Sfi1.

2 OBJECTIVES

Our primary goal is to express, isolate, purify, and characterize *Hs* Sfi1p$_{1-2}$ which will aid in future studies of *Hs* Sfi1p$_{1-2}$-*Hs* centrin complex studies and the design of novel inhibitors for this protein-protein interaction.

1. To over express *Hs* Sfi1p$_{1-2}$, using bacterial cells (*E. coli* BL21 (DE3) RIL) in a 5L bench scale fermentor.

2. To isolate and purify recombinant peptide *Hs* Sfi1p$_{1-2.}$

3. To perform the biochemical characterization of *Hs* Sfi1p$_{1-2.}$

3 LITERATURE REVIEW

3.1 The Centrosome: Structural Features and Duplication

The centrosome is an organelle in animal cells that nucleate microtubules and plays a critical role in mitotic spindle orientation and in genome stability [1]. This cell organelle is mainly characterized by being a microtubule organizing center (MTOC) in higher eukaryotes which usually establishes the number, direction and polarity of the microtubules [2]. It lacks a membrane that separates it from the rest of the cytoplasm and it has a specific status in cell organization and evolution. The centrosome was discovered by Edward Van Beneden in 1876, during the study of the anatomy and development of a group of marine parasites called the Dicyemidae [3]. Theodore Boveri described it in 1888, which was in agreement with Van Beneden past works, suggested that the centrosome is a permanent cell organelle endowed with the property of self-replication [3]. This organelle has remained enigmatic and the subject of extensive study ever since. In yeasts, the equivalent organelle is the half bridge of the spindle pole body (SPB). The SPB is a multilayered structure embedded in the nuclear envelope and is responsible for the organization of the spindle and cytoplasmic microtubules (Figure 3) [4]. The half bridge is a specialized area of the nuclear envelope and it has a critical role during SPB duplication [4].

Centrosomes are composed of two perpendicular arranged centrioles surrounded by an amorphous mass of protein termed the pericentriolar material (PCM) comprised of an estimated 150 proteins. These two microtubule based-cylinders are of defined length and diameter with a 9+0 microtubule symmetry arrangement. They are linked

together by a matrix consisting of coiled-coil proteins of the pericentrin family which anchor other matrix components. The centriole pair displays structural and functional asymmetry due to the generational difference between each member of the pair: The old, fully mature, mother centriole is distinguished by two sets of nine appendages at its distal end while the young, immature, daughter centriole, assembled during the previous cell cycle, is about 80% the length of the mother centriole as shown in Figure 2 [1]. In resting cells, the mother centriole can turn into a basal body, by docking to the plasma membrane through the distal appendages, where it templates a non-motile primary cilium that serves as a sensory organelle [1,5]. In general, each centriole structure is based on a nine triplet microtubule assembled in a cartwheel, and contains several proteins including: centrin, Sfi1, cenexin, α-, β-, γ-tubulin, dynein, ninein, centriolin, cep170, Sas6, HPOC5, galetin3, and tektin [3, 6-10]. Recently, the structural basis of the highly conserved ninefold radial symmetry of the centriole/basal body has been elucidated [6, 7]. It rests on the oligomerization of a single coiled-coil protein, SAS-6/Bld12p, which forms a cartwheel structure acting as a scaffold for centriole/basal body assembly [7].

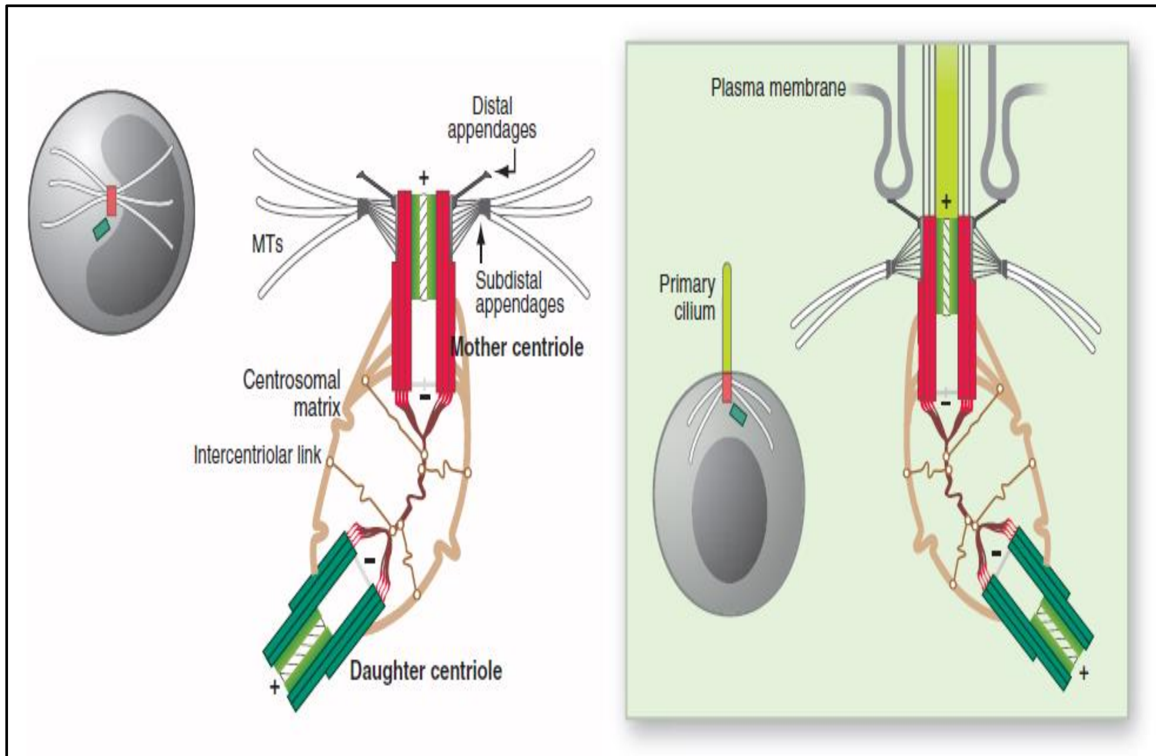


Figure 2. The centrosome of human cells. It contains a structurally and functionally asymmetric pair of centrioles, the mother centriole (red) and a daughter centriole (green). The mother centriole is distinguished by two sets of nine appendages at its distal end, which are required for anchoring microtubules (MTs) and for docking the mother centriole at the plasma membrane during ciliogenesis in quiescent cells. The daughter centriole is about 80% the length of the mother centriole (*Adapted from* Bornens et al. 2012 [1]).

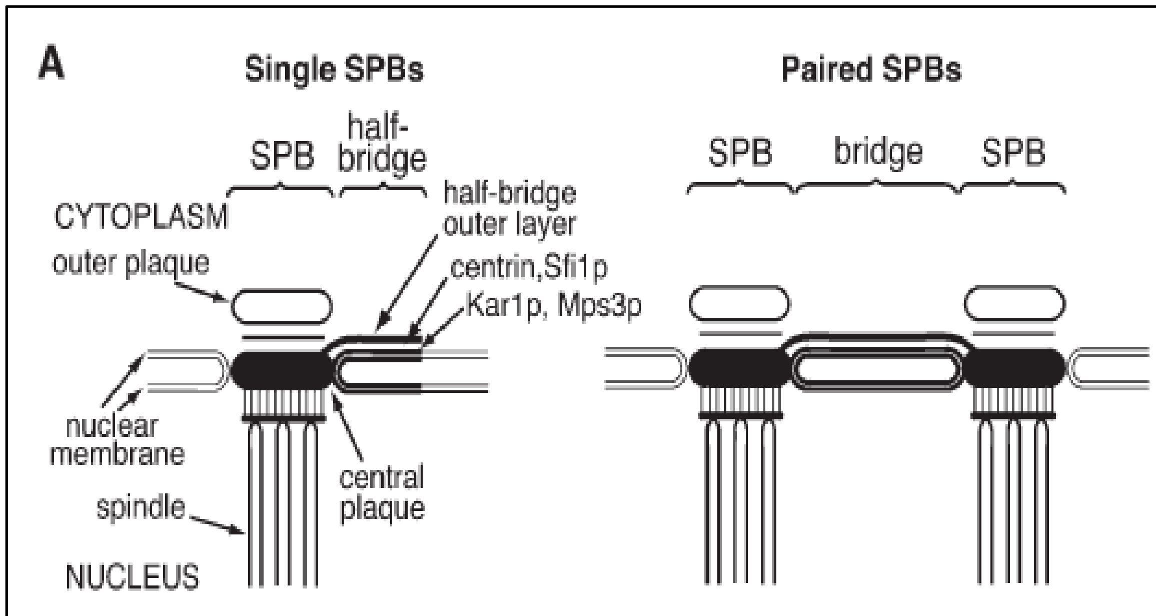


Figure 3. The spindle pole body (SPB) in yeasts. A single and paired SPB multilayered structure showing the localization of the half bridge (*Adapted from* Kilmartin, 2003 [4]).

In each cell cycle, the centrosome is duplicated and the resultant two centrosomes, organize the microtubule array of the mitotic spindle, allowing equal segregation of sister chromatids into each of the two daughter cells.

In higher eukaryote cells, centrosome reproduction consists of four morphological events: 1) centriole splitting, 2) centriole duplication, 3) centrosome disjunction, and 4) daughter centrosome separation [11]. The direct relation of the cell cycle and the centrosome duplication has been described by Salisbury as seen in Figure 4 [11]. During G_1 phase the two centrioles are oriented with the proximal end of the daughter centriole positioned along the lateral proximal wall of the mature centriole such that the two centrioles are in orthogonal arrangement. After passing the G_1 restriction point and commit to the S phase DNA replication, the two centrioles separates a short distance from one another and procentrioles begin to form along the lateral wall of the proximal end of each existing centriole. In G_2 phase the newly forming centrioles elongate and the pre-existing daughter centriole acquire molecular and structural features characteristic of a mature centriole including acquisition of a halo of pericentriolar and its associated microtubule nucleation capacity. Thus, during G_2/M, three generations of centrioles are present in the same somatic cell: the mature centriole, the daughter centriole and the two new nascent centrioles. At the G_2/M cell cycle transition the two pairs of centrioles migrate to opposite sides of the prophase nucleus and serve as the mitotic spindle [11]. At the end of mitosis, each daughter cell inherits a single centrosome.

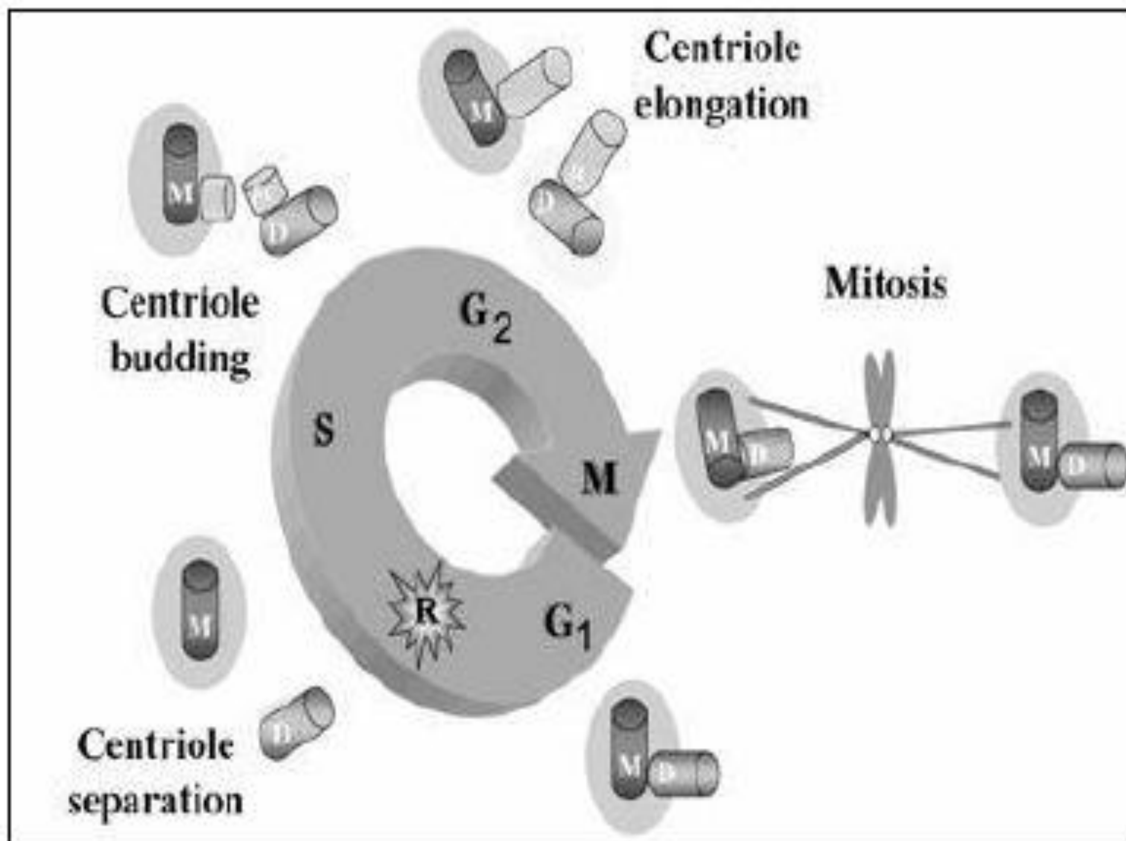


Figure 4. Duplication of the centrosome and the cell cycle. Once during each cell cycle the centrosome doubles from one-to-two in a process that is initiated by centriole duplication. In G_1 the mother centriole (M) and the daughter centriole (D) appeared in their typical orthogonal arrangement. In phase S they separate for the duplication, followed by the procentriole appearance. The procentrioles elongate in phase G_2. At mitosis, the two pair of centrioles moves to opposite sides of the nucleus. (*Adapted from* Salisbury et al. 2001[11]).

3.2 Functions of the Centrosome

The centrosome is considered to be responsible for nucleation of microtubule polymerization and anchoring the microtubules to create arrays that separate the chromatids during cell division [7]. Organization of the microtubule array of the mitotic spindle, allows equal segregation of sister chromatids into each of the two daughter cells. Moreover, it provides an important structural context for coordinating cell cycle regulation. During interphase, the centrosome organizes an astral array of microtubules (MTs) that participate in fundamental cellular functions such as intracellular trafficking, cell motility, cell adhesion and cell polarity [13]. In proliferating cells, the centrosome starts duplicating just before, or at, the onset of S phase and the two newly formed centrosomes participate in the assembly and the organization of the mitotic spindle, its orientation with respect to cortical cues, and the events of cytokinesis. The centrosome plays a role in determining the position, orientation and completion of the cytokinesis process [12]. Loss of a functional centrosome has been shown to lead to cell cycle arrest [14].

3.3 Centrosome Role in Disease

Centrosome defects have been implicated in disease processes, particularly in the origin of mitotic abnormalities and the development of aneuploidy in cancer [12]. In fact, the cells of many lethal human tumors are genetically unstable and have abnormally high number of centrosomes [10]. They are frequently amplified in cancer cells where an increase number of them can give rise to multipolar spindles in mitosis leading to the formation of aneuploid cells [14]. In aneuploid cells, failure to separate chromosomes represents a hallmark of most human carcinomas. An elevated number of centrosomes can induce the formation of additional SPB's and segregate chromosomes to an extra pole [14]. When cytokinesis occurs, daughter cells are produced that are missing the full complement of their corresponding chromosomes. These cells could not be viable, if they have lost essential genetic information. In the case that only one single chromosome is lost, a homologous chromosome could compensate. However, loss in heterozygosity can be critical in cases where the remaining chromosome carries mutations in tumor suppressor genes [14].

Recent discoveries have revealed that a modified centrosome and the primary cilium, have crucial roles in an increasing number of cellular and developmental processes, establishing a link between dysfunctional cilia and several genetic diseases [15]. Genetic diseases known as ciliopathies consist of mutations affecting the primary cilium. The primary cilium, a hair like cellular organelle, consists of a microtubule based structure found in almost all vertebrate cells and it originates from the modified mother centriole or basal body (Figure 5). The primary cilium senses a wide variety of extracellular signals and transmits the signal to the interior of the cell.

Thus, the initiation of signal transduction occurs within the primary cilium. They regard to proliferation, polarity, nerve growth, differentiation or tissue maintenance [16]. Primary cilia are structurally similar to motile cilia, the best known for lining the trachea, clearing mucus from the lungs, and generating flow. Until now, primary cilia were viewed only as vestigial organelles.

The list of disorders known as ciliopathies is constantly expanding and their phenotypes are well characterized resulting in several organs being most severely affected. Frequent cilia-related diseases are polycystic kidney disease, nephronophthisis, retinal degeneration, mental retardation, Bardet-Biedl syndrome, the Joubert syndrome and the Meckel syndrome [15-17]. The role that the cilium-centrosome complex plays in the normal function of most tissues appears to account for the involvement of multiple organ systems in ciliopathies. They were previously considered distinct disorders, but now evidence suggests they have one thing in common: defects of the primary cilium. Most of the emphasis has been placed on the function of ciliopathy proteins as potential key modulators of specific signaling cascades. Cilia may even play a role in cancer biology given their fundamental function in several developmental signaling pathways that are often misregulated in cancer [16]. Indeed, HEF1 and Aurora A, two proteins involved in cancer cell proliferation and metastasis, have been found to regulate cilia stability [16, 18]. The primary cilium is increasingly being identified as a novel regulator of a variety of cell biological processes, from development to homeostasis to cancer progression [18]. The primary cilia play a role in cell cycle regulation responsible for the coordination of cancer-related signaling molecules. Since most proteins that are altered in

ciliopathies function at the level of the cilium-centrosome complex, biomedical research is directing more efforts towards the study of the modulation of subcellular cascades at various stages of development and adult homeostasis.

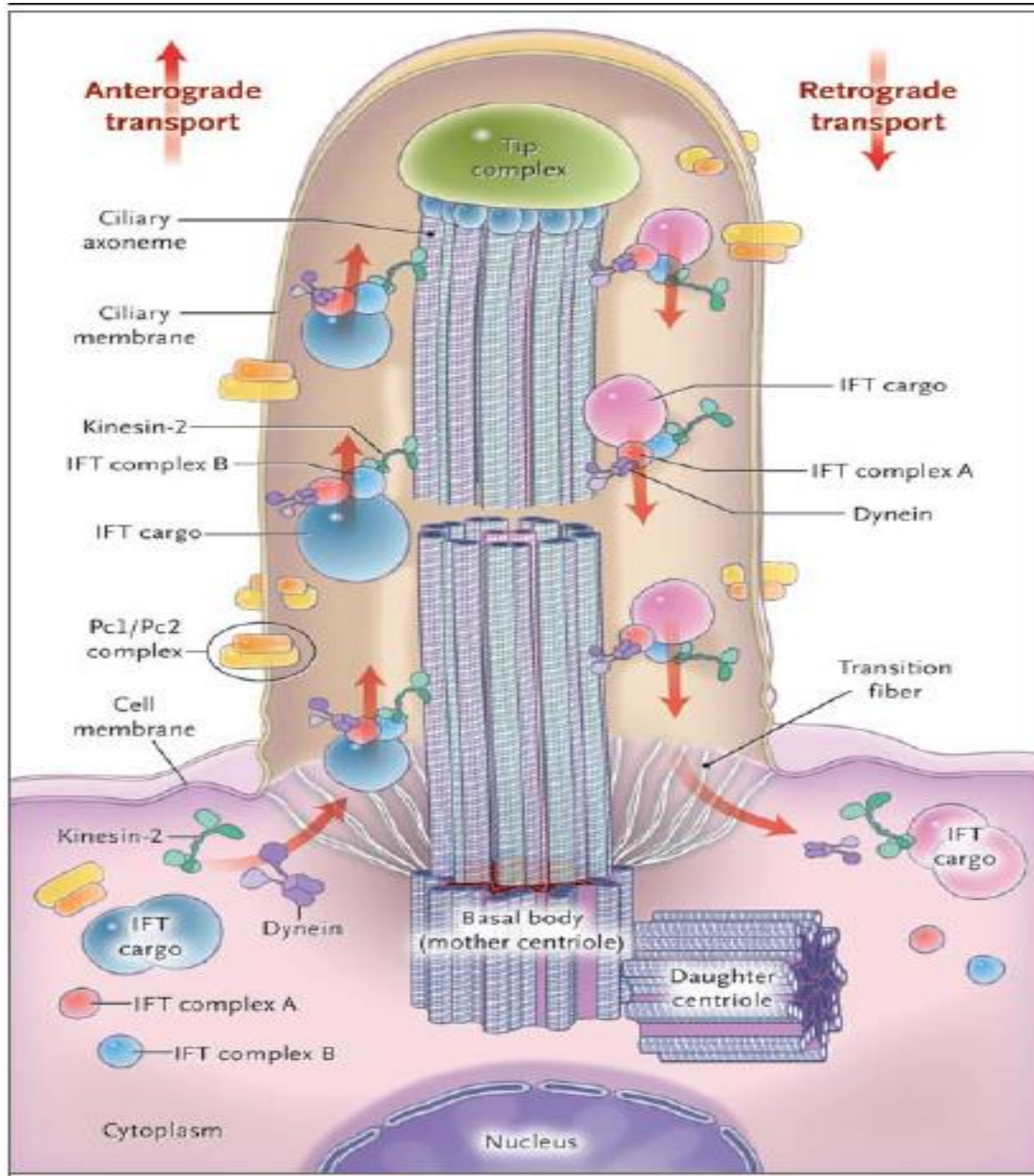


Figure 5. The structure of the primary cilium and intraflagellar transport. (Adapted from Hildebrandt, 2011[15]).

3.6 Centrin Protein: Localization and Structural Features

Centrin is an acidic protein with a molecular weight of approximately 20 kDa. It was discovered in 1984 by Dr. Salisbury, in the organism *Tetraselmis striata* [22]. This protein belongs to the EF-hand superfamily of small calcium binding proteins and it is a ubiquitous component to the centrosomes, centrioles and mitotic spindle poles [23]. In the yeast *Saccharomyces cerevisiae*, a centrin homolog (Cdc31), is crucial in the cell cycle via its regulation of the duplication of the spindle pole body (SPB) [24, 25]. X-ray crystallographic data for yeast, *chlamydomonas* and human centrins, has revealed a molecular structure comprised of four EF hands, which is the prototypical metal ion binding helix-loop-helix motif [20, 26-27]. Figure 6 shows the coordination of the calcium ions in the helix-loop-helix motif of centrin. The N-terminal domain, especially the first 20 amino acids, shows the most variable region, and the C-terminal half presents the most conserved region of the centrin sequence [23, 28]. This protein is found in eukaryotic cells but have no significant homology to proteins in archaea and bacteria. It is considered critical for the structure and function of the eukaryotic cell [23]. In particular, centrins are components of microtubule organizing centers (MTOCs), and are often located within different parts of the MTOC suggesting multiple roles [29, 30]. Centrins are components of both centrioles, of an assortment of fibers that link the two centrioles to each other, to the surrounding pericentriolar material and, in some organisms, to the cell and nuclear membranes. Centrin-containing fibers play a role in the dynamic behavior of centrosomes, through control of the position and orientation of centrosomal structures, and also in the control of centriole duplication [31-33]. There are four centrin isoforms: *Hs*cen1 has been

detected only in the basal body of the human sperm flagella and ciliated cells, while *Hs*cen2 and *Hs*cen3 are both constitutively expressed and localized to the centrosome and *Hscen4,* a pseudogene, is found in the neuronal cells of the brain [23, 34-36]. Amino acid analysis reveals that centrin is a highly conserved protein, showing an 80-90% sequence identity among vertebrates. Like troponin C, centrin is thought to be involved in a fiber-based calcium-induced contractile behavior [37]. Two clear roles have been established for centrin: the duplication of the MTOC and as constituents of contractile fibers within and to the MTOC, which can contract in response to changes in Ca^{+2} concentration [33, 38]. Centrin can form a complex with human Sfi1, a protein co-localized to the centriole. This centrin-Sfi1 complex takes place in a 23:1 molar ratio, so one molecule of human Sfi1 can bind multiple molecules of centrin. In yeast studies, Cdc31 binds directly to individual *Sacharomyces cerevisiae* Sfi1(*Sc* Sfi1) binding sites in a 14:1 molar ratio, that is, a single *Sc* Sfi1 molecule binds multiple Cdc31 proteins [20].

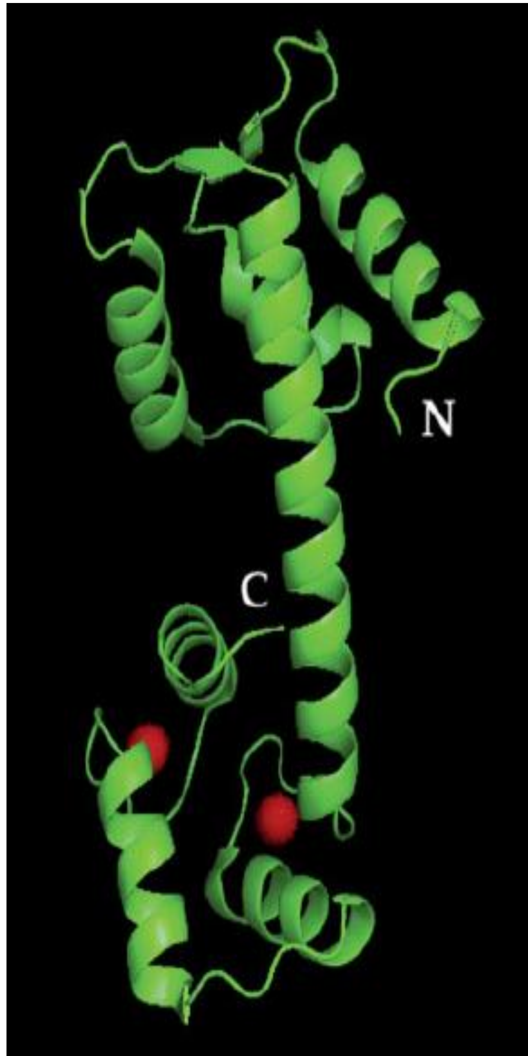


Figure 6. Full length centrin (*Hs*cen2) comprised of four EF-hands. Two calcium ions coordinated at the helix-loop-helix motif in the C-terminal region. (*Adapted from* Thompson et al. 2006 [27]).

3.5 Sfi1 Protein: Localization and Structural Features

Saccharomyces cerevisiae Sfi1 (*Sc*Sfi1) was first identified in 1999 by Ma et. al. within the half-bridge structure of the SPB [39]. The human homolog, *Homo sapiens* Sfi1 (*Hs* Sfi1) is localized to the centriole in HeLa cells. *Sc*Sfi1 has been found to be a very divergent protein with only 24.5% sequence identity compared to other yeast species. Early studies showed that there is a lack of homology in the N-terminal and C-terminal domains of Sfi1 in different species [40]. *Hs*Sfi1 and *Sc*Sfi1 are comprised of 23 and 17 tandem centrin binding sites (CBS) respectively. Recently, a family of thirteen proteins (Sfr1-13) in the organism *Tetrahymena thermophila*, has been identified containing centrin binding repeats localized at the basal body contributing to its organization and stabilization [41]. Further analysis of these CBS sequence revealed the presence of a consensus sequence of AX7LLX3F/LX2WK/R in different species (Figure 7) [19-20, 23]. Two isoforms (NCBI accession numbers: NP_001007468.1 and NP_055590) of human Sfi1 have been identified with almost identical sequences containing twenty three internal consensus repeats, but with stop codons after either 968 or 1242 amino acids [40,42]. Centrin binding sites are separated by gaps of 10 amino acids long for *Hs*Sfi1 and 23-35 amino acids long for *Sc*Sfi1 [19,40]. In *Hs*Sfi1, the short distance gap between centrin binding sites allows centrins to interact with one another [20, 40]. Sfi1 has a random coil structure and adopts a helical conformation upon binding to centrin when free [Ca^{+2}] increases (Figure 8) [19-20]. Sfi1 includes low proline content in its sequences within the repeat regions and a lack of homology in the amino- and carboxy-terminal domains from different species [19].

It is an essential protein whose depletion can cause a G_2/M arrest in the cell cycle progression, with failure to form a mitotic spindle [19].

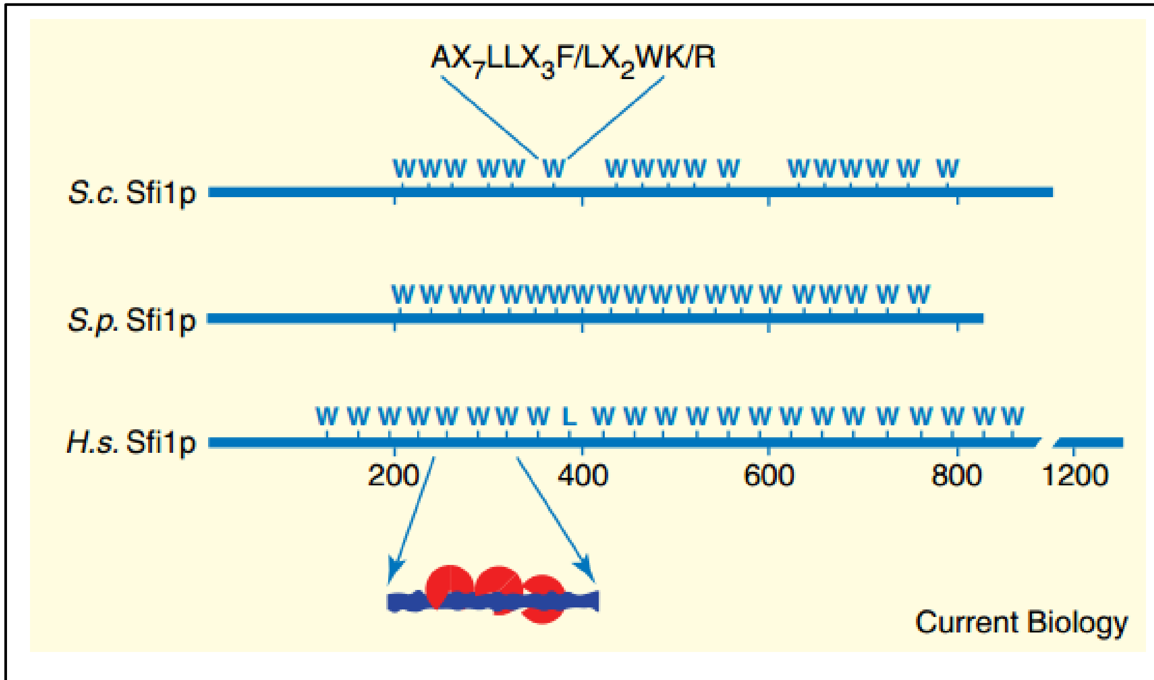


Figure 7. The consensus sequence of Sfi1 in different species. Sfi1 orthologs of the yeasts *Saccharomyces cerevisiae* (*S.c.* Sfip) and *Schizosaccharomyces pombe* (*S.p.* Sfi1p) and of *Homo sapiens (H.s.* Sfi1p).The consensus sequences includes an X indicating any other amino acid in that position of the sequence. The W indicates the position of consensus repeat and a model representation of Sfip with three centrin molecules bound to each repeat. (*Adapted from* Salisbury, 2004 [40]).

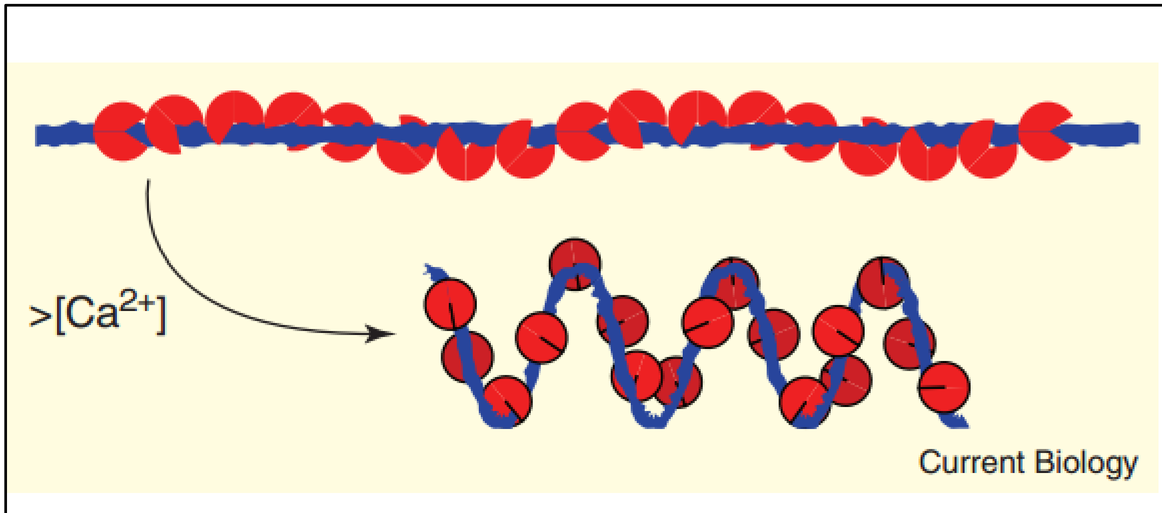


Figure 8. A model for Sfi1-centrin contraction. Conformational change upon increase of [Ca^{+2}] (*Adapted from* Salisbury, 2004 [40]).

3.6 Studies of Sfi1-Centrin and Sfi1-Cdc31 Complexes

Kilmartin [19], performed studies in budding yeasts to proove that the internal repeats in Sfi1 were centrin binding sites. GST-fusions with the amino- or carboxy-terminal regions of Sfi1, lacking the consensus repeats were used and they failed to bind Cdc31. In the other hand, Sfi1 GST-fusions that contained even only a single repeat sequence bound to Cdc31 with a molar ratio close to 1.These findings clearly demonstrated that Sfi1 binds multiple molecules of Cdc31 through its conserved repeat sequences. Moreover, a comparison of Sfi1 in different species and with similar gaps between repeats were accessed by Blast searching (Figure 9).The results showed those amino acids conserved between different organisms and the low proline content within the repeat regions for the three proteins. Kilmartin [19] also showed that the Sfi1–centrin complex is concentrated at the centrosome and the SPB. Light microscopy of yeast cells expressing recombinant Sfi1 fused to the green fluorescent protein (GFP) and parallel immuno-labeling studies at electron micro-scope resolution confirmed that Sfi1 occurs at the SPB half bridge where Cdc31 is also found [19,43]. Likewise, recombinant GFP–hSfi1 showed centrosome localization in HeLa cells, and co-localization with centrin, as a marker for centrioles. These experiments established that Sfi1-Centrin and Sfi1-Cdc31 complexes occur at the centrosome and the spindle pole body of both cell types repectively, suggesting interactions between the two proteins (Figure 10). Also, the studies suggested that Cdc31p and Sfi1p interact functionally and play an important role in SPB duplication.

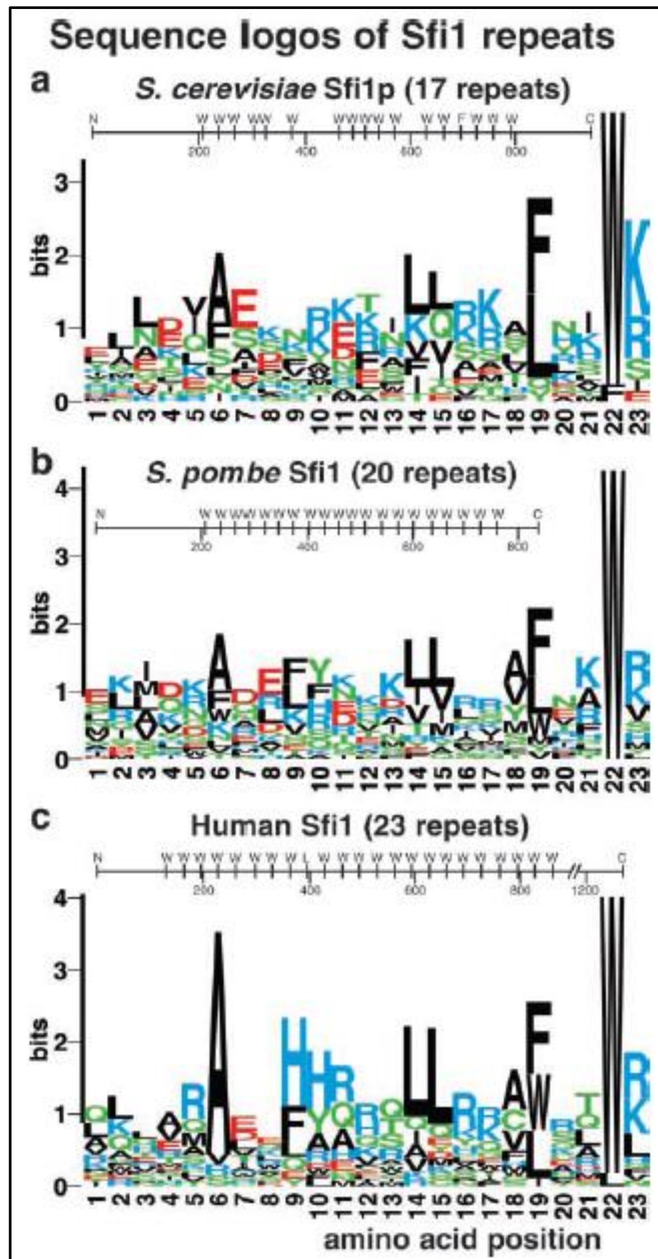


Figure 9. Sequence comparison of Sfi1 repeats in three organisms.

(*Adapted from* Kilmartin, 2003 [19]).

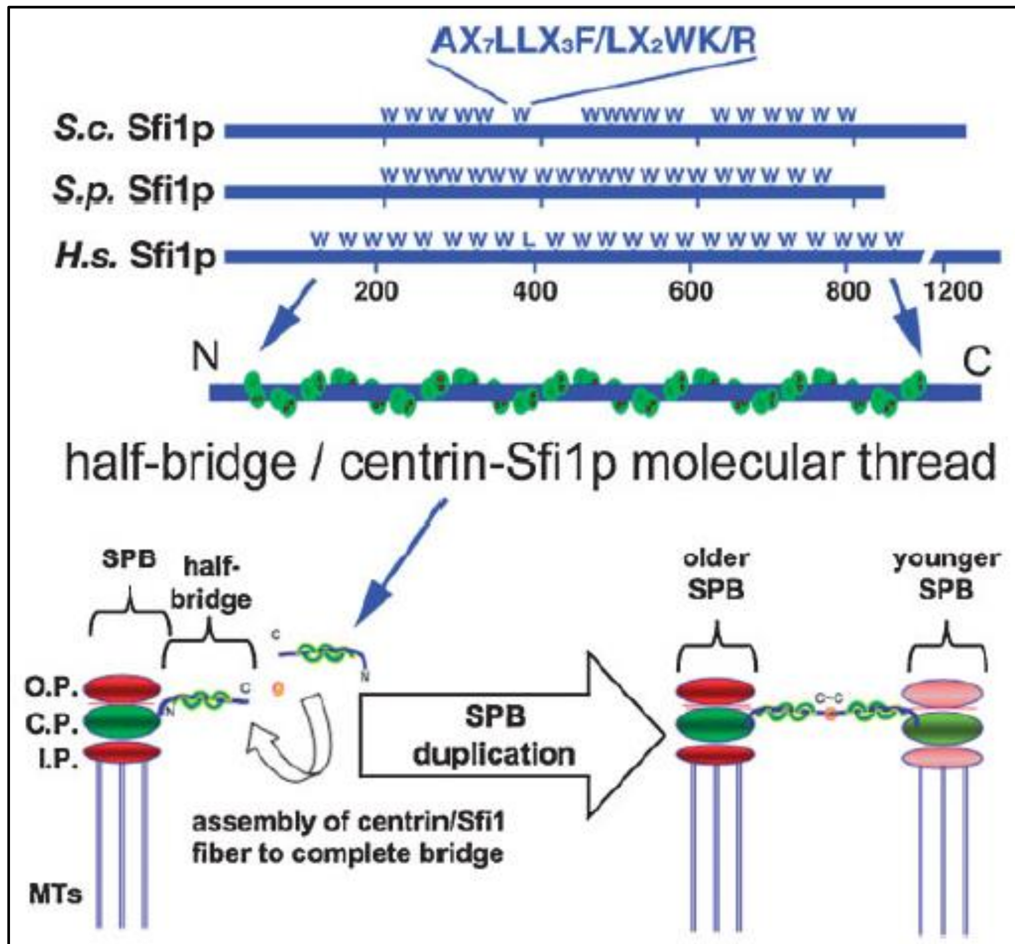


Figure 10. Duplication of the yeast SPB regulated by Sfi1-centrin complex. Consensus sequence for the centrin binding sites (W) of *S.cerevisiae*, *S.pombe* and *H.sapiens* in Sfi1. The fundamental unit of duplication for the SPB is the half-bridge (*Adapted from* Salisbury 2007 [23]).

The crystal structure of Cdc31-*Sc*Sfi1 complex was published by Li et al. in 2006 (Figure 11) [20]. This structure shows the *Sc*Sfi1 fragment as a α-helix with Cdc31 in an extended conformation bound to each repeat. On binding calcium ions, the Cdc31 opens up, exposing hydrophobic surfaces of amino acids that are able to bind *Sc*fi1. The Cdc31 N-terminal domains bind to the N-terminal half of the *Sc*Sfi1 repeat, whereas the Cdc31 C-terminal domains bind to the more conserved C-terminal half of the *Sc*Sfi 1 repeat [20].

Based on evidence from Electron Microscopy (EM) the N terminus of Sfi1p is placed at the SPB, whereas the C terminus is at the center of the bridge proposing that an *Sc*Sfi1/Cdc31 filament could span the length of the half bridge[44]. Kilmartin and coworkers [20] suggested that during SPB duplication, the half bridge could double in length through association of two, end-to-end *Sc*Sfi1 C termini, providing a new *Sc*Sfi1 N terminus as an assembly site for the new SPB (Figure 12) [44]. This suggests a model for SPB duplication where the half-bridge doubles in length by association of the *Sc*Sfi1 C to the half bridge, thereby providing a new *Sc*Sfi1 N terminus to initiate SPB assembly.

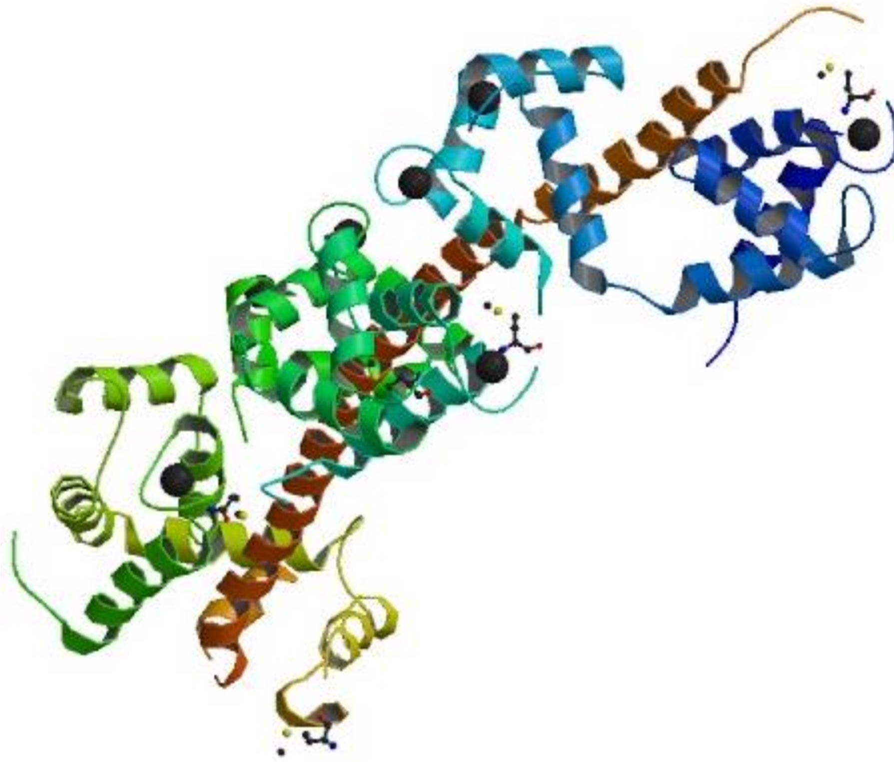


Figure 11. Crystal structure of Sfi1p-Cdc31 complex (PDB ID 2DOQ). Sfi1 (orange)appears as a helix with three centrins (blue-green) wrapped around it and coordinated with calcium ions (black) (*Adapted from* Li et al.2006 [20]).

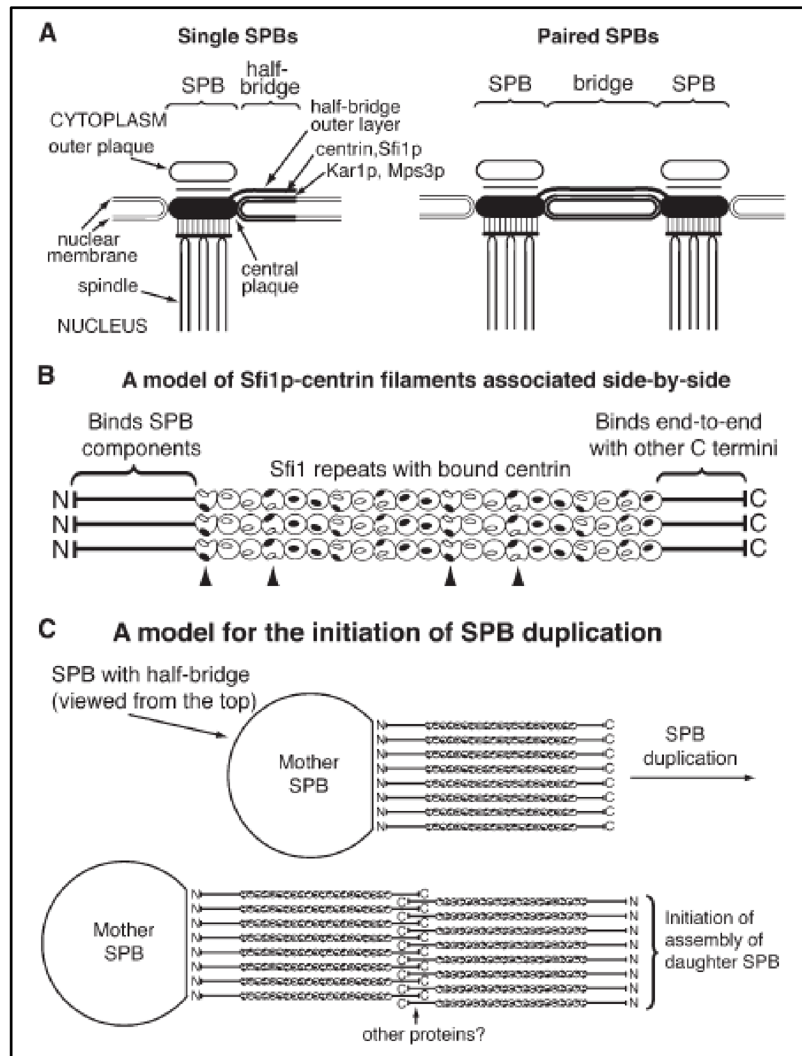


Figure 12. Model for the initiation of SPB duplication. Single and paired SPBs showing the location of half-bridge and bridge components (A), Model of *Sc*Sfi-Cdc31 filaments associated side-by-side. Arrowheads show Cdc31s interacting between filaments (B), and the *Sc*Sfi N terminus binds SPB components, whereas the C termini can associate end-to-end in an antiparallel way with or without other proteins. This provides a *Sc*Sfi 1 N terminus capable of binding SPB components and thereby initiating SPB assembly (C). (*Adapted from* Li et al.2006 [20]).

4. EXPERIMENTAL METHODS

4.1 Bacterial Protein Expression of *Hs* Sfi1p$_{1-2}$

A bacterial stock culture of *E. coli* BL21-(DE3) RIL (~150-250 µL) competent cells was transformed with pET 100-*Hs* Sfi1p$_{1-2}$. The newly transformed bacterial cells were used to inoculate 250 mL of sterile Terrific broth (MP Biomedicals, Solon, Ohio). Then 50 µg/mL of ampicillin were added and overnight incubation in an orbital shaker at 37^{0}C and 250 rpm was performed. After 12 hours, this culture was inoculated in a previously sterilized 5L BIOFLO 3000 fermentor from (New Brunswick Scientific, Edison, NJ) containing 3L of sterile Terrific broth and 50 µg/mL of ampicillin. The running conditions for the fermentor were the following: a temperature of 37^{0}C, 350 rpm of agitation and acid-base adjustments to obtain a pH value of 7.0. Cell biomass was monitored hourly, taking 1.0 mL aliquots from the fermentor run in a VoluPac tube from (Stedim Sartorius Biotech, Aubagne, France) and centrifuging for 1 minute at 7 rcf (relative centrifuge force). The expression of His-*Hs* Sfi1p$_{1-2}$ was induced, by adding 0.5 mM isopropylthio-ß-galactoside (IPTG), when the bacterial cell growth achieved the mid logarithmic phase of the bacterial growth curve. When the cell culture reached the stationary phase, the bacterial culture was harvested, by centrifugation for 30 minutes at 2,465 xg (3,500 rpm) and a temperature of 4^{0}C using a J-10 rotor and a Beckman J2-MC centrifuge. The pellets obtained from this step, were stored at -80^{0}C for further purification. The expression process is summarized in Figure 13.

4.2 Isolation of *Hs* Sfi1p$_{1-2}$

The pelleted cells were thawed and weighed. They were suspended in four times the amount (w/v) of cold lysis buffer comprised of 50 mM Tris, 500 mM NaCl, 0.04% NaN_3 and 0.1% IGEPAL, pH 7.4 and one tablet of a cocktail of protease inhibitors (2.0 mg/mL aprotinin, 0.5 mg/mL leupeptin, and 1.0 mg/mL pepstatin A) was added, to inhibit protease degradation of the His-*Hs* Sfi1p$_{1-2}$ protein. Then, the resulting cell solution was sonicated with a Branson sonifier model 450 (Branson Ultrasonics Co., Danbury, CT) approximately 3-6 times for 30 seconds, with cooling periods of 1 minute in an ice bath and until the solution was free flowing. The resulting cell lysate was centrifuged at 9,615 xg (10,000 rpm) for 15 minutes at 4°C using a JA-14 rotor and a Beckman J2-MC centrifuge. After that the supernatant (S_1) was collected and the resulting pellet (P_1) was stored at -20°C. Then, the supernatant (S_1) was subjected to a second ultracentrifuge step at 70,588 xg (30,000 rpm) for 30 minutes at 4°C using a TI-70 rotor and a Beckman L-80 ultracentrifuge. The resulting pellet (P_2) was labeled and stored at -20°C and the supernatant (S_2) was then incubated at 50^0C for 30 min. in a water bath and centrifuged at 9,615 xg (10,000 rpm) for 15 minutes at 4°C using a JA-14 rotor and a Beckman J2-MC centrifuge. The resulting pellet (P_3) was labeled and stored at -20°C. The supernatant (S_3) however, was subjected to several tangential flow filtrations (TFF) as follows: first using a hollow fiber cartridge membrane (GE Healthcare, New Jersey) of 0.1 µm pore size to clarify the sample. The clarified sample was subjected to tangential flow filtration (Pall Corporation, New York) using a 50 kDa cut off membrane to remove high molecular host cell proteins.

The filtered sample was subjected to a third TFF process using a membrane with a 3 kDa cut off, in order to perform a buffer exchange and concentrate the sample containing the desired recombinant peptide. The isolation process is summarized in Figure 13.

4.3 Purification by Affinity and Ion Exchange Chromatography

His-tag affinity chromatography (Figure 13) was performed using ÄKTA automated protein purification system (GE Healthcare, New Jersey). The chromatography column used was a 5.0 mL HiTrap Chelating HP column (GE Healthcare, Sweden) which was packed with cobalt ions, in order to form coordination bonds with His-tagged *Hs* Sfi1$p_{1-2.}$ The three essential buffers for this chromatographic step included: (**Wash Buffer 1**) 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole adjusted at pH 7.4; (**Wash Buffer 2**) 50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole adjusted at pH 7.4; and finally (**Elution Buffer**) 50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole at pH 7.4. The equipment was run at a standard flow rate of 2.5 mL/min. The collected fractions were concentrated using a Millipore centrifugal device with a 5,000 MW cut off, that has a cellulose low protein binding membrane for concentration of protein samples. After use the column was washed with eight column volumes of a buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, at pH 7.4 in order to remove any imidazole from the column. A 4-20% (Bis-Tris) gradient SDS-PAGE was run with rehydrated aliquots of the collected fractions to identify the fractions containing the His-tagged *Hs* Sfi1$p_{1-2.}$ The fractions identified as containing the recombinant protein of interest were pooled and further purified using two different anion exchange chromatograpy using the ÄKTA automated system for low volume purification and BioRad manual system with a peristaltic pump for large volume purification (Figure 13). The chromatography columns used were: a 1.0 mL Acrosep High Q Strong Anion-Exchange (Pall Life Sciences, Michigan) equilibrated and run at a flow rate of 0.5 mL/min and a 40 mL BioRad High Q Strong Anion-Exchange column

(Bio Rad Laboratories, California) equilibrated and run at a flow rate of 2.00 mL/min with an AUFS of 0.5 at a wavelength of 280 nm. The elution was obtained using buffers containing 20 mM Tris, 1 mM $CaCl_2$, 0.04% NaN_3, and 0-0.70 M NaCl or 0-1.0 M step gradients at pH 7.4. At the end of the process, the columns were washed for next use by eluting any tightly bound protein to the matrix with four column volumes of the buffer containing 1.0 M NaCl and then it was regenerated with the same buffer without NaCl. UV/Vis analysis was performed using a JASCO spectrophotometer in order to verify protein purity and determination of the concentration of the desired recombinant protein. The collected fractions were pooled, further concentrated and subjected to a 4-20% (Bis-Tris) gradient SDS-PAGE in order to identify the fractions containing His-*Hs* Sfi1$p_{1\text{-}2}$.

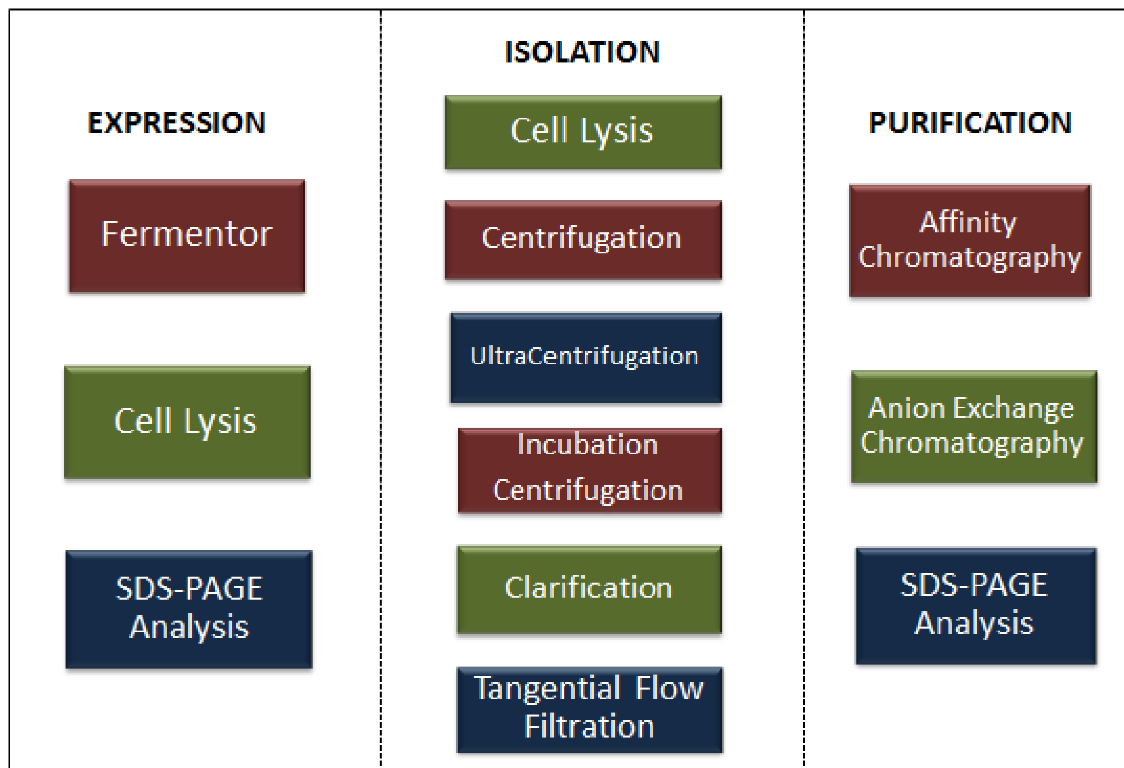


Figure 13. Bacterial expression, isolation and purification of His-*Hs* Sfi1p$_{1-2}$.

4.4 Alternate Isolation by Extraction Method

Upon closer examination the His tag was being proteolitically cleaved during its expression within the bacterial cell, and as a result the sequence of the peptide had a hydrophobic character which would then allow for the recombinant peptide to be within the pellet. In order to improve the isolation of *Hs* Sfi1p_{1-2} an established protocol by Dr. Pastrana-Rios was used [47-48]. A newly harvested pellet was subjected to lysis as described in section 4.2 and subjected to a standard extraction method as summarized in Figure 14. Typically, 10 mL of cell lysate were added to a separatory funnel and multiple extractions with $CHCl_3$:CH_3OH (2:1, v/v) and $CHCl_3$:CH_3OH (1:1, v/v) were carried out. In addition, a 0.1M potassium chloride (KCl) solution was added to improve the separation of the phases. The organic phase and the aqueous phase were concentrated by rotary evaporation until obtaining a dry sample. Then, each phase sample was re-suspended in a buffer solution of 8 mM HEPES, 50 mM NaCl, 2 mM $CaCl_2$ and 2 mM $MgCl_2$ at pH 7.4 and a 4-20% (Bis-Tris) gradient SDS-PAGE was run to verify the purity and relative mobility of the proteins comprised within each phase. After completing the electrophoresis, the resolved protein bands were then transfered by electro-blotting to a PVDF membrane (BioRad, California) with running conditions of 90 volts for 1 hour. The PVDF membrane was then coomasie blue stained and the corresponding low molecular weight bands of interest were sent for partial amino acid sequencing at the proteomics facility at Tufts University (Medford, Massachusetts).

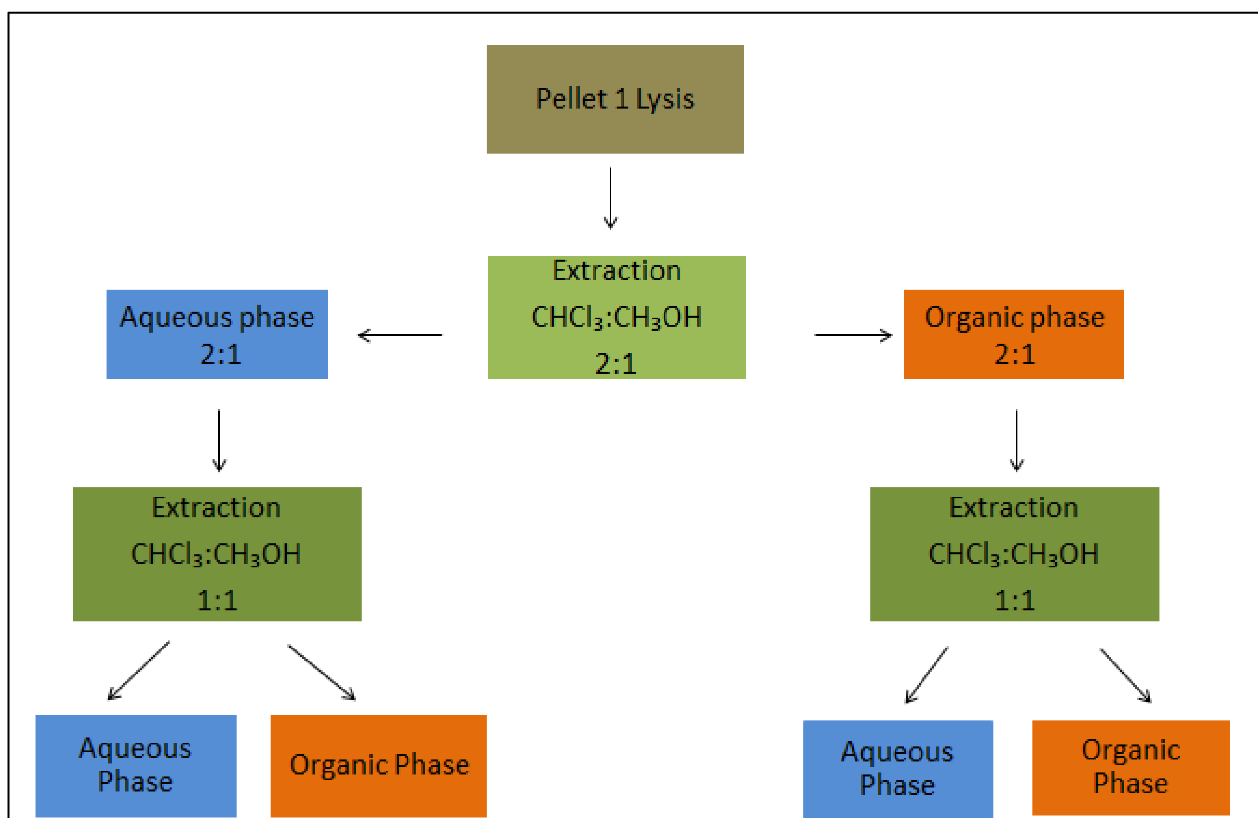


Figure 14. Alternate extraction method for the isolation of hydrophobic recombinant proteins. The solvent extraction ratios referred to above are (v/v ratios).

4.5 Alternate Purification by Size Exclusion Chromatography

The identified sample containing *Hs* Sfi1p$_{1-2}$ by partial amino acid sequencing was dialyzed against a buffer solution of 8 mM HEPES, 50 mM NaCl, 2 mM $CaCl_2$, and 2 mM $MgCl_2$ at pH 7.4. Then, it was subjected to size exclusion chromatography using ÄKTA system. The chromatography column used was a Superdex 75 (GE Healthcare, Sweden) equilibrated and run at a flow rate of 0.5 mL/min and an AUFS of 0.5 at 280 nm wavelength. The collected fractions were subjected to a 4-20% (Bis-Tris) gradient SDS-PAGE, in order to identify the fractions containing *Hs* Sfi1p$_{1-2}$. The fractions identified to contain highly pure *Hs* Sfi1p$_{1-2}$ were lyophilized and stored in a desiccator at room temperature.

5 RESULTS AND DISCUSSION

5.1 Bacterial Protein Expression of *Hs* Sfi1p_{1-2}

The over expression of *His-Hs* Sfi1p_{1-2} using the fermentor BioFlo 3000 resulted in a high yield cell culture, obtaining a pellet of 55.42 g. This was accomplished after variations of the current protein expression protocol [45], by using Terrific broth media and adding a fresh inoculum to the fermentor. As expected, no deviations in fermentation parameters were observed: agitation (350 rpm), temperature (37.0 ^{0}C), pH (7.0) and dissolved oxygen (100%). As shown in Figure 15, the fermentation last approximately 7 hours from inoculation to harvest, where the cells reached the stationary phase. An increase in cell growth was observed at the beginning of the log phase, after addition of a 0.5 mM isopropylthio-ß-galactoside (IPTG) solution. Aliquots were taken every 60 minutes during the course of the fermentation process for monitoring the cell growth. Typically, the maximum value of packed cell volume (PCV) was 16 µL.

The aliquots were later lysed and subjected to a 5% stacking, 12% separating SDS-PAGE and stained with coomasie blue to determine the extent of over-expression of the recombinant protein. As shown in Figure 16, there is a prominent band around 12 kDa initially suspected to be His-*Hs* Sfi1p_{1-2} which increased in intensity after IPTG induction. The increasing amount of this band indicates that IPTG was able to activate the transcription of T7 RNA polymerase, which in turn transcribes the His-*Hs* Sfi1p_{1-2} DNA in the plasmid under the control of the T7 promoter. This protein band shows very low basal level of expression 3 hours before induction and was over expressed after 4 hours.

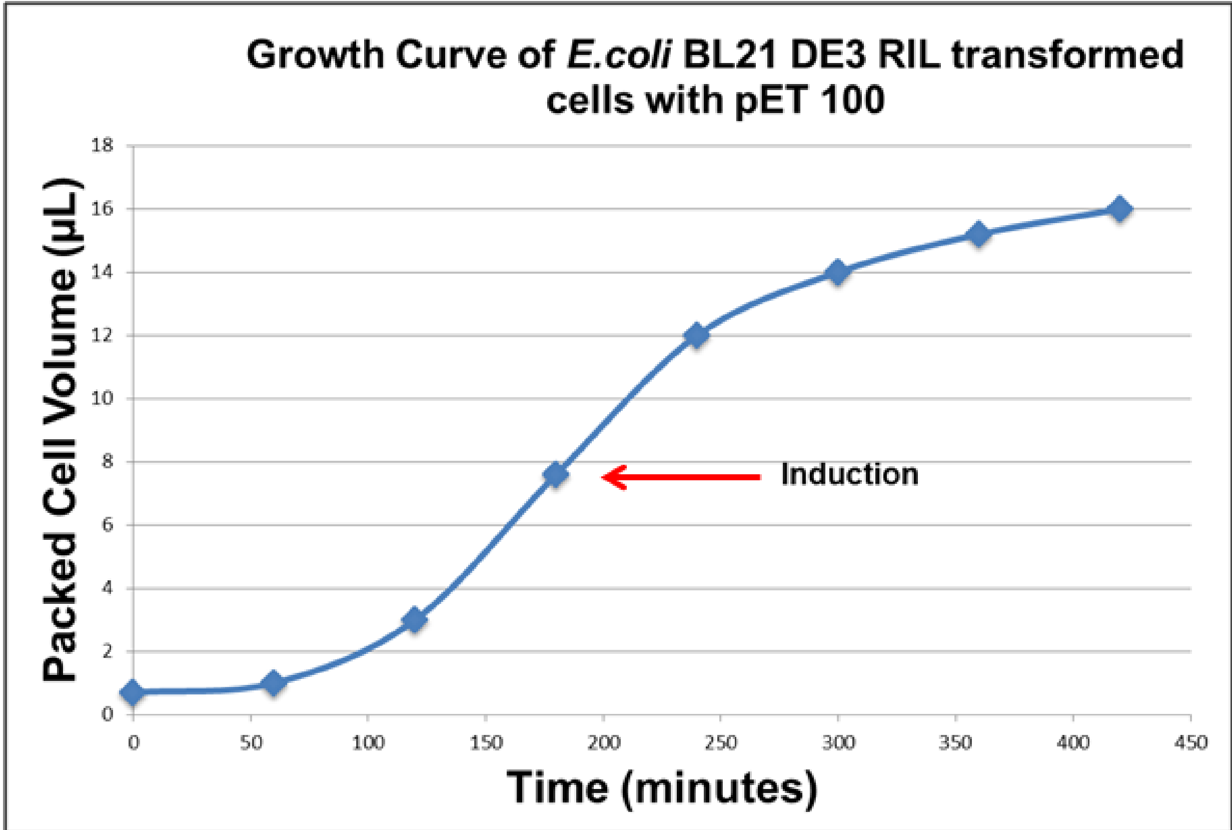


Figure 15. Growth curve of *E. coli* transformed cells with pET 100. The arrow shows the induction point 3 hours after inoculation of the fermentor.

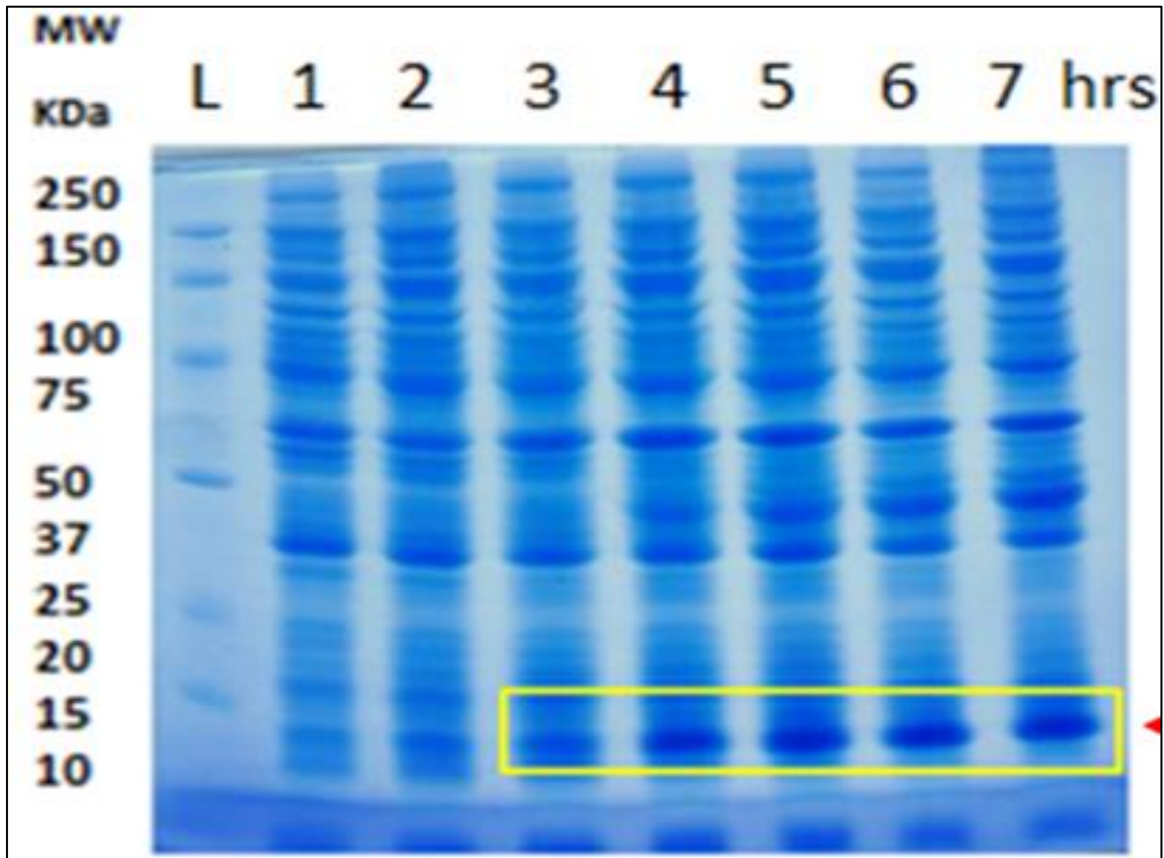


Figure 16. 5% stacking, 12% separating SDS-PAGE for the bacterial pellet lysate after high level expression. Molecular weight standards (L), cell lysates of progressive expression prior to induction (lanes 1-3), and cell lysates of progressive expression after induction (lanes 4-7).

5.2 Isolation of *Hs* Sfi1p$_{1-2}$

The protocol used to isolate the recombinant peptide was similar to that established by Pastrana et al. 2002 [45]. The pellet was lysed in batches of 10-20 grams, undergoes two centrifugation steps that removes a considerable amount of cell debris and the supernatant resulting from the second centrifugation (S_2) was heated up to 50^{0}C. Since, His-*Hs* Sfi1p$_{1-2}$ has a random coil structure the risk of thermally denaturing the desired protein sample is minimal, yet the advantage of actually removing unwanted host proteins and its associated proteases; thus avoiding the unnecessary exposure of the recombinant peptide to proteolytic cleavage. After that, a third centrifugation was executed and the presence of a cream color paste, presumable denatured host proteins and proteases was observed as a pellet in the centrifuge tube. This third supernatant (S_3) was then subjected to a 0.1 µm hollow fiber membrane in order to clarify the sample. The clarification process, allowed for a clearer supernatant sample and succeeding in removing more of the cellular debris and viscosity. Then, a second modification was performed by using two different membranes of 50 kDa and 3 kDa in a TFF system. The 50 kDa membrane was used in order to remove higher molecular weight host cell proteins, since His-*Hs* Sfi1p$_{1-2}$ molecular weight is 12 kDa and can be recovered in the filtrate. The 3 kDa membrane was used to concentrate the peptide of interest and to perform a buffer exchange that will bring the sample to the buffer conditions used for the subsequent purification process. This saved experiment time, since it only consumed about 2 hours instead of using a centrifugal filter device of 3 kDa cut off that regularly would take approximately 8 hours. After the isolation process, a 4-20% (Bis-Tris) SDS-PAGE

was run confirming the efficiency of the process in removing the unwanted host cell proteins and showing the presence of a protein band near 12 kDa (Figure 17).

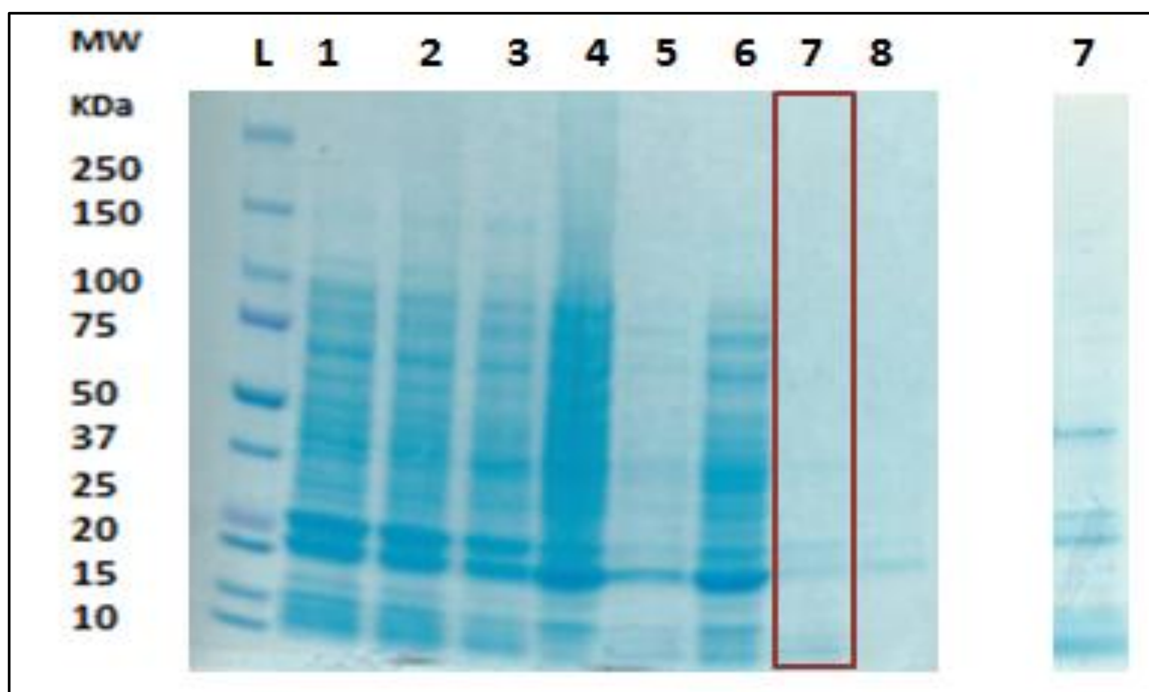


Figure 17. 4-20% (Bis-Tris) gradient SDS-PAGE for the isolation process. Molecular Weight Ladder (L), centrifugation step Supernatants (1-3), 0.1 µm membrane retentate (4), 0.1 µm membrane filtrate (5), 50 kDa TFF membrane filtrate (6), 3 kDa TFF membrane retentate (7-8). At the right of the image, lane 7 was further concentrated and stained to improve the visualization of the protein bands.

5.3 Purification by Affinity and Anion Exchange Chromatography

The retentate resulting from the 3 kDa membrane was loaded into a cobalt affinity column. Two main peaks highly symmetric were obtained as observed from the chromatogram (Figure 18). The first peak (A) corresponds to a void volume of proteins that did not interact with the column matrix and the second peak (B) corresponds to the eluted proteins that did interact with the column matrix. It is in peak (B) where His-*Hs* Sfi1p$_{1-2}$ would be present since the recombinant peptide was designed to contain a sequence of histidines residues that has affinity for the cobalt matrix. However, it is not exclusive only for His-*Hs* Sfi1p$_{1-2}$, because there is the probability of other histidine rich proteins binding, and in that way this served as a preparatory chromatography that reduced the presence of contaminating host proteins. These results were confirmed by performing a 4-20% (Bis-Tris) gradient SDS-PAGE for both peak fractions (Figure 19). Although several sample batches were purified, the void volume and the elution peak always contained other proteins, in addition to the expected recombinant peptide near 12 kDa. Moreover, the recombinant peptide appears to be also present in the void volume, suggesting that performing additional purification of the void volume sample would be required in case of a protein overload of the column. Additional purification exercises were executed purifying the void volume and using low sample volumes to prevent column overloading obtaining the same results. As mentioned earlier, an explanation for the matrix affinity of these other contaminating proteins in the elution peak was that they could be rich in histidine residues as was the case for our desired recombinant, His-*Hs* Sfi1p$_{1-2}$ allowing them to bind at the surface of the cobalt matrix. These results indicated that further

purification steps must be executed to obtain a pure sample of His-*Hs* Sfi1p$_{1-2}$ since it was partially purified.

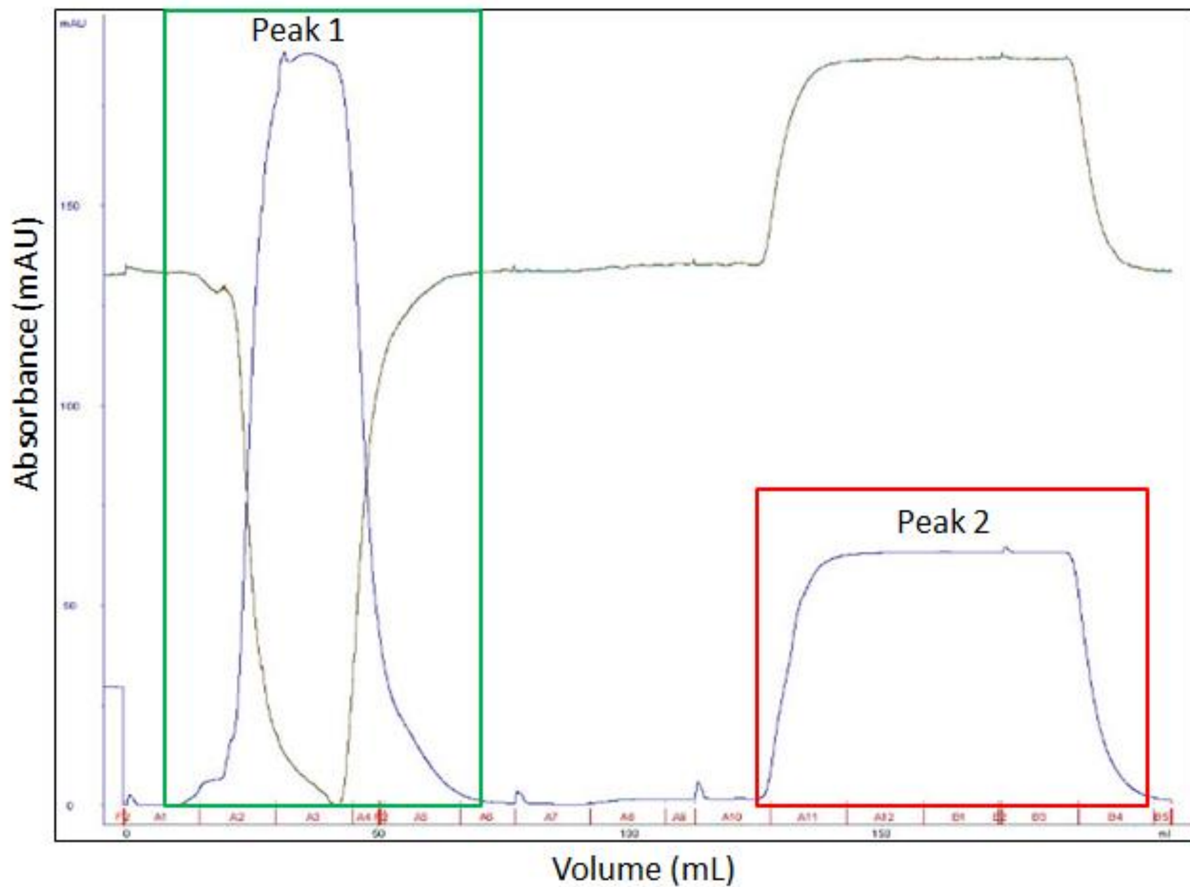


Figure 18. Cobalt affinity chromatogram. The blue line corresponds to the UV-Vis Spectrum, whereas the brown line corresponds to the conductivity of the protein sample. Host cell proteins that did not interact with the cobalt matrix (Peak 1). Eluted proteins that did interact with the cobalt matrix (Peak 2).

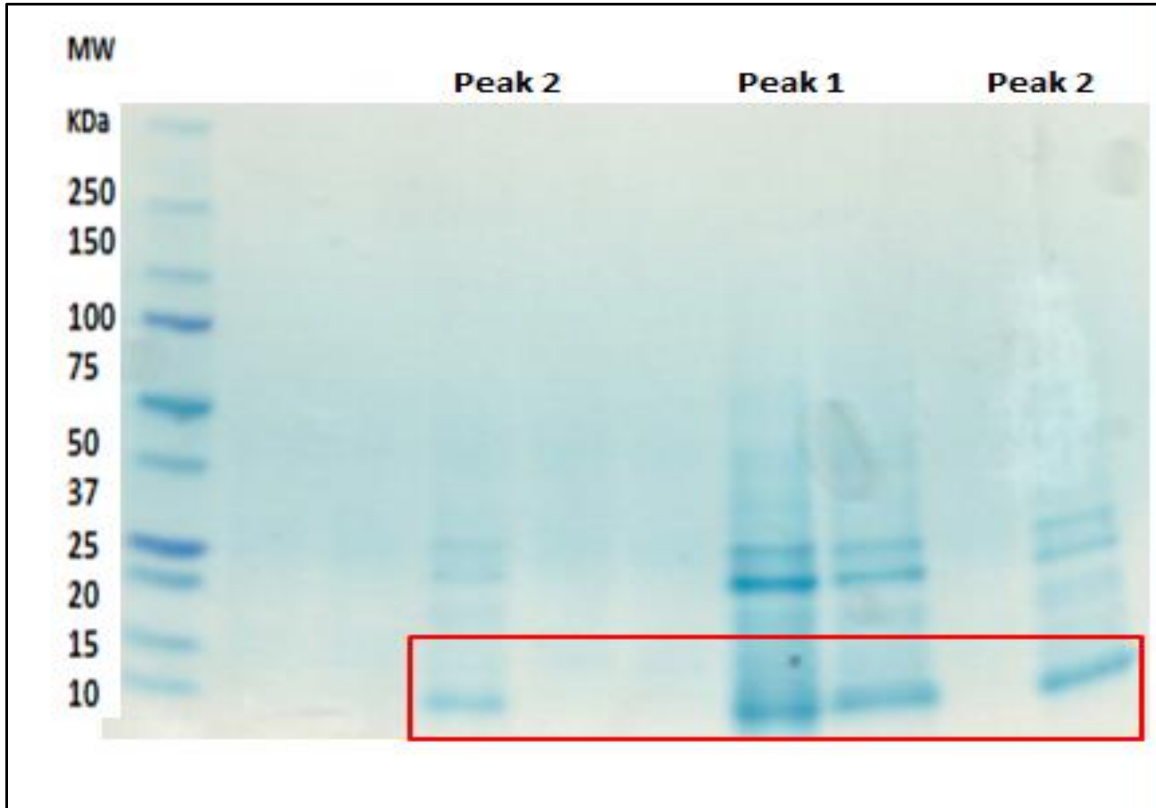


Figure 19. 4-20% (Bis-Tris) gradient SDS-PAGE for affinity chromatography fractions. The polyacrylamide gel shows a protein near 12 kDa suspected to be His-*Hs* Sfi1p$_{1-2}$ for both peaks.

The principle of ion exchange chromatography explains that a charged molecule will bind to an oppositely charged matrix. If the pH of the buffer is below the pI of the target molecule, the net charge on the target molecule will be positive and will therefore bind to a cation exchanger. Otherwise, if the pH of the buffer is above the pI of the target molecule, the net charge on the target molecule will be negative and will therefore bind to an anion exchanger. After doing a sequence analysis on Expasy ProtParam software [46] a pI = 8.57 was obtained for His-*Hs* Sfi1p_{1-2}, and the pH of the protein's buffer was 7.4, making it a candidate for a cation exchanger. However, an anion exchange chromatography step was selected in order to remove contaminating host proteins with lower isoelectric points. The pooled fractions obtained after the affinity chromatography step were loaded onto a 1.0 mL High Q Strong Anion exchange column in ÄKTA system for a low sample volume purification. The step gradient used was between 0-1.0 M NaCl (Figure 20). Three peaks were observed at 0%, 80% and 100% NaCl and after performing a 4-20% (Bis-Tris) gradient SDS-PAGE analysis a band near 12 kDa was observed for these three peaks. A UV/Vis spectroscopic analysis of the three fractions was performed and all the spectra showed the highest absorbance near 260 nm (Figure 21). Examining closely the sequence of His-*Hs* Sfi1p_{1-2}, it has five tryptophan (W) residues in its sequence, therefore a strong absorption signal at 280 nm, should have been observed. This was not the case, thus providing evidence suggesting that it was not the desired recombinant peptide.

A second anion exchange chromatography was performed, loading a large volume sample onto a 40 mL BioRad High Q Strong Anion-Exchange column (manual system). A linear ionic gradient was used between 0-0.70 M NaCl. The elution profile showed two small peaks and one major peak (Figure 22). The first peak (Peak 1) appeared before the gradient was started and corresponds to those proteins positively charged that did not interact with the positively charged column. The second peak (Peak 2) represented less positive proteins binding to the column using an elution buffer of 20% NaCl. A third peak (Peak 3) showing a sharp band or peak which had the highest absorbance of all, suggesting the presence of proteins with the lowest isoelectric points interacting with the positively charged matrix. It was observed that although the isoelectric point of His -*Hs* Sfi1p$_{1-2}$ is 8.57, and was supposed to elute only in the Void Volume, it would not interact with the positively charge column, suggesting that the protein eluted was in fact a host protein and not the desired recombinant protein as shown in the 4-20% (Bis-Tris) gradient SDS-PAGE (Figure 23).

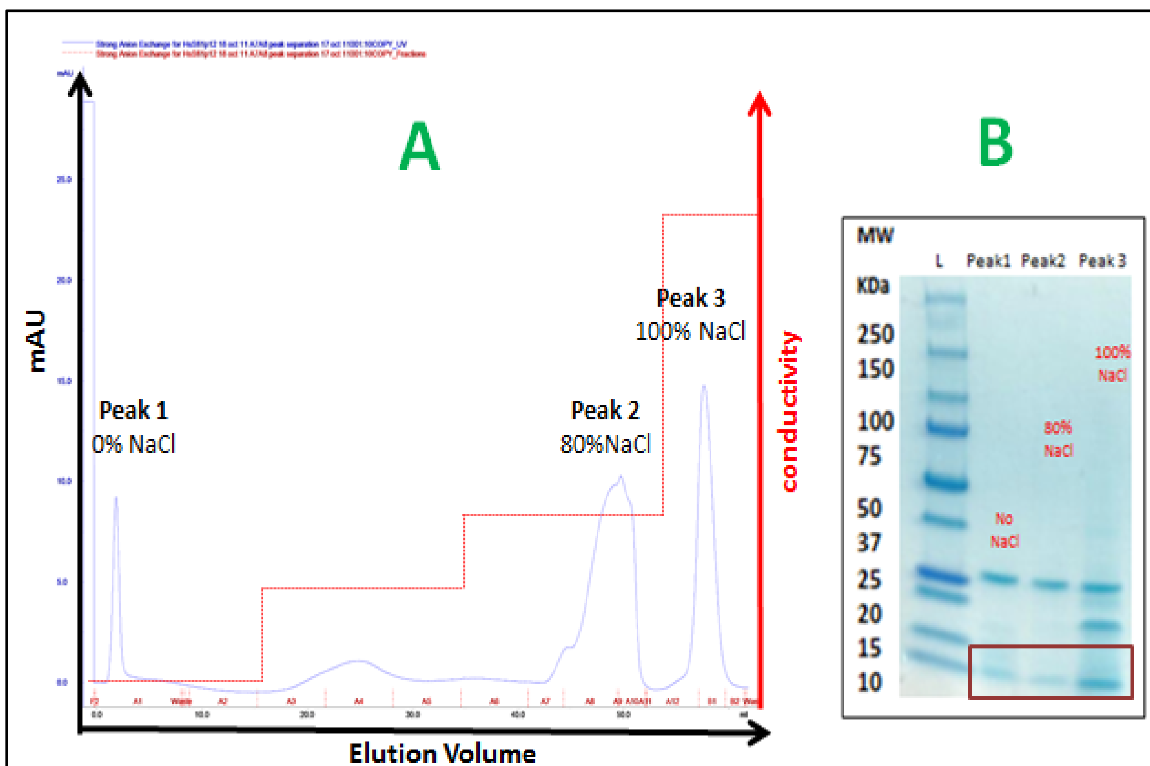


Figure 20. Anion exchange chromatogram and a 4-20% (Bis-Tris) gradient SDS-PAGE for a step gradient of 0-1.0 M NaCl. Purification of a low volume sample using the ÄKTA system. Elution of proteins that did not interact with the matrix column (Peak 1), Elution of proteins that did interact with the matrix column (Peaks 2-3).

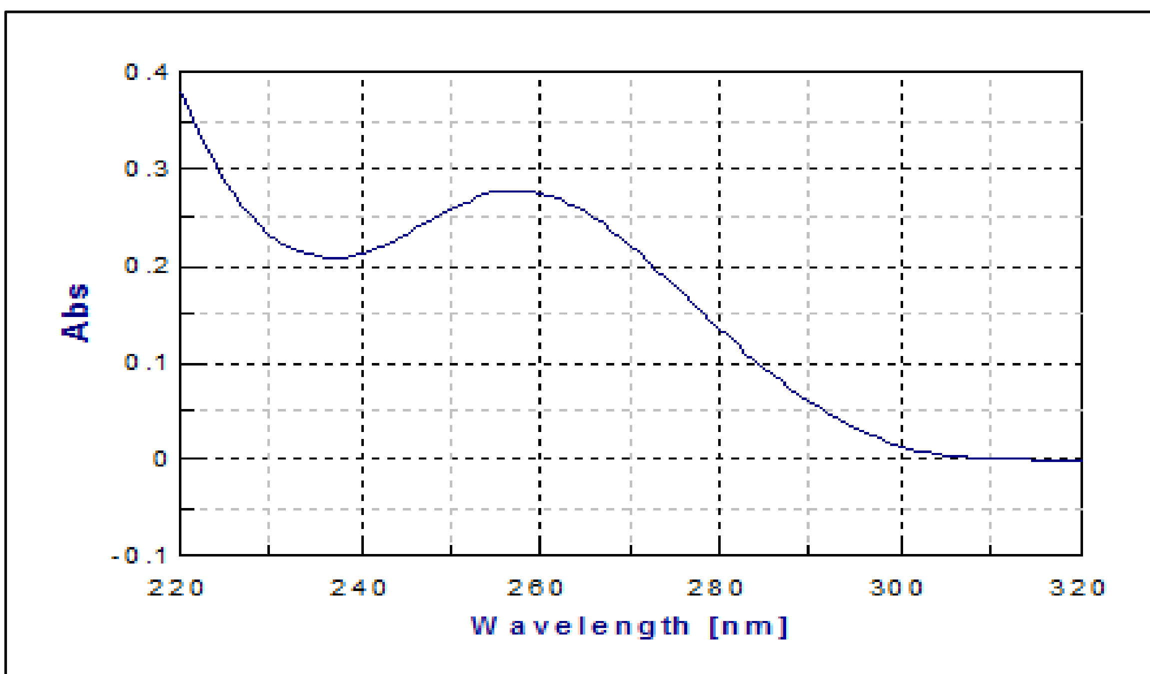


Figure 21. Characteristic UV/Vis spectrum of *Hs* Sfi1p$_{1-2}$ after anion exchange chromatography.

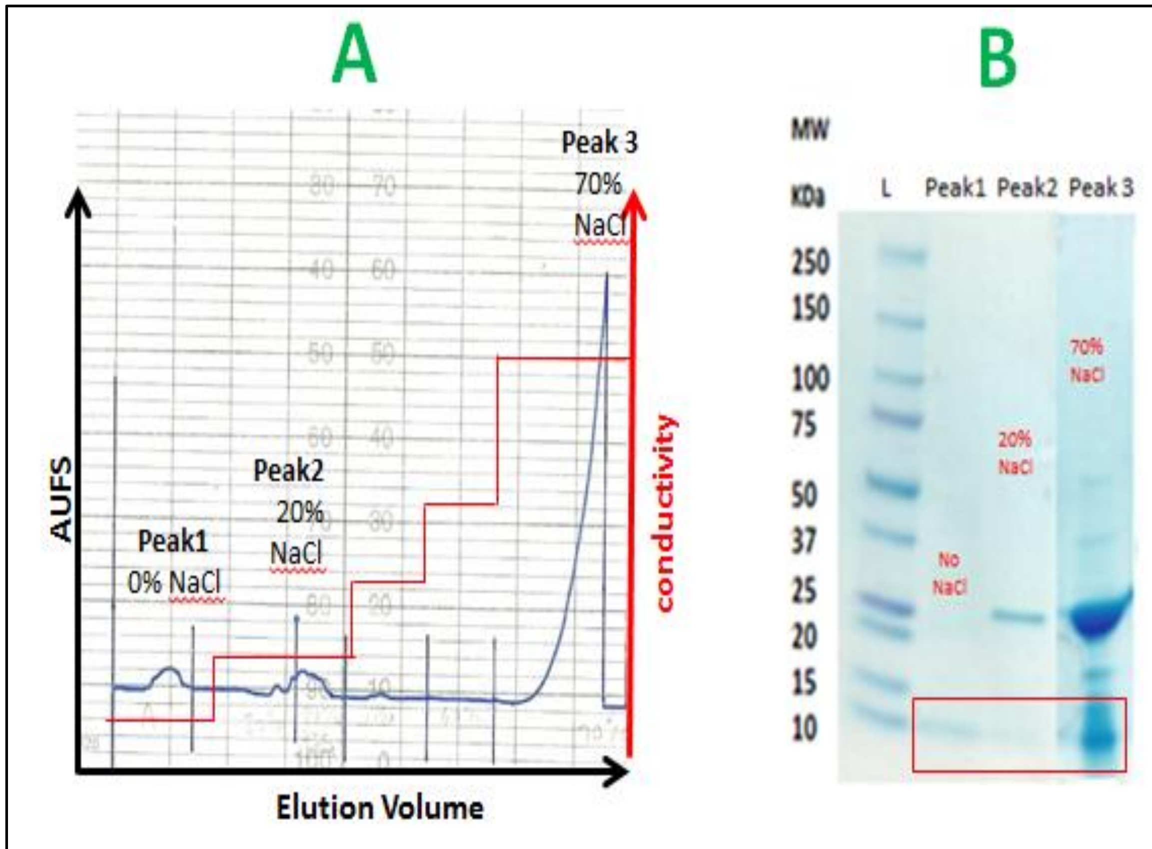


Figure 22. Anion exchange chromatogram and 4-20% (Bis-Tris) gradient SDS-PAGE for a step gradient from 0-0.70 M NaCl. Purification of a large volume sample. Elution of proteins that did not interact with the column (Peaks 1-2), Elution of proteins that did interact with the column matrix (Peak 3).

5.4 Isolation by Pellet Extraction Method

Suspecting that the protein band observed near 12 kDa was not His-*Hs* Sfi1p$_{1-2}$, the following hypothesis was stated: "***Proteolytic cleavage of the Histidine tag will result in less solubility of Hs Sfi1p$_{1-2}$ and as a consequence the peptide will stay in the pellet***". In fact, a comparison of the peptide's amino acid sequence with the histidine tag and without it using Expasy ProtParam [46] showed different characteristics (Figure 23). As observed in Figure 16, the sequence of the peptide with the histidine tag showed a considerable percent of charged amino acids that can increase the solubility of the peptide. The sequence of the peptide without the histidine tag demonstrated a higher percent of hydrophobic amino acids governing the sequence and as a consequence solubility of the peptide can be reduced. A second hypothesis is that: "***the hydrophobic residues in the protein dominate the biochemical behavior***." This would result in the protein interacting with other hydrophobic proteins during the isolation process and as a result remaining in the pellet, but during the solvent extraction the desired recombinant protein would be observed in the aqueous phase. Then, the following step performed was to run a 4-20% (Bis-Tris) gradient SDS-PAGE of diluted pellets including the three centrifugal steps, looking for the presence of a protein band near 12 kDa. In fact, Pellet 1 that corresponded to the first centrifugal step, showed a much defined band near 12 kDa (Figure 24). The small aliquots of Pellet 2 and Pellet 3 were also analyzed, which were shown to have little or no amount of the recombinant peptide. This means that these pellets were comprised mainly of contaminating proteins and degraded components of cellular membranes, assumed by its viscous nature. After finding this

evidence, an extraction method with organic solvents was use in order to isolate His-*Hs* Sfi1p$_{1-2}$ from the pellet. The solvent system used consists of a sequence of extractions with $CHCl_3$:CH_3OH (2:1, v/v) and a $CHCl_3$:CH_3OH (1:1, v/v) ratio and then evaporating the solvent system from the extracted samples. The sample fractions containing the organic and the aqueous phase $CHCl_3$:CH_3OH (2:1, v/v) were subjected to a 4-20% (Bis-Tris) gradient SDS-PAGE (Figure 25). The presence of a protein band near 12 kDa was observed for each phase.

Hs Sfi1p$_{1-2}$ with His-tag

Amino acid composition			
Ala	(A)	3	3.1%
Arg	(R)	8	8.3%
Asn	(N)	2	2.1%
Asp	(D)	6	6.2%
Cys	(C)	2	2.1%
Gln	(Q)	5	5.2%
Glu	(E)	6	6.2%
Gly	(G)	5	5.2%
His	(H)	9	9.4%
Ile	(I)	0	0.0%
Leu	(L)	6	6.2%
Lys	(K)	6	6.2%
Met	(M)	5	5.2%
Phe	(F)	6	6.2%
Pro	(P)	2	2.1%
Ser	(S)	3	3.1%
Thr	(T)	4	4.2%
Trp	(W)	5	5.2%
Tyr	(Y)	8	8.3%
Val	(V)	5	5.2%

96 amino acids
pI= 8.57
MW=12,119 Da

Hs Sfi1p$_{1-2}$ without His-tag

Amino acid composition			
Ala	(A)	2	3.1%
Arg	(R)	6	9.2%
Asn	(N)	1	1.5%
Asp	(D)	2	3.1%
Cys	(C)	2	3.1%
Gln	(Q)	3	4.6%
Glu	(E)	6	9.2%
Gly	(G)	0	0.0%
His	(H)	3	4.6%
Ile	(I)	0	0.0%
Leu	(L)	5	7.7%
Lys	(K)	5	7.7%
Met	(M)	1	1.5%
Phe	(F)	6	9.2%
Pro	(P)	2	3.1%
Ser	(S)	1	1.5%
Thr	(T)	3	4.6%
Trp	(W)	5	7.7%
Tyr	(Y)	7	10.8%
Val	(V)	5	7.7%

65 amino acids
pI= 8.98
MW=8,592 Da

Figure 23. Expasy ProtParam comparison of parameters for His-*Hs* Sfi1p$_{1-2}$ and *Hs* Sfi1p$_{1-2}$.The recombinant peptide with the histidine tag (left) and without it (right).

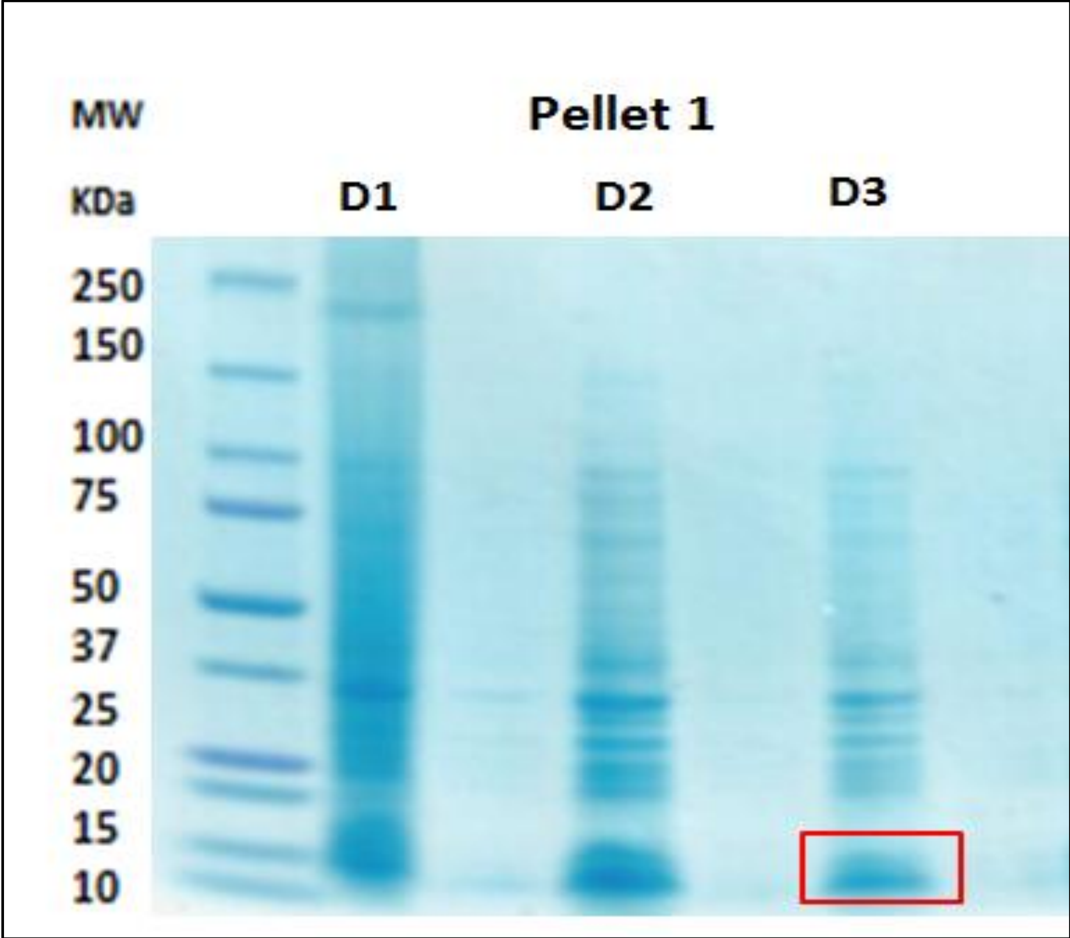


Figure 24. Pellet 1 dilutions (D1-D3) to identify a protein band at 12 kDa.

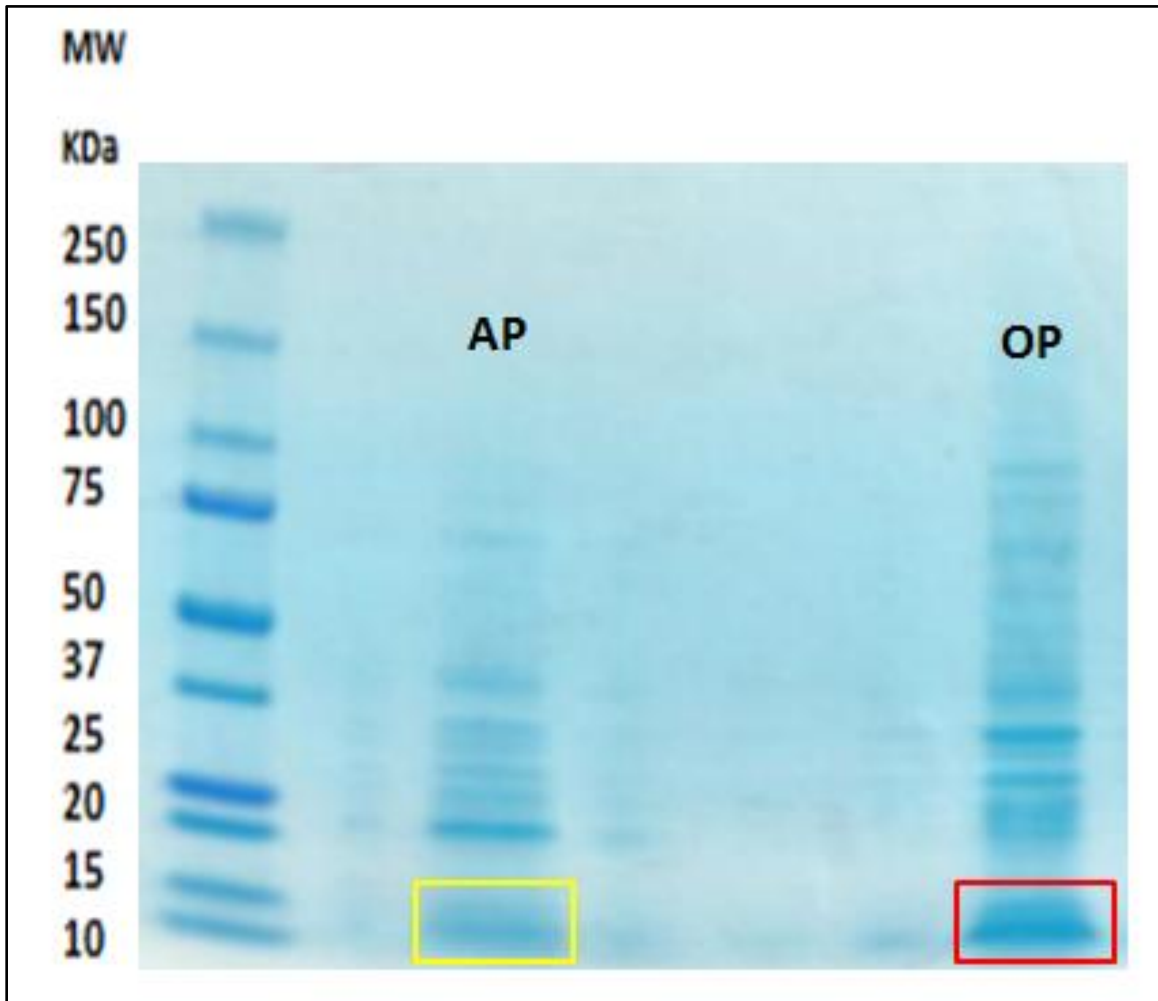


Figure 25. SDS-PAGE of the Aqueous Phase (AP) and the Organic Phase (OP) for $CHCl_3$:CH_3OH (2:1, v/v).

In addition, another 4-20% (Bis-Tris) gradient SDS-PAGE analysis was performed for both phases obtained from the $CHCl_3$:CH_3OH (1:1, v/v) ratio extraction and the presence of the protein band near 12 kDa was observed again (Figure 26). In the same analysis, the supernatant and the pellet of a lysed sample of Pellet 1 was run to make a comparison with the supernatant processed through a Mustang Q membrane that works in an analogous way to a High Q Strong Anion Exchange column. As seeing in Figure 19, after passing the supernatant through the membrane there is no visible band near 12 kDa. This represents additional evidence indicating that the protein isolated from the supernatant after centrifugation did not correspond to the recombinant peptide His-*Hs* Sfi1p_{1-2}. After consecutive extractions, the samples obtained for the aqueous and organic phases were compared in various 4-20% (Bis-Tris) gradient SDS-PAGE (Figure 27). At the bottom of Figure 27, one can notice a difference in relative mobility, suggesting that they could be possibly two different proteins. Therefore, further analysis based on the protein characteristics will confirm the identity of those protein samples. In fact, results from partial amino acid sequencing at the N-terminal confirmed the presence of partially His tagged *Hs* Sfi1p_{1-2} at the aqueous phase of the extraction (Figure 28). However, no hits were found for the organic phase, indicating that it is a different protein (Figure 28).

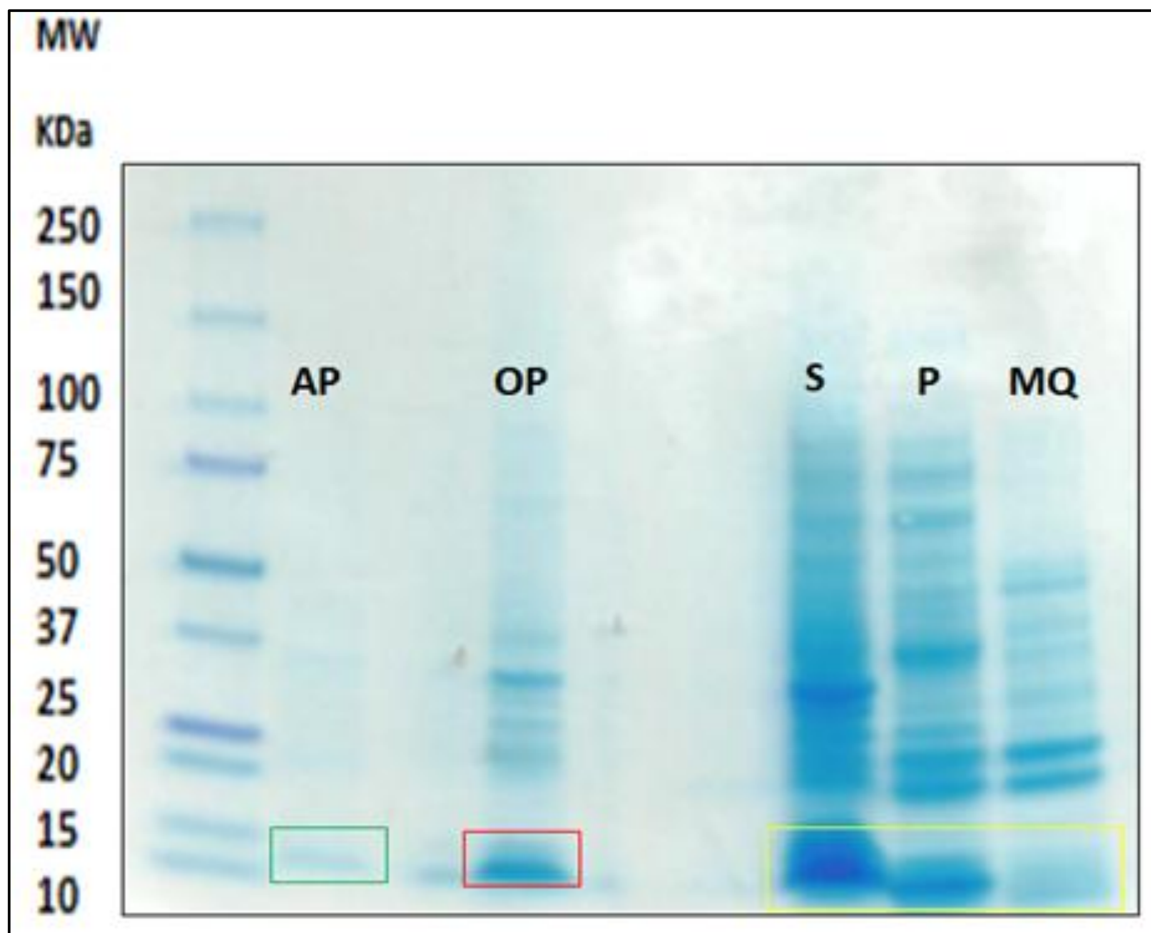


Figure 26. SDS-PAGE of the Aqueous phase (AP) and the Organic phase (OP) for the extraction $CHCl_3$:CH_3OH (1:1, v/v). At the left, comparison of the supernatant (S) and the pellet (P) with a sample treated with a Mustang Q Membrane (MQ).

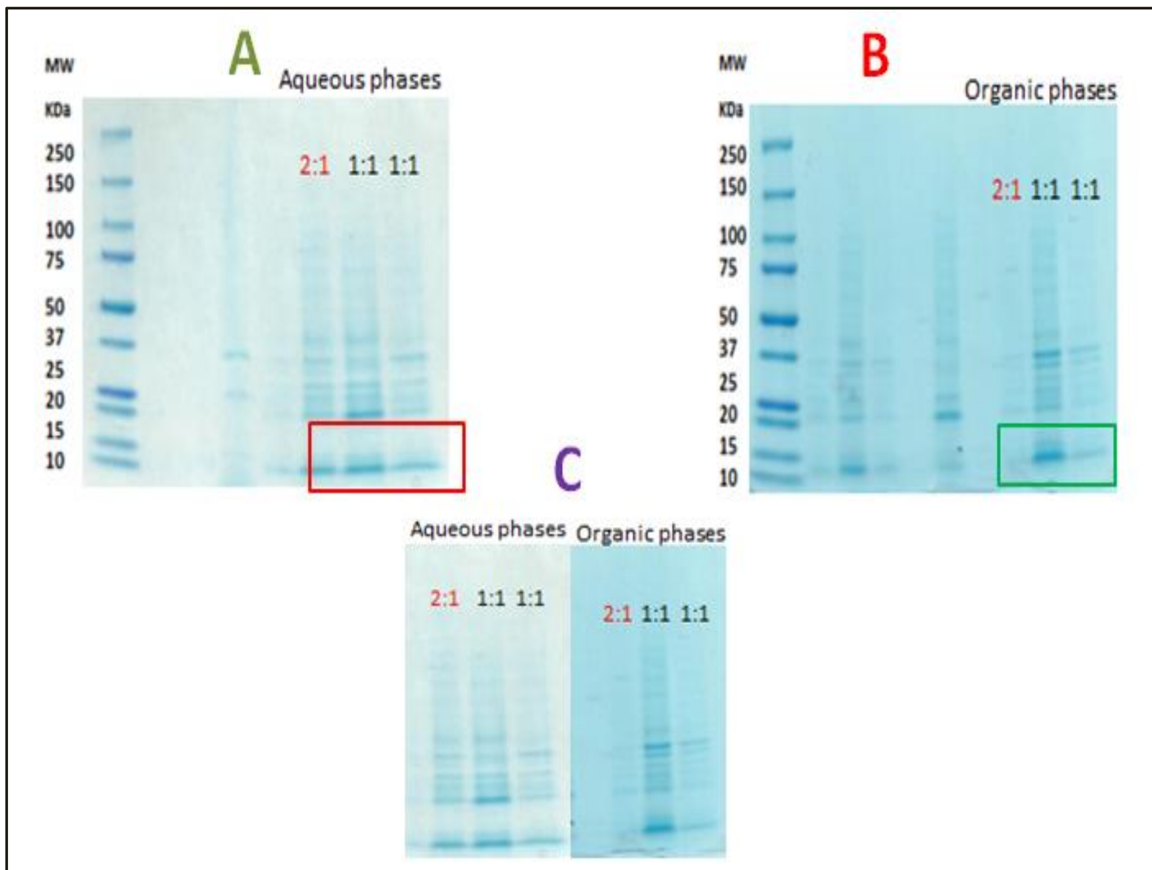


Figure 27. SDS-PAGE of consecutive extractions for the Aqueous phases (A), the Organic phases (B) and a comparison of relative mobility for both phases (C). The solvent extraction ratios referred to above are $CHCl_3$:CH_3OH (v/v ratios).

His-*Hs* Sfi1p$_{1-2}$	1						7
	M	**R**	**G**	**S**	**H**	**H**	**H**
SEQ RESULTS	M	R	G	S	H	H	V/I
12,119 Da							

not His-Hs Sfi1p1-2	1						7
	M	**R**	**G**	**S**	**H**	**H**	**H**
SEQ RESULTS	S	L	S	T	E	A	T

Figure 28. Partial amino acid sequencing results from the proteomics facility at Tufts University. Amino acid sequence for the aqueous phase sample (Top) and the organic phase sample (bottom).

5.5 Purification by Size Exclusion Chromatography

Following the successful identification of *Hs* $Sfi1p_{1-2}$ at the aqueous phase, a size exclusion chromatography was performed in order to remove contaminant proteins previously observed. Several aqueous sample batches were purified in order to obtain sufficient protein for the biochemical characterization and a similar chromatogram for each run was obtained as the one observed in (Figure 29). Size exclusion column's specifications established that the recombinant peptide can be eluted near 20 mL in the chromatogram. After performing a 4-20% (Bis-Tris) gradient SDS-PAGE of the different fractions pooled, a band was observed corresponding to the volume interval of 18-20 mL (Figure 30). However, the band was not completely pure and almost two less visible impurities were present. Finally, a UV-Vis spectroscopy confirmed a spectrum with absorbance maxima near 280 nm, as expected for *Hs* $Sfi1p_{1-2}$ (Figure 31).

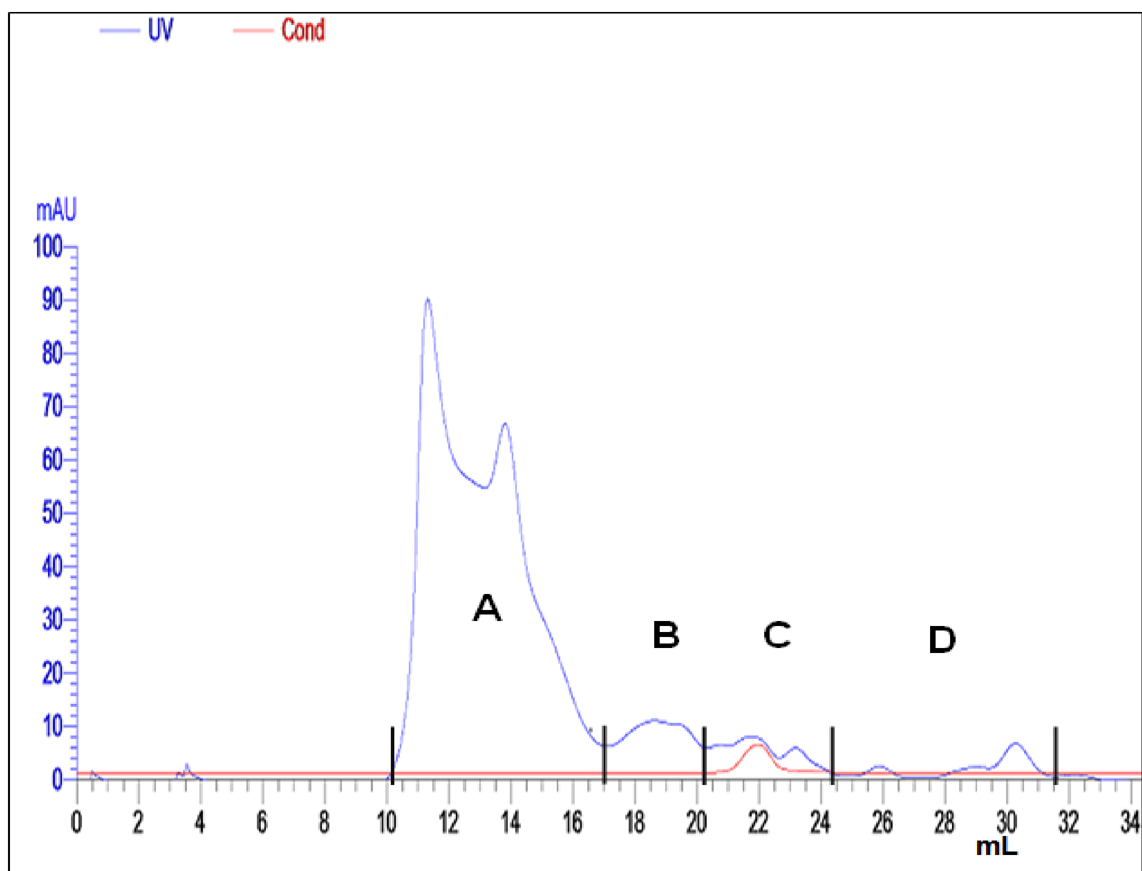


Figure 29. Size exclusion chromatogram of the aqueous phase sample.

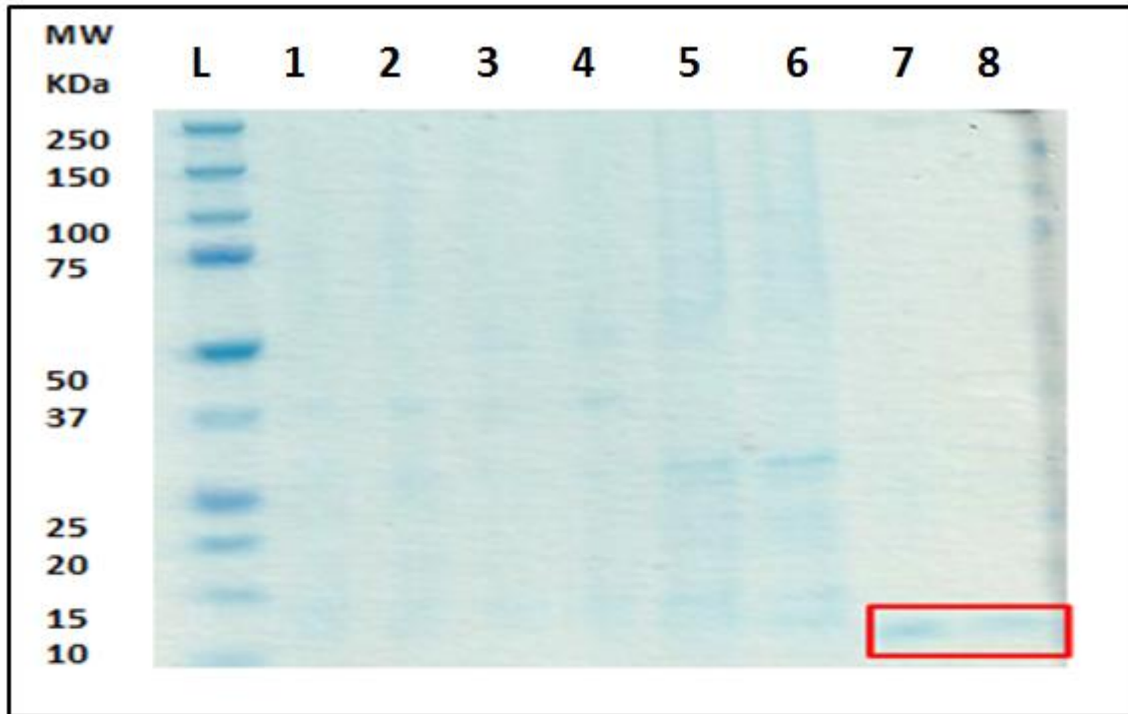


Figure 30. SDS-PAGE of size exclusion pooled fractions indicating the presence of *Hs* Sfi1$p_{1\text{-}2}$.

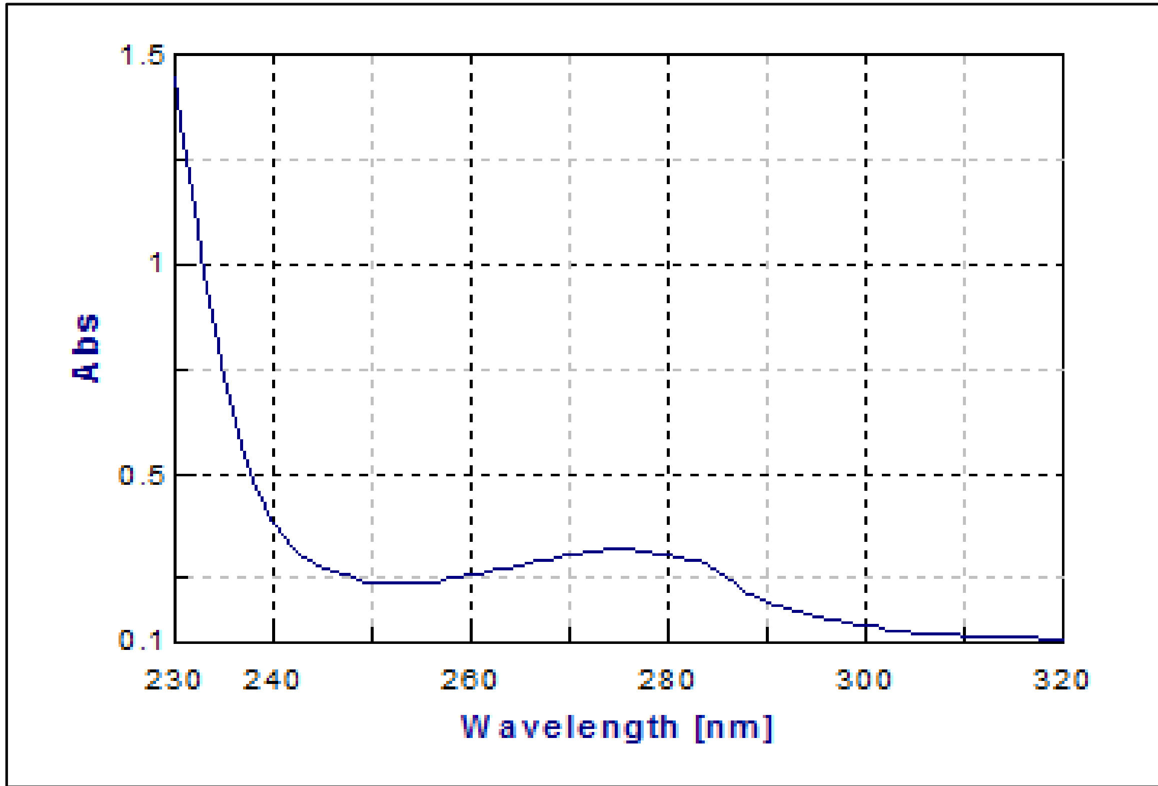


Figure 31. Characteristic UV/Vis spectrum of *Hs* Sfi1p$_{1-2}$ fractions after size exclusion chromatography.

6 CONCLUSIONS AND FUTURE WORK

The expression, isolation, purification, and biochemical characterization of recombinant *Hs* $Sfi1p_{1-2}$ have been described in this work. The over expression of *Hs* $Sfi1p_{1-2}$ was achieved using the BioFlo 3000 bench scale fermentor, obtaining a high yield of bacterial pellet and protein expression for the subsequent purification process. The isolation and purification strategies described here were performed based on the biochemical characteristics of the recombinant peptide and after a successful modification of the protocol previously established by Pastrana et al. [45]. A preliminary analysis by 4-20% (Bis-Tris) gradient SDS-PAGE of the isolation and purification process showed a prominent band at 12 kDa, however further analysis was required in order to confirm its identity. Another 4-20% (Bis-Tris) gradient SDS-PAGE analysis of lysed pellet 1 confirmed the presence of a 12 kDa protein, suspected to be His-*Hs* $Sfi1p_{1-2,}$ although it was supposed to be at the supernatant. The extraction method with organic solvents allowed further isolation of *Hs* $Sfi1p_{1-2}$ from pellet 1 sample. Moreover, the partial amino acid sequencing analysis confirmed the presence of the recombinant peptide with a Histidine tag at the aqueous phase. These findings suggests that although the His-tagged will aid in allowing the peptide to be more soluble there could be other two possible scenarios: 1) The Histidine tag could be degraded by proteolytic cleavage and the protein will remain at the pellet, 2) The hydrophobic nature of the protein could predominate and therefore the tag can represent only an alternative for proteins with little or no hydrophobic character. Results from partial amino acid sequencing for the aqueous phase protein sample supports the hypothesis that the hydrophobic residues in the protein domains the

biochemical behavior, and this could result in the protein interacting with other hydrophobic proteins during the isolation process and as a result remaining in the pellet.

It is highly recommended to optimize the process describe in this work for large scale isolation and purification, in order to obtain a higher *Hs* Sfi1p_{1-2} peptide yield for subsequent analysis. Further investigation will involve the study of the interaction of *Hs* Sfi1p_{1-2}- *Hs* centrin, using different biophysical techniques in order to elucidate the thermodynamics governing binding, complex stability, and the conformational changes involved in complex formation.

REFERENCES

1. Bornens, M. The centrosome in cells and organisms.Science.**2012**, *335*, 422-426.

2. Salisbury, J. L.; Whitehead, C. M.; Lingle, W. L.; Barret, S. L. Centrosomes and Cancer. J. Cell Biol. **1999,** *91,* 451-460.

3. Nigg, E. A. Centrosomes in development and disease. Wiley-VCH: Weinheim, **2004.**

4. Kilmartin, J.V. Sfi1 has conserved centrin-binding sites and an essential function in budding yeast spindle body duplication. J. Cell Biol. **2003,** *162,* 1211-1221.

5. Carvalho-Santos, Z.; Azimzadeh, J.; Pereira-Leal, J.B.; Bettencourt-Dias, M.Tracing the origins of centrioles,cilia and flagella. J. Cell Biol. **2011,** *194,* 165-175.

6. Kitagawa, D.; Vakonakis, I.; Olieric, N.; Hilbert, M.; Keller, D.; Olieric V.; Bortfeld, M.; Erat, M.C.; Fluckiger, I.; Gonczy, P.; Steinmetz, M.O. Structural basis of the 9-fold symmetry of centrioles. Cell. **2011,** *144,* 364-375.

7. Breugel, M.V.; Hirono, M.; Andreeva, A.; Yanagisawa, H.; Yamaguchi, S.; Nakazawa, Y.; Morgner, N.; Petrovich, M.; Ebong, I.A.; Robinson, C.V.; Johnson, C.M.; Veprintsev, D.; Zuber, B. Structure of sas-6 suggests its organization in centrioles. Science. **2011,** *331,* 1196-1199.

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

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

10. Hincheliffe, E. H.; Sluder, G. It Takes Two to Tango: understanding how centrosome duplication is regulated throughout the cell cycle. Genes Dev. **2001,** *15,* 1167-1181.

11. Salisbury, J.L. The contribution of epigenetic changes to abnormal centrosomes and genomic instability in breast cancer. J. Mammary Gland Biol. Neoplasia. 2001, 6, 203-212.

12. Salisbury, J.L. Centrosomes: Coiled-coils organize the cell center. Curr. Biol.**2003,** *13,* 88-90.

13. Azimzadeh, J.; Bornens, M. Structure and duplication of the centrosome. J. Cell Sci. 2007, 120. 2139-2142.

14. Vlastimil, S.; Merdes, A. The centrosome and cell proliferation. Cell Div. **2006,** *26,* 1-5.

15. Hildebrandt, F.; Benzing, T.; Katsanis, N. Ciliopathies. N. Engl. J. Med **2011,** *364,* 1533-1543.

16. Lancaster, M.; Gleeson, J.G. The primary cilium as a cellular signaling center: lessons from disease. Curr. Opin. Genet. Dev. **2009,** *19,* 220-229.

17. Bergmann, C. Ciliopathies. Eur. J. Pediatr. **2011,** *171,* 1285-3000.

18. Pugacheva E.N.; Jablonski, S.A.; Hartman, T.R.; Henske, E.P.; Golemis, E.A. HEF1-dependent Aurora A activation induces disassembly of the primary cilium. Cell. **2007,** *129,* 1351-1353.

19. Kilmartin, J.V. Sfi1 has conserved centrin-binding sites and an essential function in budding yeast spindle body duplication. J. Cell Biol. **2003,** *162,* 1211-1221.

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

21. Pecorino, L. M. Molecular Biology of Cancer. Oxford University Press; Oxford, NY, **2008**.

22. Salisbury, J.L.; Baron, A.; Surek, B.;Melkoniana, M. Striated flagellar roots: isolation and partial characterization of a calcium-modulated contractile organelle. J. Cell Biol. **1984,** *99,* 962-970.

23. Salisbury, J.L. A Mechanistic View on the Evolutionary Origin for Centrin-Based Control of Centriole Duplication. J.Cell. Physiol. **2007,** *213,* 420-428.

24. Trojan, P.; Krauss, N.; Choe, H-W.; Gießl, A.; Pulvermüller, A.;Wolfrum, U. Centrins in retinal photoreceptor cells: Regulators in the connecting cilium. Progr. Ret. Eye Research. **2008,** *27,* 237-259.

25. Gießl, A.; Trojan, P.; Pulvermüller, A.; Wolfrum, U. Centrins, potential regulators of transducing translocation in photoreceptor cells.**2004,** Williams

DS (ed.) In: *Cell biology and related disease of the outer retina.* World Scientific. 195-222.

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

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

28. Hart, P.E.; Glantz, J.N.; Orth, J.D.; Poynter, G.M.; and Salisbury, J.L. Testis-specific murine centrin, Cen1: genomic characterization and evidence for retroposition of a gene encoding a centrosome protein. Genomics.**1999,** *60*, 111-120.

29. Levy, Y.Y.; Lai, E.Y.; Remillard, S.P.; Heintzelman, M.B.; Fulton, C. Centrin is a conserved protein that forms diverse associations with centrioles and MTOCs in Naegleria and other organisms. Cell. Motil. Cytoskel. **1996,** *33,* 298-323.

30. Gogendeau, D.; Klotz, C.; Arnaiz, O.; Malinowska, A.; Dadlez, M.; Garreau de Loubresse, N.; Ruiz, F.; Koll, F.; Beisson, J. Functional diversification of centrins and cell morphological complexity. J.Cell Biol. **2007,** *121,* 65-74.

31. Baum, P.; Yip, C.; Goetsch, L.; Byers, B. A yeast gene essential for regulation of spindle pole duplication. Mol.Cell.Biol. **1988**, *8,* 5386-5397.

32. Marshall, W.F.; Vucica, Y.; Rosenbaum, J.L. Kinetics and regulation of de novo centriole assembly: Implications for the mechanism of centriole duplication. Curr.Biol. **2001**, *11,* 308-317.

33. Salisbury, J.L.; Suino, K.M.; Busby, R.; Springett, M. Centrin-2 is required for centriole duplication in mammalian. *Curr. Biol.* **2002,** *12,* 1287-1292.

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

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

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

37. Mazia, D.The chromosome cycle and centrosome cycle in the mitotic cycle. Int Review Cytol. **1987**,*100,* 49-9.

38. Middendorp, S.T.; Kuntziger, Y.; Abraham, S.; Holmes, N.; Bordes, M.; Paintrand, A. Paoletti, J.; Bornens, M. A role for centrin 3 in centrosome reproduction. J. Cell Biol. **2003,** *148,* 405–416.

39. Ma, P.; Winderickx, J.; Nauwelaers, D.; Dumortier, F.; De Doncker, A.; Thevelein, J.M.; Van Dijck, P. Deletion of SFI1, a novel suppressor of partial Ras-cAMP pathway deficiency in the yeast *Saccharomyces cerevisiae*, causes G2 arrest. Yeast.**1999**, *15, 1097*-1109.

40. Salisbury, J.L. Centrosomes: Sfi1p and centrin unravel a structural riddle. *Curr. Biol.* **2004**, *14*, 27-29.

41. Stemm-Wolf, A.J.; Fox, J.M.; Winey, M. Sfr13 is a member of a large family of asymmetrically localized Sfi1-repeat proteins and is important for basal body separation and stability in *Tetrahymena thermophile*.J Cell Sci. **2013**, *126*, 1477-1523.

42. http://www.ncbi.nlm.nih.gov. (Accessed on January, 2011)

43. Spang, A.; Courtney, I.; Grein, K.; Matzner, M.; Schiebel, E. The Cdc31p-binding protein Kar1p is a component of the half bridge of the yeast spindle pole body. J Cell Biol.**1995**, *128*, 863-877.

44. Anderson, E.A.; Prudden, J.; Prochnik, S.; Giddings, T.H.; Hardwick, K.G. Novel Sfi1 alleles uncover additional functions for Sfi1p in bipolar spindle assembly and function. Mol.Biol.Cell. **2007**, *18*, 2047-2056.

45. Pastrana-Rios, B.; Ocaña,W.; Rios, M.; Vargas, G.L.; Ysa, Y.; Poynter, G.; Tapia, J.; Salisbury,J.L. Centrin: Its Secondary Structure in the presence and absence of cations. Biochemistry. **2002,** *41*, 6911-6919.

46. http://www.expasy.org/tools/protparam.html. (Accessed on April, 2012)

47. Pastrana, B. (**1992**). M.Sc. Thesis Rutgers, The State University of New Jersey, New Jersey.

48. Pastrana-Rios, B. (**1995**). Ph.D. Thesis Rutgers, The State University of New Jersey, New Jersey.

49. Díaz-Casas,A. (**2012**). M.Sc. Thesis, University of Puerto Rico, Mayagüez.

APPENDIX A

QUIM 8995: BIOINFORMATICS, PROTEINS & PROTEIN-PROTEIN INTERACTIONS (PPI'S)
SEQUENCE ANALYSIS CLUSTALW

INTRODUCTION

The study of proteins is not limited to the acquisition of data directly from laboratory experiments. It is highly desirable to perform first a computational analysis to gain some previous knowledge about the protein. Bioinformatics can provide us the necessary tools, because it is the application of computer science and information technology to fields like biology, chemistry and related areas. Therefore the use of Bioinformatics for the study of proteins in Biology and Chemistry is crucial. Particularly for protein analysis, sequence alignment of amino acids helps to identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships between different organisms. ClustalW is a widely used multiple sequence alignment computer program, that can perform a pairwise analysis and create a phylogenetic tree. It uses an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. The program enables the researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. Alignments can reflect a degree of evolutionary change between sequences descended from a common ancestor or convergent evolution can occur to produce apparent similarity between proteins that are evolutionarily unrelated but perform similar functions and have similar structures. For the purpose of this project, Sfi1 protein 23 binding sites and Sfi1 entire sequence from diverse organisms was analyzed by using ClustalW.

ORGANISMS FOR SFI1 SEQUENCE ANALYSIS

Organism	Protein name	Protein Database	Accession Number	Length
Homo sapiens (human)				
Isoform 1	*HsSFI1_1*	UniProt	A8K8P3-1	1242 aa
Isoform 2	*HsSFI1_2*	UniProt	A8K8P3-2	1211 aa
Isoform 3	*HsSFI1_3*	UniProt	A8K8P3-3	1160 aa
Isoform 4	*HsSFI1_4*	UniProt	A8K8P3-4	1031 aa
Isoform 5	*HsSFI1_5*	UniProt	A8K8P3-5	991 aa
Isoform 9	*HsSFI1_9*	UniProt	A8K8P3-9	1187 aa
Isoform 10	*HsSFI1_10*	UniProt	A8K8P3-10	1148 aa
Mus musculus (mouse)				
Isoform 1	*MmSFI1_1*	UniProt	Q3UZY0-1	1216 aa
Isoform 2	*MmSFI1_2*	UniProt	Q3UZY0-2	1184 aa
Isoform 3	*MmSFI1_3*	UniProt	Q3UZY0-3	717 aa
Isoform 4	*MmSFI1_4*	UniProt	Q3UZY0-4	120 aa
Isoform 5	*MmSFI1_5*	UniProt	Q3UZY0-5	485 aa
Sus Scrofa (pig)	*SsFI1*	UniProt	F1RLV5	1213 aa
Canis familiaris (canine)	*CfSFI1*	UniProt	E2RNU8	1246 aa
Rattus norvegicus (rata)	*RnSFI1*	UniProt	D3ZTG6	1209 aa
Chlamydomonas reinhardtii (alga)	*CrSFI1*	UniProt	A8J1N9	2988 aa
Saccharomyces cerevisiae (levadura)	*ScSFI1*	UniProt	Q12369	946 aa
Danio rerio (pez Cebra)	*DrSFI1*	UniProt	F1QNA2	921 aa
Pan troglodytes (chimpance)				
P.troglodytes Sfi1 homolog	*PtSfi1*	NCBI	XP_003317259.1	1186aa
P.troglodytes Sfi1 homolog 1	*PtSfi1_1*	NCBI	XP_003317260.1	1147 aa
Isoform 3	*PtSfi1_3*	NCBI	XP_001147952.2	1159 aa
Isoform 8	*PtSfi1_8*	NCBI	XP_001148452.1	1241 aa
Isoform 10	*PtSfi1_10*	NCBI	XP_001148585.1	1210 aa
Gorilla gorilla	*Not available*	N/A	N/A	N/A
Escherichia coli (bacteria)	*Not available*	N/A	N/A	N/A
Scherffelia dubia (alga)	*Not available*	N/A	N/A	N/A
Caenorhabditis elegans (nematode)	*Not available*	N/A	N/A	N/A

ORGANISMS AMINO ACID SEQUENCES

>HsSFI1_1

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RYFKDGAVKK PYSAKTLSNK KSSASFGIRR
ELPSTSHLVQ YRGTHTCTRQ GRLRELRIRC VARKFLYLWI RMTFGRVFPS KARFYYEQRL
LRKVFEEWKE EWWVFQHEWK LCVRADCHYR YYLYNLMFQT WKTYVRQQQE MRNKYIRAEV
HDAKQKMRQA WKSWLIYVVV RRTKLQMQTT ALEFRQRIIL RVWWSTWRQR LGQVRVSRAL
HASALKHRAL SLQVQAWSQW REQLLYVQKE KQKVVSAVKH HQHWQKRRFL KAWLEYLQVR
RVKRQQNEMA ERFHHVTVLQ IYFCDWQQAW ERRESLYAHH AQVEKLARKM ALRRAFTHWK
HYMLLCAEEA AQFEMAEEHH RHSQLYFCFR ALKDNVTHAH LQQIRRNLAH QQHGVTLLHR
FWNLWRSQIE QKKERELLPL LHAAWDHYRI ALLCKCIELW LQYTQKRRYK QLLQARADGH
FQQRALPAAF HTWNRLWRWR HQENVLSARA TRFHRETLEK QVFSLWRQKM FQHRENRLAE
RMAILHAERQ LLYRSWFMWH QQAAARHQEQ EWQTVACAHH RHGRLKKAFC LWRESAQGLR
TERTGRVRAA EFHMAQLLRW AWSQWRECLA LRGAERQKLM RADLHHQHSV LHRALQAWVT
YQGRVRSILR EVAARESQHN RQLLRGALRR WKENTMARVD EAKKTFQAST HYRRTICSKV
LVQWREAVSV QMYYRQQEDC AIWEAQKVLD RGCLRTWFQR WWDCSRRSAQ QRLQLERAVQ
HHHRQLLLEG LARWKTHHLQ CVRKRLLHRQ STQLLAQRLS RTCFRQWRQQ LAARRQEQRA
TVRALWFWAF SLQAKVWATW LAFVLERRRK KARLQWALQA YQGQLLQEGA TRLLRFAASM
KASRQQLQAQ QQVQAAHSLH RAVRRCATLW KQKVLGRGGK PQPLAAIAPS RKVTFEGPLL
NRIAAGAGDG TLETKRPQAS RPLGALGRLA AEEPHALELN TAHSARKQPR RPHFLLEPAQ
SQRPQKPQEH GLGMAQPAAP SLTRPFLAEA PTALVPHSPL PGALSSAPGP KQPPTASTGP
ELLLLPLSSF MPCGAAAPAR VSAQRATPRD KPPVPSSLAS VPDPHLLLPG DFSATRAGPG
LSTAGSLDLE AELEEIQQQL LHYQTTKQNL WSCRRQASSL RRWLELNREE PGPEDQEVEQ
 QVQKELEQVE MQIQLLAEEL QAQRQPIGAC VARIQALRQA LC
```

>HsSFI1_2

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RYFKDGAVKK PYSAKTLSNK KSSASFGIRR
ELPSTSHLVQ YRGTHTCTRQ GRLRELRIRC VARKFLYLWI RMTFGRVFPS KARFYYEQRL
LRKVFEEWKE EWWVFQHEWK LCVRADCHYR YYLYNLMFQT WKTYVRQQQE MRNKYIRAEV
HDAKQKMRQA WKSWLIYVVV RRTKLQMQTT ALEFRQRIIL RVWWSTWRQR LGQVRVSRAL
HASALKHRAL SLQVQAWSQW REQLLYVQKE KQKVVSAVKH HQHWQKRRFL KAWLEYLQVR
RVKRQQNEMA ERFHHVTVLQ IYFCDWQQAW ERRESLYAHH AQVEKLARKM ALRRAFTHWK
HYMLLCAEEA AQFEMAEEHH RHSQLLLHRF WNLWRSQIEQ KKERELLPLL HAAWDHYRIA
LLCKCIELWL QYTQKRRYKQ LLQARADGHF QQRALPAAFH TWNRLWRWRH QENVLSARAT
```

```
RFHRETLEKQ VFSLWRQKMF QHRENRLAER MAILHAERQL LYRSWFMWHQ QAAARHQEQE
WQTVACAHHR HGRLKKAFCL WRESAQGLRT ERTGRVRAAE FHMAQLLRWA WSQWRECLAL
RGAERQKLMR ADLHHQHSVL HRALQAWVTY QGRVRSILRE VAARESQHNR QLLRGALRRW
KENTMARVDE AKKTFQASTH YRRTICSKVL VQWREAVSVQ MYYRQQEDCA IWEAQKVLDR
GCLRTWFQRW WDCSRRSAQQ RLQLERAVQH HHRQLLLEGL ARWKTHHLQC VRKRLLHRQS
TQLLAQRLSR TCFRQWRQQL AARRQEQRAT VRALWFWAFS LQAKVWATWL AFVLERRRKK
ARLQWALQAY QGQLLQEGAT RLLRFAASMK ASRQQLQAQQ QVQAAHSLHR AVRRCATLWK
QKVLGRGGKP QPLAAIAPSR KVTFEGPLLN RIAAGAGDGT LETKRPQASR PLGALGRLAA
EEPHALELNT AHSARKQPRR PHFLLEPAQS QRPQKPQEHG LGMAQPAAPS LTRPFLAEAP
TALVPHSPLP GALSSAPGPK QPPTASTGPE LLLLPLSSFM PCGAAAPARV SAQRATPRDK
PPVPSSLASV PDPHLLLPGD FSATRAGPGL STAGSLDLEA ELEEIQQQLL HYQTTKQNLW
SCRRQASSLR RWLELNREEP GPEDQEVEQQ VQKELEQVEM QIQLLAEELQ AQRQPIGACV
 ARIQALRQAL C
```

>HsSFI1_3

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RFYYEQRLLR KVFEEWKEEW WVFQHEWKLC

VRADCHYRYY LYNLMFQTWK TYVRQQQEMR NKYIRAEVHD AKQKMRQAWK SWLIYVVVRR

TKLQMQTTAL EFRQRIILRV WWSTWRQRLG QVRVSRALHA SALKHRALSL QVQAWSQWRE

QLLYVQKEKQ KVVSAVKHHQ HWQKRRFLKA WLEYLQVRRV KRQQNEMAER FHHVTVLQIY

FCDWQQAWER RESLYAHHAQ VEKLARKMAL RRAFTHWKHY MLLCAEEAAQ FEMAEEHHRH

SQLYFCFRAL KDNVTHAHLQ QIRRNLAHQQ HGVTLLHRFW NLWRSQIEQK KERELLPLLH

AAWDHYRIAL LCKCIELWLQ YTQKRRYKQL LQARADGHFQ QRALPAAFHT WNRLWRWRHQ

ENVLSARATR FHRETLEKQV FSLWRQKMFQ HRENRLAERM AILHAERQLL YRSWFMWHQQ

AAARHQEQEW QTVACAHHRH GRLKKAFCLW RESAQGLRTE RTGRVRAAEF HMAQLLRWAW

SQWRECLALR GAERQKLMRA DLHHQHSVLH RALQAWVTYQ GRVRSILREV AARESQHNRQ

LLRGALRRWK ENTMARVDEA KKTFQASTHY RRTICSKVLV QWREAVSVQM YYRQQEDCAI

WEAQKVLDRG CLRTWFQRWW DCSRRSAQQR LQLERAVQHH HRQLLLEGLA RWKTHHLQCV

RKRLLHRQST QLLAQRLSRT CFRQWRQQLA ARRQEQRATV RALWFWAFSL QAKVWATWLA

FVLERRRKKA RLQWALQAYQ GQLLQEGATR LLRFAASMKA SRQQLQAQQQ VQAAHSLHRA
```

```
VRRCATLWKQ KVLGRGGKPQ PLAAIAPSRK VTFEGPLLNR IAAGAGDGTL ETKRPQASRP
LGALGRLAAE EPHALELNTA HSARKQPRRP HFLLEPAQSQ RPQKPQEHGL GMAQPAAPSL
TRPFLAEAPT ALVPHSPLPG ALSSAPGPKQ PPTASTGPEL LLLPLSSFMP CGAAAPARVS
AQRATPRDKP PVPSSLASVP DPHLLLPGDF SATRAGPGLS TAGSLDLEAE LEEIQQQLLH
YQTTKQNLWS CRRQASSLRR WLELNREEPG PEDQEVEQQV QKELEQVEMQ IQLLAEELQA
QRQPIGACVA RIQALRQALC
```

>HsSFI1_4

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RFYYEQRLLR KVFEEWKEEW WVFQHEWKLC
VRADCHYRVW WSTWRQRLGQ VRVSRALHAS ALKHRALSLQ VQAWSQWREQ LLYVQKEKQK
VVSAVKHHQH WQKRRFLKAW LEYLQVRRVK RQQNEMAERF HHVTVLQIYF CDWQQAWERR
ESLYAHHAQV EKLARKMALR RAFTHWKHYM LLCAEEAAQF EMAEEHHRHS QLYFCFRALK
DNVTHAHLQQ IRRNLAHQQH GVTLLHRFWN LWRSQIEQKK ERELLPLLHA AWDHYRIALL
CKCIELWLQY TQKRRYKQLL QARADGHFQQ RALPAAFHTW NRLWRWRHQE NVLSARATRF
HRETLEKQVF SLWRQKMFQH RENRLAERMA ILHAERQLLY RSWFMWHQQA AARHQEQEWQ
TVACAHHRHG RLKKAFCLWR ESAQGLRTER TGRVRAAEFH MAQLLRWAWS QWRECLALRG
AERQKLMRAD LHHQHSVLHR ALQAWVTYQG RVRSILREVA ARESQHNRQL LRGALRRWKE
NTMARVDEAK KTFQASTHYR RTICSKVLVQ WREAVSVQMY YRQQEDCAIW EAQKVLDRGC
LRTWFQRWWD CSRRSAQQRL QLERAVQHHH RQLLLEGLAR WKTHHLQCVR KRLLHRQSTQ
LLAQRLSRTC FRQWRQQLAA RRQEQRATVR ALWFWAFSLQ AKVWATWLAF VLERRRKKAR
LQWALQAYQG QLLQEGATRL LRFAASMKAS RQQLQAQQQV QAAHSLHRAV RRCATLWKQK
VLGRGGKPQP LAAIAPSRKV TFEGPLLNRI AAGAGDGTLE TKRPQASRPL GALGRLAAEE
PHALELNTAH SARKQPRRPH FLLEPAQSQR PQKPQEHGLG MAQPAAPSLT RPFLAEAPTA
LVPHSPLPGA LSSAPGPKQP PTASTGPELL LLPLSSFMPC GAAAPARVSA QRATPRDKPP
VPSSLASVPD PHLLLPGDFS ATRAGPGLST AAWTLRLNLR RSSSNYCTTR PPSRTSGPVG
GKRAACAGGW S
```

>HsSFI1_5

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RYFKDGAVKK PYSAKTLSNK KSSASFGIRR
ELPSTSHLVQ YRGTHTCTRQ GRLRELRIRF YYEQRLLRKV FEEWKEEWWV FQHEWKLCVR
ADCHYRYYLY NLMFQTWKTY VRQQQEMRNK YIRAEVHDAK QKMRQAWKSW LIYVVVRRTK
LQMQTTALEF RQRIILRVWW STWRQRLGQV RVSRALHASA LKHRALSLQV QAWSQWREQL
LYVQKEKQKV VSAVKHHQHW QKRRFLKAWL EYLQVRRVKR QQNEMAERFH HVTVLQIYFC
DWQQAWERRE SLYAHHAQVE KLARKMALRR AFTHWKHYML LCAEEAAQFE MAEEHHRHSQ
LYFCFRALKD NVTHAHLQQI RRNLAHQQHG VTLLHRFWNL WRSQIEQKKE RELLPLLHAA
WDHYRIALLC KCIELWLQYT QKRRYKQLLQ ARADGHFQQR ALPAAFHTWN RLWRWRHQEN
VLSARATRFH RETLEKQVFS LWRQKMFQHR ENRLAERMAI LHAERQLLYR SWFMWHQQAA
ARHQEQEWQT VACAHHRHGR LKKAFCLWRE SAQGLRTERT GRVRAAEFHM AQLLRWAWSQ
WRECLALRGA ERQKLMRADL HHQHSVLHRA LQAWVTYQGR VRSILREVAA RESQQLQAQQ
QVQAAHSLHR AVRRCATLWK QKVLGRGGKP QPLAAIAPSR KVTFEGPLLN RIAAGAGDGT
LETKRPQASR PLGALGRLAA EEPHALELNT AHSARKQPRR PHFLLEPAQS QRPQKPQEHG
LGMAQPAAPS LTRPFLAEAP TALVPHSPLP GALSSAPGPK QPPTASTGPE LLLLPLSSFM
PCGAAAPARV SAQRATPRDK PPVPSSLASV PDPHLLLPGD FSATRAGPGL STAGSLDLEA
ELEEIQQQLL HYQTTKQNLW SCRRQASSLR RWLELNREEP GPEDQEVEQQ VQKELEQVEM
QIQLLAEELQ AQRQPIGACV ARIQALRQAL C
```

>HsSFI1_9

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RYFKDGAVKK PYSAKTLSNK KSSASFGIRR
ELPSTSHLVQ YRGTHTCTRQ GRLRELRIRF YYEQRLLRKV FEEWKEEWWV FQHEWKLCVR
ADCHYRYYLY NLMFQTWKTY VRQQQEMRNK YIRAEVHDAK QKMRQAWKSW LIYVVVRRTK
LQMQTTALEF RQRIILRVWW STWRQRLGQV RVSRALHASA LKHRALSLQV QAWSQWREQL
LYVQKEKQKV VSAVKHHQHW QKRRFLKAWL EYLQVRRVKR QQNEMAERFH HVTVLQIYFC
DWQQAWERRE SLYAHHAQVE KLARKMALRR AFTHWKHYML LCAEEAAQFE MAEEHHRHSQ
LLLHRFWNLW RSQIEQKKER ELLPLLHAAW DHYRIALLCK CIELWLQYTQ KRRYKQLLQA
```

```
RADGHFQQRA LPAAFHTWNR LWRWRHQENV LSARATRFHR ETLEKQVFSL WRQKMFQHRE
NRLAERMAIL HAERQLLYRS WFMWHQQAAA RHQEQEWQTV ACAHHRHGRL KKAFCLWRES
AQGLRTERTG RVRAAEFHMA QLLRWAWSQW RECLALRGAE RQKLMRADLH HQHSVLHRAL
QAWVTYQGRV RSILREVAAR ESQHNRQLLR GALRRWKENT MARVDEAKKT FQASTHYRRT
ICSKVLVQWR EAVSVQMYYR QQEDCAIWEA QKVLDRGCLR TWFQRWWDCS RRSAQQRLQL
ERAVQHHHRQ LLLEGLARWK THHLQCVRKR LLHRQSTQLL AQRLSRTCFR QWRQQLAARR
QEQRATVRAL WFWAFSLQAK VWATWLAFVL ERRRKKARLQ WALQAYQGQL LQEGATRLLR
FAASMKASRQ QLQAQQQVQA AHSLHRAVRR CATLWKQKVL GRGGKPQPLA AIAPSRKVTF
EGPLLNRIAA GAGDGTLETK RPQASRPLGA LGRLAAEEPH ALELNTAHSA RKQPRRPHFL
LEPAQSQRPQ KPQEHGLGMA QPAAPSLTRP FLAEAPTALV PHSPLPGALS SAPGPKQPPT
ASTGPELLLL PLSSFMPCGA AAPARVSAQR ATPRDKPPVP SSLASVPDPH LLLPGDFSAT
RAGPGLSTAG SLDLEAELEE IQQQLLHYQT TKQNLWSCRR QASSLRRWLE LNREEPGPED
QEVEQQVQKE LEQVEMQIQL LAEELQAQRQ PIGACVARIQ ALRQALC
```

>HsSFI1_10

```
MKNLLTEKCI SSHNFHQKVI KQRMEKKVDS RFYYEQRLLR KVFEEWKEEW WVFQHEWKLC
VRADCHYRYY LYNLMFQTWK TYVRQQQEMR NKYIRAEVHD AKQKMRQAWK SWLIYVVVRR
TKLQMQTTAL EFRQRIILRV WWSTWRQRLG QVRVSRALHA SALKHRALSL QVQAWSQWRE
QLLYVQKEKQ KVVSAVKHHQ HWQKRRFLKA WLEYLQVRRV KRQQNEMAER FHHVTVLQIY
FCDWQQAWER RESLYAHHAQ VEKLARKMAL RRAFTHWKHY MLLCAEEAAQ FEMAEEHHRH
SQLYFCFRAL KDNVTHAHLQ QIRRNLAHQQ HGVTLLHRFW NLWRSQIEQK KERELLPLLH
AAWDHYRIAL LCKCIELWLQ YTQKRRYKQL LQARADGHFQ QRALPAAFHT WNRLWRWRHQ
ENVLSARATR FHRETLEKQV FSLWRQKMFQ HRENRLAERM AILHAERQLL YRSWFMWHQQ
AAARHQEQEW QTVACAHHRH GRLKKAFCLW RESAQGLRTE RTGRVRAAEF HMAQLLRWAW
SQWRECLALR GAERQKLMRA DLHHQHSVLH RALQAWVTYQ GRVRSILREV AARESQHNRQ
LLRGALRRWK ENTMARVDEA KKTFQVLVQW REAVSVQMYY RQQEDCAIWE AQKVLDRGCL
```

```
RTWFQRWWDC SRRSAQQRLQ LERAVQHHHR QLLLEGLARW KTHHLQCVRK RLLHRQSTQL
LAQRLSRTCF RQWRQQLAAR RQEQRATVRA LWFWAFSLQA KVWATWLAFV LERRRKKARL
QWALQAYQGQ LLQEGATRLL RFAASMKASR QQLQAQQQVQ AAHSLHRAVR RCATLWKQKV
LGRGGKPQPL AAIAPSRKVT FEGPLLNRIA AGAGDGTLET KRPQASRPLG ALGRLAAEEP
HALELNTAHS ARKQPRRPHF LLEPAQSQRP QKPQEHGLGM AQPAAPSLTR PFLAEAPTAL
VPHSPLPGAL SSAPGPKQPP TASTGPELLL LPLSSFMPCG AAAPARVSAQ RATPRDKPPV
PSSLASVPDP HLLLPGDFSA TRAGPGLSTA GSLDLEAELE EIQQQLLHYQ TTKQNLWSCR
RQASSLRRWL ELNREEPGPE DQEVEQQVQK ELEQVEMQIQ LLAEELQAQR QPIGACVARI
QALRQALC
```

>MmSFI1_1

```
MEKKIGSRSF RDGVVKKPCS PKTLPLKKSS AFSGIQREPS RSCHSIYYHA SQNWTRYRLQ
ELRIRCVARK FLYLWIRVTF GRVTPSRARI FHEQKILQKV FGEWREEWWV SQREWKLCVR
ADCHYRYYLY NLIFQNWKTF VHQQREMRKR FRIAEHHDTK QKMCQAWKSW LIYMVSRRTK
LHMKTTALEF RRQSVLCFWW SKWRWRLGQA HAEHALHAVA VKHRALSLQL QGWLRWQEQL
LISQRDRRKE ATAVQHYQHW QKQRSLKAWL KYLQICRVKR WQNEMAVQFH RATVLQIHFC
DWQWAWEWRQ SLSAHQALVV KLAGRMVLRR AFTHWKHYML LQAEEAAQRE AAAEHRQHYL
LYSCFRAFKD NVTQARLQQT RKKLAQQLRD TTLLHRFWNL WQSRIEQREE RVQTPSLHAA
LSHYRVTVLH KCVRVWLRYV HKRQWQQLLR ARADGHFQQR ALPAAFYTWY RGWLWHQQRR
ILHTKAVRFH RGTLEKQVFA LWRQKMSQHR ENCLAERMAI LQAEQQLLRR FWFVWHQQAA
VCQLERQQQA MAIAHHHSGL LRRAFCIWKE STQGFRIERM GRAQAAHFHS AQLLSRAWSM
WRECLALRLE EQQKLKCAAL HSQCILLRRA LQKWLVYQNR VRSVLREVAA RERQHNRQLL
WWALHLWREN TMARLDGAKK TSQARVHYSR TLCSKVLVQW REVTSVQIYY RQKEAAALRE
ARKALDRGRL QNWFQHWRFC SQRAAQQRFQ LGQAAQHHHW QLLMEAMARW KAHHLGCIRK
KFLQRQAAQL LAQRLSRACF CQWRKQLAVR KQEQWGTARA LWLWAFSLQA KVWTAWLGFV
LERRRKKARL ERAMQAYQQQ LLQEGATRLL RFTAGTKAFR QQLQAQQQVQ AAHSLHCAVR
HCAELWKKKV LGPGKTSQPP APTTFSKRVT FKDSFLSGHA AEAGDATQET KKLRAPPSQG
VLGSLAGAAG EPCHLDLNAA RSSRKQPRRP SFLLERLGSQ RSPEWYSLGE QQLEKPPEEE
STALLGGSSL TRPFLPGVLP NVPGPKLPPT ASPGLELLPP SSIMPHAAGG TARVSAKPSI
PGPQPWGCPS LPRDLDPQLL PGDSISTRTE PVYGSEATGH TELEAELEGI QQQLQHYQTT
KQNLWSCQRQ ANSLRRWLEL SQEEPKSEDL HLEEQVKTEL EEVELQVQQL AKELEAQRQP
VGTCIARVRA LRRALC
```

>MmSFI1_2

```
MEKKIGSRSF RDGVVKKPCS PKTLPLKKSS AFSGIQREPS RSCHSIYYHA SQNWTRYRLQ
ELRIRCVARK FLYLWIRVTF GRVTPSRARI FHEQKILQKV FGEWREEWWV SQREWKLCVR
ADCHYRYYLY NLIFQNWKTF VHQQREMRKR FRIAEHHDTK QKMCQAWKSW LIYMVSRRTK
LHMKTTALEF RRQSVLCFWW SKWRWRLGQA HAEHALHAVA VKHRALSLQL QGWLRWQEQL
LISQRDRRKE ATAVQHYQHW QKQRSLKAWL KYLQICRVKR WQNEMAVQFH RATVLQIHFC
DWQWAWEWRQ SLSAHQALVV KLAGRMVLRR AFTHWKHYML LQAEEAAQRE AAAEHRQHYL
LYSCFRAFKD NVTQARLQQT RKKLAQQLRD TTLLHRFWNL WQSRIEQREE RVQTPSLHAA
LSHYRVTVLH KCVRVWLRYV HKRQWQQLLR ARADGHFQQR ALPAAFYTWY RGWLWHQQRR
ILHTKAVRFH RGTLEKQVFA LWRQKMSQHR ENCLAERMAI LQAEQQLLRR FWFVWHQQAA
VCQLERQQQA MAIAHHHSGL LRRAFCIWKE STQGFRIERM GRAQAAHFHS AQLLSRAWSM
WREVYQNRVR SVLREVAARE RQHNRQLLWW ALHLWRENTM ARLDGAKKTS QARVHYSRTL
CSKVLVQWRE VTSVQIYYRQ KEAAALREAR KALDRGRLQN WFQHWRFCSQ RAAQQRFQLG
QAAQHHHWQL LMEAMARWKA HHLGCIRKKF LQRQAAQLLA QRLSRACFCQ WRKQLAVRKQ
EQWGTARALW LWAFSLQAKV WTAWLGFVLE RRRKKARLER AMQAYQQQLL QEGATRLLRF
TAGTKAFRQQ LQAQQQVQAA HSLHCAVRHC AELWKKKVLG PGKTSQPPAP TTFSKRVTFK
DSFLSGHAAE AGDATQETKK LRAPPSQGVL GSLAGAAGEP CHLDLNAARS SRKQPRRPSF
LLERLGSQRS PEWYSLGEQQ LEKPPEEEST ALLGGSSLTR PFLPGVLPNV PGPKLPPTAS
PGLELLPPSS IMPHAAGGTA RVSAKPSIPG PQPWGCPSLP RDLDPQLLPG DSISTRTEPV
YGSEATGHTE LEAELEGIQQ QLQHYQTTKQ NLWSCQRQAN SLRRWLELSQ EEPKSEDLHL
```

EEQVKTELEE VELQVQQLAK ELEAQRQPVG TCIARVRALR RALC

>MmSFI1_3

```
MEKKIGSRSF RDGVVKKPCS PKTLPLKKSS AFSGIQREPS RSCHSIYYHA SQNWTRYRLQ
ELRIRCVARK FLYLWIRVTF GRVTPSRARI FHEQKILQKV FGEWREEWWV SQREWKLCVR
ADCHYRYYLY NLIFQNWKTF VHQQREMRKR FRIAEHHDTK QKMCQAWKSW LIYMVSRRTK
LHMKTTALEF RRQSVLWVTV LHKCVRVWLR YVHKRQWQQL LRARADGHFQ QRALPAAFYT
```

```
WYRGWLWHQQ RRILHTKAVR FHRGTLEKQV FALWRQKMSQ HRENCLAERM AILQAEQQLL
RRFWFVWHQQ AAVCQLERQQ QAMAIAHHHS GLLRRAFCIW KESTQGFRIE RMGRAQAAHF
HSAQLLSRAW SMWREVYQNR VRSVLREVAA RERQHNRQLL WWALHLWREN TMARLDGAKK
TSQARVHYSR TLCSKVLVQW REVTSVQIYY RQKEAAALRE ARKALDRGRL QNWFQHWRFC
SQRAAQQRFQ LGQAAQHHHW QLLMEAMARW KAHHLGCIRK KFLQRQAAQL LAQRLSRACF
CQWRKQLAVR KQEQWGTARA LWLWAFSLQA KVWTAWLGFV LERRRKKARL ERAMQAYQQQ
LLQEGATRLL RFTAGTKAFR QQLQAQQQVQ AAHSLHCAVR HCAELWKKKV LGPGKTSQPP
APTTFSKRVT FKDSFLSGHA AEAGDATQET KKLRAPPSQG VLGSLAGAAG EPCHLDL
```

>MmSFI1_4

```
MTAEVNGSTS GNHRSFRDGV VKKPCSPKTL PLKKSSAFSG IQREPSRSCH SIYYHASQNW
TRYRLQELRI RCVARKFLYL WIRVTFGRVT PSRARIFHEQ KILQKVFGEW REEWWVSQRE
```

>MmSFI1_5

```
MEKKIGSRSF RDGVVKKPCS PKTLPLKKSS AFSGIQREPS RSCHSIYYHA SQNWTRYRLQ
ELRIRCVARK FLYLWIRVTF GRVTPSRARI FHEQKILQKV FGEWREEWWV SQREWKLCVR
ADCHYRYYLY NLIFQNWKTF VHQQREMRKR FRIAEHHDTK QKMCQAWKSW LIYMVSRRTK
LHMKTTALEF RRQSVLCFWW SKWRWRLGQA HAEHALHAVA VKHRALSLQL QGWLRWQEQL
LISQRDRRKE ATAVQHYQHW QKQRSLKAWL KYLQICRVKR WQNEMAVQFH RATVLQIHFC
DWQWAWEWRQ SLSAHQALVV KLAGRMVLRR AFTHWKHYML LQAEEAAQRE AAAEHRQHYL
LVRAWSELEG IQQQLQHYQT TKQNLWSCQR QANSLRRWLE LSQEEPKSED LHLEEQVKTE
LEEVRPGQPR ASPWLSFLSA CLVPPSRPCP QVELQVQQLA KELEAQRQPV GTCIARVRAL
RRALC
```

>ScSFI1

```
MGKFGTTNKS TENLLRDKFV PETSPTNIPT DVLIKQGQIT DSTESLIHGG AERYIVNALK
```

```
PIELNKTEGF FEDPPFHLPS PPVDSTNLEY EDVTDLPKNG LRYDLNDISV EVIEDLYRQI
EAFLVHFKLS RSFLQIFKNY VNILIQEGIN PLRDEYFTIL EDELKGFFTF NSVIEEILEI
FLIHPRNKFI ALSLAEYTYA KNKIRRHFNH WKTVCELNEE ANRFANQAKL RVQEAVFYIW
SDKTLKYSQM ANDEAESFRN TWLLFRSFQQ WITLTQTLKE QSRLADQAFL NKMFRKILKA
QEHWKHLETV NTDNIKKIFL RTTFHIWKLR HKEINYHGLE RRIFERIKQK VINYEYNKSI
AEKVRSFSLQ RKYLNKWEKK NIENEDKLGA LYELENKFIK QKFFRKLNRS FQHSQQEAIA
KSKLNQTLLR CVFEKMWLKR FEDHLHLYSI VSLKEANLVK RIFHSWKKLL YIDLKASDYS
RTNLLKSSLR SWKLEVKLKI FEQKCKKSIQ ASAYRTWRKR IQYGKISSEH VKTAFCAKYL
GVWKRRMLQM NSMNDEASKF YEEGLVNECL AIWKERLIKT KELEDRYNFL CKTHAILTVK
RTLMHIDNVH LLYTKLAPSM DRVKLSKAFL KWRKATRFKV RHKLNDILHV YEKSKERELQ
SQLFNAWRNR FCFYTEECNI QAISKRNYQL EKMVLKKFRE RLLEIVKSEE LADEVREEFV
LVKTFYIWKT HLDEIFYMST LLEQSEANKQ FIITSKFLKM WSLRFLKIKR NDETVEVFRH
RWDRATVRGL LLLWKNRSDS SPKRRKDFNL KHELKTPIRS DSQNASTIPG SERIKQHRME
AMKSHYSRAR RAIPSPVKSS SVLDSTAKKQ INLESTTGLN GSPTRGKPLR YSPRRTTRNM
 PSKVDHIDFG RIPAVPFSLS ANSPKIDQDM DYIREHDKSP LSRKRQ
```

>CrSFI1

```
MTINNPLRLS SEVSPLRHVL DITARSQGKD LTAYETYRAL PSPTRPYSGY DYAVVAEIIR
AAEVTVDKAH KKGLGSGIVT LQSVLQAYEH VLPRHGVKAD EDTYYYRLLL KLSLDPALDW
WIKLNRETGT TGGRAAFFSN AGSDVTSMGR SHSFYRGDTS PTRVQPRTSG AGASDYSASH
SAAPTFRSSI PGNGAADGGS RVPGRNSPYR STASLSSMPG YQQQQETAAQ AAAREAAARA
NAAAQAARDA IEAARQAAAT AHLPSSRFMF NTPPMTPAMG PVPPPSARTE GGASGWGDGS
MGGAGSIVPY SVSEGGATGP GGASQYGPGS RYGPGSPHGT TAPSSVRGTY GISEHDPLSP
DASAYGAGSR GTRSGGGADG GSMGGAVGPG ATRASSGGGG GPGRVSAAQL AELDEPVLSE
EAEAYLDYNR LARAFRMWRK NTFARKASRF ERDHALKNWA VATSFWAVHL LRRCFARWRA
EGPSKTALAL QLWAGHSLSA CFWRWLSLTR YLRGKGTECD GHWRVRTLRR CLRMWRLYTR
SQQLKADNVE KALGFWVGRS LHICFRVWRQ FAANQLRKNE AVQKAVGHWA GKSLRACFEV
WREFTMLQQH KSVATARALR HWTGTTLRSC FALWRLLVED AHSNWDQATQ HLMRRMLLKW
RIAAKQLTRM SAAYHAIADA TARRVMSQAM RTLLSAHAFR HETANRLRAA VARMRNSLLS
AAFNKWSEAR QAGQMKRERM ALALKWWSSR GLARSWNTWR EAVAASQRAT KAARANAQRI
LRPLLRDTFY AWLDWHDVRL QKEVRAVAAR RTMQLRQMRL VVLGWRGLVL ERRHSLVSFQ
RIMRNEVQRI MSTAFQRWYR EHDNLRRAAE ARLTVKRNAT KRCLRHWRAL AQAQRAHREW
RLERLRLASE QRQRRLLVAW QVVAGLLAEH NRLVKSSLFK LHRRQAHSVL AAWRDRVLHS
VAKRMKLLAG LLYWEQQVKG RAWLAWRQRQ HGWRIKAAKG ELAGRHWRTR RLALCFAVFL
ALWQAYQGRL TRALLFRSRR NDIIKADVLD EWRRLVGWLQ WKNDKIRRAL AHHRSKLGGG
```

AFYGWYAVTR HSIATKRAVA AAVLHWVRRR RAAAFVWWRN WTAYRRDLRV RGAGYALVRA
ARAKEKVVLA FKANVIRKAR IAGAVAHRQR YMTRLALSGW LLRTLLAADF VRRLKAVGAQ
LNVQLVVDGL EAWREYTIYR RTKKGMLHAA LRYWRLSRQR AAMDAMRWYT TRRQLKRAVL
LRGRTGAAAR ALRAWRDEAE YRRRLRDRYA MLASGSRHAA LRRGLDMWRD YIANRASKKG
LREQALRHWL RGTAGRVFRS WRDYQRHLRH AEERAGDIVQ RWRKRDAAEA LAAFAEYAAH
RRRKRTADAM SRRSRARTAL QAWREYMKLR TAEGFWSRSR LNDAIGAWRH YANHRRRLKE
LWGRTRRGRL ASALSSWRDY ATYSARLKTV SNAAKQMAAL KALLERVLVR TAALAFYGWR
EVVADAKRWR AVQAVTQERM QQRPELAEVA MYAATRMRNW QLALAFYTWY DNARESRYLK
NRASQAILLY ANRLLYNAWS VWLAHTLRNR HLRKKLERCL NSTRAKTSRG VFDAWLAAAK
YLGRLRKVSD LVKDKLLRGT LVGAFTSWRR TAIQFRRLRG ILTRVMSRAL STAWEAWRDA
VAERHERMQS LAVFVGHWRN LHLSAAFNAW VAHVHKRAAA RRLCLRVLGR LLEWAWYGWR
EAVFEAGVSR GADVHEARLL ARSWRGWRWA TSEGRKAAAL TAALNQAAAL TAMGVWRRQW
LARAAYRRLY CRRALLGWHS RTQELRAMRE RLWYAARFLL NGCLLRSFSA WWQYTQAMSI
KRAVWVRKQR ALAEALRRGG ELVAARNAEL MEIAFRGWRM QTGLLREVTR RLRSIQGRTL
ATAFGQWREY AAVKRARLEK QAAVIRSRLL ARPFRTWRAA ARGAASELSA KSAAALMQRR
GVLLAATFRV WRELHGVVGS RREALRRALG VVLRAEAGVL KQAVWGAWRW AVARRGDLLD
LVSAVSAGRR RRVLGEVFDV WLQYTQAMRR GGIDPGSPYM SPRERGTERR LITRMAALAG
NDQALAPAAH AGGASVAATD ALLDGLYPST GMQHASERRQ ELSAFLDFAR RATSTGRAGR
NGRSAAQQAA PPVYERAESI SSDLSRASTG LSSGAGGADL TERAMFSVRA MPGMISPGKG
RGPASGLTPA HIQLATAGRA APAPLSARRP SGPQDFAGLS SPFDDRLDLR ELYGAADQAA
AAAPAPFGGF HTTAGYGARG AQLASGAASP MPLMPPPAAR GGLAMATPLT ASVPIAPRRI
DFTASRPGSG AGAGPATGGA YGYGGPAGRG PGTPRSQQFA QPAEPAYPDD SDGASTTSSD
FLPPSEARAR AQQARAAAVA TPRIALGVAA IHGTFIQANR PEHFHAVHLP IAAGGQWLVI
GGWGWSDHWG ETDVLGLLGA LQRHCIQRRW PAPAYVLLAP PTTVEHQQGL GFPFGIPVPF
DLVQTGVGVH HACSLYTNLA TAEHLTAVFR EMRPLPQRRL PQLVAPGVRL PVAAQVEPLP
HVPVHAPGQR LAALPAGATC IPADANAPAQ PLSPRDWEEL LGHVPGASSG LPPSAALAAL
RSAPPPLALS AVLAAAVVLQ RHYLTPHDTH TPPPQVLAPA DAGPVLPLGG SDDAEDTELA
RMLIQRQPGS DLLVFATMTW GPEQQLALDT VKAEFAREGL ILRAYDPSRP LILHTDWCED
GVSGVLGQLD DDAREYMVAC VSRTCNVHER RYGSYKGELL AAVWAIQTLR PYLHATPFTL
VTDHAPLEWL MSQPELTGQA ARWAMILQQY SFSRHHNACR AAAAPGAALG ACVTPAVRVF
AGHAHAGRRH HGTQPRPAPR LRLRTSALHG SVVGVSTFLR QHALIFAAET DAISVLEVPG
GLAPGLQACL RLGWRIRRYC NLDSSGTVTD SAYNRQAAPA SVIQHPCL

>CfSFI1

```
MKNLLTKKYI SSHWPNFHQK RRKLRMQRKV DSRSFRGGAV KKPYPPKISA NQKSSTLLGI
RSEIPSASHP VQCHASHGWT RKNRLRELRI RCVARKFLYL WIRMTFGRVF PSKARFYYEQ
RILRKVFEEW KEEWWVFQRE WKLCVRADCH YRYYLYNLMF QTWRTYVYQQ QEMRNKYIRA
ENHDAKQKMR QAWKSWLIYV VFRRTKLHMQ TTALEFRQQS ILWLWWRRWR QQLGQVCVGR
ALHASAVKHR ALSLQQQAWS QWWEQLLHVQ RERQKMVSAV KHHQHWQKWK SLKAWLKYVQ
VRRVKRQLNE MAERFHRVTV LQTHFCDWQW AWERKERLSA QQARVEGLAR RMTLRRAFTH
WKHYVLLCAE EAAQWKVAED HSRCSLLHFC FRALKENVTH AHLQQIRRNL AHQQRDITLL
HRFWNLWQSR IEQREEREQL PSLCAAWAHC RMTLLRKCIK SWLQYTQKRR YKQRLRSRAD
SHFQQRALPA AFNTWRRFWR WHQQESVLNA RAARFHRETL EKQVFAIWWQ KMFQHRENRL
AERMAILHSE RQLLQRSWSM WHQQAAARHR EWQRQAVACA HHRHRQLRKA FCIWRETARG
LRTEKMGLVL AAEFQAARLL HWAWSRWREC LALHAAEQRK LMRASLHHQQ ALLHGTLQTW
ATYQSQVRSI LQEVAARESQ HKRRLLRSVF RRWRENTLAR VNEAKQTSRA STHYRRTLCS
KILVQWQEAA SVQIYYRQQE DCAIKEAQKV LERGCLRTWF RHWRDQGQRA AQQRVQLRRA
AGHHGRRLLL QTMARWKAHH FGCVRKRLLQ RQGAQLLAQT RSRACFHQWK QQLANRRREQ
QGTARALWFW SFSLQAKVWA AWLGFVLERR RKKVRLERAG QAYHQQLLRE GATRLLRFAG
SMKAFRQQLH AQQQEQAAHS LHRAVHRCAM LWKQKVLGQG KERRPLPTMP RRRVTFDSPL
PTCVAAGAGD ATLETRRPGD PHVPWGALGS LSLAAGDPQL LELNATRLAR KQPRRPHFLL
EPVQSQRPQA PQECDRGLVW PAGPSLMRPF PPEAWMALVP SSPSPQPRAL LGRPGLKLPP
TLKTGPELLP PSSFMPCGVE APARASAQPT TPGLTLQASP SPASVPQSRL LLPGDFMGTG
RSSGTITAGH DDLEAELESI QQQLQDYQTM KQNLSSCQRQ ARSLRRWLEL SQEEPRPEDQ
 EAEQQVQEEL QEVELQIQQL ASELQAQRQP IRACIARVQA LRQALC
```

>SsFI1

```
SNSRSFRDGA VKKPYSPKIL SNKKCSAFPG MRSEIPKAHR PIQFLPSLAR PRRGRLRELR
LRCVARKFLY LWIRMTFGRV FPSKARFYYE QRILRKVFEE WKEEWWVSHR EWKLCVRADC
HYRYYLYNLT FQTWKAYVHQ QQEMRNKYMR AEDHDAKHKM RRAWKSWLIY LVFRKTKLQM
QTTALEFRQH RVLWVWWSEW RRQLGQVRLG HALQASAVKH RVQSLQLQAW SRWREQLLHV
WRERQKVVSA VQHHQCRQKW RFLKAWLEYL HVRRVKRQRN TMAERFHYVT VLQAHFCDWQ
WAWERRQSLY AHQARVKGLA RRMALRRAFT HWKHYVLLCA EAVIWRKGAE EHRRRCLLHF
CFRALKDNVT RAHLQRMRRN LAHQQHHVML LRRFWNVWQS RIEQREEREH LPSLCAAWDH
YRVTLLRQCF KSWLQYTRKR RSQQLLQARA DSHFQQRALP AAFQAWRRLW QWRQQEQVLS
ARAMCFHREM LEKQVFAVWC QKMSQRRDYR LAERMAILHA ERQLLRRHWS TWCQQAAACC
LEQLWQAVAR AHHRLERLRK AFCVWRERTQ GLRTERMGRS WAARFHSTRL LRWAWTRWRE
GLALRSAERQ QLMRADLHRQ HILLHRVLQT WRIHQSHVRS VLQEVAAREA WHRRQRLRAA
FRQWRENAVA QADKAKKTSQ AAHHYRRTLC FKVLMEWREA ASVQVYYRQQ EDDAVREAQV
```

```
VLSRGHLRTW FWRWWDHSQR AAWQRVQLER AAQHYRRRLL LQAVAQWKAH HLGCIRKRLL
QRQAARLLAQ TLCQACFHQW RRQLEDRRRE RQETARALWF WAFSLQAKVW AAWLGFVLER
RRKKARQERA VQAYHQQLLR EGVTRLLRFA AGTKAFRQRL HAQQQVQAAH SLQRTVRRCA
TLWKQKALGP GRGPQPPTSA VLSRRVTFED PPLSSVAAGT GDASLETKRP PAPRGLWGAL
GSPVSAAGEP QLLELNAARW ARKQPRCPDF LLEPEASQTP LGGQGPEALW GRDPALVTPF
LAKAKAWAAL DPSSPLPSAP GLKPPPSVGP GPELLPPSSF TPCGARDGWL LSSGPAGPGP
DTVGVYLGPG IRDPWVQPGP SQEEGQDQAQ FSAATGHTTL VAELEGIQQQ LQDYQTMKQN
LRSCQRQASS LRRWLELSRE EPRPEDQEAE RQVQQELQEL EMQIQQLSSE LQAQRQPIRA
CIARVQALRQ TLC
```

>DrSFI1

```
DRMKHACALK EKESKAHHHY CLQLLKQTLH GWIRHAENRQ AKNKKIAVAQ DVRRLSVLER
FWWQWFEVLQ SRRVEKDRGQ RADKLAQHGS QRLVFSHWRH YVSICSQKAK REKSAMHHRQ
RYLLHLGLKG FALNVTQSKT YRLNKNISVQ HRHQNLLAMC WNVWQLRLDR EENRRLQPQM
SVAQNHHKFS VLRFNLHHWK KRLAEHKKCQ ELELRADACF AQRIFPQCVN TWMEFTAQRT
EKREQKERAE QQYRQQTFSW VFYTWWRNLE ARRDRRLAER TAVLHAEHVC YVRVWSKWYS
GAVQRREERI KQAVADSLYK HTLQQKNLNH WKNRFINIQT SQKRFKQAEG HHEQRCLKRA
MTDWHQYVDH RRLKKEKIAE MEKHYHNKLL KHALDAWKSY HLQTQAISQL VEYSYQEHEQ
HLVRRMFCLW RINVLQLVEE REKENRATCL SQKHLVSQVF LAWQQRTVSA HLHHHKIYQV
MKKRSFELHK LRTCQHFFIC WKTQLQSRRR EAEQTETALW HWSLNLQAKV FCMWRIWIAE
RQRKQKRLTE AAQFYRDELL REGVTHILTH TAHMNAFSTN IAQHSYEQSC RQLQEVVRRC
ALRWKQRALC KPVKDKSTPS KDSQAKKSVS FFLPGDPMPT LCPGRIQSPE QRQKDSVIHQ
MWSVRASRPQ PRRPGDLLES PAKHLSRDLK ISRGQRVFVS ETAFPSVPST FPSPLPSVPL
THNPLNSPAM VLQPKTLLLK PDSPGCIATT SQDILLPPST FTLSTAHSKQ KYLRDGGPLL
HSSHFTSRFF PTQALGLHER IPEDEEEDVD VEQTENLTKE LLDIRLEMQR YQQDRKQLQT
WRKLQKVLGN WLETTGTEGE TDERDSILKE LTELESRISS LSMRIKKQKP SMIRHAARVN
 TIQSQLLPSE GTTNLPTEYG I
```

>RnSFI1

```
MEKKIGFRSF RDGVVKKPCS PKALPLKKSS AFSGIQRELT RSCHSSYYQS SQSWTPQRLR
ELRVRCVARK FLYLWMRVTF GRVTPSRARF FHEQKILRKV FGEWREEWWV SQREWKLCVR
ADCHYRYYLY NLIFQNWKSF VYQQREMRKR FQRAEHHDMK QKMRHAWKSW LIYMVARRTK
FRMKSTALEF RRRSVLCFWW NKWRWRLGRV HAEQALHAAA VKHRALSLQL QAWLRWQEQL
LISQRERRKV ATAVRHHQRW QKQRTLKAWL KYLRICRVKR WQNEMAVQFH RITVLQIHFC
DWQWAWEWRQ SLSAHQALVA KLARKMVLRR AFAHWKHYML LQEEEAAQRE AAAEHHQHCL
LHSCFRAFKD NVTRARLKRM RKSLAHQLRD TTLLHRFWNL WQSRVEQREE RVQPPSLLAA
QSHYRMTVLR KCVRVWLQYV HKRRCLQSLQ ARALGHFQQR ALPAAFYTWY RVWRWHQQRR
TLHTRAVRFH REALEKQVFA LWRQKVSQHR EIRLAERMAM LQAEQQLLRR FWFVWHQRAA
ACHQERQRQA MAIAHHHSGL LRRAFCIWRE STQGFRMERM GRAQASHFHS VRLLCWAWSK
WREGLALRME EQQKLKRAAL HSQRTLLHRA LQKWLVYQDR MRSVLQEVAA RERQHNRQLL
WWVLHRWREN TMARLDGAKK TSQARVHYNR TLCSKVLVQW REVTSVKIYY RQKEAAALRE
ARKALDRGRL RTWLQRWQVC RQREAQQTFR LQQAAQHHRR QLLAGAMARW KAYHLCCVRK
KFLQRQAAQL LAQRLGRACF CQWRKQLAVR KQEQWGTARA LWFWAFSLQA KVWAAWLGFV
LEKRRKKARL EQAMQAYHQQ LLQEGATRLL RFTAGMKALR QQLQAQQQVQ AAHSLHCAVR
HCAELWKKKV LGPDKTSQPP APIAFSRRVT FKDFSLSGLA AEAGDAPLSQ GVLGSLAVAA
```

```
GEPFIPELNE ARPSRKQPRR PSFLLERVQS QRSPERCTLG EQQLQKTPEK GQSMAPPGGS
SLTRPFLPVV RPNVSGPKLP PTASPGLELL PPSSFMPHGV RDAARVSTKP SISGPQPWGC
PSLTRDPDPH LLLPGDSTST RTGPGYGSET TGHTELEAEL EGIQQQLQHY QTTKQNLWSC
QRQANSLRRW LELSQEEPSC EDLHLEEQVK TELEEVELQI QQLAKELEAQ RQPVGTCIAR
 VRALRRALC
```

>PtSFI1

```
MKNLLTEKCI SSHNFHQKVI KQRMEKVDSR YFKDGAVKKP YSAKTLSNKK SSASFGIRRE
LPSTSHLVQY RGTHTCTRQG RLRELRIRFY YEQRLLQKVF EEWKEEWWVF QHEWKLCVRA
DCHYRYYLYN LMFQTWKTYV RQQQEMRNKY IRAEVHDAKQ KMRQAWKSWL IYVVVRRTKL
QMQTTALEFR QRIILRVWWS TWRQRLGQVR VSRALHASAL KHRALSLQVQ AWSQWREQLL
YVQKEKQKVV SAVKHHQHWQ KRRFLKAWLE YLQVRRVKRQ QDEMAERFHH VTVLQIHFCD
WQQAWERRES LYAHHAQVEK LARKMALRRA FTHWKHYMLL CAEEAAQFEM AEEHHRHSQL
LLHRFWNLWR SQIEQKKERE LLPLLHAAWD HYRIALLCKC IKLWVQYTQK RRYKQLLQAR
ADGHFQQRAL PAAFHTWNRL WRWCHQENVL SARATRFHRE TLEKQVFSLW RQKMFQHREN
HLAERMAILH AERQLLYRSW FMWHQQAAAR HQEQEWQTVA CAHHRHGRLK KAFCLWRESA
QGLRTERTGR VRAAEFHMAQ LLRWAWSQWR ECLALRGAER QKLMRADLHH QHSVLHRALQ
AWVTYQGRVR SILREVAARE SQHNRQLLRG ALRRWKENTM ARVDEAKKTF QASTHYRRTI
CSKVLVQWWE AVSVQIYYRQ QEDCAIWEAQ KVLDRGCLRT WFQRWWDCSR RSAQQRLQLE
RAVQHHRRQL LLEGLARWKM HHLQCVRKRL LHRQSTQLLA QRLSRTCFRQ WRQQLAARRQ
EQRATVRALW FWAFSLQAKV WATWLAFVLE RRRKKARLQR ALQAYQGQLL QEGATRLLRF
AASMKASRQQ LQAQQQVQAA HSLHRAVHRC ATLWKQKVLG RGGKPQPLAA IAPSRKVTFE
GPLLNRIAAG AGDATLETKR PQASRPPGAL GRLAAEEPHA LELNTAHSAR KQPRRPHFLL
EPAQSQRPQK PQEHGLGMAQ PAAPSLTRPF LAEAPTALVP HSPLPGALSS APGPKQPPTA
STGLELLLLP PSSFMPCGAA APARMSAQRA TPRDKPPVPS SLASVPDPHL LLPGDFSATR
AGPGLSTAGS LDLEAELEEI QQQLLHYQTT KQNLWSCRRQ ASSLRRWLEL NREEPGPEDQ
 EVEQQVQKEL EQVEMQIQLL AEELQAQRQP IGACVARIQA LRQALC
```

>PtSFI1_1

```
MKNLLTEKCI SSHNFHQKVI KQRMEKVDSR FYYEQRLLQK VFEEWKEEWW VFQHEWKLCV
RADCHYRYYL YNLMFQTWKT YVRQQQEMRN KYIRAEVHDA KQKMRQAWKS WLIYVVVRRT
KLQMQTTALE FRQRIILRVW WSTWRQRLGQ VRVSRALHAS ALKHRALSLQ VQAWSQWREQ
LLYVQKEKQK VVSAVKHHQH WQKRRFLKAW LEYLQVRRVK RQQDEMAERF HHVTVLQIHF
CDWQQAWERR ESLYAHHAQV EKLARKMALR RAFTHWKHYM LLCAEEAAQF EMAEEHHRHS
QLYFCFRALK DNVTHAHLQQ IRRNLAHQQH GVTLLHRFWN LWRSQIEQKK ERELLPLLHA
AWDHYRIALL CKCIKLWVQY TQKRRYKQLL QARADGHFQQ RALPAAFHTW NRLWRWCHQE
NVLSARATRF HRETLEKQVF SLWRQKMFQH RENHLAERMA ILHAERQLLY RSWFMWHQQA
AARHQEQEWQ TVACAHHRHG RLKKAFCLWR ESAQGLRTER TGRVRAAEFH MAQLLRWAWS
```

QWRECLALRG AERQKLMRAD LHHQHSVLHR ALQAWVTYQG RVRSILREVA ARESQHNRQL
LRGALRRWKE NTMARVDEAK KTFQVLVQWW EAVSVQIYYR QQEDCAIWEA QKVLDRGCLR
TWFQRWWDCS RRSAQQRLQL ERAVQHHRRQ LLLEGLARWK MHHLQCVRKR LLHRQSTQLL
AQRLSRTCFR QWRQQLAARR QEQRATVRAL WFWAFSLQAK VWATWLAFVL ERRRKKARLQ
RALQAYQGQL LQEGATRLLR FAASMKASRQ QLQAQQQVQA AHSLHRAVHR CATLWKQKVL
GRGGKPQPLA AIAPSRKVTF EGPLLNRIAA GAGDATLETK RPQASRPPGA LGRLAAEEPH
ALELNTAHSA RKQPRRPHFL LEPAQSQRPQ KPQEHGLGMA QPAAPSLTRP FLAEAPTALV
PHSPLPGALS SAPGPKQPPT ASTGLELLLL PPSSFMPCGA AAPARMSAQR ATPRDKPPVP
SSLASVPDPH LLLPGDFSAT RAGPGLSTAG SLDLEAELEE IQQQLLHYQT TKQNLWSCRR
QASSLRRWLE LNREEPGPED QEVEQQVQKE LEQVEMQIQL LAEELQAQRQ PIGACVARIQ
ALRQALC

>PtSFI1_3

MKNLLTEKCI SSHNFHQKVI KQRMEKVDSR FYYEQRLLQK VFEEWKEEWW VFQHEWKLCV
RADCHYRYYL YNLMFQTWKT YVRQQQEMRN KYIRAEVHDA KQKMRQAWKS WLIYVVVRRT
KLQMQTTALE FRQRIILRVW WSTWRQRLGQ VRVSRALHAS ALKHRALSLQ VQAWSQWREQ
LLYVQKEKQK VVSAVKHHQH WQKRRFLKAW LEYLQVRRVK RQQDEMAERF HHVTVLQIHF
CDWQQAWERR ESLYAHHAQV EKLARKMALR RAFTHWKHYM LLCAEEAAQF EMAEEHHRHS
QLYFCFRALK DNVTHAHLQQ IRRNLAHQQH GVTLLHRFWN LWRSQIEQKK ERELLPLLHA
AWDHYRIALL CKCIKLWVQY TQKRRYKQLL QARADGHFQQ RALPAAFHTW NRLWRWCHQE
NVLSARATRF HRETLEKQVF SLWRQKMFQH RENHLAERMA ILHAERQLLY RSWFMWHQQA
AARHQEQEWQ TVACAHHRHG RLKKAFCLWR ESAQGLRTER TGRVRAAEFH MAQLLRWAWS
QWRECLALRG AERQKLMRAD LHHQHSVLHR ALQAWVTYQG RVRSILREVA ARESQHNRQL
LRGALRRWKE NTMARVDEAK KTFQASTHYR RTICSKVLVQ WWEAVSVQIY YRQQEDCAIW
EAQKVLDRGC LRTWFQRWWD CSRRSAQQRL QLERAVQHHR RQLLLEGLAR WKMHHLQCVR
KRLLHRQSTQ LLAQRLSRTC FRQWRQQLAA RRQEQRATVR ALWFWAFSLQ AKVWATWLAF
VLERRRKKAR LQRALQAYQG QLLQEGATRL LRFAASMKAS RQQLQAQQQV QAAHSLHRAV
HRCATLWKQK VLGRGGKPQP LAAIAPSRKV TFEGPLLNRI AAGAGDATLE TKRPQASRPP
GALGRLAAEE PHALELNTAH SARKQPRRPH FLLEPAQSQR PQKPQEHGLG MAQPAAPSLT
RPFLAEAPTA LVPHSPLPGA LSSAPGPKQP PTASTGLELL LLPPSSFMPC GAAAPARMSA
QRATPRDKPP VPSSLASVPD PHLLLPGDFS ATRAGPGLST AGSLDLEAEL EEIQQQLLHY
QTTKQNLWSC RRQASSLRRW LELNREEPGP EDQEVEQQVQ KELEQVEMQI QLLAEELQAQ
RQPIGACVAR IQALRQALC

>PtSFI1_8

MKNLLTEKCI SSHNFHQKVI KQRMEKVDSR YFKDGAVKKP YSAKTLSNKK SSASFGIRRE
LPSTSHLVQY RGTHTCTRQG RLRELRIRCV ARKFLYLWIR MTFGRVFPSK ARFYYEQRLL
QKVFEEWKEE WWVFQHEWKL CVRADCHYRY YLYNLMFQTW KTYVRQQQEM RNKYIRAEVH
DAKQKMRQAW KSWLIYVVVR RTKLQMQTTA LEFRQRIILR VWWSTWRQRL GQVRVSRALH
ASALKHRALS LQVQAWSQWR EQLLYVQKEK QKVVSAVKHH QHWQKRRFLK AWLEYLQVRR
VKRQQDEMAE RFHHVTVLQI HFCDWQQAWE RRESLYAHHA QVEKLARKMA LRRAFTHWKH

YMLLCAEEAA QFEMAEEHHR HSQLYFCFRA LKDNVTHAHL QQIRRNLAHQ QHGVTLLHRF
WNLWRSQIEQ KKERELLPLL HAAWDHYRIA LLCKCIKLWV QYTQKRRYKQ LLQARADGHF
QQRALPAAFH TWNRLWRWCH QENVLSARAT RFHRETLEKQ VFSLWRQKMF QHRENHLAER
MAILHAERQL LYRSWFMWHQ QAAARHQEQE WQTVACAHHR HGRLKKAFCL WRESAQGLRT
ERTGRVRAAE FHMAQLLRWA WSQWRECLAL RGAERQKLMR ADLHHQHSVL HRALQAWVTY
QGRVRSILRE VAARESQHNR QLLRGALRRW KENTMARVDE AKKTFQASTH YRRTICSKVL
VQWWEAVSVQ IYYRQQEDCA IWEAQKVLDR GCLRTWFQRW WDCSRRSAQQ RLQLERAVQH
HRRQLLLEGL ARWKMHHLQC VRKRLLHRQS TQLLAQRLSR TCFRQWRQQL AARRQEQRAT
VRALWFWAFS LQAKVWATWL AFVLERRRKK ARLQRALQAY QGQLLQEGAT RLLRFAASMK
ASRQQLQAQQ QVQAAHSLHR AVHRCATLWK QKVLGRGGKP QPLAAIAPSR KVTFEGPLLN
RIAAGAGDAT LETKRPQASR PPGALGRLAA EEPHALELNT AHSARKQPRR PHFLLEPAQS
QRPQKPQEHG LGMAQPAAPS LTRPFLAEAP TALVPHSPLP GALSSAPGPK QPPTASTGLE
LLLLPPSSFM PCGAAAPARM SAQRATPRDK PPVPSSLASV PDPHLLLPGD FSATRAGPGL
STAGSLDLEA ELEEIQQQLL HYQTTKQNLW SCRRQASSLR RWLELNREEP GPEDQEVEQQ
VQKELEQVEM QIQLLAEELQ AQRQPIGACV ARIQALRQAL C

>PtSFI110

MKNLLTEKCI SSHNFHQKVI KQRMEKVDSR YFKDGAVKKP YSAKTLSNKK SSASFGIRRE
LPSTSHLVQY RGTHTCTRQG RLRELRIRCV ARKFLYLWIR MTFGRVFPSK ARFYYEQRLL
QKVFEEWKEE WWVFQHEWKL CVRADCHYRY YLYNLMFQTW KTYVRQQQEM RNKYIRAEVH
DAKQKMRQAW KSWLIYVVVR RTKLQMQTTA LEFRQRIILR VWWSTWRQRL GQVRVSRALH
ASALKHRALS LQVQAWSQWR EQLLYVQKEK QKVVSAVKHH QHWQKRRFLK AWLEYLQVRR
VKRQQDEMAE RFHHVTVLQI HFCDWQQAWE RRESLYAHHA QVEKLARKMA LRRAFTHWKH
YMLLCAEEAA QFEMAEEHHR HSQLLLHRFW NLWRSQIEQK KERELLPLLH AAWDHYRIAL
LCKCIKLWVQ YTQKRRYKQL LQARADGHFQ QRALPAAFHT WNRLWRWCHQ ENVLSARATR
FHRETLEKQV FSLWRQKMFQ HRENHLAERM AILHAERQLL YRSWFMWHQQ AAARHQEQEW
QTVACAHHRH GRLKKAFCLW RESAQGLRTE RTGRVRAAEF HMAQLLRWAW SQWRECLALR
GAERQKLMRA DLHHQHSVLH RALQAWVTYQ GRVRSILREV AARESQHNRQ LLRGALRRWK
ENTMARVDEA KKTFQASTHY RRTICSKVLV QWWEAVSVQI YYRQQEDCAI WEAQKVLDRG
CLRTWFQRWW DCSRRSAQQR LQLERAVQHH RRQLLLEGLA RWKMHHLQCV RKRLLHRQST
QLLAQRLSRT CFRQWRQQLA ARRQEQRATV RALWFWAFSL QAKVWATWLA FVLERRRKKA
RLQRALQAYQ GQLLQEGATR LLRFAASMKA SRQQLQAQQQ VQAAHSLHRA VHRCATLWKQ
KVLGRGGKPQ PLAAIAPSRK VTFEGPLLNR IAAGAGDATL ETKRPQASRP PGALGRLAAE
EPHALELNTA HSARKQPRRP HFLLEPAQSQ RPQKPQEHGL GMAQPAAPSL TRPFLAEAPT
ALVPHSPLPG ALSSAPGPKQ PPTASTGLEL LLLPPSSFMP CGAAAPARMS AQRATPRDKP
PVPSSLASVP DPHLLLPGDF SATRAGPGLS TAGSLDLEAE LEEIQQQLLH YQTTKQNLWS
CRRQASSLRR WLELNREEPG PEDQEVEQQV QKELEQVEMQ IQLLAEELQA QRQPIGACVA
RIQALRQALC

SEQUENCE ALIGNMENT RESULTS FOR HOMO SAPIENS, MUS MUSCULUS AND PAN TROGLODYTES

```
HsSFI1_1        MKNLLTEKCIS-SHNFHQKVIKQRMEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIR
59
HsSFI1_5        MKNLLTEKCIS-SHNFHQKVIKQRMEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIR
59
HsSFI1_9        MKNLLTEKCIS-SHNFHQKVIKQRMEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIR
59
HsSFI1_3        MKNLLTEKCIS-SHNFHQKVIKQRMEKK--------------------------------
27
HsSFI1_4        MKNLLTEKCIS-SHNFHQKVIKQRMEKK--------------------------------
27
HsSFI1_10       MKNLLTEKCIS-SHNFHQKVIKQRMEKK--------------------------------
27
PtSFI1_1        MKNLLTEKCIS-SHNFHQKVIKQRMEK---------------------------------
26
PtSFI1_3        MKNLLTEKCIS-SHNFHQKVIKQRMEK---------------------------------
26
PtSFI1_8        MKNLLTEKCIS-SHNFHQKVIKQRMEKVDSR-YFKDGAVKKPYSAKTLSNKKSSASFGIR
58
PtSFI1          MKNLLTEKCIS-SHNFHQKVIKQRMEKVDSR-YFKDGAVKKPYSAKTLSNKKSSASFGIR
58
HsSFI1_2        MKNLLTEKCIS-SHNFHQKVIKQRMEKKVDSRYFKDGAVKKPYSAKTLSNKKSSASFGIR
59
PtSFI110        MKNLLTEKCIS-SHNFHQKVIKQRMEKVDSR-YFKDGAVKKPYSAKTLSNKKSSASFGIR
58
MmSFI1_4        MTAEVNGSTSGNHRSFRDGVVKKPCSPK--------------------------------
28
MmSFI1_5        MEKKIG------SRSFRDGVVKKPCSPK--------------------------------
22
MmSFI1_2        MEKKIG------SRSFRDGVVKKPCSPK--------------------------------
22
MmSFI1_1        MEKKIG------SRSFRDGVVKKPCSPK--------------------------------
22
MmSFI1_3        MEKKIG------SRSFRDGVVKKPCSPK--------------------------------
22
                *   :        :.*:: *:*:  .

HsSFI1_1        RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIRCVARKF
95
HsSFI1_5        RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIR------
89
HsSFI1_9        RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIR------
89
HsSFI1_3        --------------------------------------------------VDSR------
31
HsSFI1_4        --------------------------------------------------VDSR------
31
```

```
HsSFI1_10       --------------------------------------------------VDSR------
31
PtSFI1_1        --------------------------------------------------VDSR------
30
PtSFI1_3        --------------------------------------------------VDSR------
30
PtSFI1_8        RELPSTSHLVQYRGTHTCTRQGRLRELRIRCVARKFLYLWIRMTFGRVFPSKAR------
112
PtSFI1          RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIR------
88
HsSFI1_2        RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIRCVARKF
95
PtSFI110        RELPSTSHLVQYRG------------------------THTCTRQGRLRELRIRCVARKF
94
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        ------------------------------------------------------------
MmSFI1_1        ------------------------------------------------------------
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        LYLWIRMTFGRVFPSKARFYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
155
HsSFI1_5        ------------------FYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
131
HsSFI1_9        ------------------FYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
131
HsSFI1_3        ------------------FYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
73
HsSFI1_4        ------------------FYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYR-----
68
HsSFI1_10       ------------------FYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
73
PtSFI1_1        ------------------FYYEQRLLQKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
72
PtSFI1_3        ------------------FYYEQRLLQKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
72
PtSFI1_8        ------------------FYYEQRLLQKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
154
PtSFI1          ------------------FYYEQRLLQKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
130
HsSFI1_2        LYLWIRMTFGRVFPSKARFYYEQRLLRKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
155
PtSFI110        LYLWIRMTFGRVFPSKARFYYEQRLLQKVFEEWKEEWWVFQHEWKLCVRADCHYRYYLYN
154
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        ------------------------------------------------------------
MmSFI1_1        ------------------------------------------------------------
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
215
HsSFI1_5        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
191
```

```
HsSFI1_9        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
191
HsSFI1_3        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
133
HsSFI1_4        ------------------------------------------------------------
HsSFI1_10       LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
133
PtSFI1_1        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
132
PtSFI1_3        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
132
PtSFI1_8        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
214
PtSFI1          LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
190
HsSFI1_2        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
215
PtSFI110        LMFQTWKTYVRQQQEMRNKYIRAEVHDAKQKMRQAWKSWLIYVVVRRTKLQMQTTALEFR
214
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        ------------------------------------------------------------
MmSFI1_1        ------------------------------------------------------------
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
275
HsSFI1_5        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
251
HsSFI1_9        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
251
HsSFI1_3        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
193
HsSFI1_4        ------VWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
122
HsSFI1_10       QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
193
PtSFI1_1        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
192
PtSFI1_3        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
192
PtSFI1_8        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
274
PtSFI1          QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
250
HsSFI1_2        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
275
PtSFI110        QRIILRVWWSTWRQRLGQVRVSRALHASALKHRALSLQVQAWSQWREQLLYVQKEKQKVV
274
MmSFI1_4        ----------TLPLKKSSAFSGIQREPSRSCHSIYYHASQNWTRYRLQELRIR-------
71
MmSFI1_5        ----------TLPLKKSSAFSGIQREPSRSCHSIYYHASQNWTRYRLQELRIR-------
65
MmSFI1_2        ----------TLPLKKSSAFSGIQREPSRSCHSIYYHASQNWTRYRLQELRIR-------
65
```

```
MmSFI1_1        ----------TLPLKKSSAFSGIQREPSRSCHSIYYHASQNWTRYRLQELRIR-------
65
MmSFI1_3        ----------TLPLKKSSAFSGIQREPSRSCHSIYYHASQNWTRYRLQELRIR-------
65
                          *   : ...  .   ..*   *       * *:::* * * ::

HsSFI1_1        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
335
HsSFI1_5        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
311
HsSFI1_9        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
311
HsSFI1_3        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
253
HsSFI1_4        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
182
HsSFI1_10       SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
253
PtSFI1_1        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQDEMAERFHHVTVLQIHFCDWQQAWERRES
252
PtSFI1_3        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQDEMAERFHHVTVLQIHFCDWQQAWERRES
252
PtSFI1_8        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQDEMAERFHHVTVLQIHFCDWQQAWERRES
334
PtSFI1          SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQDEMAERFHHVTVLQIHFCDWQQAWERRES
310
HsSFI1_2        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQNEMAERFHHVTVLQIYFCDWQQAWERRES
335
PtSFI110        SAVKHHQHWQKRRFLKAWLEYLQVRRVKRQQDEMAERFHHVTVLQIHFCDWQQAWERRES
334
MmSFI1_4        --------CVARKFLYLWIR-VTFGRVTPSR---ARIFHEQKILQKVFGEWREEWWVSQR
119
MmSFI1_5        --------CVARKFLYLWIR-VTFGRVTPSR---ARIFHEQKILQKVFGEWREEWWVSQR
113
MmSFI1_2        --------CVARKFLYLWIR-VTFGRVTPSR---ARIFHEQKILQKVFGEWREEWWVSQR
113
MmSFI1_1        --------CVARKFLYLWIR-VTFGRVTPSR---ARIFHEQKILQKVFGEWREEWWVSQR
113
MmSFI1_3        --------CVARKFLYLWIR-VTFGRVTPSR---ARIFHEQKILQKVFGEWREEWWVSQR
113
                           *:**  *:. : . **. .:   *. **. .:**  * :*:: *   :

HsSFI1_1        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
395
HsSFI1_5        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
371
HsSFI1_9        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQL----------
361
HsSFI1_3        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
313
HsSFI1_4        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
242
HsSFI1_10       LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
313
PtSFI1_1        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
312
```

```
PtSFI1_3        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
312
PtSFI1_8        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQLYFCFRALKDN
394
PtSFI1          LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQL----------
360
HsSFI1_2        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQL----------
385
PtSFI110        LYAHHAQVEKLARKMALRRAFTHWKHYMLLCAEEAAQFEMAEEHHRHSQL----------
384
MmSFI1_4        E-----------------------------------------------------------
120
MmSFI1_5        EWKLCVRADCHYRYYLYNLIFQNWKTFVHQQREMRKRFRIAEHHDTKQKMCQAWKS----
169
MmSFI1_2        EWKLCVRADCHYRYYLYNLIFQNWKTFVHQQREMRKRFRIAEHHDTKQKMCQAWKS----
169
MmSFI1_1        EWKLCVRADCHYRYYLYNLIFQNWKTFVHQQREMRKRFRIAEHHDTKQKMCQAWKS----
169
MmSFI1_3        EWKLCVRADCHYRYYLYNLIFQNWKTFVHQQREMRKRFRIAEHHDTKQKMCQAWKS----
169


HsSFI1_1        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
455
HsSFI1_5        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
431
HsSFI1_9        ---------------------LLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
400
HsSFI1_3        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
373
HsSFI1_4        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
302
HsSFI1_10       VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
373
PtSFI1_1        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
372
PtSFI1_3        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
372
PtSFI1_8        VTHAHLQQIRRNLAHQQHGVTLLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
454
PtSFI1          ---------------------LLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
399
HsSFI1_2        ---------------------LLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
424
PtSFI110        ---------------------LLHRFWNLWRSQIEQKKERELLPLLHAAWDHYRIALLCK
423
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        --------------------------WLIYMVSRRTKLHMKTTALEFRRQ-----SVLCF
198
MmSFI1_2        --------------------------WLIYMVSRRTKLHMKTTALEFRRQ-----SVLCF
198
MmSFI1_1        --------------------------WLIYMVSRRTKLHMKTTALEFRRQ-----SVLCF
198
MmSFI1_3        --------------------------WLIYMVSRRTKLHMKTTALEFRRQSVLWVTVLHK
203
```

```
HsSFI1_1        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
515
HsSFI1_5        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
491
HsSFI1_9        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
460
HsSFI1_3        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
433
HsSFI1_4        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
362
HsSFI1_10       CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
433
PtSFI1_1        CIKLWVQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWCHQENVLSARATRFHR
432
PtSFI1_3        CIKLWVQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWCHQENVLSARATRFHR
432
PtSFI1_8        CIKLWVQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWCHQENVLSARATRFHR
514
PtSFI1          CIKLWVQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWCHQENVLSARATRFHR
459
HsSFI1_2        CIELWLQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWRHQENVLSARATRFHR
484
PtSFI110        CIKLWVQYTQKRRYKQLLQARADGHFQQRALPAAFHTWNRLWRWCHQENVLSARATRFHR
483
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        WWSKWRWRLGQAHAEHALHAVAVKHRALSLQLQGWLRWQEQLLISQRDRRKEATAVQHYQ
258
MmSFI1_2        WWSKWRWRLGQAHAEHALHAVAVKHRALSLQLQGWLRWQEQLLISQRDRRKEATAVQHYQ
258
MmSFI1_1        WWSKWRWRLGQAHAEHALHAVAVKHRALSLQLQGWLRWQEQLLISQRDRRKEATAVQHYQ
258
MmSFI1_3        CVRVWLRYVHKRQWQQLLRARADGHFQQRALPAAFYTWYRGWLWHQQRRILHTKAVRFHR
263


HsSFI1_1        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
575
HsSFI1_5        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
551
HsSFI1_9        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
520
HsSFI1_3        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
493
HsSFI1_4        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
422
HsSFI1_10       ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
493
PtSFI1_1        ETLEKQVFSLWRQKMFQHRENHLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
492
PtSFI1_3        ETLEKQVFSLWRQKMFQHRENHLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
492
PtSFI1_8        ETLEKQVFSLWRQKMFQHRENHLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
574
PtSFI1          ETLEKQVFSLWRQKMFQHRENHLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
519
HsSFI1_2        ETLEKQVFSLWRQKMFQHRENRLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
544
```

```
PtSFI110        ETLEKQVFSLWRQKMFQHRENHLAERMAILHAERQLLYRSWFMWHQQAAARHQEQEWQTV
543
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        HWQKQRSLKAWLKYLQICRVKRWQNEMAVQFHRATVLQIHFCDWQWAWEWRQSLSAHQAL
318
MmSFI1_2        HWQKQRSLKAWLKYLQICRVKRWQNEMAVQFHRATVLQIHFCDWQWAWEWRQSLSAHQAL
318
MmSFI1_1        HWQKQRSLKAWLKYLQICRVKRWQNEMAVQFHRATVLQIHFCDWQWAWEWRQSLSAHQAL
318
MmSFI1_3        GTLEKQVFALWRQKMSQHRENCLAERMAILQAEQQLLRRFWFVWHQQAAVCQLERQQQAM
323


HsSFI1_1        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
635
HsSFI1_5        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
611
HsSFI1_9        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
580
HsSFI1_3        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
553
HsSFI1_4        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
482
HsSFI1_10       ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
553
PtSFI1_1        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
552
PtSFI1_3        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
552
PtSFI1_8        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
634
PtSFI1          ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
579
HsSFI1_2        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
604
PtSFI110        ACAHHRHGRLKKAFCLWRESAQGLRTERTGRVRAAEFHMAQLLRWAWSQWRECLALRGAE
603
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        VVKLAGRMVLRRAFTHWKHYMLLQAEEAAQREAAAEHRQHYLLVRAWSELEGIQQQLQHY
378
MmSFI1_2        VVKLAGRMVLRRAFTHWKHYMLLQAEEAAQREAAAEHRQHYLLYSCFRAFKDNVTQARLQ
378
MmSFI1_1        VVKLAGRMVLRRAFTHWKHYMLLQAEEAAQREAAAEHRQHYLLYSCFRAFKDNVTQARLQ
378
MmSFI1_3        AIAHHHSGLLRRAFCIWKESTQGFRIERMGRAQAAHFHSAQLLSRAWSMWREVYQNRVRS
383


HsSFI1_1        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
695
HsSFI1_5        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARES------------------
653
HsSFI1_9        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
640
HsSFI1_3        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
613
```

```
HsSFI1_4        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
542
HsSFI1_10       RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
613
PtSFI1_1        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
612
PtSFI1_3        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
612
PtSFI1_8        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
694
PtSFI1          RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
639
HsSFI1_2        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
664
PtSFI110        RQKLMRADLHHQHSVLHRALQAWVTYQGRVRSILREVAARESQHNRQLLRGALRRWKENT
663
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        QTTKQNLWSCQRQANSLRRWLELSQEEPKSEDLHLEEQVKTELEEVRPGQPRASP-----
433
MmSFI1_2        QTRKKLAQQLRDTTLLHRFWNLWQSRIEQREERVQTPSLHAALSHYRVTVLHKCVRVWLR
438
MmSFI1_1        QTRKKLAQQLRDTTLLHRFWNLWQSRIEQREERVQTPSLHAALSHYRVTVLHKCVRVWLR
438
MmSFI1_3        VLREVAARERQHNRQLLWWALHLWRENTMARLDGAKKTSQARVHYSR-------------
430


HsSFI1_1        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWREAVSVQMYYRQQ
737
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWREAVSVQMYYRQQ
682
HsSFI1_3        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWREAVSVQMYYRQQ
655
HsSFI1_4        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWREAVSVQMYYRQQ
584
HsSFI1_10       MARVDEAKKTFQ------------------------------VLVQWREAVSVQMYYRQQ
643
PtSFI1_1        MARVDEAKKTFQ------------------------------VLVQWWEAVSVQIYYRQQ
642
PtSFI1_3        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWWEAVSVQIYYRQQ
654
PtSFI1_8        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWWEAVSVQIYYRQQ
736
PtSFI1          MARVDEAKKTFQASTHY------------------RRTICSKVLVQWWEAVSVQIYYRQQ
681
HsSFI1_2        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWREAVSVQMYYRQQ
706
PtSFI110        MARVDEAKKTFQASTHY------------------RRTICSKVLVQWWEAVSVQIYYRQQ
705
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------WLSFLSACLVPPSRPCPQVELQVQQ
458
MmSFI1_2        YVHKRQWQQLLRARADGHFQQRALPAAFYTWYRGWLWHQQRRILHTKAVRFHRGTLEKQV
498
MmSFI1_1        YVHKRQWQQLLRARADGHFQQRALPAAFYTWYRGWLWHQQRRILHTKAVRFHRGTLEKQV
498
```

```
MmSFI1_3        -------------------------------------TLCSKVLVQWREVTSVQIYYRQK
453


HsSFI1_1        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
767
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
712
HsSFI1_3        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
685
HsSFI1_4        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
614
HsSFI1_10       EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
673
PtSFI1_1        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
672
PtSFI1_3        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
684
PtSFI1_8        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
766
PtSFI1          EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
711
HsSFI1_2        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
736
PtSFI110        EDCAIWEAQKVLDRGCLRTWFQRWWDCSRR------------------------------
735
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        LAKELEAQRQPVGTCIARVRALRRALC---------------------------------
485
MmSFI1_2        FALWRQKMSQHRENCLAERMAILQAEQQLLRRFWFVWHQQAAVCQLERQQQAMAIAHHHS
558
MmSFI1_1        FALWRQKMSQHRENCLAERMAILQAEQQLLRRFWFVWHQQAAVCQLERQQQAMAIAHHHS
558
MmSFI1_3        EAAALREARKALDRGRLQNWFQHWRFCSQR------------------------------
483


HsSFI1_1        ------------------------------------------------------------
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        ------------------------------------------------------------
HsSFI1_3        ------------------------------------------------------------
HsSFI1_4        ------------------------------------------------------------
HsSFI1_10       ------------------------------------------------------------
PtSFI1_1        ------------------------------------------------------------
PtSFI1_3        ------------------------------------------------------------
PtSFI1_8        ------------------------------------------------------------
PtSFI1          ------------------------------------------------------------
HsSFI1_2        ------------------------------------------------------------
PtSFI110        ------------------------------------------------------------
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        GLLRRAFCIWKESTQGFRIERMGRAQAAHFHSAQLLSRAWSMWREVYQNRVRSVLREVAA
618
MmSFI1_1        GLLRRAFCIWKESTQGFRIERMGRAQAAHFHSAQLLSRAWSMWRECLALRLEEQQKLKCA
618
MmSFI1_3        ------------------------------------------------------------
```

```
HsSFI1_1        ------------------------------------------------------------
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        ------------------------------------------------------------
HsSFI1_3        ------------------------------------------------------------
HsSFI1_4        ------------------------------------------------------------
HsSFI1_10       ------------------------------------------------------------
PtSFI1_1        ------------------------------------------------------------
PtSFI1_3        ------------------------------------------------------------
PtSFI1_8        ------------------------------------------------------------
PtSFI1          ------------------------------------------------------------
HsSFI1_2        ------------------------------------------------------------
PtSFI110        ------------------------------------------------------------
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        RERQHNRQLLWWALHLWRENTMARLDGAKKTSQARVHYSRTLCSKVLVQWREVTSVQIYY
678
MmSFI1_1        ALHSQCILLRRALQKWLVYQNRVRSVLREVAARERQHNRQLLWWALHLWRENTMARLDGA
678
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        ------------------------------------------------------------
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        ------------------------------------------------------------
HsSFI1_3        ------------------------------------------------------------
HsSFI1_4        ------------------------------------------------------------
HsSFI1_10       ------------------------------------------------------------
PtSFI1_1        ------------------------------------------------------------
PtSFI1_3        ------------------------------------------------------------
PtSFI1_8        ------------------------------------------------------------
PtSFI1          ------------------------------------------------------------
HsSFI1_2        ------------------------------------------------------------
PtSFI110        ------------------------------------------------------------
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        RQKEAAALREARKALDRGRLQNWFQHWRFCSQRAAQQRFQLGQAAQHHHWQLLMEAMARW
738
MmSFI1_1        KKTSQARVHYSRTLCSKVLVQWREVTSVQIYYRQKEAAALREARKALDRGRLQNWFQHWR
738
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
822
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
767
HsSFI1_3        -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
740
HsSFI1_4        -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
669
HsSFI1_10       -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
728
PtSFI1_1        -----SAQQRLQLERAVQHHRRQLLLEGLARWKMHHLQCVRKRLLHRQSTQLLAQRLSRT
727
```

```
PtSFI1_3        -----SAQQRLQLERAVQHHRRQLLLEGLARWKMHHLQCVRKRLLHRQSTQLLAQRLSRT
739
PtSFI1_8        -----SAQQRLQLERAVQHHRRQLLLEGLARWKMHHLQCVRKRLLHRQSTQLLAQRLSRT
821
PtSFI1          -----SAQQRLQLERAVQHHRRQLLLEGLARWKMHHLQCVRKRLLHRQSTQLLAQRLSRT
766
HsSFI1_2        -----SAQQRLQLERAVQHHHRQLLLEGLARWKTHHLQCVRKRLLHRQSTQLLAQRLSRT
791
PtSFI110        -----SAQQRLQLERAVQHHRRQLLLEGLARWKMHHLQCVRKRLLHRQSTQLLAQRLSRT
790
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        KAHHLGCIRKKFLQRQAAQLLAQRLSRACFCQWRKQLAVRKQEQWGTARALWLWAFSLQA
798
MmSFI1_1        FCSQRAAQQRFQLGQAAQHHHWQLLMEAMARWKAHHLGCIRKKFLQRQAAQLLAQRLSRA
798
MmSFI1_3        -----AAQQRFQLGQAAQHHHWQLLMEAMARWKAHHLGCIRKKFLQRQAAQLLAQRLSRA
538


HsSFI1_1        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
882
HsSFI1_5        ------------------------------------------------------------
HsSFI1_9        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
827
HsSFI1_3        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
800
HsSFI1_4        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
729
HsSFI1_10       CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
788
PtSFI1_1        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQRALQAYQ
787
PtSFI1_3        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQRALQAYQ
799
PtSFI1_8        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQRALQAYQ
881
PtSFI1          CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQRALQAYQ
826
HsSFI1_2        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQWALQAYQ
851
PtSFI110        CFRQWRQQLAARRQEQRATVRALWFWAFSLQAKVWATWLAFVLERRRKKARLQRALQAYQ
850
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        KVWTAWLGFVLERRRKKARLERAMQAYQQQLLQEGATRLLRFTAGTKAFRQQLQAQQQVQ
858
MmSFI1_1        CFCQWRKQLAVRKQEQWGTARALWLWAFSLQAKVWTAWLGFVLERRRKKARLERAMQAYQ
858
MmSFI1_3        CFCQWRKQLAVRKQEQWGTARALWLWAFSLQAKVWTAWLGFVLERRRKKARLERAMQAYQ
598


HsSFI1_1        GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
942
HsSFI1_5        ----------------------QQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
691
```

```
HsSFI1_9          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
887
HsSFI1_3          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
860
HsSFI1_4          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
789
HsSFI1_10         GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
848
PtSFI1_1          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVHRCATLWKQKVLGRGGKPQ
847
PtSFI1_3          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVHRCATLWKQKVLGRGGKPQ
859
PtSFI1_8          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVHRCATLWKQKVLGRGGKPQ
941
PtSFI1            GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVHRCATLWKQKVLGRGGKPQ
886
HsSFI1_2          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVRRCATLWKQKVLGRGGKPQ
911
PtSFI110          GQLLQEGATRLLRFAASMKASRQQLQAQQQVQAAHSLHRAVHRCATLWKQKVLGRGGKPQ
910
MmSFI1_4          ------------------------------------------------------------
MmSFI1_5          ------------------------------------------------------------
MmSFI1_2          AAHSLHCAVRHCAELWKKKVLGPGKTSQPPAPTTFSKRVTFKDSFLSGHAAEAGDATQET
918
MmSFI1_1          QQLLQEGATRLLRFTAGTKAFRQQLQAQQQVQAAHSLHCAVRHCAELWKKKVLGPGKTSQ
918
MmSFI1_3          QQLLQEGATRLLRFTAGTKAFRQQLQAQQQVQAAHSLHCAVRHCAELWKKKVLGPGKTSQ
658


HsSFI1_1          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
1002
HsSFI1_5          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
751
HsSFI1_9          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
947
HsSFI1_3          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
920
HsSFI1_4          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
849
HsSFI1_10         PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
908
PtSFI1_1          PLAAIAPSRKVTFEGPLLNRIAAGAGDATLETKRPQASRPPGALGRLAAEEPHALELNTA
907
PtSFI1_3          PLAAIAPSRKVTFEGPLLNRIAAGAGDATLETKRPQASRPPGALGRLAAEEPHALELNTA
919
PtSFI1_8          PLAAIAPSRKVTFEGPLLNRIAAGAGDATLETKRPQASRPPGALGRLAAEEPHALELNTA
1001
PtSFI1            PLAAIAPSRKVTFEGPLLNRIAAGAGDATLETKRPQASRPPGALGRLAAEEPHALELNTA
946
HsSFI1_2          PLAAIAPSRKVTFEGPLLNRIAAGAGDGTLETKRPQASRPLGALGRLAAEEPHALELNTA
971
PtSFI110          PLAAIAPSRKVTFEGPLLNRIAAGAGDATLETKRPQASRPPGALGRLAAEEPHALELNTA
970
MmSFI1_4          ------------------------------------------------------------
MmSFI1_5          ------------------------------------------------------------
```

```
MmSFI1_2        KKLRAPPSQGVLGSLAGAAGEPCHLDLNAARSSRKQPRRPSFLLERLGSQRSPEWYSLGE
978
MmSFI1_1        PPAPTTFSKRVTFKDSFLSGHAAEAGDATQETKKLRAPPSQGVLGSLAGAAGEPCHLDLN
978
MmSFI1_3        PPAPTTFSKRVTFKDSFLSGHAAEAGDATQETKKLRAPPSQGVLGSLAGAAGEPCHLDL-
717


HsSFI1_1        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1062
HsSFI1_5        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
811
HsSFI1_9        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1007
HsSFI1_3        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
980
HsSFI1_4        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
909
HsSFI1_10       HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
968
PtSFI1_1        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
967
PtSFI1_3        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
979
PtSFI1_8        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1061
PtSFI1          HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1006
HsSFI1_2        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1031
PtSFI110        HSARKQPRRPHFLLEPAQSQRPQKPQEHGLGMAQPAAPSLTRPFLAEAPTALVPHSPLPG
1030
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        QQLEKPPEEESTALLGGSSLTRPFLPGVLPNVPGPKLPPTASPGLELLPPSSIMPHAAGG
1038
MmSFI1_1        AARSSRKQPRRPSFLLERLGSQRSPEWYSLGEQQLEKPPEEESTALLGGSSLTRPFLPGV
1038
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
1122
HsSFI1_5        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
871
HsSFI1_9        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
1067
HsSFI1_3        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
1040
HsSFI1_4        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
969
HsSFI1_10       ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
1028
PtSFI1_1        ALSSAPGPKQPPTASTGLELLLLPPSSFMPCGAAAPARMSAQRATPRDKPPVPSSLASVP
1027
PtSFI1_3        ALSSAPGPKQPPTASTGLELLLLPPSSFMPCGAAAPARMSAQRATPRDKPPVPSSLASVP
1039
```

```
PtSFI1_8        ALSSAPGPKQPPTASTGLELLLLPPSSFMPCGAAAPARMSAQRATPRDKPPVPSSLASVP
1121
PtSFI1          ALSSAPGPKQPPTASTGLELLLLPPSSFMPCGAAAPARMSAQRATPRDKPPVPSSLASVP
1066
HsSFI1_2        ALSSAPGPKQPPTASTGPELLLLPLSSFMPCGAAAPARVSAQRATPRDKPPVPSSLASVP
1091
PtSFI110        ALSSAPGPKQPPTASTGLELLLLPPSSFMPCGAAAPARMSAQRATPRDKPPVPSSLASVP
1090
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        TARVSAKPSIPGPQPWGCPSLPRDLDPQLLPGDSISTRTEPVYGSEATGHTELEAELEGI
1098
MmSFI1_1        LPNVPGPKLPPTASPGLELLPPSSIMPHAAGGTARVSAKPSIPGPQPWGCPSLPRDLDPQ
1098
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1182
HsSFI1_5        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
931
HsSFI1_9        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1127
HsSFI1_3        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1100
HsSFI1_4        DPHLLLPGDFSATRAGPGLSTAA-----------WTLRLNLRRSSSNYCTTRPPS-----
1013
HsSFI1_10       DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1088
PtSFI1_1        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1087
PtSFI1_3        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1099
PtSFI1_8        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1181
PtSFI1          DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1126
HsSFI1_2        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1151
PtSFI110        DPHLLLPGDFSATRAGPGLSTAGSLDLEAELEEIQQQLLHYQTTKQNLWSCRRQASSLRR
1150
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        QQQLQHYQTTKQNLWSCQRQANSLRRWLELSQEEPKSEDLHLEEQVKTELEEVELQVQQL
1158
MmSFI1_1        LLPGDSISTRTEPVYGSEATGHTELEAELEGIQQQLQHYQTTKQNLWSCQRQANSLRRWL
1158
MmSFI1_3        ------------------------------------------------------------


HsSFI1_1        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1242
HsSFI1_5        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
991
HsSFI1_9        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1187
```

```
HsSFI1_3        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1160
HsSFI1_4        ------RTSGP--------------------------VGGKR---AACAGGWS-------
1031
HsSFI1_10       WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1148
PtSFI1_1        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1147
PtSFI1_3        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1159
PtSFI1_8        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1241
PtSFI1          WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1186
HsSFI1_2        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1211
PtSFI110        WLELNREEPGPEDQEVEQQVQKELEQVEMQIQLLAEELQAQRQPIGACVARIQALRQALC
1210
MmSFI1_4        ------------------------------------------------------------
MmSFI1_5        ------------------------------------------------------------
MmSFI1_2        AKELEAQRQPVGTCIARVRALRRALC----------------------------------
1184
MmSFI1_1        ELSQEEPKSEDLHLEEQVKTELEEVELQVQQLAKELEAQRQPVGTCIARVRALRRALC--
1216
MmSFI1_3        ------------------------------------------------------------
```

SCORE RESULTS FOR HOMO SAPIENS, MUS MUSCULUS AND PAN TROGLODYTES

SeqA	Name	Length	SeqB	Name	Length	Score
1	HsSFI1_1	1242	2	HsSFI1_2	1211	100.0
1	HsSFI1_1	1242	3	HsSFI1_3	1160	100.0
1	HsSFI1_1	1242	4	HsSFI1_4	1031	96.0
1	HsSFI1_1	1242	5	HsSFI1_5	991	100.0
1	HsSFI1_1	1242	6	HsSFI1_9	1187	100.0
1	HsSFI1_1	1242	7	HsSFI1_10	1148	100.0
1	HsSFI1_1	1242	8	MmSFI1_1	1216	64.0
1	HsSFI1_1	1242	9	MmSFI1_2	1184	64.0
1	HsSFI1_1	1242	10	MmSFI1_3	717	66.0
1	HsSFI1_1	1242	11	MmSFI1_4	120	56.0
1	HsSFI1_1	1242	12	MmSFI1_5	485	63.0
1	HsSFI1_1	1242	13	PtSFI1	1186	98.0
1	HsSFI1_1	1242	14	PtSFI1_1	1147	98.0
1	HsSFI1_1	1242	15	PtSFI1_3	1159	98.0
1	HsSFI1_1	1242	16	PtSFI1_8	1241	98.0
1	HsSFI1_1	1242	17	PtSFI110	1210	98.0
2	HsSFI1_2	1211	3	HsSFI1_3	1160	97.0
2	HsSFI1_2	1211	4	HsSFI1_4	1031	93.0
2	HsSFI1_2	1211	5	HsSFI1_5	991	96.0
2	HsSFI1_2	1211	6	HsSFI1_9	1187	100.0
2	HsSFI1_2	1211	7	HsSFI1_10	1148	97.0
2	HsSFI1_2	1211	8	MmSFI1_1	1216	62.0
2	HsSFI1_2	1211	9	MmSFI1_2	1184	62.0
2	HsSFI1_2	1211	10	MmSFI1_3	717	66.0
2	HsSFI1_2	1211	11	MmSFI1_4	120	56.0

2	HsSFI1_2	1211	12	MmSFI1_5	485	64.0
2	HsSFI1_2	1211	13	PtSFI1	1186	98.0
2	HsSFI1_2	1211	14	PtSFI1_1	1147	95.0
2	HsSFI1_2	1211	15	PtSFI1_3	1159	95.0
2	HsSFI1_2	1211	16	PtSFI1_8	1241	98.0
2	HsSFI1_2	1211	17	PtSFI110	1210	98.0
3	HsSFI1_3	1160	4	HsSFI1_4	1031	96.0
3	HsSFI1_3	1160	5	HsSFI1_5	991	94.0
3	HsSFI1_3	1160	6	HsSFI1_9	1187	97.0
3	HsSFI1_3	1160	7	HsSFI1_10	1148	100.0
3	HsSFI1_3	1160	8	MmSFI1_1	1216	63.0
3	HsSFI1_3	1160	9	MmSFI1_2	1184	61.0
3	HsSFI1_3	1160	10	MmSFI1_3	717	59.0
3	HsSFI1_3	1160	11	MmSFI1_4	120	13.0
3	HsSFI1_3	1160	12	MmSFI1_5	485	53.0
3	HsSFI1_3	1160	13	PtSFI1	1186	95.0
3	HsSFI1_3	1160	14	PtSFI1_1	1147	98.0
3	HsSFI1_3	1160	15	PtSFI1_3	1159	98.0
3	HsSFI1_3	1160	16	PtSFI1_8	1241	98.0
3	HsSFI1_3	1160	17	PtSFI110	1210	95.0
4	HsSFI1_4	1031	5	HsSFI1_5	991	77.0
4	HsSFI1_4	1031	6	HsSFI1_9	1187	93.0
4	HsSFI1_4	1031	7	HsSFI1_10	1148	94.0
4	HsSFI1_4	1031	8	MmSFI1_1	1216	59.0
4	HsSFI1_4	1031	9	MmSFI1_2	1184	57.0
4	HsSFI1_4	1031	10	MmSFI1_3	717	54.0

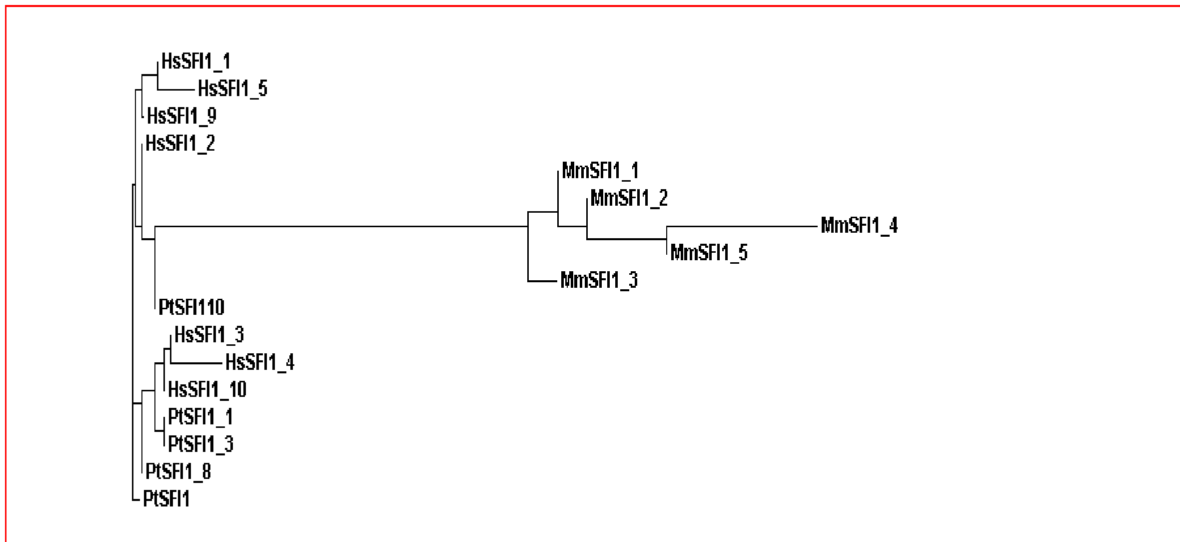
4	HsSFI1_4	1031	11	MmSFI1_4	120	13.0
4	HsSFI1_4	1031	12	MmSFI1_5	485	34.0
4	HsSFI1_4	1031	13	PtSFI1	1186	90.0
4	HsSFI1_4	1031	14	PtSFI1_1	1147	93.0
4	HsSFI1_4	1031	15	PtSFI1_3	1159	94.0
4	HsSFI1_4	1031	16	PtSFI1_8	1241	93.0
4	HsSFI1_4	1031	17	PtSFI110	1210	90.0
5	HsSFI1_5	991	6	HsSFI1_9	1187	96.0
5	HsSFI1_5	991	7	HsSFI1_10	1148	94.0
5	HsSFI1_5	991	8	MmSFI1_1	1216	60.0
5	HsSFI1_5	991	9	MmSFI1_2	1184	58.0
5	HsSFI1_5	991	10	MmSFI1_3	717	40.0
5	HsSFI1_5	991	11	MmSFI1_4	120	36.0
5	HsSFI1_5	991	12	MmSFI1_5	485	58.0
5	HsSFI1_5	991	13	PtSFI1	1186	95.0
5	HsSFI1_5	991	14	PtSFI1_1	1147	92.0
5	HsSFI1_5	991	15	PtSFI1_3	1159	92.0
5	HsSFI1_5	991	16	PtSFI1_8	1241	98.0
5	HsSFI1_5	991	17	PtSFI110	1210	95.0
6	HsSFI1_9	1187	7	HsSFI1_10	1148	97.0
6	HsSFI1_9	1187	8	MmSFI1_1	1216	62.0
6	HsSFI1_9	1187	9	MmSFI1_2	1184	60.0
6	HsSFI1_9	1187	10	MmSFI1_3	717	63.0
6	HsSFI1_9	1187	11	MmSFI1_4	120	36.0
6	HsSFI1_9	1187	12	MmSFI1_5	485	59.0
6	HsSFI1_9	1187	13	PtSFI1	1186	98.0

6	HsSFI1_9	1187	14	PtSFI1_1	1147	95.0
6	HsSFI1_9	1187	15	PtSFI1_3	1159	95.0
6	HsSFI1_9	1187	16	PtSFI1_8	1241	98.0
6	HsSFI1_9	1187	17	PtSFI110	1210	98.0
7	HsSFI1_10	1148	8	MmSFI1_1	1216	63.0
7	HsSFI1_10	1148	9	MmSFI1_2	1184	61.0
7	HsSFI1_10	1148	10	MmSFI1_3	717	58.0
7	HsSFI1_10	1148	11	MmSFI1_4	120	13.0
7	HsSFI1_10	1148	12	MmSFI1_5	485	53.0
7	HsSFI1_10	1148	13	PtSFI1	1186	95.0
7	HsSFI1_10	1148	14	PtSFI1_1	1147	98.0
7	HsSFI1_10	1148	15	PtSFI1_3	1159	98.0
7	HsSFI1_10	1148	16	PtSFI1_8	1241	98.0
7	HsSFI1_10	1148	17	PtSFI110	1210	95.0
8	MmSFI1_1	1216	9	MmSFI1_2	1184	100.0
8	MmSFI1_1	1216	10	MmSFI1_3	717	99.0
8	MmSFI1_1	1216	11	MmSFI1_4	120	89.0
8	MmSFI1_1	1216	12	MmSFI1_5	485	93.0
8	MmSFI1_1	1216	13	PtSFI1	1186	62.0
8	MmSFI1_1	1216	14	PtSFI1_1	1147	63.0
8	MmSFI1_1	1216	15	PtSFI1_3	1159	63.0
8	MmSFI1_1	1216	16	PtSFI1_8	1241	64.0
8	MmSFI1_1	1216	17	PtSFI110	1210	62.0
9	MmSFI1_2	1184	10	MmSFI1_3	717	99.0
9	MmSFI1_2	1184	11	MmSFI1_4	120	89.0
9	MmSFI1_2	1184	12	MmSFI1_5	485	93.0

8	MmSFI1_1	1216	13	PtSFI1	1186	62.0
8	MmSFI1_1	1216	14	PtSFI1_1	1147	63.0
8	MmSFI1_1	1216	15	PtSFI1_3	1159	63.0
8	MmSFI1_1	1216	16	PtSFI1_8	1241	64.0
8	MmSFI1_1	1216	17	PtSFI110	1210	62.0
9	MmSFI1_2	1184	10	MmSFI1_3	717	99.0
9	MmSFI1_2	1184	11	MmSFI1_4	120	89.0
9	MmSFI1_2	1184	12	MmSFI1_5	485	93.0
9	MmSFI1_2	1184	13	PtSFI1	1186	60.0
9	MmSFI1_2	1184	14	PtSFI1_1	1147	61.0
9	MmSFI1_2	1184	15	PtSFI1_3	1159	61.0
9	MmSFI1_2	1184	16	PtSFI1_8	1241	64.0
9	MmSFI1_2	1184	17	PtSFI110	1210	62.0
10	MmSFI1_3	717	11	MmSFI1_4	120	89.0
10	MmSFI1_3	717	12	MmSFI1_5	485	51.0
10	MmSFI1_3	717	13	PtSFI1	1186	62.0
10	MmSFI1_3	717	14	PtSFI1_1	1147	58.0
10	MmSFI1_3	717	15	PtSFI1_3	1159	59.0
10	MmSFI1_3	717	16	PtSFI1_8	1241	66.0
10	MmSFI1_3	717	17	PtSFI110	1210	66.0
11	MmSFI1_4	120	12	MmSFI1_5	485	89.0
11	MmSFI1_4	120	13	PtSFI1	1186	37.0
11	MmSFI1_4	120	14	PtSFI1_1	1147	15.0
11	MmSFI1_4	120	15	PtSFI1_3	1159	15.0
11	MmSFI1_4	120	16	PtSFI1_8	1241	57.0
11	MmSFI1_4	120	17	PtSFI110	1210	57.0

12	MmSFI1_5	485	13	PtSFI1	1186	58.0
12	MmSFI1_5	485	14	PtSFI1_1	1147	53.0
12	MmSFI1_5	485	15	PtSFI1_3	1159	53.0
12	MmSFI1_5	485	16	PtSFI1_8	1241	63.0
12	MmSFI1_5	485	17	PtSFI110	1210	63.0
13	PtSFI1	1186	14	PtSFI1_1	1147	97.0
13	PtSFI1	1186	15	PtSFI1_3	1159	97.0
13	PtSFI1	1186	16	PtSFI1_8	1241	100.0
13	PtSFI1	1186	17	PtSFI110	1210	100.0
14	PtSFI1_1	1147	15	PtSFI1_3	1159	100.0
14	PtSFI1_1	1147	16	PtSFI1_8	1241	100.0
14	PtSFI1_1	1147	17	PtSFI110	1210	97.0
15	PtSFI1_3	1159	16	PtSFI1_8	1241	100.0
15	PtSFI1_3	1159	17	PtSFI110	1210	97.0
16	PtSFI1_8	1241	17	PtSFI110	1210	100.0

DENDOGRAM FOR HOMO SAPIENS, MUS MUSCULUS AND PAN TROGLODYTES



SEQUENCE ALIGNMENT RESULTS FOR SUS SCROFA, CANIS FAMILIARIS, RATTUS NOVERGICUS, CHLAMYDOMONAS REINHARDTII, SACCHAROMYCES CEREVISIAE AND DANIO RERIO

```
CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          MTINNPLRLSSEVSPLRHVLDITARSQGKDLTAYETYRALPSPTRPYSGYDYAVVAEIIR
60
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          AAEVTVDKAHKKGLGSGIVTLQSVLQAYEHVLPRHGVKADEDTYYYRLLLKLSLDPALDW
120
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          WIKLNRETGTTGGRAAFFSNAGSDVTSMGRSHSFYRGDTSPTRVQPRTSGAGASDYSASH
180
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          SAAPTFRSSIPGNGAADGGSRVPGRNSPYRSTASLSSMPGYQQQQETAAQAAAREAAARA
240
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          NAAAQAARDAIEAARQAAATAHLPSSRFMFNTPPMTPAMGPVPPPSARTEGGASGWGDGS
300
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
```

```
CrSFI1          MGGAGSIVPYSVSEGGATGPGGASQYGPGSRYGPGSPHGTTAPSSVRGTYGISEHDPLSP
360
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          DASAYGAGSRGTRSGGGADGGSMGGAVGPGATRASSGGGGGPGRVSAAQLAELDEPVLSE
420
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          EAEAYLDYNRLARAFRMWRKNTFARKASRFERDHALKNWAVATSFWAVHLLRRCFARWRA
480
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          EGPSKTALALQLWAGHSLSACFWRWLSLTRYLRGKGTECDGHWRVRTLRRCLRMWRLYTR
540
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          SQQLKADNVEKALGFWVGRSLHICFRVWRQFAANQLRKNEAVQKAVGHWAGKSLRACFEV
600
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          WREFTMLQQHKSVATARALRHWTGTTLRSCFALWRLLVEDAHSNWDQATQHLMRRMLLKW
660
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          RIAAKQLTRMSAAYHAIADATARRVMSQAMRTLLSAHAFRHETANRLRAAVARMRNSLLS
720
ScSFI1          ------------------------------------------------------------
```

```
CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          AAFNKWSEARQAGQMKRERMALALKWWSSRGLARSWNTWREAVAASQRATKAARANAQRI
780
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          LRPLLRDTFYAWLDWHDVRLQKEVRAVAARRTMQLRQMRLVVLGWRGLVLERRHSLVSFQ
840
ScSFI1          ------------------------------------------------------------

CfSFI1          -MKNLLTKKYISSHWPNFHQKRRKLRMQRKVDSRSFRGGAVKKPYPPKISANQKSSTLLG
59
SsFI1           ------------------------------SNSRSFRDGAVKKPYSPKILSNKKCSAFPG
30
RnSFI1          --------------------------MEKKIGFRSFRDGVVKKPCSPKALPLKKSSAFSG
34
DrSFI1          ------------------------------------------------------------
CrSFI1          RIMRNEVQRIMSTAFQRWYREHDNLRRAAEARLTVKRNATKRCLRHWRALAQAQRAHREW
900
ScSFI1          ------------------------------------------------------------

CfSFI1          IRSEIPSASHPVQCHASHGWTRKNRLRELRIRCVARKFLYLWIRMTFGRVFPSKARFYY-
118
SsFI1           MRSEIPKAHRPIQFLPSLARPRRGRLRELRLRCVARKFLYLWIRMTFGRVFPSKARFYY-
89
RnSFI1          IQRELTRSCHSSYYQSSQSWTP-QRLRELRVRCVARKFLYLWMRVTFGRVTPSRARFFH-
92
DrSFI1          ------------------------------------------------------------
CrSFI1          RLERLRLASEQRQRRLLVAWQVVAGLLAEHNRLVKSSLFKLHRRQAHSVLAAWRDRVLHS
960
ScSFI1          ------------------------------------------------------------

CfSFI1          --------------EQRILRKVFEEWKEEWW----------VFQREWKLCVRADCHYRYY
154
SsFI1           --------------EQRILRKVFEEWKEEWW----------VSHREWKLCVRADCHYRYY
125
RnSFI1          --------------EQKILRKVFGEWREEWW----------VSQREWKLCVRADCHYRYY
128
DrSFI1          ------------------------------------------------------------
CrSFI1          VAKRMKLLAGLLYWEQQVKGRAWLAWRQRQHGWRIKAAKGELAGRHWRTRRLALCFAVFL
1020
ScSFI1          ------------------------------------------------------------
```

```
CfSFI1          LYNLMFQTWRTYVYQQQEMRNKYIRAENHDAKQKMR------------------------
190
SsFI1           LYNLTFQTWKAYVHQQQEMRNKYMRAEDHDAKHKMR------------------------
161
RnSFI1          LYNLIFQNWKSFVYQQREMRKRFQRAEHHDMKQKMR------------------------
164
DrSFI1          ------------------------------------------------------------
CrSFI1          ALWQAYQGRLTRALLFRSRRNDIIKADVLDEWRRLVGWLQWKNDKIRRALAHHRSKLGGG
1080
ScSFI1          ------------------------------------------------------------


CfSFI1          --QAWKSWLIYVVFRRTKLHMQTTALEFRQQSILWLWWRRWRQQLGQVCVG---------
239
SsFI1           --RAWKSWLIYLVFRKTKLQMQTTALEFRQHRVLWVWWSEWRRQLGQVRLG---------
210
RnSFI1          --HAWKSWLIYMVARRTKFRMKSTALEFRRRSVLCFWWNKWRWRLGRVHAE---------
213
DrSFI1          ------------------------------------------------------------
CrSFI1          AFYGWYAVTRHSIATKRAVAAAVLHWVRRRRAAAFVWWRNWTAYRRDLRVRGAGYALVRA
1140
ScSFI1          ----------------------MGKFGTTNKSTENLLRDKFVPETSPTNIP---------
29


CfSFI1          -----------------RALHASAVKHRALSLQQQAWSQWWEQLLHVQRERQKMVSAVKH
282
SsFI1           -----------------HALQASAVKHRVQSLQLQAWSRWREQLLHVWRERQKVVSAVQH
253
RnSFI1          -----------------QALHAAAVKHRALSLQLQAWLRWQEQLLISQRERRKVATAVRH
256
DrSFI1          -----------------------------------------DRMKHACALKEKESKAHHH
19
CrSFI1          ARAKEKVVLAFKANVIRKARIAGAVAHRQRYMTRLALSGWLLRTLLAADFVRRLKAVGAQ
1200
ScSFI1          ----------------------------TDVLIKQGQITDSTESLIHGGAERYIVNALKP
61
                                                          .        .    .

CfSFI1          HQHWQKWKSLKAWLKYVQVRRVKRQLNEMAERFHRVT-----------------------
319
SsFI1           HQCRQKWRFLKAWLEYLHVRRVKRQRNTMAERFHYVT-----------------------
290
RnSFI1          HQRWQKQRTLKAWLKYLRICRVKRWQNEMAVQFHRIT-----------------------
293
DrSFI1          YCLQLLKQTLHGWIRHAENRQAKNKKIAVAQDVRRLS-----------------------
56
CrSFI1          LNVQLVVDGLEAWREYTIYRRTKKGMLHAALRYWRLSRQRAAMDAMRWYTTRRQLKRAVL
1260
ScSFI1          IELN----KTEGFFEDPPFHLPSPPVDSTNLEYEDVT-----------------------
94
                          ..: .       .            ::

CfSFI1          ------VLQTHFCDWQWAWERKERLSAQQARVEGLARRMTL-RRAFTHWKHYVLLCAEEA
372
SsFI1           ------VLQAHFCDWQWAWERRQSLYAHQARVKGLARRMAL-RRAFTHWKHYVLLCAEAV
343
```

```
RnSFI1          ------VLQIHFCDWQWAWEWRQSLSAHQALVAKLARKMVL-RRAFAHWKHYMLLQEEEA
346
DrSFI1          ------VLERFWWQWFEVLQSRRVEKDRGQRADKLAQHGSQ-RLVFSHWRHYVSICSQKA
109
CrSFI1          LRGRTGAAARALRAWRDEAEYRRRLRDRYAMLASGSRHAAL-RRGLDMWRDYIANRASKK
1319
ScSFI1          ------DLPKNGLRYDLNDISVEVIEDLYRQIEAFLVHFKLSRSFLQIFKNYVNILIQEG
148
                              :       .              :    *  :  ::.*:    .

CfSFI1          AQWKVAEDHSRCSLLHFCFR-------------------------------ALKENVT
399
SsFI1           IWRKGAEEHRRRCLLHFCFR-------------------------------ALKDNVT
370
RnSFI1          AQREAAAEHHQHCLLHSCFR-------------------------------AFKDNVT
373
DrSFI1          KREKSAMHHRQRYLLHLGLK-------------------------------GFALNVT
136
CrSFI1          GLREQALRHWLRGTAGRVFRSWRDYQRHLRHAEERAGDIVQRWRKRDAAEALAAFAEYAA
1379
ScSFI1          INPLRDEYFTILEDELKGFFTFN------------------------SVIEEILEIFLI
183
                        .         :                                   :

CfSFI1          HAHLQQIRRNLAHQQRDITLLHRFWNLWQSRIEQ---------------REEREQLPSLC
444
SsFI1           RAHLQRMRRNLAHQQHHVMLLRRFWNVWQSRIEQ---------------REEREHLPSLC
415
RnSFI1          RARLKRMRKSLAHQLRDTTLLHRFWNLWQSRVEQ---------------REERVQPPSLL
418
DrSFI1          QSKTYRLNKNISVQHRHQNLLAMCWNVWQLRLDR---------------EENRRLQPQMS
181
CrSFI1          HRRRKRTADAMSRRSRARTALQAWREYMKLRTAEGFWSRSRLNDAIGAWRHYANHRRRLK
1439
ScSFI1          HPRNKFIALSLAEYTYAKNKIRRHFNHWKTVCELN---------------------EEAN
222
                : :       ::        :    :  :

CfSFI1          AAWAHCRMTLLRKCIKSWLQYTQKRRYKQ-------------------RLRSRADSHFQQ
485
SsFI1           AAWDHYRVTLLRQCFKSWLQYTRKRRSQQ-------------------LLQARADSHFQQ
456
RnSFI1          AAQSHYRMTVLRKCVRVWLQYVHKRRCLQ-------------------SLQARALGHFQQ
459
DrSFI1          VAQNHHKFSVLRFNLHHWKKRLAEHKKCQ-------------------ELELRADACFAQ
222
CrSFI1          ELWGRTRRGRLASALSSWRDYATYSARLKTVSNAAKQMAALKALLERVLVRTAALAFYGW
1499
ScSFI1          RFANQAKLRVQEAVFYIWSDKTLKYSQMAN---------------------DEAESFRNT
261
                    : :       .  * .                                 * .

CfSFI1          RALPAAFNTWRRFWRWHQQESVLNA-------RAARFHRETLEKQVFAIWWQKMFQHREN
538
SsFI1           RALPAAFQAWRRLWQWRQQEQVLSA-------RAMCFHREMLEKQVFAVWCQKMSQRRDY
509
```

```
RnSFI1          RALPAAFYTWYRVWRWHQQRRTLHT-------RAVRFHREALEKQVFALWRQKVSQHREI
512
DrSFI1          RIFPQCVNTWMEFTAQRTEKREQKE-------RAEQQYRQQTFSWVFYTWWRNLEARRDR
275
CrSFI1          REVVADAKRWRAVQAVTQERMQQRPELAEVAMYAATRMRNWQLALAFYTWYDNARESRYL
1559
ScSFI1          WLLFRSFQQWITLTQTLKEQSRLAD---------QAFLNKMFRKILKAQEHWKHLETVNT
312
                  .      *  .     :.                  .:            :

CfSFI1          RLAERMAILHSERQLLQRSWSMWHQQAAARHREWQRQAVACAHHRHRQLRKAFCIWRETA
598
SsFI1           RLAERMAILHAERQLLRRHWSTWCQQAAACCLEQLWQAVARAHHRLERLRKAFCVWRERT
569
RnSFI1          RLAERMAMLQAEQQLLRRFWFVWHQRAAACHQERQRQAMAIAHHHSGLLRRAFCIWREST
572
DrSFI1          RLAERTAVLHAEHVCYVRVWSKWYSGAVQRREERIKQAVADSLYKHTLQQKNLNHWKNRF
335
CrSFI1          KNRASQAILLYANRLLYNAWSVWLAHTLRNRHLRKKLERCLNSTRAKTSRGVFDAWLAAA
1619
ScSFI1          DNIKKIFLRTTFHIWKLRHKEINYHGLERRIFERIKQKVINYEYNKSIAEKVRSFSLQRK
372
                       :    .    .                          .    .

CfSFI1          RGLRTEKMG--------------------LVLAAEFQAAR---------LLHWAWSRWRE
629
SsFI1           QGLRTERMG--------------------RSWAARFHSTR---------LLRWAWTRWRE
600
RnSFI1          QGFRMERMG--------------------RAQASHFHSVR---------LLCWAWSKWRE
603
DrSFI1          INIQTSQKR--------------------FKQAEGHHEQR---------CLKRAMTDWHQ
366
CrSFI1          KYLGRLRKVSDLVKDKLLRGTLVGAFTSWRRTAIQFRRLRGILTRVMSRALSTAWEAWRD
1679
ScSFI1          YLNKWEKKNIE----------------NEDKLGALYELEN-----------KFIKQKFFR
405
                      :                         .  ..  .                 :

CfSFI1          CLALHAAEQRKLMRASLHHQQALLHGTLQTWATYQSQVRSILQEVAARE-----------
678
SsFI1           GLALRSAERQQLMRADLHRQHILLHRVLQTWRIHQSHVRSVLQEVAARE-----------
649
RnSFI1          GLALRMEEQQKLKRAALHSQRTLLHRALQKWLVYQDRMRSVLQEVAARE-----------
652
DrSFI1          YVDHRRLKKEKIAEMEKHYHNKLLKHALDAWKSYHLQTQAISQLVEYSY-----------
415
CrSFI1          AVAERHERMQSLAVFVGHWRNLHLSAAFNAWVAHVHKRAAARRLCLRVLGRLLEWAWYGW
1739
ScSFI1          KLNRSFQHSQQEAIAKSKLNQTLLRCVFEKMWLKRFEDHLHLYSIVSLK-----------
454
                 :     . ..      : ..  *  .::       .

CfSFI1          -------------SQHKRRLLRSVFRRWREN------------------TLARVNEAKQT
707
SsFI1           -------------AWHRRQRLRAAFRQWREN------------------AVAQADKAKKT
678
```

```
RnSFI1          -------------RQHNRQLLWWVLHRWREN------------------TMARLDGAKKT
681
DrSFI1          -------------QEHEQHLVRRMFCLWRIN------------------VLQLVEEREKE
444
CrSFI1          REAVFEAGVSRGADVHEARLLARSWRGWRWATSEGRKAAALTAALNQAAALTAMGVWRRQ
1799
ScSFI1          ----------------EANLVKRIFHSWKKL-------------------------LYID
473
                                . . :      *:

CfSFI1          SRASTHYRRTLCSKILVQWQEAASVQIYYRQQEDCAIKEAQKVLERGCLRTWFRHWRDQG
767
SsFI1           SQAAHHYRRTLCFKVLMEWREAASVQVYYRQQEDDAVREAQVVLSRGHLRTWFWRWWDHS
738
RnSFI1          SQARVHYNRTLCSKVLVQWREVTSVKIYYRQKEAAALREARKALDRGRLRTWLQRWQVCR
741
DrSFI1          NRATCLSQKHLVSQVFLAWQQRTVSAHLHHHK----------------------------
476
CrSFI1          WLARAAYRRLYCRRALLGWHSRTQELRAMRERLWYAARFLLNGCLLRSFSAWWQYTQAMS
1859
ScSFI1          LKASDYSRTNLLKSSLRSWKLEVKLKIFEQKCKKS-------------------------
508
                  *    .       :  *:  .      :.

CfSFI1          QRAAQQRVQLRRAAGHHGRRLLLQTMARWKAHHFGCVRKRLLQ------RQGAQLLAQTR
821
SsFI1           QRAAWQRVQLERAAQHYRRRLLLQAVAQWKAHHLGCIRKRLLQ------RQAARLLAQTL
792
RnSFI1          QREAQQTFRLQQAAQHHRRQLLAGAMARWKAYHLCCVRKKFLQ------RQAAQLLAQRL
795
DrSFI1          -------------------------------------IYQVMK------KRSFELHKLRT
493
CrSFI1          IKRAVWVRKQRALAEALRRGGELVAARNAELMEIAFRGWRMQTGLLREVTRRLRSIQGRT
1919
ScSFI1          ----------------------------IQASAYRTWRKRIQYG-----KISSEHVKTAF
535
                                                       :.            .

CfSFI1          SRACFHQWKQQLANRRREQQGTARALWFWSFS----------------------------
853
SsFI1           CQACFHQWRRQLEDRRRERQETARALWFWAFS----------------------------
824
RnSFI1          GRACFCQWRKQLAVRKQEQWGTARALWFWAFS----------------------------
827
DrSFI1          CQHFFICWKTQLQSRRREAEQTETALWHWSLN----------------------------
525
CrSFI1          LATAFGQWREYAAVKRARLEKQAAVIRSRLLARPFRTWRAAARGAASELSAKSAAALMQR
1979
ScSFI1          CAKYLGVWKRRMLQMNSMNDEASKFYEEGLVN----------------------------
567
                    :  *:      .              .

CfSFI1          ----------------------------------------LQAKVWAAWLGFVLER----
869
SsFI1           ----------------------------------------LQAKVWAAWLGFVLER----
840
```

```
RnSFI1          ----------------------------------------LQAKVWAAWLGFVLEK----
843
DrSFI1          ----------------------------------------LQAKVFCMWRIWIAER----
541
CrSFI1          RGVLLAATFRVWRELHGVVGSRREALRRALGVVLRAEAGVLKQAVWGAWRWAVARRGDLL
2039
ScSFI1          -------------------------------------------ECLAIWKERLIKT----
580
                                                                *   : .

CfSFI1          ----------RRKKVRLERAGQAYHQQLLREG--------------------ATRLLRFA
899
SsFI1           ----------RRKKARQERAVQAYHQQLLREG--------------------VTRLLRFA
870
RnSFI1          ----------RRKKARLEQAMQAYHQQLLQEG--------------------ATRLLRFT
873
DrSFI1          ----------QRKQKRLTEAAQFYRDELLREG--------------------VTHILTHT
571
CrSFI1          DLVSAVSAGRRRRVLGEVFDVWLQYTQAMRRGGIDPGSPYMSPRERGTERRLITRMAALA
2099
ScSFI1          ------------KELEDRYNFLCKTHAILTVK----------------------RTLMHI
606
                           :               :                         :

CfSFI1          GSMKAFRQQLHAQ------------------------QQEQAAHSLHRAVHRCAMLWKQK
935
SsFI1           AGTKAFRQRLHAQ------------------------QQVQAAHSLQRTVRRCATLWKQK
906
RnSFI1          AGMKALRQQLQAQ------------------------QQVQAAHSLHCAVRHCAELWKKK
909
DrSFI1          AHMNAFSTNIAQH------------------------SYEQSCRQLQEVVRRCALRWKQR
607
CrSFI1          GNDQALAPAAHAGGASVAATDALLDGLYPSTGMQHASERRQELSAFLDFARRATSTGRAG
2159
ScSFI1          DNVHLLYTKLAPS-----------------------MDRVKLSKAFLKWRKATRFKVRHK
643
                  : :                                .  :    :    :      :

CfSFI1          VLGQGK-ERRPLPTMPRRRVTFDSPL--PTCVAAGAGDATLETRRPGDPHVPWGALGSLS
992
SsFI1           ALGPGRGPQPPTSAVLSRRVTFEDPP--LSSVAAGTGDASLETKRPPAPRGLWGALGSPV
964
RnSFI1          VLGPDKTSQPPAPIAFSRRVTFKDFS--LSGLAAEAGDAPLSQ----------GVLGSLA
957
DrSFI1          ALCKPVKDKSTPSKDSQAKKSVSFFL--PGDPMPTLCP---------------GRIQSPE
650
CrSFI1          RNGRSAAQQAAPPVYERAESISSDLSRASTGLSSGAGGADLTERAMFSVRAMPGMISPGK
2219
ScSFI1          LNDILHVYEKSKERELQSQLFNAWRN--RFCFYTEECNIQAISKRN-------YQLEKMV
694
                        . .      .              .                     :

CfSFI1          LAAGDP-------QLLELNATRLARKQPRRPHFLLEPVQSQRPQAPQECDRGLVWPAGPS
1045
SsFI1           SAAGEP-------QLLELNAARWARKQPRCPDFLLEPEASQTPLGGQGPE--ALWGRDPA
1015
```

```
RnSFI1          VAAGEP-------FIPELNEARPSRKQPRRPSFLLERVQSQRSPERCTLGEQQLQKTPEK
1010
DrSFI1          QRQKDS-------VIHQMWSVRASRPQPRRPGDLLESPAKHLSRDLKISRGQRVFVSETA
703
CrSFI1          GRGPASGLTPAHIQLATAGRAAPAPLSARRPSGPQDFAGLSSPFDDRLDLRELYGAADQA
2279
ScSFI1          LKKFRER-------LLEIVKSEELADEVREEFVLVKTFYIWKTHLDEIFYMSTLLEQSEA
747
                              :           . *      .      .

CfSFI1          LMRPFPP--EAWMALVPSSPSPQPRALLGRPGLKLPPTLKTGPELLPPS--SFMPCGVEA
1101
SsFI1           LVTPFLAKAKAWAALDPSSPLP------SAPGLKPPPSVGPGPELLPPS--SFTPCGARD
1067
RnSFI1          GQSMAPPGGSSLTRPFLPVVRP------NVSGPKLPPTASPGLELLPPS--SFMPHGVRD
1062
DrSFI1          FPSVPSTFPSPLPSVPLTHNPLNSPAMVLQPKTLLLKPDSPGCIATTSQDILLPPSTFTL
763
CrSFI1          AAAAPAPFGGFHTTAGYGARGAQLASGAASPMPLMPPPAARGGLAMATPLTASVPIAPRR
2339
ScSFI1          NKQFIITSKFLKMWSLRFLKIKRNDETVEVFRHRWDRATVRGLLLLWKNRSDSSPKRRKD
807
                      .                              .   *            *

CfSFI1          PARASAQPTTPG-LTLQAS-PSPASVPQSRLLLPGDFM----GTGRSSGTITAGHDDLEA
1155
SsFI1           GWLLSSGPAGPGPDTVGVY-LGPGIRDPWVQPGPSQEE----GQDQAQFSAATGHTTLVA
1122
RnSFI1          AARVSTKPSISGPQPWGCPSLTRDPDPHLLLPGDSTST----RTGPGYGSETTGHTELEA
1118
DrSFI1          STAHSKQKYLRDGGPLLHSSHFTSRFFPTQALGLHERI----PEDEEEDVDVEQTENLTK
819
CrSFI1          IDFTASRPGSGAGAGPATGGAYGYGGPAGRGPGTPRSQQFAQPAEPAYPDDSDGASTTSS
2399
ScSFI1          FNLKHELKTPIRSDSQNASTIPGSERIKQHRMEAMKSH-------YSRARRAIPSPVKSS
860


CfSFI1          ELESIQQQLQDYQTMKQNLSSCQRQAR--------------------------SLRRWLE
1189
SsFI1           ELEGIQQQLQDYQTMKQNLRSCQRQAS--------------------------SLRRWLE
1156
RnSFI1          ELEGIQQQLQHYQTTKQNLWSCQRQAN--------------------------SLRRWLE
1152
DrSFI1          ELLDIRLEMQRYQQDRKQLQTWRKLQK--------------------------VLGNWLE
853
CrSFI1          DFLPPSEARARAQQARAAAVATPRIALGVAAIHGTFIQANRPEHFHAVHLPIAAGGQWLV
2459
ScSFI1          SVLDSTAKKQINLESTTGLNGSPTRGK------------------------------PLR
890
                ..                                                        *

CfSFI1          LS----------------------------------------------------------
1191
SsFI1           LS----------------------------------------------------------
1158
```

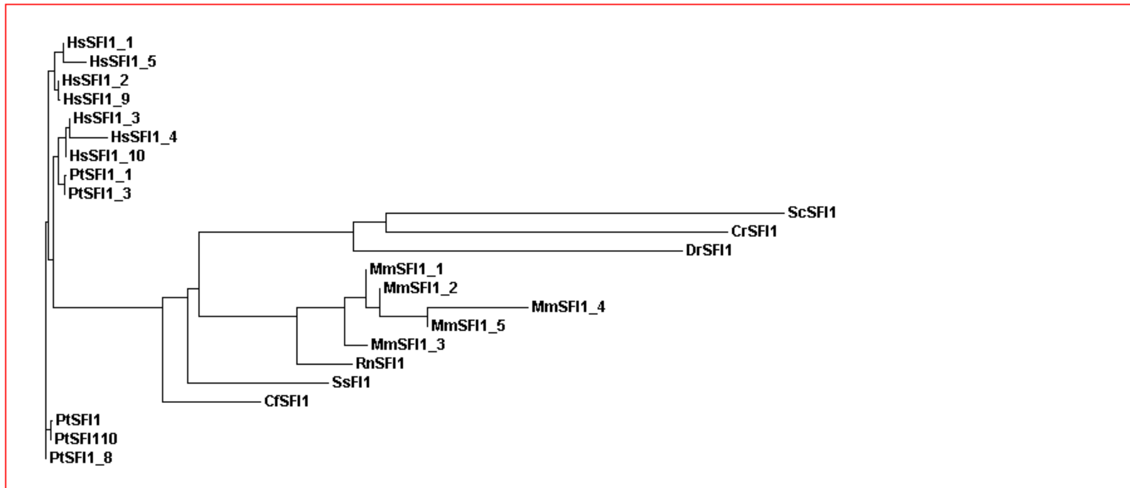
```
RnSFI1          LS----------------------------------------------------------
1154
DrSFI1          TT----------------------------------------------------------
855
CrSFI1          IGGWGWSDHWGETDVLGLLGALQRHCIQRRWPAPAYVLLAPPTTVEHQQGLGFPFGIPVP
2519
ScSFI1          YS----------------------------------------------------------
892

CfSFI1          -------------------------QEEPRPEDQEAEQQVQEELQEVELQIQQLASELQA
1226
SsFI1           -------------------------REEPRPEDQEAERQVQQELQELEMQIQQLSSELQA
1193
RnSFI1          -------------------------QEEPSCEDLHLEEQVKTELEEVELQIQQLAKELEA
1189
DrSFI1          -------------------------GTEGETDER---DSILKELTELESRISSLSMRIKK
887
CrSFI1          FDLVQTGVGVHHACSLYTNLATAEHLTAVFREMRPLPQRRLPQLVAPGVRLPVAAQVEPL
2579
ScSFI1          ------------------------PRRTTRNMPSKVDHIDFGRIPAVPFSLSANSPKIDQ
928
                                                          .:      :   :

CfSFI1          QRQPIRACIARVQALRQALC----------------------------------------
1246
SsFI1           QRQPIRACIARVQALRQTLC----------------------------------------
1213
RnSFI1          QRQPVGTCIARVRALRRALC----------------------------------------
1209
DrSFI1          QKPSMIRHAARVNTIQSQLLPSEGTTNLPTEYGI--------------------------
921
CrSFI1          PHVPVHAPGQRLAALPAGATCIPADANAPAQPLSPRDWEELLGHVPGASSGLPPSAALAA
2639
ScSFI1          DMDYIREHDKSPLSRKRQ------------------------------------------
946
                    :        :

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          LRSAPPPLALSAVLAAAVVLQRHYLTPHDTHTPPPQVLAPADAGPVLPLGGSDDAEDTEL
2699
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          ARMLIQRQPGSDLLVFATMTWGPEQQLALDTVKAEFAREGLILRAYDPSRPLILHTDWCE
2759
ScSFI1          ------------------------------------------------------------

CfSFI1          ------------------------------------------------------------
```

```
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          DGVSGVLGQLDDDAREYMVACVSRTCNVHERRYGSYKGELLAAVWAIQTLRPYLHATPFT
2819
ScSFI1          ------------------------------------------------------------


CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          LVTDHAPLEWLMSQPELTGQAARWAMILQQYSFSRHHNACRAAAAPGAALGACVTPAVRV
2879
ScSFI1          ------------------------------------------------------------


CfSFI1          ------------------------------------------------------------
SsFI1           ------------------------------------------------------------
RnSFI1          ------------------------------------------------------------
DrSFI1          ------------------------------------------------------------
CrSFI1          FAGHAHAGRRHHGTQPRPAPRLRLRTSALHGSVVGVSTFLRQHALIFAAETDAISVLEVP
2939
ScSFI1          ------------------------------------------------------------


CfSFI1          -------------------------------------------------
SsFI1           -------------------------------------------------
RnSFI1          -------------------------------------------------
DrSFI1          -------------------------------------------------
CrSFI1          GGLAPGLQACLRLGWRIRRYCNLDSSGTVTDSAYNRQAAPASVIQHPCL 2988
ScSFI1          -------------------------------------------------
```

SCORE RESULTS FOR SUS SCROFA, CANIS FAMILIARIS, RATTUS NOVERGICUS, CHLAMYDOMONAS REINHARDTII, SACCHAROMYCES CEREVISIAE AND DANIO RERIO

SeqA	Name	Length	SeqB	Name	Length	Score
1	ScSFI1	946	2	CrSFI1	2988	11.0
1	ScSFI1	946	3	CfSFI1	1246	9.0
1	ScSFI1	946	4	SsFI1	1213	9.0
1	ScSFI1	946	5	DrSFI1	921	10.0
1	ScSFI1	946	6	RnSFI1	1209	9.0
2	CrSFI1	2988	3	CfSFI1	1246	15.0
2	CrSFI1	2988	4	SsFI1	1213	17.0
2	CrSFI1	2988	5	DrSFI1	921	14.0
2	CrSFI1	2988	6	RnSFI1	1209	18.0
3	CfSFI1	1246	4	SsFI1	1213	69.0
3	CfSFI1	1246	5	DrSFI1	921	26.0
3	CfSFI1	1246	6	RnSFI1	1209	63.0
4	SsFI1	1213	5	DrSFI1	921	26.0
4	SsFI1	1213	6	RnSFI1	1209	61.0
5	DrSFI1	921	6	RnSFI1	1209	24.0

DENDOGRAM RESULTS FOR ALL ORGANISMS



HOMO SAPIENS SFI1'S CENTRIN BINDING SITES BELOW SEQUENCE ANALYSIS

Sfi1's Centrin binding site sequence name	Amino acids chain length
Bd_Site 1	27
Bd_Site 2	32
Bd_Site 3	33
Bd_Site 4	33
Bd_Site 5	27
Bd_Site 6	38
Bd_Site 7	34
Bd_Site 8	33
Bd_Site 9	48
Bd_Site 10	53
Bd_Site 11	35
Bd_Site 12	33
Bd_Site 13	33
Bd_Site 14	33
Bd_Site 15	33
Bd_Site 16	33
Bd_Site 17	33
Bd_Site 18	33
Bd_Site 19	33
Bd_Site 20	33
Bd_Site 21	25
Bd_Site 22	54
Bd_Site 23	82

AMINO ACID SEQUENCE FOR HSSFI1 CENTRIN BINDING SITES

>Bd_Site1

VFPS KARFYYEQRL LRKVFEEWKE EWW

>Bd_Site2

EWWVFQHEWK LCVRADCHYR YYLYNLMFQT WK

>Bd_Site3

TYVRQQQE MRNKYIRAEV HDAKQKMRQA WKSWL

>Bd_Site4

IYVVV RRTKLQMQTT ALEFRQRIIL RVWWSTWR

>Bd_Site5

QR LGQVRVSRAL HASALKHRAL SLQVQ

>Bd_Site6

AWSQW REQLLYVQKE KQKVVSAVKH HQHWQKRRFL KAW

>Bd_Site7

LEYLQVR RVKRQQNEMA ERFHHVTVLQ IYFCDWQ

>Bd_Site8

QAW ERRESLYAHH AQVEKLARKM ALRRAFTHWK

>Bd_Site9

HYMLLCAEEA AQFEMAEEHH RHSQLYFCFR ALKDNVTHAH LQQIRRNL

>Bd_Site10

AH QQHGVTLLHR FWNLWRSQIE QKKERELLPL LHAAWDHYRI ALLCKCIELW L

>Bd_Site11

SQIE QKKERELLPL LHAAWDHYRI ALLCKCIELW L

>Bd_Site12

QYTQKRRYK QLLQARADGH FQQRALPAAF HTWN

>Bd_Site13

RLWRWR HQENVLSARA TRFHRETLEK QVFSLWR

>Bd_Site14

QKM FQHRENRLAE RMAILHAERQ LLYRSWFMWH

>Bd_Site15

QQAAARHQEQ EWQTVACAHH RHGRLKKAFC LWR

>Bd_Site16

ESAQGLR TERTGRVRAA EFHMAQLLRW AWSQWR

>Bd_Site17

ECLA LRGAERQKLM RADLHHQHSV LHRALQAWV

>Bd_Site18

T YQGRVRSILR EVAARESQHN RQLLRGALRR WK

>Bd_Site19

ENTMARVD EAKKTFQAST HYRRTICSKV LVQWR

>Bd_Site20

EAVSV QMYYRQQEDC AIWEAQKVLD RGCLRTWF

>Bd_Site21

QRLQLERAVQ HHHRQLLLEG LARWK

>Bd_Site22

THHLQ CVRKRLLHRQ STQLLAQRLS RTCFRQWRQQ LAARRQEQRA TVRALWFWA

>Bd_Site23

F SLQAKVWATW LAFVLERRRK KARLQWALQA YQGQLLQEGA TRLLRFAASM KASRQQLQAQ
QQVQAAHSLH RAVRRCATLW K

SEQUENCE ALIGNMENT RESULTS FOR HOMO SAPIENS SFI1'S CENTRIN BINDING SITES BELOW SEQUENCE ANALYSIS

```
Bd_Site14       -------------------------------------------------QKMFQHRENRL
11
Bd_Site21       -------------------------------------------------QRLQLER----
7
Bd_Site6        -------------------------------------------AWSQWREQLLYVQKEKQ
17
Bd_Site17       -------------------------------------------------ECLALRGAERQ
11
Bd_Site4        -------------------------------------------------IYVVVRRTKLQ
11
Bd_Site16       -------------------------------------------------ESAQGLRTERT
11
Bd_Site7        ------------------------------------------------LEYLQVRRVKRQ
12
Bd_Site9        -------------------------------------------------HYMLLCAEEAA
11
Bd_Site10       -------------------------------AHQQHGVTLLHRFWNLWRSQIEQKKEREL
29
Bd_Site11       -------------------------------------------------SQIEQKKEREL
11
Bd_Site13       -------------------------------------------------RLWRWRHQENV
11
Bd_Site15       -------------------------------------------------QQAAARHQEQE
11
Bd_Site3        -------------------------------------------------TYVRQQQEMRN
11
Bd_Site12       -------------------------------------------------QYTQKRRYKQL
11
Bd_Site18       -------------------------------------------------TYQGRVRSILR
11
Bd_Site19       -------------------------------------------------ENTMARVDEAK
11
Bd_Site23       FSLQAKVWATWLAFVLERRRKKARLQWALQAYQGQLLQEGATRLLRFAASMKASRQQLQA
60
Bd_Site8        -------------------------------------------------QAWERRESLYA
11
Bd_Site1        -----------------------------------------------------------V
1
Bd_Site2        -------------------------------------------------EWWVFQH-EWK
10
Bd_Site5        -------------------------------------------------QRLGQVRVSRA
11
Bd_Site22       -------------------------------------------------THHLQCVRKRL
11
Bd_Site20       -------------------------------------------------EAVSVQMYYRQ
11
```

```
Bd_Site14       AERMAILH--AERQLLYRSWFMWH--------------------- 33
Bd_Site21       ----AVQH--HHRQLLLEGLARWK--------------------- 25
Bd_Site6        KVVSAVKH--HQHWQKRRFLKAW---------------------- 38
Bd_Site17       KLMRADLH--HQHSVLHRALQAWV--------------------- 33
Bd_Site4        MQTTALEF--RQRIILRVWWSTWR--------------------- 33
Bd_Site16       GRVRAAEF--HMAQLLRWAWSQWR--------------------- 33
Bd_Site7        QNEMAERF--HHVTVLQIYFCDWQ--------------------- 34
Bd_Site9        QFEMAEEH--HRHSQLYFCFRALKDNVTHAHLQQIRRNL------ 48
Bd_Site10       LPLLHAAWDHYRIALLCKCIELWL--------------------- 53
Bd_Site11       LPLLHAAWDHYRIALLCKCIELWL--------------------- 35
Bd_Site13       LSARATRF--HRETLEKQVFSLWR--------------------- 33
Bd_Site15       WQTVACAH--HRHGRLKKAFCLWR--------------------- 33
Bd_Site3        KYIRAEVH--DAKQKMRQAWKSWL--------------------- 33
Bd_Site12       LQARADGH--FQQRALPAAFHTWN--------------------- 33
Bd_Site18       EVAARESQ--HNRQLLRGALRRWK--------------------- 33
Bd_Site19       KTFQASTH--YRRTICSKVLVQWR--------------------- 33
Bd_Site23       QQQVQAAH--SLHRAVRRCATLWK--------------------- 82
Bd_Site8        HHAQVEKL--ARKMALRRAFTHWK--------------------- 33
Bd_Site1        FPSKARFY--YEQRLLRKVFEEWKEEWW----------------- 27
Bd_Site2        LCVRADCH--YRYYLYNLMFQTWK--------------------- 32
Bd_Site5        LHASALKH-----RALSLQVQ------------------------ 27
Bd_Site22       LHRQSTQL--LAQRLSRTCFRQWRQQLAARRQEQRATVRALWFWA 54
Bd_Site20       QEDCAIWE--AQKVLDRGCLRTWF--------------------- 33
```

SCORE RESULTS FOR HOMO SAPIENS SFI1'S CENTRIN BINDING SITES BELOW SEQUENCE ANALYSIS

SeqA	Name	Length	SeqB	Name	Length	Score
1	Bd_Site1	27	2	Bd_Site2	32	22.0
1	Bd_Site1	27	3	Bd_Site3	33	11.0
1	Bd_Site1	27	4	Bd_Site4	33	11.0
1	Bd_Site1	27	5	Bd_Site5	27	7.0
1	Bd_Site1	27	6	Bd_Site6	38	7.0
1	Bd_Site1	27	7	Bd_Site7	34	11.0
1	Bd_Site1	27	8	Bd_Site8	33	18.0
1	Bd_Site1	27	9	Bd_Site9	48	11.0
1	Bd_Site1	27	10	Bd_Site10	53	22.0
1	Bd_Site1	27	11	Bd_Site11	35	22.0
1	Bd_Site1	27	12	Bd_Site12	33	22.0
1	Bd_Site1	27	13	Bd_Site13	33	22.0
1	Bd_Site1	27	14	Bd_Site14	33	14.0
1	Bd_Site1	27	15	Bd_Site15	33	18.0
1	Bd_Site1	27	16	Bd_Site16	33	18.0
1	Bd_Site1	27	17	Bd_Site17	33	11.0
1	Bd_Site1	27	18	Bd_Site18	33	18.0
1	Bd_Site1	27	19	Bd_Site19	33	18.0
1	Bd_Site1	27	20	Bd_Site20	33	18.0
1	Bd_Site1	27	21	Bd_Site21	25	20.0
1	Bd_Site1	27	22	Bd_Site22	54	22.0
1	Bd_Site1	27	23	Bd_Site23	82	14.0
2	Bd_Site2	32	3	Bd_Site3	33	12.0
2	Bd_Site2	32	4	Bd_Site4	33	6.0
2	Bd_Site2	32	5	Bd_Site5	27	3.0

2	Bd_Site2	32	6	Bd_Site6	38	18.0
2	Bd_Site2	32	7	Bd_Site7	34	9.0
2	Bd_Site2	32	8	Bd_Site8	33	9.0
2	Bd_Site2	32	9	Bd_Site9	48	15.0
2	Bd_Site2	32	10	Bd_Site10	53	15.0
2	Bd_Site2	32	11	Bd_Site11	35	15.0
2	Bd_Site2	32	12	Bd_Site12	33	25.0
2	Bd_Site2	32	13	Bd_Site13	33	28.0
2	Bd_Site2	32	14	Bd_Site14	33	6.0
2	Bd_Site2	32	15	Bd_Site15	33	12.0
2	Bd_Site2	32	16	Bd_Site16	33	15.0
2	Bd_Site2	32	17	Bd_Site17	33	18.0
2	Bd_Site2	32	18	Bd_Site18	33	9.0
2	Bd_Site2	32	19	Bd_Site19	33	21.0
2	Bd_Site2	32	20	Bd_Site20	33	9.0
2	Bd_Site2	32	21	Bd_Site21	25	24.0
2	Bd_Site2	32	22	Bd_Site22	54	9.0
2	Bd_Site2	32	23	Bd_Site23	82	9.0
3	Bd_Site3	33	4	Bd_Site4	33	9.0
3	Bd_Site3	33	5	Bd_Site5	27	14.0
3	Bd_Site3	33	6	Bd_Site6	38	18.0
3	Bd_Site3	33	7	Bd_Site7	34	18.0
3	Bd_Site3	33	8	Bd_Site8	33	15.0
3	Bd_Site3	33	9	Bd_Site9	48	15.0
3	Bd_Site3	33	10	Bd_Site10	53	6.0
3	Bd_Site3	33	11	Bd_Site11	35	6.0

3	Bd_Site3	33	12	Bd_Site12	33	18.0
3	Bd_Site3	33	13	Bd_Site13	33	9.0
3	Bd_Site3	33	14	Bd_Site14	33	24.0
3	Bd_Site3	33	15	Bd_Site15	33	9.0
3	Bd_Site3	33	16	Bd_Site16	33	24.0
3	Bd_Site3	33	17	Bd_Site17	33	21.0
3	Bd_Site3	33	18	Bd_Site18	33	21.0
3	Bd_Site3	33	19	Bd_Site19	33	12.0
3	Bd_Site3	33	20	Bd_Site20	33	6.0
3	Bd_Site3	33	21	Bd_Site21	25	24.0
3	Bd_Site3	33	22	Bd_Site22	54	9.0
3	Bd_Site3	33	23	Bd_Site23	82	9.0
4	Bd_Site4	33	5	Bd_Site5	27	14.0
4	Bd_Site4	33	6	Bd_Site6	38	12.0
4	Bd_Site4	33	7	Bd_Site7	34	30.0
4	Bd_Site4	33	8	Bd_Site8	33	12.0
4	Bd_Site4	33	9	Bd_Site9	48	15.0
4	Bd_Site4	33	10	Bd_Site10	53	12.0
4	Bd_Site4	33	11	Bd_Site11	35	9.0
4	Bd_Site4	33	12	Bd_Site12	33	30.0
4	Bd_Site4	33	13	Bd_Site13	33	18.0
4	Bd_Site4	33	14	Bd_Site14	33	15.0
4	Bd_Site4	33	15	Bd_Site15	33	21.0
4	Bd_Site4	33	16	Bd_Site16	33	33.0
4	Bd_Site4	33	17	Bd_Site17	33	6.0

4	Bd_Site4	33	18	Bd_Site18	33	18.0
4	Bd_Site4	33	19	Bd_Site19	33	18.0
4	Bd_Site4	33	20	Bd_Site20	33	9.0
4	Bd_Site4	33	21	Bd_Site21	25	12.0
4	Bd_Site4	33	22	Bd_Site22	54	9.0
4	Bd_Site4	33	23	Bd_Site23	82	9.0
5	Bd_Site5	27	6	Bd_Site6	38	22.0
5	Bd_Site5	27	7	Bd_Site7	34	14.0
5	Bd_Site5	27	8	Bd_Site8	33	11.0
5	Bd_Site5	27	9	Bd_Site9	48	11.0
5	Bd_Site5	27	10	Bd_Site10	53	18.0
5	Bd_Site5	27	11	Bd_Site11	35	14.0
5	Bd_Site5	27	12	Bd_Site12	33	18.0
5	Bd_Site5	27	13	Bd_Site13	33	14.0
5	Bd_Site5	27	14	Bd_Site14	33	11.0
5	Bd_Site5	27	15	Bd_Site15	33	11.0
5	Bd_Site5	27	16	Bd_Site16	33	22.0
5	Bd_Site5	27	17	Bd_Site17	33	18.0
5	Bd_Site5	27	18	Bd_Site18	33	22.0
5	Bd_Site5	27	19	Bd_Site19	33	14.0
5	Bd_Site5	27	20	Bd_Site20	33	7.0
5	Bd_Site5	27	21	Bd_Site21	25	16.0
5	Bd_Site5	27	22	Bd_Site22	54	14.0
5	Bd_Site5	27	23	Bd_Site23	82	18.0
6	Bd_Site6	38	7	Bd_Site7	34	17.0
6	Bd_Site6	38	8	Bd_Site8	33	15.0

6	Bd_Site6	38	9	Bd_Site9	48	15.0
6	Bd_Site6	38	10	Bd_Site10	53	10.0
6	Bd_Site6	38	11	Bd_Site11	35	14.0
6	Bd_Site6	38	12	Bd_Site12	33	24.0
6	Bd_Site6	38	13	Bd_Site13	33	12.0
6	Bd_Site6	38	14	Bd_Site14	33	6.0
6	Bd_Site6	38	15	Bd_Site15	33	18.0
6	Bd_Site6	38	16	Bd_Site16	33	18.0
6	Bd_Site6	38	17	Bd_Site17	33	42.0
6	Bd_Site6	38	18	Bd_Site18	33	18.0
6	Bd_Site6	38	19	Bd_Site19	33	21.0
6	Bd_Site6	38	20	Bd_Site20	33	18.0
6	Bd_Site6	38	21	Bd_Site21	25	24.0
6	Bd_Site6	38	22	Bd_Site22	54	21.0
6	Bd_Site6	38	23	Bd_Site23	82	26.0
7	Bd_Site7	34	8	Bd_Site8	33	12.0
7	Bd_Site7	34	9	Bd_Site9	48	32.0
7	Bd_Site7	34	10	Bd_Site10	53	17.0
7	Bd_Site7	34	11	Bd_Site11	35	8.0
7	Bd_Site7	34	12	Bd_Site12	33	27.0
7	Bd_Site7	34	13	Bd_Site13	33	24.0
7	Bd_Site7	34	14	Bd_Site14	33	12.0
7	Bd_Site7	34	15	Bd_Site15	33	21.0
7	Bd_Site7	34	16	Bd_Site16	33	21.0
7	Bd_Site7	34	17	Bd_Site17	33	30.0
7	Bd_Site7	34	18	Bd_Site18	33	18.0

7	Bd_Site7	34	19	Bd_Site19	33	12.0
7	Bd_Site7	34	20	Bd_Site20	33	6.0
7	Bd_Site7	34	21	Bd_Site21	25	32.0
7	Bd_Site7	34	22	Bd_Site22	54	11.0
7	Bd_Site7	34	23	Bd_Site23	82	20.0
8	Bd_Site8	33	9	Bd_Site9	48	9.0
8	Bd_Site8	33	10	Bd_Site10	53	9.0
8	Bd_Site8	33	11	Bd_Site11	35	12.0
8	Bd_Site8	33	12	Bd_Site12	33	21.0
8	Bd_Site8	33	13	Bd_Site13	33	18.0
8	Bd_Site8	33	14	Bd_Site14	33	18.0
8	Bd_Site8	33	15	Bd_Site15	33	21.0
8	Bd_Site8	33	16	Bd_Site16	33	12.0
8	Bd_Site8	33	17	Bd_Site17	33	12.0
8	Bd_Site8	33	18	Bd_Site18	33	27.0
8	Bd_Site8	33	19	Bd_Site19	33	6.0
8	Bd_Site8	33	20	Bd_Site20	33	12.0
8	Bd_Site8	33	21	Bd_Site21	25	16.0
8	Bd_Site8	33	22	Bd_Site22	54	18.0
8	Bd_Site8	33	23	Bd_Site23	82	27.0
9	Bd_Site9	48	10	Bd_Site10	53	8.0
9	Bd_Site9	48	11	Bd_Site11	35	11.0
9	Bd_Site9	48	12	Bd_Site12	33	12.0
9	Bd_Site9	48	13	Bd_Site13	33	12.0
9	Bd_Site9	48	14	Bd_Site14	33	21.0
9	Bd_Site9	48	15	Bd_Site15	33	21.0

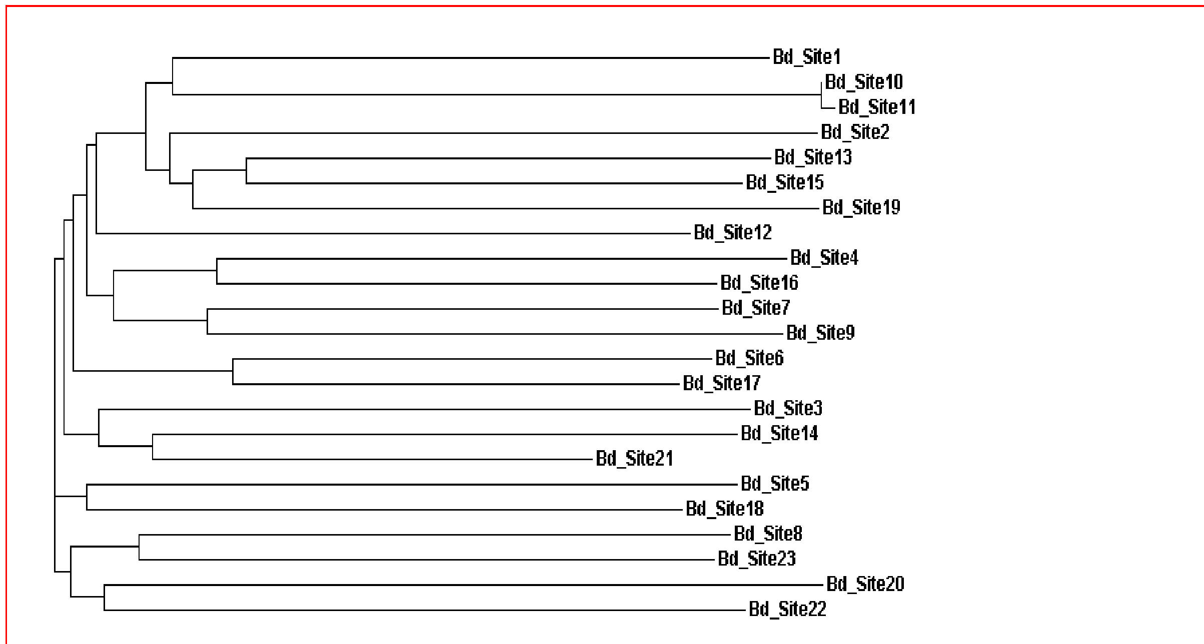
9	Bd_Site9	48	16	Bd_Site16	33	15.0
9	Bd_Site9	48	17	Bd_Site17	33	27.0
9	Bd_Site9	48	18	Bd_Site18	33	9.0
9	Bd_Site9	48	19	Bd_Site19	33	6.0
9	Bd_Site9	48	20	Bd_Site20	33	12.0
9	Bd_Site9	48	21	Bd_Site21	25	24.0
9	Bd_Site9	48	22	Bd_Site22	54	12.0
9	Bd_Site9	48	23	Bd_Site23	82	8.0
10	Bd_Site10	53	11	Bd_Site11	35	100.0
10	Bd_Site10	53	12	Bd_Site12	33	15.0
10	Bd_Site10	53	13	Bd_Site13	33	21.0
10	Bd_Site10	53	14	Bd_Site14	33	15.0
10	Bd_Site10	53	15	Bd_Site15	33	18.0
10	Bd_Site10	53	16	Bd_Site16	33	21.0
10	Bd_Site10	53	17	Bd_Site17	33	9.0
10	Bd_Site10	53	18	Bd_Site18	33	21.0
10	Bd_Site10	53	19	Bd_Site19	33	15.0
10	Bd_Site10	53	20	Bd_Site20	33	6.0
10	Bd_Site10	53	21	Bd_Site21	25	24.0
10	Bd_Site10	53	22	Bd_Site22	54	7.0
10	Bd_Site10	53	23	Bd_Site23	82	15.0
11	Bd_Site11	35	12	Bd_Site12	33	9.0
11	Bd_Site11	35	13	Bd_Site13	33	12.0
11	Bd_Site11	35	14	Bd_Site14	33	9.0
11	Bd_Site11	35	15	Bd_Site15	33	21.0
11	Bd_Site11	35	16	Bd_Site16	33	18.0

11	Bd_Site11	35	16	Bd_Site16	33	18.0
11	Bd_Site11	35	17	Bd_Site17	33	9.0
11	Bd_Site11	35	18	Bd_Site18	33	21.0
11	Bd_Site11	35	19	Bd_Site19	33	15.0
11	Bd_Site11	35	20	Bd_Site20	33	6.0
11	Bd_Site11	35	21	Bd_Site21	25	24.0
11	Bd_Site11	35	22	Bd_Site22	54	5.0
11	Bd_Site11	35	23	Bd_Site23	82	8.0
12	Bd_Site12	33	13	Bd_Site13	33	21.0
12	Bd_Site12	33	14	Bd_Site14	33	18.0
12	Bd_Site12	33	15	Bd_Site15	33	30.0
12	Bd_Site12	33	16	Bd_Site16	33	15.0
12	Bd_Site12	33	17	Bd_Site17	33	24.0
12	Bd_Site12	33	18	Bd_Site18	33	18.0
12	Bd_Site12	33	19	Bd_Site19	33	9.0
12	Bd_Site12	33	20	Bd_Site20	33	6.0
12	Bd_Site12	33	21	Bd_Site21	25	24.0
12	Bd_Site12	33	22	Bd_Site22	54	21.0
12	Bd_Site12	33	23	Bd_Site23	82	21.0
13	Bd_Site13	33	14	Bd_Site14	33	21.0
13	Bd_Site13	33	15	Bd_Site15	33	36.0
13	Bd_Site13	33	16	Bd_Site16	33	24.0
13	Bd_Site13	33	17	Bd_Site17	33	18.0
13	Bd_Site13	33	18	Bd_Site18	33	9.0
13	Bd_Site13	33	19	Bd_Site19	33	24.0
13	Bd_Site13	33	20	Bd_Site20	33	6.0

13	Bd_Site13	33	21	Bd_Site21	25	20.0
13	Bd_Site13	33	22	Bd_Site22	54	18.0
13	Bd_Site13	33	23	Bd_Site23	82	9.0
14	Bd_Site14	33	15	Bd_Site15	33	12.0
14	Bd_Site14	33	16	Bd_Site16	33	24.0
14	Bd_Site14	33	17	Bd_Site17	33	15.0
14	Bd_Site14	33	18	Bd_Site18	33	15.0
14	Bd_Site14	33	19	Bd_Site19	33	9.0
14	Bd_Site14	33	20	Bd_Site20	33	9.0
14	Bd_Site14	33	21	Bd_Site21	25	36.0
14	Bd_Site14	33	22	Bd_Site22	54	15.0
14	Bd_Site14	33	23	Bd_Site23	82	24.0
15	Bd_Site15	33	16	Bd_Site16	33	24.0
15	Bd_Site15	33	17	Bd_Site17	33	30.0
15	Bd_Site15	33	18	Bd_Site18	33	12.0
15	Bd_Site15	33	19	Bd_Site19	33	27.0
15	Bd_Site15	33	20	Bd_Site20	33	15.0
15	Bd_Site15	33	21	Bd_Site21	25	20.0
15	Bd_Site15	33	22	Bd_Site22	54	12.0
15	Bd_Site15	33	23	Bd_Site23	82	33.0
16	Bd_Site16	33	17	Bd_Site17	33	24.0
16	Bd_Site16	33	18	Bd_Site18	33	21.0
16	Bd_Site16	33	19	Bd_Site19	33	18.0
16	Bd_Site16	33	20	Bd_Site20	33	6.0
16	Bd_Site16	33	21	Bd_Site21	25	28.0
16	Bd_Site16	33	22	Bd_Site22	54	15.0

16	Bd_Site16	33	21	Bd_Site21	25	28.0
16	Bd_Site16	33	22	Bd_Site22	54	15.0
16	Bd_Site16	33	23	Bd_Site23	82	6.0
17	Bd_Site17	33	18	Bd_Site18	33	15.0
17	Bd_Site17	33	19	Bd_Site19	33	21.0
17	Bd_Site17	33	20	Bd_Site20	33	6.0
17	Bd_Site17	33	21	Bd_Site21	25	36.0
17	Bd_Site17	33	22	Bd_Site22	54	18.0
17	Bd_Site17	33	23	Bd_Site23	82	24.0
18	Bd_Site18	33	19	Bd_Site19	33	9.0
18	Bd_Site18	33	20	Bd_Site20	33	18.0
18	Bd_Site18	33	21	Bd_Site21	25	36.0
18	Bd_Site18	33	22	Bd_Site22	54	12.0
18	Bd_Site18	33	23	Bd_Site23	82	15.0
19	Bd_Site19	33	20	Bd_Site20	33	6.0
19	Bd_Site19	33	21	Bd_Site21	25	20.0
19	Bd_Site19	33	22	Bd_Site22	54	12.0
19	Bd_Site19	33	23	Bd_Site23	82	9.0
20	Bd_Site20	33	21	Bd_Site21	25	24.0
20	Bd_Site20	33	22	Bd_Site22	54	15.0
20	Bd_Site20	33	23	Bd_Site23	82	9.0
21	Bd_Site21	25	22	Bd_Site22	54	24.0
21	Bd_Site21	25	23	Bd_Site23	82	36.0
22	Bd_Site22	54	23	Bd_Site23	82	20.0

DENDOGRAM RESULTS FOR HSSFI1 CENTRIN BINDING SITES



SUMMARY OF RESULTS

Sequence alignment results for centrin binding sites suggested sequence identity for most of their sequences, but there were low score results (3-36). Dendogram analysis suggested strong evolutionary relations between all the binding sites, but especially marked between binding sites 10 and 11. Sequence alignment results cannot be analyzed by all organisms in ClustalW, so they were divided into two groups. The first group included *Homo sapiens (human)*, *Mus musculus (mouse)* and *Pan troglodytes* (chimpance). In general, the alignment suggested the presence of sequence identity, with a few conserved substitutions at 21 sites, involving positive charge and non-polar amino acids. Non conservative substitutions were found at 15 sites mostly of polar and non-polar amino acids. *Homo sapiens* and *Pan troglodytes* sequences were more similar between them, when compared with *Mus musculus* which presented amino acid substitutions. The score reports where higher between isoforms of the same organism: *Homo sapiens* (93-100), *Pan troglodytes* (97-100) and *Mus musculus* (89-100). In fact, scores reflected the similarity obtained before by sequence alignment, were *Homo sapiens* and *Pan troglodytes* have scores between 95 and 100. *Mus musculus* compared with the two organisms presented the lowest scores (13-66) being more related to *Pan troglodytes*. The dendogram for these organisms showed that the most evolutionary related are *Homo sapiens* and *Pan troglodytes*, were *Mus musculus* is more distant form them. These results makes sense, because it is of popular knowledge that humans are more related to simians than to rodents, but in this project it was confirmed at the molecular level, not exclusively by phenotypes.

The second group of organisms analyzed included *Sus scrofa* (pig), *Canis familiaris* (dog), *Rattus novergicus* (rat), *Chlamydomonas reinhardtii* (algae), *Saccharomyces cerevisiae (yeast)* and *Danio rerio* (fish). In this particular analysis the presence of a lot of gaps at the beginning and at the end of the alignment suggested more divergence between these organisms. In spite of that, there were several conserved substitutions at 54 sites in positive and aromatic amino acids and non-conserved sequences at 54 sites too, but between several types of amino acids. Score results

showed only a limited analysis were most similarity were presented by *Canis familiaris* with *Sus scrofa* (69), *Canis familiaris* with *Rattus novergicus* (63) and *Sus scrofa* with *Rattus novergicus* (61). For the rest of organisms scores were lower (9-26) and less similars. ClustalW cannot established an dendogram for this second group of organisms, so a dendogram containing all the organisms were performed and like the analysis of the first group of organisms, it showed *Homo sapiens* and *Pan troglodytes* as the more evolutionary rand then *Canis familiaris* with *Sus scrofa*. *Rattus novergicus* was more related to *Mus musculus* and *Saccharomyces cerevisiae* with *Chlamydomonas reinhardtii*. Interestingly, *Danio rerio*, was related to *Saccharomyces cerevisiae* and *Chlamydomonas reinhardtii*, that were more evolutionary distant form the rest of the organisms.

CONCLUSIONS

ClustalW permitted the analysis of Sfi1 protein between diverse organisms. The alignment suggested sequence homology for most part of each sequence, and it is true for those with similar length and for very closes organisms. In fact sequence homology is particularly related to similar protein functions. Scores showed how similar or different organisms are by giving a number, that indicates sequence identity. These results were confirmed by dendograms where organisms can be classified in distant or more evolutionary related.

REFERENCES

1) Kilmartin, J.V. Sfi1 has conserved centrin-binding sites and an essential function in budding yeast spindle body duplication. J. Cell Biol. **2003,** *162,* 1211-1221.

2) Salisbury, J. L. Centrosomes: Sfi1p and unravel a structural riddle.Curr. Biol. **2004,** *14,* 27-29.

3) http://www.**uniprot**.org/

4) http://www.ebi.ac.uk/clustalw/index.html?

BIOGRAPHICAL SKETCH

Publication

- Del Valle-Sosa, L.; Alfaro, E.; Santiago, J.; Narváez, D.; Rosado, M.C.; Rodríguez, A.; Gómez, A.M.; Schreiter, E.R.; Pastrana-Ríos, Belinda. The structure, molecular dynamics, and energetics of centrin-melittin complex. Proteins. **2011,** *79,* 3132-3143.

Presentations

- Poster Presentation. XVI Sigma Xi Symposium. Mayagüez, Puerto Rico. April, 12, 2011.
- Poster Presentation. The Lilly Academy Technical Forum. Carolina, Puerto Rico. April, 15, 2011.
- Oral Presentation and Poster Presentation. IFPAC/PAT Summer Summit. San Juan, Puerto Rico. June 12-13, 2012.
- Poster Presentation. The Lilly Academy Technical Forum. Carolina, Puerto Rico. May, 3, 2013.

Internship

- COOP Program. PR05 Plant Lilly del Caribe. Carolina, Puerto Rico. August 2012-May 2013.