University of Puerto Rico

Rio Piedras Campus

Department of Biology

Transcript, Structural, and Histologic Characterization of a Novel EF-Hand

Protein and its Isoform Specific to the Ambulacraria Clade

By

Arisnel Soto Acabá

A dissertation submitted to the Biology Intercampus doctoral program in partial

fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

February 2020

Table of Contents

List of Tables

List of Figures

List of Supplementary Data

List of Abbreviations

CaBP	Calcium-binding protein
CDD	Conserved domain database
cDNA	Complementary DNA
CV	Column volume
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleoside triphosphate
dpe	Days post evisceration
DTT	Dithiothreitol
ECM	Extracellular matrix
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid
EMP	Exponential mega priming PCR
EST	Expressed sequence tag
***E*-value**	Expected value
FT	Flow-through
Gal	Galactosidase
GG	Glycylglycine
Glc	Glucose
GRAVY	Grand Average of Hydropathicity
GST	Glutathione *S*-transferase
His	Histidine
HSD	Honestly significant difference

IB	Inclusion body
IPTG	Isopropyl b-D-1-thiogalactopyranoside
LB	Luria Bertani
MAFFT	Multiple Alignment using Fast Fourier Transform
NADH	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
NC	Nitrocellulose
NI	Normal intestine
Ni-NTA	Nickel-nitrilotriacetic acid
NM	Normal mesentery
OB	Orpin B
OBS	Orpin B, signal less
OBSm	Orpin B, signal less, and mutated
OD	Optical density
ORF	Open reading frame
Orpin	*Orpin* gene or RNA transcript
Orpin	*Orpin* encoding protein or translated amino acid sequence
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
RF	Restriction free
RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Reverse transcriptase PCR
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography

SLS	Spindle-like structures
Sp	Supernatant
SPARC	Secreted, acidic, and rich in cysteine
sq-RT-PCR	semi-quantitative RT-PCR
TB	Terrific Broth
Tm	Melting temperature
UTR	untranslated region
UV	Ultraviolet

This dissertation has been accepted by the faculty of the

Biology Intercampus Doctoral Program

University of Puerto Rico

Rio Piedras Campus and Medical Sciences Campus

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In the Subject of BIOLOGY

_____

José E. García Arrarás, Ph.D.

Thesis Advisor

_____

Abel Baerga, Ph.D.

Committee Member

_____

Irvin Vega, Ph.D.

Committee Member

_____

José Rodríguez Medina, Ph.D.

Committee Member

_____

Carlos González, Ph.D.

Committee Member

Abstract

Transcriptomic databases have become one of the main sources for protein discovery. In our studies of transcripts from normal echinoderms, we have identified several transcripts that have attracted our attention. One of these is a previously unidentified transcript (*Orpin*) that appeared to be upregulated during intestinal regeneration. In Chapter 2, using bioinformatics tools we: (1) identified a second *Orpin* sequence (2) describe their motifs and domains, and perform phylogenetic analyses that suggest that Orpins might comprise a novel subfamily of EF-hand containing proteins specific to the Ambulacraria clade. Semi-quantitative RT-PCR analyses revealed that *Orpin* mRNAs are expressed in various tissues but no significant differential expression was found in regenerating tissues.

In Chapter 3 we expressed and purified genetically modified versions of Orpin to further characterize this EF-hand protein. We developed two protocols:

The first protocol consisted in the production of His-OBSm, which is His tagged with two additional genetic modifications: the deletion of the signal peptide encoding sequence and the addition of a 25 residues peptide from the pET200 plasmid vector. These modifications made possible the purification of a soluble recombinant version of Orpin B.

The second protocol was developed to produce a soluble recombinant Orpin that best resembled the original protein. This protocol consisted of the expression of a GST-Orpin lacking the signal peptide region. For this protocol, we tested different parameters that affect protein expression, including, additives, host cells, growth media and supplementation, and protein extraction methods. The best parameters were identified and used to obtain a soluble Orpin form. We propose that the developed strategies can be used to increase the soluble expression and purification of other EF-hand proteins that are difficult to express by standard methods.

In chapter 4 we produced antibodies against His-OBSm and used these to identify the cells expressing the protein in *H. glaberrima* tissues. Antibody specificity was tested through Western Blots.

Finally, we investigated the effect of the Wnt signaling pathway on *Orpin* mRNA expression in regenerating gut explants and found an apparent decrease expression of an *Orpin B* isoform following Wnt/B-catenin pathway activation.

Transcript, Structural, and Histologic Characterization of a Novel EF-Hand

Protein and its Isoform Specific to the Ambulacraria Clade

By

Arisnel Soto Acabá

Acknowlegments

I want to express my gratitude and appreciation to all of the people that enabled this research to be possible.

First of all, I wish to express my sincere appreciation to my supervisor, Dr. José García Arrarás, who has the substance of a true scientist and mentor: he convincingly guided and encouraged me to be professional and do the right thing even when the road got tough. Without his persistent help, the goal of this project would not have been realized. Our meetings and conversations were vital in inspiring me to think outside the box, from multiple perspectives to form a comprehensive and objective critique.

I would like to express my gratitude and appreciation to Dr. Abel Baerga, for providing access to his lab in RCM and MSRC, without all of that people, instrumentation, equipment, and materials, the research would not have been possible. Furthermore, I am extremely grateful to Vilmarie Mercado, Dr. Carlos Rullán, Dr. Ramón Gómez, and all of Dr. Baerga's laboratory group for our friendly and scientific conversations. Your support and contributions have been invaluable throughout this study.

I would like to say big thank you for Dr. Carlos González and all of his research group members for their energy, understanding and help throughout my project, especially to Dr. Josean González and Dr. Dina Paola Bracho for the help and support throughout RE-free cloning, Western Blots, Mass Spectroscopy, and insightful ideas.

I wish to acknowledge and thank Dr. Irving Vega and Dr. José Rodríguez Medina for their overall insights in the protein biochemistry and molecular biology fields and knowledge that have made this an inspiring experience for me.

Furthermore, I would like to recognize the invaluable assistance that Dr. Vanessa Torres provided on the immunohistochemistry experiments and Dr. Samir Bello during his ex vivo cell culture experiments that provided me the opportunity to discover new things on the Development Laboratory.

I also wish to show my gratitude to Dr. Adolfo Plazaola for inspiration, considerate guidance and believe in me. It is whole-heartedly appreciated that his great advice for my studies proved significant towards the success of this goal and professional development. I would also like to thank Dr. Victor Febo for providing advice, guidance, and support on statistical analyses.

I would like to express my sincere gratitude to the RISE program, for research funding and for letting me share my results and ideas through travels, meetings, and conferences that help me connect and collaborate with the scientific community worldwide.

And my biggest thanks for the unconditional support and great love of my family, my mother, Nelsa Acabá; my father, Baltazar Soto; and my brother, Lcdo. Zarel Soto. They kept me going on and this work and life goal would not have been possible without their encouragement and comprehension. As well, I wish to express my deepest gratitude to María del Mar for her unconditional support and encouragement through the hardest times, which kept me motivated until the culmination of the Ph.D. I also greatly appreciate the unconditional love and companionship of my loyal dog, Dalí who was also there with me.

Chapter 1 Background Information

1.1 Prologue

1.1.1 The use of genomics and transcriptomics for the identification of proteins of significant interest for humans.

The elucidation and construction of the genomic and transcriptomic libraries facilitated the identification of proteins of significant interest for humanity. One of the earlier examples of these proteins is green fluorescent protein (GFP), which emits green fluorescence light upon exposure to ultraviolet or blue light. The gene encoding GFP was identified in transcriptomic and genomic cDNA libraries of the hydromedusan jellyfish, *Aequorea victoria* (Prasher, Eckenrode, Ward, Prendergast, & Cormier, 1992) and was successfully cloned and expressed independently by Inouye and Tsuji's group (Inouye & Tsuji, 1994) and Chalfie's group (Chalfie, Tu, Euskirchen, Ward, & Prasher, 1994) almost simultaneously. This facilitated enhancements and modifications to the recombinant version of GFP (rGFP) that led to the development of applications as a molecular tracer. The use of rGFP to study gene expression, identifying and localizing proteins and organelles (Delagrave, Hawtin, Silva, Yang, & Youvan, 1995; Rizzuto, Brini, Pizzo, Murgia, & Pozzan, 1995; Webb, Decatur, Teleman, & Losick, 1995) and as a reporter for detecting viral infection (Baulcombe, Chapman, & Cruz, 1995; Wu, Zou, Koothan, & Cline, 1995) are only a few early examples of the vast number of applications that brought the GFP revolution.

Since then, genomic, transcriptomic, and proteomic technologies advances have paved the way for the identification and discovery of genes or proteins with

significant impact in a wide range of scientific fields. Among the high-throughput technologies, DNA analysis by microarrays (Alhamdani, Schröder, & Hoheisel, 2009) and Nextgen sequencing (NGS) (Baker, 2012; Ulahannan, Kovac, Mulholland, Cazier, & Tomlinson, 2013) have become powerful discovery and identification tools. The oncology biomedical field have immensely taken advantage of these approaches to develop more accurate cancer diagnostic tools and therapies based on the discovery of novel biomarkers. NGS analysis have been used to identify GATA3 gene (encoded transcription factor) and MLL3 gene (encoded histone trimethyltransferase), as molecular targets for detection and treatment therapies for ER$^+$ breast cancer (Banerji et al., 2012; Curtis et al., 2012; Ellis et al., 2012; Koboldt et al., 2012; Stephens et al., 2012). In other aspects, the frequent amplification of *cyclin D1* detected in luminal-type breast cancer were shown in The Cancer Genome Atlas (TCGA) Network breast cancer study (Koboldt et al., 2012). These results coupled to other several studies indicate that CDK4 and CDK6 associated to cell-cycle activation represent valuable therapeutic targets of ER$^+$ advanced breast cancer (Harbour, Luo, Santi, Postigo, & Dean, 1999; Leung & Potter, 1987; Sherr, 1996; Sherr & Roberts, 1999; van den Heuvel & Harlow, 1993; Weinberg, 1995).

Curiosity-driven science coupled to high-throughput technology has also led to the identification of novel proteins with many interesting applications. Integrating RNA-seq (Metzker, 2010) with proteomics, biophysical and mechanical characterization tools has accelerated the biomimetic engineering research area.

A study of molluscan tissues reported that identification of novel proteins combined with genetic engineering generated a wide range of high-performance materials including self-healing bioelastomeric membranes, natural adhesives and enzymes involved in their bioprocessing, and robust protein-based polymers (Guerette et al., 2013).

The first case correspond to studies of the tropical marine snail *Pugilina cochlidium,* related to *Busycon canaliculus,* whose egg cases have unique shock-absorbing properties and long-range elasticity (Miserez & Guerette, 2013; Miserez, Wasko, Carpenter, & Waite, 2009). The group generated a transcriptome database from the egg case-secreting nidamental gland and identified two proteins, PECP-1 and PECP-2, with strong homology to the egg case sequences of *B. canaliculus*. The characterization of these energy-absorbing proteins including sequence design, membrane assembly and novel cross-linking mechanisms are of particular interest for the biomimetic engineering of protein-based membranes with precisely tuned mechanical properties (Miserez & Guerette, 2013).

The second studied organism was the green mussel (*Perna viridis*), which firmly attaches to underwater structures, through the production of a multiprotein, fiber-plaque assembly. The key component enabling the adherence to a variety of different surfaces is the post-translational residue dihydroxphenylalanine (DOPA) that arises from tyrosine hydroxylation (Lee, Messersmith, Israelachvili, & Waite, 2011). After RNA-seq from green mussel foot coupled to mass spectroscopy, the

investigators identified the thread's self-healing properties protein preColD, and the adhesive proteins PVFP-3, PVFP5, and PVFP-6 (Guerette et al., 2013). In addition, they identified five transcripts encoding proteins with homology to tyranosinases that convert tyrosine into adhesive DOPA. Mussel foot protein tyrosinases could be used in the engineering of biomimetic high-performance adhesives from green approaches.

The group also investigated the molecular design of the jumbo squid (*Dosidicus gigas*) sucker ring teeth (SRT) from the tentacles used for predatory attacks. Through the generation of a transcriptome library of the sucker tissue surrounding the SRT they identified a structural protein, named suckerin-39 among other homologs. The suckerin proteins exhibited unusual thermoplastic behavior after processed into nanopatterned surfaces and fibers associated to its silk-like protein architecture held by weak interactions (Guerette et al., 2013). Even further, they demonstrated that cross-linked films of the recombinant version of (rec)suckerin-39 exhibited higher modulus (stiffness) and hardness than many engineered polymers (Zok & Miserez, 2007) and a predicted wear resistance superior to that of synthetic structural polymers. The potential for other medical research applications of this protein was demonstrated by showing that human mesenchymal cells (hMSC) grown on (rec)suckerin-39 films displayed high cell viability for all conditions (biocompatibility and cell growth support).

Taken together, these results demonstrated that RNA-seq provided significant insights that could be translated into the processing of structural and functional eco-friendly biomaterials. Advances in high-throughput sequencing technologies have facilitated the identification of novel proteins with a broad range of applications. The knowledge achieved through the integration of high-throughput sequence to other science fields offer opportunities of innovation that continues to improve human life in many aspects.

In our case, we focused on proteins associated with regenerative processes. For this we used a novel model system, the sea cucumber *Holothuria glaberrima*. Although as usually occurs in the scientific process, our results differed from those expected, the research took us across previously unsuspected paths. Nonetheless, in the introduction to this thesis I will explain the rationale that led to the initial experiments and will keep the focus on the regenerative process that we aimed to target and the model system that provided the source of the putative proteins of interest.

1.2 Project Description

About 60 to 70 million people are affected by digestive diseases each year in the USA alone, and approximately 8 million die from them worldwide (Everhart & Ruhl, 2009; Peery et al., 2015, 2012). Disrupted cellular mechanisms for homeostatic control and deficient repair mechanisms are the main culprits. Repair mechanisms play an important role in healing after natural events that damage the digestive tract arise and after gastrointestinal surgery. Accordingly, the ability to regenerate intestinal tissue could be an important therapeutic tool. However, it is astonishing how little is known about the underpinnings of regeneration. This is mostly because the regenerative capacity of vertebrates is limited. Among deuterostomes, the sea cucumber is one of the closest relative to vertebrates that has the capacity for digestive tract regeneration. *Holothuria glaberrima*, a sea cucumber that has received extensive attention as a model, contains digestive tissues and structures that closely resemble those found in vertebrates. In addition, these animals possess the innate ability to regenerate their entire digestive tract after it is self-eviscerated. Thus, this model is attractive for studying the molecular basis of regeneration.

Although we are still far from understanding the cellular mechanisms that underlie digestive tract regeneration in the sea cucumber, molecular analyses have been initiated. Based on an existing expressed sequence tag (EST) database, we found that several genes are overexpressed at 3 and 7 days post evisceration (dpe). One of these genes codes for a novel putative calcium-binding

protein (CaBP), named Orpin. Unlike most calcium-binding proteins, Orpin contains a putative signal peptide and is predicted to be secreted (Ortiz-Pineda et al., 2009). *Orpin* was found to be overexpressed during the early stages of regeneration. The increase in *Orpin* transcript levels is concurrent with the appearance of spindle-like structures (SLS), a characteristic of myocyte dedifferentiation (García-Arrarás et al., 2011). In addition, *Orpin* overexpression occurs at the same time as extracellular matrix remodeling in the regenerating structure (García-Arrarás et al., 2011; Mashanov & García-Arrarás, 2011). This work focuses on understanding the function of Orpin with the aim of characterizing the protein structure and expression profile to determine its role during the regeneration process.

1.2.1 Aim 1: Characterize *Orpin* isoform transcript and amino acid encoding sequences, and their expression in tissues from normal and regeneration animals.

I have now found an isoform to the original Orpin sequence from two transcript databases generated by high throughput sequencing. This new sequence shares a 93.2% identity with Orpin. Both putative proteins have a predicted calcium-binding domain composed of two EF-hand motifs.

We experimentally verified *Orpin* sequences, including untranslated regions (UTRs), to compare the structural relationship between the two isoforms. We now hypothesize that there are two *Orpin* isoforms and that they share similar

spatial/temporal expression patterns. To test this we 1) performed a bioinformatics comparison of the putative calcium-binding domains and additional differences in primary and secondary structures, 2) built an expression profile of both *Orpin* isoforms during the first 10 days of regeneration, 3) determined the tissue specificity of the corresponding transcripts sequences, and 4) determine their phylogenetic relationship to other known EF-Hand proteins.

1.2.2 Aim 2: Expression and purification of a recombinant version of Orpin.

Previous studies show that Orpin mRNAs encode two putative EF-hand motifs based on the helix-loop-helix topology and amino acid conservation, a characteristic of calcium-binding proteins. In addition, these sequences encode a 20 amino acid signal peptide characteristic of secreted proteins. Several EF-Hand proteins are targeted outside the cell. A recombinant version of Orpin is needed to enable structural studies that will facilitate the future determination of calcium-binding activity and secretion. Accordingly, we hypothesized that a recombinant version of Orpin can be expressed and purified in a soluble form through genetic engineering and determination of proper expression conditions. To test this we 1) expressed and purified various recombinant protein variants and 2) determine how calcium ions affect its structural organization and solubility. The development of different strategies to clone, express, and purify the recombinant Orpin will provide useful insights that can facilitate the soluble expression and purification of other EF-Hand proteins that are difficult to express. The recombinant version of Orpin will facilitate future studies for the characterization of Orpin structural features and will pave the way into the elucidation of its physiological role.

1.2.3 Aim 3: Determine the Orpin protein-expressing cells and insights to its possible association with other known molecular pathways in *H. glaberrima*.

Determining which tissues and cells express Orpin proteins will help to gain insights into their function, particularly when presented in the context of the available cellular and histological knowledge from previous studies in *Holothuria glaberrima.* The Wnt signaling pathway, which is associated to regenerative processes such as cell proliferation and stem cell maintenance and differentiation (Clevers & Nusse, 2012; van der Flier & Clevers, 2009), has been thoroughly studied in this sea cucumber. A non-canonical β-catenin-independent Wnt pathway, Wnt/PCP, is associated to calcium release. Thus, we wanted to study the effect of *Orpin* isoforms expression during regulatory processes of these pathways. To test this we induced the canonical Wnt pathway with the agonist LiCl in tissue explants from the regenerating intestines of *H. glaberrima.* In order to study the effect of *Orpin* variants expression during activation of the Wnt pathway, we 1) extracted RNA from cultured regenerating tissues and 2) measure expression levels of *Orpin* variants through semi-quantitative RT-PCR. In other aspects, we wanted to identify Orpin protein in tissues from *H. glaberrima* to determine tissue localization and enable future characterization studies of Orpin physiological role. Thus, we hypothesized that a recombinant version of Orpin can be used as an antigen to produce antibodies that are able to identify Orpin expressing cells in the sea cucumber. To test this, we 1) produced α-Orpin

antibody sera and 2) determine tissues and cells expressing Orpin through immunohistochemistry.

1.3 Background and Significance

1.3.1 Regeneration mechanisms

Regeneration is a series of coordinated events through which a tissue, organ or body part is replaced after it has been injured or lost. There are two known mechanisms for regeneration, morphallaxis, and epimorphosis. Morphallactic regeneration is mediated by the redistribution and remodeling of existing cells to replace a functional body part (Sánchez Alvarado, 2000, 2004, 2006). Although some cell division occurs, organ replacement is principally made by the rearrangement and differentiation of remaining tissues. Unlike morphallaxis, epimorphic regeneration relies on the formation of a blastema by the proliferation of new cells (Carlson, 2007).

There are many animals with some capacity for epimorphic regeneration, although their capabilities vary dramatically. Urodeles, such as newts, are capable of regenerating their lens, retina, jaws, heart, limbs, and tails (Echeverri, Clarke, & Tanaka, 2001; S Ghosh, Thorogood, & Ferretti, 1994; Maden & Wallace, 1975; Oberpriller & Oberpriller, 1974; Tsonis & Del Rio-Tsonis, 2004; Wallace, Maden, & Wallace, 1974). Zebrafish can regenerate their fins, optic nerves, heart, lateral line, axons of the CNS and parts of the cerebellum (C. G. Becker & Becker, 2007; T. Becker, Wullimann, Becker, Bernhardt, & Schachner, 1998; Dufourcq et al.,

2006; Hata, Namae, & Nishina, 2007; Kästner & Wolburg, 1982; Liu, Azodi, Kerstetter, & Wilson, 2004; Nakatani, Kawakami, & Kudo, 2007). Compared to invertebrates, however, vertebrates' regenerative capacities are limited. Planarians are able to completely regenerate every tissue and organ (Saló et al., 2009; Sánchez Alvarado, 2006). Hydra is capable of regenerating a complete organism from a small fraction of its body (Gierer et al., 1972; Lenhoff & Lenhoff, 1986; Noda, 1971). Echinoderms can rapidly regenerate most body parts. Within the echinoderms, Holothurians exhibit the ability to regenerate their viscera completely after its ejection (Díaz-Miranda, Blanco, & García-Arrarás, 1995; García-Arrarás et al., 1999, 1998; Quiñones, Rosa, Ruiz, & García-Arrarás, 2002).

1.3.2 *Holothuria glaberrima* as a model

The sea cucumber *Holothuria glaberrima* have proven to be an excellent model for intestinal regeneration research (Díaz-Miranda et al., 1995; García-Arrarás et al., 1999, 1998; Quiñones et al., 2002). Phylogenetically, it is a deuterostome, which are animals that are closely related to vertebrates. Also, its digestive tract morphology and cellular composition resembles its vertebrate counterparts. This echinoderm can regenerate its entire intestine along with other internal organs after the natural process of autotomy. This process of evisceration occurs naturally but also can be chemically induced in the laboratory (García-Arrarás et al., 1998). The entire digestive tract regeneration takes about five weeks to complete, thus providing tissue samples relatively fast. Additionally, the organism's maintenance is easy and cost-efficient.

Our laboratory has been studying the cellular and molecular mechanisms underlying intestinal regeneration in *H. glaberrima* for over a decade. A recently constructed differential gene expression profile from regenerating tissues has provided insights into the molecular mechanisms related to the process (Ortiz-Pineda et al., 2009).

1.3.3 Cellular mechanisms of digestive tract regeneration in *Holothuria glaberrima*

The cellular and morphological events that occur during digestive tract regeneration in H. glaberrima have been extensively studied (García-Arrarás et al., 1998; García-Arrarás & Greenberg, 2001). Following evisceration, the animal proceeds to regenerate its lost internal organs. The newly regenerated intestine starts to form from surrounding mesentery tissue where the previous organ was attached (Fig. 1.1). Initial studies showed that several biological processes take place during digestive tract regeneration. These include cell proliferation (García-Arrarás et al., 1998), cell migration (García-Arrarás et al., 2006), extracellular matrix remodeling (Quiñones et al., 2002), and cell dedifferentiation (Candelaria, Murray, File, & García-Arrarás, 2006). The regeneration initiates as a thickening of the tip of the mesentery where the wound is healing. This blastema-like structure starts to grow in size and in the number of cells during the first week. A rod-like structure begins to form and extends from the cloaca to the esophagus. At the end of the second week, a lumen is noticeable and initial differentiation is evident since various tissue layers can be observed. During the last weeks, the luminal

epithelium is formed by cell migration from both ends of the regenerating intestine, leaving a functional intestine that is smaller but proportionally similar to the original organ.
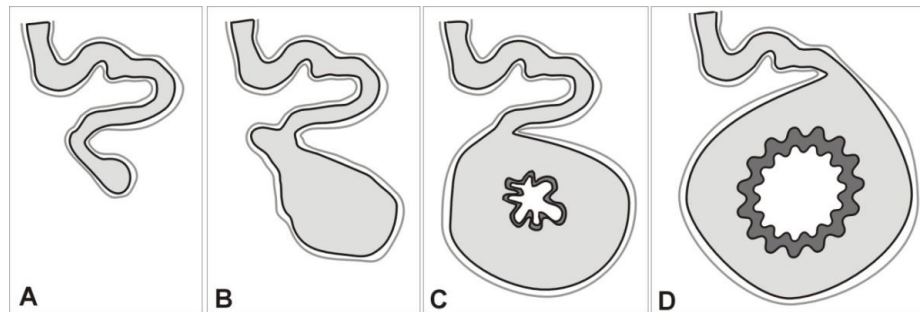
Figure 1.1. Diagram of the regeneration stages in the sea cucumber intestine
Wound healing stage (3 dpe) where the future intestine primordium is initially observed at the tip of the remaining mesentery where the wound has healed (A). First week of regeneration (7dpe) where the intestinal rudiment forms by a thickening of the mesentery (B). In the second week (14dpe), the lumen has already formed and the basic layout of the intestine is established (C). By the third week (21dpe), the different tissue layers grow and the intestine enlarges (D).

1.3.4 Cellular dedifferentiation during intestinal regeneration in *Holothuria glaberrima*

Dedifferentiation is a regulated biological process where cells undergo a regression from a specialized differentiated state to a precursor state reminiscent of stem cells. This process is observed during regeneration of different tissues and organs in various organisms with different regenerative capacities. Collectively, these events resemble a reverse developmental process at different levels.

During the initial stages of tissue healing and reorganization of the mesentery, extensive cellular dedifferentiation of the mesothelium starts to take place at the distal free margin (Candelaria et al., 2006; Mashanov, Dolmatov, & Heinzeller,

2005). Dedifferentiated mesothelium undergoes a drastic simplification in its organization, involving the formation of a simple epithelial layer of irregularly shaped cells (Mashanov & García-Arrarás, 2011). These morphological changes occur concomitantly with changes in the cellular cytoskeleton. This process has been mostly studied in the regeneration of the holothurian body wall muscles. It is characterized by the condensation of filaments into SLS which are then discarded to the extracellular milieu (Dolmatov & Ginanova, 2001; García-Arrarás & Dolmatov, 2010; Mashanov et al., 2005; San Miguel-Ruiz & García-Arrarás, 2007). The SLS formation described during the regeneration of the body wall is also seen during the initial formation of the blastema-like structure. Muscle dedifferentiation initiates from the free end of the mesentery as soon as 24 hours after evisceration and peaks at 5 dpe with a subsequent gradual decrease (Mashanov & García-Arrarás, 2011). A loss of organization and posterior disappearance of muscle fibers is connected to the appearance of SLSs. Therefore, SLSs have been used as dedifferentiation markers.

Muscle dedifferentiation has also been observed in other regenerative processes, including those that take place in vertebrates such as urodeles. Additional studies revealed that muscle fiber dedifferentiation plays a key role in the formation of the blastema-like structure (Echeverri et al., 2001; Kumar, Velloso, Imokawa, & Brockes, 2000; Lo, Allen, & Brockes, 1993). These biological processes are mediated through the action of orchestrated molecular interactions. There is evidence that supports the role of several extracellular signals triggering

cellular dedifferentiation (Echeverri & Tanaka, 2002). Such is the case of SPARC, a CaBP that is directly implicated in the reorganization and remodeling of ECM through interaction with collagen fibers (Sage, Vernon, Funk, Everitt, & Angello, 1989; Q. Yan & Sage, 1999).

1.3.5 Molecular studies of regeneration

Our laboratory has investigated molecular events occurring during intestinal regeneration. The use of gene-by-gene strategies, high throughput sequencing and microarrays has made possible the identification of multiple genes associated with the regenerative processes of the intestine. Such is the case of survivin and mortalin genes (Mashanov, Zueva, Rojas-Catagena, & Garcia-Arraras, 2010). In addition, homologs to Calbindin-D32k and parvalbumin were identified using other molecular techniques such as bioinformatics, Western Blot, immunohistochemistry, and molecular cloning (Díaz-Balzac, Lázaro-Peña, García-Rivera, González, & García-Arrarás, 2012). Currently, there are two available libraries of transcripts that were used in the identification of multiple overexpressed genes associated with the regeneration of the intestine and nerve.

1.3.6 Calcium and calcium-binding proteins role during regeneration

Calcium is an essential signaling molecule that controls physiological processes such as neurotransmission, fertilization, smooth muscle cell contraction and differentiation (Blundon, Wright, Brodwick, & Bittner, 1995; Eisen & Reynolds, 1985; Hovav, Parnas, & Parnas, 1992). The variety of functions mediated by

calcium is mirrored by the diversity of calcium-binding proteins, key elements to most cell biological processes. CaBPs with EF-hand domains have been used as markers for various biological events such as proliferation, differentiation, and regeneration. The following information provides four examples of the various roles of CaBPs that are related to regenerative processes: (1) Parvalbumin has been associated with motoneuron survival and cell differentiation after nerve injury (Dekkers et al., 2004). (2) Oncomodulins have been implicated in axonal regeneration as signal molecules derived from macrophages (Yin et al., 2006). (3) SPARC (Secreted, Acidic, and Rich in Cysteine) proteins have been linked to extracellular matrix remodeling (ECM), cell turnover, and repair (Sage et al., 1989; Q. Yan & Sage, 1999). Finally, (4) calmodulin has been shown to affect nerve cell growth in the sciatic nerve of frogs (Remgård, Ekström, Wiklund, & Edström, 1995). It is interesting to notice that the few CaBPs found in the ECM are associated with similar cellular events occurring during digestive tract regeneration. If Orpin is proven to be influencing the ECM it might also be associated with the ECM remodeling events that take place during regeneration.

1.3.7 *Orpin*

ESTs coding for the *Orpin* gene were initially found in cDNA libraries from regenerating intestine (Rojas-Cartagena et al., 2007). Moreover, in gene expression microarray experiments, *Orpin* ESTs were found to be overexpressed with high significance at 3 dpe and with lower significance at 7 dpe (Ortiz-Pineda

et al., 2009). At the end of the second week of regeneration, there was no significant overexpression of *Orpin* transcripts.

The microarray results were validated using a semi-quantitative RT-PCR normalized against the quantification of the housekeeping gene, *NADH dehydrogenase subunit 5*. The evidence supported the significant increase of *Orpin* transcript levels in early intestine regeneration stages. This up-regulation peaked on the third day after evisceration. The over-expression of *Orpin* progressively decreased until the second week of regeneration in which no differential expression was detected when compared against normal tissues (Fig. 1.2).
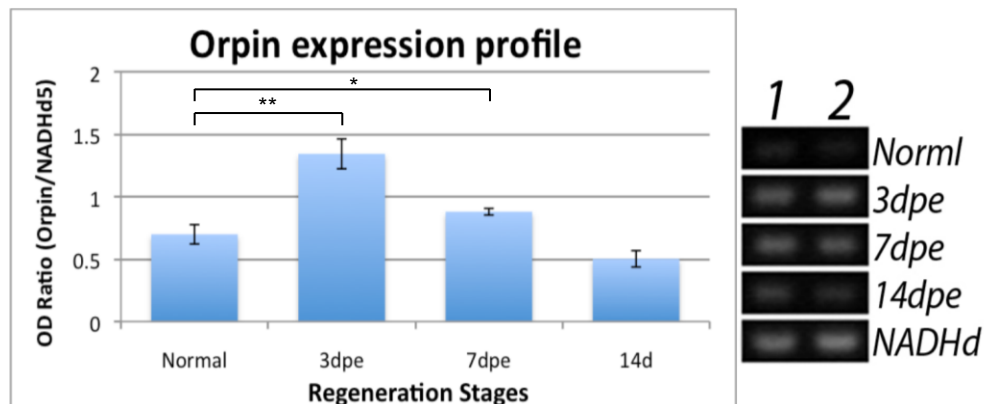
Figure 1.2. RT-PCR validation of the expression profile of the *Orpin* gene
Bars correspond to averages of the normalized values for each stage plus the standard error of the means (left). Statistical significance was considered under the null hypothesis that no difference is expected between normal averages and the 3 regeneration time points tested (3, 7, and 14 dpe). Statistical significance was found when comparing 3 dpe vs. normal averages (t-test: $P < .001$). At 7 dpe significance was lower than 3 dpe but still overexpressed vs. normal ($P = .04$). No difference was noticed at 14dpe ($P = .08$). Lanes 1 and 2 (right) represent two different samples from normal, 3, 7, and 14dpe. Samples from *NADH dehydrogenase* are also shown. Significance threshold: * = $P < .05$; ** = $P < .01$ (Adapted with permission from Ortíz-Pineda, 2010).

Orpin has an ORF region consisting of 366 bp and encoding a 122 amino acid protein. After a search for homology through BLAST, it shows no significant similarity with any available sequence from the database. Later on, two additional sequences that exhibited high similarity with Orpin were added to the database, both from the hemichordate *Saccoglossus kowalevskii*. These showed 50% and 51% identity shared when compared to Orpin. Bioinformatics analyses were performed in order to retrieve relevant information about protein composition, properties, and putative function. A transmembrane domain was found to be present using TMHMM (3.0) server. In addition, the sequence showed a high probability to be targeted outside the cell when analyzed through Neural Networks and Hidden Markov Models. The SignalIP server was used to determine whether the sequence is post-translationally modified. This algorithm predicted that Orpin has a signal peptide in the first 20 amino acids along with a cleavage site in the amino acids 20-21. This feature implies that the protein is secreted.

In addition, Orpin contains a calcium-binding domain composed of two EF-hands. The topology of these EF-hands varies very little between them but conserve the key amino acids of the canonical domain structure. The EF-hand domains are the hallmark of every calcium-binding proteins (CaBPs). Several EF-Hand subfamilies consist of less than four EF-hands like the thoroughly known S100s. The low containing EF-hand proteins usually act as calcium signal transducers, rather than function as calcium modulators or buffers such as the higher number in EF-Hand motifs such as parvalbumin. They do so by a

conformational change upon metal-binding in order to interact with their corresponding partners (Grabarek, 2006; Ikura, 1996; Ikura & Ames, 2006; Kretsinger, Tolbert, Nakayama, & Pearson, 1991; Yanyi, Shenghui, Yubin, & Jie, 2010). Furthermore, several are secreted to the outside of the cell as some S100 (Bianchi, Kastrisianaki, Giambanco, & Donato, 2011; R Donato et al., 2013; Rosario Donato, 1999) and oncomodulin (Yin et al., 2006). The amino acid composition of these domains determines its metal-binding potential. However, there are a number of EF-hand-containing proteins which activation is independent and does not bind calcium ions. If it were experimentally proven that Orpin actually binds calcium ions, this novel peptide would be classified as a new type of secreted calcium-binding protein.

1.4 Significance

Cellular dedifferentiation might be used in the future to enhance the regenerative capacity of vertebrates. This may allow the replacement of entire body structures and the repair of injured organs. From a scientific perspective, it is important to identify the genes and proteins that regulate cellular dedifferentiation in order to increase our understanding of the molecular machinery that guides vertebrate regeneration. From a medical perspective, identification of these molecules could lead to therapies aimed at improving the regenerative processes in vertebrates.

If no association with the cellular processes discussed before, was found whatsoever, the characterization of this novel EF-Hand proteins would still be relevant to the metal-binding protein chemistry field. The fact that this protein presented no homology to other known EF-Hand CaBPs, its few EF-hand domains, its particular motif architecture, and its apparent secretion, could be the result of a new EF-Hand CaBP subfamily. The amino acid composition of its EF-hand domains, as well as its location in the peptide sequence, could shed some light on the evolution of these motifs. Thus, the present investigation proves to be very interesting to studies of evolution of EF-hand proteins

1.5 References

Alhamdani, M. S., Schröder, C., & Hoheisel, J. D. (2009). Oncoproteomic profiling with antibody microarrays. *Genome Medicine*, *1*(7), 68. https://doi.org/10.1186/gm68

Baker, M. (2012). Digital PCR hits its stride. *Nature Methods*, *9*(6), 541–544. https://doi.org/10.1038/nmeth.2027

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., … Meyerson, M. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, *486*(7403), 405–409. https://doi.org/10.1038/nature11154

Baulcombe, D. C., Chapman, S., & Cruz, S. (1995). Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant Journal*, *7*(6), 1045–1053. https://doi.org/10.1046/j.1365-313X.1995.07061045.x

Becker, C. G., & Becker, T. (2007). Growth and pathfinding of regenerating axons in the optic projection of adult fish. *Journal of Neuroscience Research*, *85*(12), 2793–2799. https://doi.org/10.1002/jnr.21121

Becker, T., Wullimann, M. F., Becker, C. G., Bernhardt, R. R., & Schachner, M. (1998). Axonal regrowth after spinal cord transection in adult zebrafish. *The Journal of Comparative Neurology*, *377*(4), 577–595.

Bianchi, R., Kastrisianaki, E., Giambanco, I., & Donato, R. (2011). S100B protein stimulates microglia migration via rage-dependent up-regulation of chemokine expression and release. *Journal of Biological Chemistry*, *286*(9), 7214–7226. https://doi.org/10.1074/jbc.M110.169342

Blundon, J. A., Wright, S. N., Brodwick, M. S., & Bittner, G. D. (1995). Presynaptic calcium-activated potassium channels and calcium channels at a crayfish neuromuscular junction. *Journal of Neurophysiology*, *73*(1), 178–189. https://doi.org/10.1152/jn.1995.73.1.178

Candelaria, A. G., Murray, G., File, S. K., & García-Arrarás, J. E. (2006). Contribution of mesenterial muscle dedifferentiation to intestine regeneration in the sea cucumber Holothuria glaberrima. *Cell and Tissue Research*, *325*(1), 55–65. https://doi.org/10.1007/s00441-006-0170-z

Carlson, B. M. (2007). *Principles of Regenerative Biology*. Elsevier. https://doi.org/10.1016/B978-0-12-369439-3.X5000-4

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., & Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. *Science, 263*(5148), 802–805. https://doi.org/10.1126/science.8303295

Clevers, H., & Nusse, R. (2012). Wnt/β-Catenin Signaling and Disease. *Cell*, *149*(6), 1192–1205. https://doi.org/10.1016/j.cell.2012.05.012

Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., … Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), 346–352. https://doi.org/10.1038/nature10983

Dekkers, J., Bayley, P., Dick, J. R ., Schwaller, B., Berchtold, M ., & Greensmith, L. (2004). Over-expression of parvalbumin in transgenic mice rescues motoneurons from injury-induced cell death. *Neuroscience*, *123*(2), 459–466. https://doi.org/10.1016/j.neuroscience.2003.07.013

Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M., & Youvan, D. C. (1995). Red-Shifted Excitation Mutants of the Green Fluorescent Protein. *Nature Biotechnology*, *13*(2), 151–154. https://doi.org/10.1038/nbt0295-151

Díaz-Balzac, C. A., Lázaro-Peña, M. I., García-Rivera, E. M., González, C. I., & García-Arrarás, J. E. (2012). Calbindin-D32k is localized to a subpopulation of neurons in the nervous system of the sea cucumber Holothuria glaberrima (Echinodermata). *PloS One*, *7*(3), e32689. https://doi.org/10.1371/journal.pone.0032689

Díaz-Miranda, L., Blanco, R. E., & García-Arrarás, J. E. (1995). Localization of the heptapeptide GFSKLYFamide in the sea cucumber holothuria glaberrima (echinodermata): A light and electron microscopic study. *Journal of Comparative Neurology*, *352*(4), 626–640. https://doi.org/10.1002/cne.903520410

Dolmatov, I. Y., & Ginanova, T. T. (2001). Muscle regeneration in holothurians. *Microscopy Research and Technique*, *55*(6), 452–463. https://doi.org/10.1002/jemt.1190

Donato, R, Cannon, B. R., Sorci, G., Riuzzi, F., Hsu, K., Weber, D. J., & Geczy, C. L. (2013). Functions of S100 proteins. *Current Molecular Medicine*, *13*(1), 24–57. https://doi.org/10.2174/1566524011307010024

Donato, Rosario. (1999). Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1450*(3), 191–231. https://doi.org/10.1016/S0167-4889(99)00058-0

Dufourcq, P., Roussigné, M., Blader, P., Rosa, F., Peyrieras, N., & Vriz, S. (2006). Mechano-sensory organ regeneration in adults: The zebrafish lateral line as a model. *Molecular and Cellular Neuroscience*, *33*(2), 180–187. https://doi.org/10.1016/j.mcn.2006.07.005

Echeverri, K., Clarke, J. D. W., & Tanaka, E. M. (2001). In Vivo Imaging Indicates Muscle Fiber Dedifferentiation Is a Major Contributor to the Regenerating Tail Blastema. *Developmental Biology*, *236*(1), 151–164. https://doi.org/10.1006/dbio.2001.0312

Echeverri, K., & Tanaka, E. M. (2002). Mechanisms of muscle dedifferentiation during regeneration. *Seminars in Cell & Developmental Biology*, *13*(5), 353–360. https://doi.org/10.1016/S1084952102000915

Eisen, A., & Reynolds, G. T. (1985). Source and sinks for the calcium released during fertilization of single sea urchin eggs. *The Journal of Cell Biology*, *100*(5), 1522–1527. https://doi.org/10.1083/jcb.100.5.1522

Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W., … Mardis, E. R. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, *486*(7403), 353–360. https://doi.org/10.1038/nature11143

Everhart, J. E., & Ruhl, C. E. (2009). Burden of Digestive Diseases in the United States Part I: Overall and Upper Gastrointestinal Diseases. *Gastroenterology*, *136*(2), 376–386. https://doi.org/10.1053/j.gastro.2008.12.015

García-Arrarás, J. E., Diaz-Miranda, L., Torres, I. I., File, S., Jiménez, L. B., Rivera-Bermudez, K., … Cruz, W. (1999). Regeneration of the enteric nervous system in the sea cucumberHolothuria glaberrima. *The Journal of Comparative Neurology*, *406*(4), 461–475. https://doi.org/10.1002/(SICI)1096-9861(19990419)406:4<461::AID-CNE4>3.0.CO;2-0

García-Arrarás, J. E., & Dolmatov, I. Y. (2010). Echinoderms; potential model systems for studies on muscle regeneration, *16*(8), 942–955.

García-Arrarás, J. E., Estrada-Rodgers, L., Santiago, R., Torres, I. I., Díaz-Miranda, L., & Torres-Avillán, I. (1998). Cellular mechanisms of intestine regeneration in the sea cucumber, Holothuria glaberrima Selenka (Holothuroidea:Echinodermata). *The Journal of Experimental Zoology*, *281*(4), 288–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9658592

García-Arrarás, J. E., & Greenberg, M. J. (2001). Visceral Regeneration in Holothurians. *Microscopy Research and Technique*, *55*(March), 438–451. https://doi.org/10.1002/jemt.1189

García-Arrarás, J. E., Schenk, C., Rodrígues-Ramírez, R., Torres, I. I., Valentín, G., & Candelaria, A. G. (2006). Spherulocytes in the echinoderm Holothuria glaberrima and their involvement in intestinal regeneration. *Developmental Dynamics*, *235*(12), 3259–3267. https://doi.org/10.1002/dvdy.20983

García-Arrarás, J. E., Valentín-Tirado, G., Flores, J. E., Rosa, R. J., Rivera-Cruz, A., San Miguel-Ruiz, J. E., & Tossas, K. (2011). Cell dedifferentiation and epithelial to mesenchymal transitions during intestinal regeneration in H. glaberrima. *BMC Developmental Biology*, *11*(1), 61. https://doi.org/10.1186/1471-213X-11-61

Ghosh, S., Thorogood, P., & Ferretti, P. (1994). Regenerative capability of upper and lower jaws in the newt. *The International Journal of Developmental Biology*, *38*(3), 479–490. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7531480

Gierer, A., Berking, S., Bode, H., David, C. N., Flick, K., HANSMANN, G., … Trenkner, E. (1972). Regeneration of Hydra from Reaggregated Cells. *Nature New Biology*, *239*(91), 98–101. https://doi.org/10.1038/newbio239098a0

Grabarek, Z. (2006). Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. *Journal of Molecular Biology*, *359*(3), 509–525. https://doi.org/10.1016/j.jmb.2006.03.066

Guerette, P. A., Hoon, S., Seow, Y., Raida, M., Masic, A., Wong, F. T., … Miserez, A. (2013). Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. *Nature Biotechnology*, *31*(10), 908–915. https://doi.org/10.1038/nbt.2671

Harbour, J. W., Luo, R. X., Santi, A. D., Postigo, A. A., & Dean, D. C. (1999). Cdk Phosphorylation Triggers Sequential Intramolecular Interactions that Progressively Block Rb Functions as Cells Move through G1. *Cell*, *98*(6), 859–869. https://doi.org/10.1016/S0092-8674(00)81519-6

Hata, S., Namae, M., & Nishina, H. (2007). Liver development and regeneration: From laboratory study to clinical therapy. *Development, Growth & Differentiation*, *49*(2), 163–170. https://doi.org/10.1111/j.1440-169X.2007.00910.x

Hovav, G., Parnas, H., & Parnas, I. (1992). Neurotransmitter release: Facilitation and three-dimensional diffusion of intracellular calcium. *Bulletin of Mathematical Biology*, *54*(5), 875–894. https://doi.org/10.1007/BF02459934

Ikura, M. (1996). Calcium binding and conformational response in EF-hand proteins. *Trends in Biochemical Sciences*, *21*(1), 14–17. https://doi.org/10.1016/0968-0004(96)80879-6

Ikura, M., & Ames, J. B. (2006). Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: Two ways to promote multifunctionality. *Proceedings of the National Academy of Sciences*, *103*(5), 1159–1164. https://doi.org/10.1073/pnas.0508640103

Inouye, S., & Tsuji, F. I. (1994). Aequorea green fluorescent protein. Expression of the gene and fluorescent characteristics of the recombinant protein. *FEBS Letters*, *341*(2–3), 277–280. https://doi.org/10.1016/0014-5793(94)80472-9

Kästner, R., & Wolburg, H. (1982). Functional Regeneration of the Visual System in Teleosts. Comparative Investigations after Optic Nerve Crush and Damage of the Retina. *Zeitschrift Für Naturforschung C*, *37*(11–12), 1274–1280. https://doi.org/10.1515/znc-1982-11-1229

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., & McMichael, J. F. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Kretsinger, R. H., Tolbert, D., Nakayama, S., & Pearson, W. (1991). The EF-Hand, Homologs and Analogs. *Novel Calcium-Binding Proteins*, 17–37. https://doi.org/10.1007/978-3-642-76150-8_3

Kumar, A., Velloso, C. P., Imokawa, Y., & Brockes, J. P. (2000). Plasticity of Retrovirus-Labelled Myotubes in the Newt Limb Regeneration Blastema. *Developmental Biology*, *218*(2), 125–136. https://doi.org/10.1006/dbio.1999.9569

Lee, B. P., Messersmith, P. B., Israelachvili, J. N., & Waite, J. H. (2011). Mussel-Inspired Adhesives and Coatings. *Annual Review of Materials Research*, *41*(1), 99–132. https://doi.org/10.1146/annurev-matsci-062910-100429

Lenhoff, H. M., & Lenhoff, S. G. (1986). *Hydra and the Birth of Experimental Biology, 1744: Abraham Trembley's Memoires Concerning the Polyps*. Pacific Grove, California: Boxwood Pr.

Leung, B. S., & Potter, A. H. (1987). Mode of Estrogen Action on Cell Proliferative Kinetics in CAMA-1 Cells. I. Effect of Serum and Estrogen. *Cancer Investigation*, *5*(3), 187–194. https://doi.org/10.3109/07357908709011735

Liu, Q., Azodi, E., Kerstetter, A. ., & Wilson, A. . (2004). Cadherin-2 and cadherin-4 in developing, adult and regenerating zebrafish cerebellum. *Developmental Brain Research*, *150*(1), 63–71. https://doi.org/10.1016/j.devbrainres.2004.03.002

Lo, D. C., Allen, F., & Brockes, J. P. (1993). Reversal of muscle differentiation during urodele limb regeneration. *Proceedings of the National Academy of Sciences*, *90*(15), 7230–7234. https://doi.org/10.1073/pnas.90.15.7230

Maden, M., & Wallace, H. (1975). The origin of limb regenerates from cartilage grafts. *Acta Embryologiae Experimentalis*, (2), 77–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1232722

Mashanov, V. S., Dolmatov, I. Y., & Heinzeller, T. (2005). Transdifferentiation in Holothurian Gut Regeneration. *The Biological Bulletin*, *209*(3), 184–193. https://doi.org/10.2307/3593108

Mashanov, V. S., & García-Arrarás, J. E. (2011). Gut regeneration in holothurians: A snapshot of recent developments. *Biological Bulletin*, *221*(1), 93–109.

Mashanov, V. S., Zueva, O. R., Rojas-Catagena, C., & Garcia-Arraras, J. E. (2010). Visceral regeneration in a sea cucumber involves extensive expression of survivin and mortalin homologs in the mesothelium. *BMC Developmental Biology*, *10*(1), 117. https://doi.org/10.1186/1471-213X-10-117

Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. https://doi.org/10.1038/nrg2626

Miserez, A., & Guerette, P. A. (2013). Phase transition-induced elasticity of α-helical bioelastomeric fibres and networks. *Chem. Soc. Rev.*, *42*(5), 1973–1995. https://doi.org/10.1039/C2CS35294J

Miserez, A., Wasko, S. S., Carpenter, C. F., & Waite, J. H. (2009). Non-entropic and reversible long-range deformation of an encapsulating bioelastomer. *Nature Materials*, *8*(11), 910–916. https://doi.org/10.1038/nmat2547

Nakatani, Y., Kawakami, A., & Kudo, A. (2007). Cellular and molecular processes of regeneration, with special emphasis on fish fins. *Development, Growth & Differentiation*, *49*(2), 145–154. https://doi.org/10.1111/j.1440-169X.2007.00917.x

Noda, K. (1971). Reconstitution of hydra from dissociated cells. *The Zoological Society of Japan*, *80*(3), 99–101. Retrieved from https://dl.ndl.go.jp/info:ndljp/pid/10844940

Oberpriller, J. O., & Oberpriller, J. C. (1974). Response of the adult newt ventricle to injury. *Journal of Experimental Zoology*, *187*(2), 249–259. https://doi.org/10.1002/jez.1401870208

Ortiz-Pineda, P. A., Ramírez-Gómez, F., Pérez-Ortiz, J., González-Díaz, S., Santiago-De Jesús, F., Hernández-Pasos, J., … García-Arrarás, J. E. (2009). Gene expression profiling of intestinal regeneration in the sea cucumber. *BMC Genomics*, *10*, 262. https://doi.org/10.1186/1471-2164-10-262

Ortiz-Pineda, P. A. (2010). Analysis and characterization of genes associated with intestinal regeneration in the sea cucumber (Echinodermata: Holothuroidea) (UMI No. 3412547) (Doctoral dissertation, University of Puerto Rico Río Piedras Campus, San Juan, Puerto Rico). Available from ProQuest Dissertations Publishing. https://search.proquest.com/docview/743816791

Peery, A. F., Crockett, S. D., Barritt, A. S., Dellon, E. S., Eluri, S., Gangarosa, L. M., … Sandler, R. S. (2015). Burden of Gastrointestinal, Liver, and Pancreatic Diseases in the United States. *Gastroenterology*, *149*(7), 1731-1741.e3. https://doi.org/10.1053/j.gastro.2015.08.045

Peery, A. F., Dellon, E. S., Lund, J., Crockett, S. D., McGowan, C. E., Bulsiewicz, W. J., … Shaheen, N. J. (2012). Burden of Gastrointestinal Disease in the United States: 2012 Update. *Gastroenterology*, *143*(5), 1179-1187.e3. https://doi.org/10.1053/j.gastro.2012.08.002

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, *111*(2), 229–233. https://doi.org/10.1016/0378-1119(92)90691-H

Quiñones, J. L., Rosa, R., Ruiz, D. L., & García-Arrarás, J. E. (2002). Extracellular Matrix Remodeling and Metalloproteinase Involvement during Intestine Regeneration in the Sea Cucumber Holothuria glaberrima. *Developmental Biology*, *250*(1), 181–197. https://doi.org/10.1006/dbio.2002.0778

Remgård, P., Ekström, P. A. R., Wiklund, P., & Edström, A. (1995). Calmodulin and In Vitro Regenerating Frog Sciatic Nerves: Release and Extracellular Effects. *European Journal of Neuroscience*, *7*(6), 1386–1392. https://doi.org/10.1111/j.1460-9568.1995.tb01130.x

Rizzuto, R., Brini, M., Pizzo, P., Murgia, M., & Pozzan, T. (1995). Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells. *Current Biology*, *5*(6), 635–642. https://doi.org/10.1016/S0960-9822(95)00128-X

Rojas-Cartagena, C., Ortíz-Pineda, P., Ramírez-Gómez, F., Suárez-Castillo, E. C., Matos-Cruz, V., Rodríguez, C., … García-Arrarás, J. E. (2007). Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber Holothuria glaberrima. *Physiological Genomics*, *31*(2), 203–215. https://doi.org/10.1152/physiolgenomics.00228.2006

Sage, H., Vernon, R. B., Funk, S. E., Everitt, E. A., & Angello, J. (1989). SPARC, a secreted protein associated with cellular proliferation, inhibits cell spreading in vitro and exhibits Ca+2-dependent binding to the extracellular matrix. *The Journal of Cell Biology*, *109*(1), 341–356. https://doi.org/10.1083/jcb.109.1.341

Saló, E., Abril, J. F., Adell, T., Cebria, F., Eckelt, K., Fernandez-Taboada, E., … Rodríguez-Esteban, G. (2009). Planarian regeneration: achievements and future directions after 20 years of research. *The International Journal of Developmental Biology*, *53*(8-9–10), 1317–1327. https://doi.org/10.1387/ijdb.072414es

San Miguel-Ruiz, J. E., & García-Arrarás, J. E. (2007). Common cellular events occur during wound healing and organ regeneration in the sea cucumber Holothuria glaberrima. *BMC Developmental Biology*, *7*(1), 115. https://doi.org/10.1186/1471-213X-7-115

Sánchez Alvarado, A. (2000). Regeneration in the metazoans: Why does it happen? *BioEssays*, *22*(6), 578–590. https://doi.org/10.1002/(SICI)1521-1878(200006)22:6<578::AID-BIES11>3.0.CO;2-#

Sánchez Alvarado, A. (2004). Regeneration and the need for simpler model organisms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1445), 759–763. https://doi.org/10.1098/rstb.2004.1465

Sánchez Alvarado, A. (2006). Planarian Regeneration: Its End Is Its Beginning. *Cell*, *124*(2), 241–245. https://doi.org/10.1016/j.cell.2006.01.012

Sherr, C. J. (1996). Cancer Cell Cycles. *Science*, *274*(5293), 1672–1677. https://doi.org/10.1126/science.274.5293.1672

Sherr, C. J., & Roberts, J. M. (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes & Development*, *13*(12), 1501–1512. https://doi.org/10.1101/gad.13.12.1501

Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., … Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, *486*(7403), 400–404. https://doi.org/10.1038/nature11017

Tsonis, P. A., & Del Rio-Tsonis, K. (2004). Lens and retina regeneration: transdifferentiation, stem cells and clinical applications. *Experimental Eye Research*, *78*(2), 161–172. https://doi.org/10.1016/j.exer.2003.10.022

Ulahannan, D., Kovac, M. B., Mulholland, P. J., Cazier, J.-B., & Tomlinson, I. (2013). Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *British Journal of Cancer, 109*(4), 827–835. https://doi.org/10.1038/bjc.2013.416

van den Heuvel, S., & Harlow, E. (1993). Distinct roles for cyclin-dependent kinases in cell cycle control. *Science*, *262*(5142), 2050–2054. https://doi.org/10.1126/science.8266103

van der Flier, L. G., & Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annual Review of Physiology*, *71*, 241–260. https://doi.org/10.1146/annurev.physiol.010908.163145

Wallace, H., Maden, M., & Wallace, B. M. (1974). Participation of cartilage grafts in amphibian limb regeneration. *Journal of Embryology and Experimental Morphology, 32*(2), 391–404. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4463210

Webb, C. D., Decatur, A., Teleman, A., & Losick, R. (1995). Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in Bacillus subtilis. *Journal of Bacteriology*, *177*(20), 5906–5911. https://doi.org/10.1128/JB.177.20.5906-5911.1995

Weinberg, R. A. (1995). The retinoblastoma protein and cell cycle control. *Cell*, *81*(3), 323–330. https://doi.org/10.1016/0092-8674(95)90385-2

Wu, G.-Y., Zou, D.-J., Koothan, T., & Cline, H. T. (1995). Infection of frog neurons with vaccinia virus permits in vivo expression of foreign proteins. *Neuron*, *14*(4), 681–684. https://doi.org/10.1016/0896-6273(95)90211-2

Yan, Q., & Sage, E. H. (1999). SPARC, a Matricellular Glycoprotein with Important Biological Functions. *Journal of Histochemistry & Cytochemistry*, *47*(12), 1495–1505. https://doi.org/10.1177/002215549904701201

Yanyi, C., Shenghui, X., Yubin, Z., & Jie, Y. J. (2010). Calciomics: prediction and analysis of EF-hand calcium binding proteins by protein engineering. *Science China. Chemistry*, *53*(1), 52–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2926812&tool=pmcentrez&rendertype=abstract

Yin, Y., Henzl, M. T., Lorber, B., Nakazawa, T., Thomas, T. T., Jiang, F., … Benowitz, L. I. (2006). Oncomodulin is a macrophage-derived signal for axon regeneration in retinal ganglion cells. *Nature Neuroscience*, *9*(May), 843–852. https://doi.org/10.1038/nn1701

Zok, F. W., & Miserez, A. (2007). Property maps for abrasion resistance of materials. *Acta Materialia*, *55*(18), 6365–6371. https://doi.org/10.1016/j.actamat.2007.07.042

Chapter 2 Characterization of Two Novel EF-Hand Proteins Identifies a Clade of Putative Ca^{2+}-Binding Protein Specific to the Ambulacraria

2.1 Abstract

In recent years, transcriptomic databases have become one of the main sources for protein discovery. In our studies of nervous system and digestive tract regeneration in echinoderms, we have identified several transcripts that have attracted our attention. One of these molecules corresponds to a previously unidentified transcript (*Orpin*) from the sea cucumber *Holothuria glaberrima* that appeared to be upregulated during intestinal regeneration. We have now identified a second highly similar sequence and analyzed the predicted proteins using bioinformatics tools. Both sequences have EF-hand motifs characteristic of calcium-binding proteins (CaBPs) and N-terminal signal peptides. Sequence comparison analyses such as multiple sequence alignments and phylogenetic analyses only showed significant similarity to sequences from other echinoderms or from hemichordates. Semi-quantitative RT-PCR analyses revealed that transcripts from these sequences are expressed in various tissues including muscle, haemal system, gonads, and mesentery. However, contrary to previous reports, there was no significant differential expression in regenerating tissues. Nonetheless, the identification of unique features in the predicted proteins suggests that these might comprise a novel subfamily of EF-hand containing proteins specific to the Ambulacraria clade.

2.2 Introduction

Modern genome and transcriptome studies allow for the identification and discovery of hitherto unknown sequences that code for different types of proteins. This discovery process has been possible due to the ease by which DNA and/or RNA sequences are obtained, even from non-model organisms that make available millions of sequences for comparative analyses. Our group has focused on transcriptomes obtained from normal and regenerating tissues of an echinoderm, the sea cucumber *Holothuria glaberrima* (García-Arrarás et al., 1998; García-Arrarás & Greenberg, 2001; Rojas-Cartagena et al., 2007; San Miguel-Ruiz & García-Arrarás, 2007). Studies in this model have been done to explore gene expression of intestinal and nervous systems in an attempt to expand our knowledge of the Echinodermata, a phylum which lies on the evolutionary branch of chordates (Mashanov, Zueva, & García-Arrarás, 2013; Mashanov, Zueva, & Garcia-Arraras, 2010).

In this effort, we have constructed several transcriptomic libraries using high throughput sequence analyses, including EST (expressed sequence tag) analyses (Rojas-Cartagena et al., 2007), 454 and Illumina sequencing (Mashanov, Zueva, & García-Arrarás, 2014; Ortiz-Pineda et al., 2009). Moreover, we have performed differential gene expression studies, particularly microarrays and transcriptomic comparisons between normal and regenerating tissues. The results from these experiments have been a large number of differentially expressed genes associated with the regenerating tissues.

Out of these hundreds of genes, we have focused on the study of unknown sequences that show increased expression during regeneration. One of these molecules corresponds to a previously unidentified transcript from *H. glaberrima* that was shown to be upregulated during the initial stages of intestine regeneration by microarray analyses (Ortiz-Pineda et al., 2009; Rojas-Cartagena et al., 2007). The sequence was annotated to public databases as *Orpin* (GU191018.1, ACZ73832.1) on 12-13-2009.

We now provide a full report on the putative Orpin sequence including the prediction of an N-terminal signal peptide, which is characteristic of secreted proteins. Moreover, we have discovered an additional Orpin isoform in *H. glaberrima* and provide a full description of both Orpin isoforms. Both sequences are newly discovered putative EF-hand coding proteins with structural characteristics that are evolutionarily related to this group of proteins. We have also probed other available databases and have found previously undescribed sequences whose similarities suggest they are part of the Orpin family, a protein family that appears to be restricted to the Ambulacraria clade.

2.3 Materials and Methods

2.3.1 Animals

Adult specimens (10–15 cm in length) of the sea cucumber *H. glaberrima* were collected in coastal areas of northeastern Puerto Rico and kept in indoor in aerated

seawater aquaria at room temperature (RT: 22°C ± 2°C). Evisceration was induced by 0.35 M KCl injections (3–5 mL) into the coelomic cavity (García-Arrarás et al., 1998). Eviscerated animals were let to regenerate for 3, 5, 7, 10, and 14 days before the dissection and tissue extraction. For the dissection, organisms were anesthetized by placement in ice-cold water for 1 h. (García-Arrarás et al., 1998; Rojas-Cartagena et al., 2007; Suárez-Castillo & García-Arrarás, 2007). Dissected tissues were rinsed in ice-cold filtered seawater and processed for RNA isolation.

2.3.2 RNA extraction and cDNA synthesis

RNA extraction was performed on tissue extracts of normal and 3, 5, 7, 10, and 14 dpe animals. Extracted tissues included gonads, mesentery, haemal system, respiratory tree, longitudinal muscle, and radial nerve cords. After dissection, tissues were placed in 1 mL of TRIzol reagent (Invitrogen), homogenized with a PowerGen Model 125 Homogenizer (Thermo Scientific) and incubated 30 min on ice. These samples were mixed vigorously with 200 μL of chloroform and incubated 10 min at RT. After centrifuged at 12,000 rpm at 4°C, the aqueous RNA phase was separated, mixed with 70% ethanol, and transferred to an RNeasy Mini Kit column (QIAGEN) for deoxyribonuclease (DNase) treatment (QIAGEN). Total RNA was extracted following the manufacturer's protocol. The concentration and purity of the total RNA was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The cDNA was synthesized from 1 μg of the total RNA using the ImProm-II Reverse Transcription System (Promega) and oligo (dT)$_{23}$ primers.

2.3.3 Semi-quantitative RT-PCR

RT-PCR reactions were performed using cDNAs prepared from extracted RNA. These reactions were set up in a reaction volume of 25 µL with the final concentration of the PCR primers of 100 nM. Specific primers for the most variable regions between *Orpin A* and *Orpin B* sequences were designed using OligoAnalizer tools from the Integrated DNA Technology webpage (www.idtdna.com). The primers used were: *Orpin B* forward: 5'-ACAGGGAGTACAAACAGTCGTCAA-3' and *Orpin B* reverse: 5'-CTATTTACTCTGCAACTGACACTTTCT-3'; *Orpin A* forward: 5'-ACTTCTGCAGAATCAGTTGTTAAGA-3' and *Orpin A* reverse: 5'-TTCAGTGGAGTCGCCAAC-3'. RT-PCR reactions were performed on three independent RNA samples purified from each of the regeneration stages (previously mentioned) as well as from the normal intestines. The PCR amplification was done by an initial denaturation step of 94°C (45 s), a primer annealing step of 50.2°C (45 s), and an extension step of 72°C (45 s) with a final additional 72°C (10 min) for 28 cycles for *Orpin A*, 26 cycles for *Orpin B*, and 26 cycles for *NADH*, as the amplification parameters for each pair of primers. All samples were analyzed in triplicate. Additional tissues were amplified for 35 cycles (2–4 replicates). The relative expression of *Orpin A* and *Orpin B* was normalized relative to the expression of the housekeeping gene *NADH dehydrogenase subunit 5* using ImageJ software (Schneider, Rasband, & Eliceiri, 2012) from the optical density values from electrophoresed sample bands on 1% agarose gels, using a Molecular Imager ChemiDoc XRS+ (BioRad). The primers used for the *NADH*

sequence amplification were: forward: 5'-CGGCTACTTCTGCGTTCTTC-3' and reverse: 5'-ATAGGCGCTGTCTCACTGGT-3'. The *Orpin A* and *Orpin B* sequences were confirmed by sequencing excised electrophoresed sample bands at the Sequencing and Genotyping Facility (UPR-RP).

2.3.4 Bioinformatics analyses

Homolog sequences were identified and retrieved from the NCBI GeneBank protein database (Sayers et al., 2019) using the original Orpin sequence previously identified (Ortiz-Pineda et al., 2009) as a query. BLASTp (Altschul, Gish, Miller, Myers, & Lipman, 1990; Bias & Gish, 1994) were performed against the public non-redundant protein database in GeneBank. Conserved domain identification and UTR analysis were performed using CDD (Marchler-bauer et al., 2017), RegRNA (Huang, Chien, Jen, & Huang, 2006), UTRScan (Bengert & Dandekar, 2003) and PSIPRED (Buchan & Jones, 2019; D. T. Jones, 1999), ScanProsite (Castro et al., 2006), InterProScan 5 (P. Jones et al., 2014; Quevillon et al., 2005), Phobos (Yao, Liu, Tan, & Song, 2016), SignalP 5.0 (Juan et al., 2019), and Phobius (Favio et al., 2019) on Geneious 11.1.5 software (https://www.geneious.com). Sequence alignments were carried out with MUSCLE (Edgar, Drive, & Valley, 2004) (10 iterations) and the Blosum62 matrix and edited with Geneious software 11.1.5 (https://www.geneious.com). Note: It is possible that there are N-terminal sequencing artifacts on two annotated sequences from *A. japonicus* sequences (ARI48335.1 and PIK49419.1). If we delete the residues from the predicted cytoplasmic N-terminal region from the ARI48335.1 sequence and from

PIK49419.1 up until their next methionine, they also show a predicted signal peptide of 21 residues each.

2.3.5 Phylogenetic analysis

EF-Hand proteins and other similar sequences were retrieved from literature and protein database as mentioned in results section and the multiple sequence alignment was performed on MAFFT v7.309 (Katoh & Standley, 2013) with BLOSUM62 scoring matrix, gap open penalty of 1.57, and offset value of 0.123. For the tree building, the Maximum-Likelihood analysis was done using JTT model of sequence evolution with 1000 bootstraps using PhyML 3.0 (Guindon et al., 2010) plugin using Geneious 11.1.5 software (https://www.geneious.com). The corresponding sequences are included in Supp. Table 2.1 and Supp. Table 2.2. The tree was edited for better visualization and colors in iTOL v4 online tool (Letunic & Bork, 2019). The (frog) *X. laevis,* (mouse), *M. musculus,* and (human) *H. sapiens* calcineurin A sequences were selected as outgroups and does not contain EF-Hand motifs. In addition, the Orpin homologs from *A. japonicus* ARI48335.1 and PIK49419.1 were edited for the analyses by deleting the N-terminal residues down to the second predicted methionine for the reason mentioned above.

2.3.6 Statistical analyses

Statistical significance of the resulting data was evaluated through one-way ANOVA using the JMP®, Version 12. SAS Institute Inc., Cary, NC, 1989-2019. The

multiple comparison procedure and statistical test Tukey-Kramer HSD (honestly significant difference) was used to determine significant differences between means from optical densities determined by ImageJ software as mentioned before (Schneider et al., 2012). The Tukey-Kramer results are displayed as small circles for high number of data points and large circles for low number of data points. The large red circle shows significant differences to small grey circles sample means. All values were reported as the mean ± standard mean error, including mean diamond with confidence interval ([1 – alpha] x 100), and outlier box plot from a quantiles report. While a $P < .05$ and $P < .001$ were considered to indicate statistical significance difference between groups.

2.3.7 Ethics statement

This research deals only with invertebrate animals, thus the University of Puerto Rico IACUC waives ethical approval of research performed on invertebrates. Animals were sacrificed by immersion in ice cold water for 29-30 min and then sectioning the anterior part of the animal close to the oral nerve ring, which accounts for the main component of the nervous system.

2.4 Results

2.4.1 Identification of the original *Orpin* (*Orpin A*) sequence and characterization of a second *Orpin* isoform (*Orpin B*)

The original report (Rojas-Cartagena et al., 2007) described a contig sequence (4766-1) which was later annotated as *Orpin*. This contig was used as a template to identify the remaining nucleotides upstream from the open reading frame (ORF) region through RACE-PCR analysis (Ortíz-Pineda et al., 2010). The *Orpin* sequence is composed of 106 nucleotides from the 5' UTR and 291 nucleotides from the 3' UTR with a 366 nucleotide ORF (plus stop codon) that encodes a putative 122 amino acid peptide followed by a stop codon (Figs. 2.1 and 2.2). The nucleotide composition of this gene sequence was validated by sequencing the RT-PCR products amplified from a normal intestine tissue cDNA sample (Figs. 2.2 and Supp. Figure 2.1). At the time it was annotated in the NCBI database (ACZ73832.1; 12/13/2009), there was no match with other sequences. Two similar sequences from the hemichordate *Saccoglossus kowalevskii* were later added as *Orpin*-like sequences (XP_006824981.1 and XP_002736736.1).

After performing further in-depth analyses of the available transcriptome libraries from regenerating and non-regenerating intestine and regenerating and non-regenerating radial nerve, we discovered an additional highly similar sequence that was identified as a putative *Orpin* isoform. This new putative protein shared 90% identity and 98% similarity with the original *Orpin* sequence but displayed different UTR's from the original sequence. We refer to this sequence

as *Orpin B* to differentiate it from the original *Orpin* which we refer from now on as *Orpin A.*

The sequence corresponding to *Orpin B* was also validated through RT-PCR amplification and sequencing (Figs. 2.3 and Supp. Figure 2.1). *Orpin B* mRNA sequence is composed of 103 nucleotides from the 5' UTR and 364 nucleotides from the 3' UTR (Figs. 2.1 and 2.3). Its ORF is 369 nucleotides (plus stop codon) long and encodes a putative 123 amino acid protein.
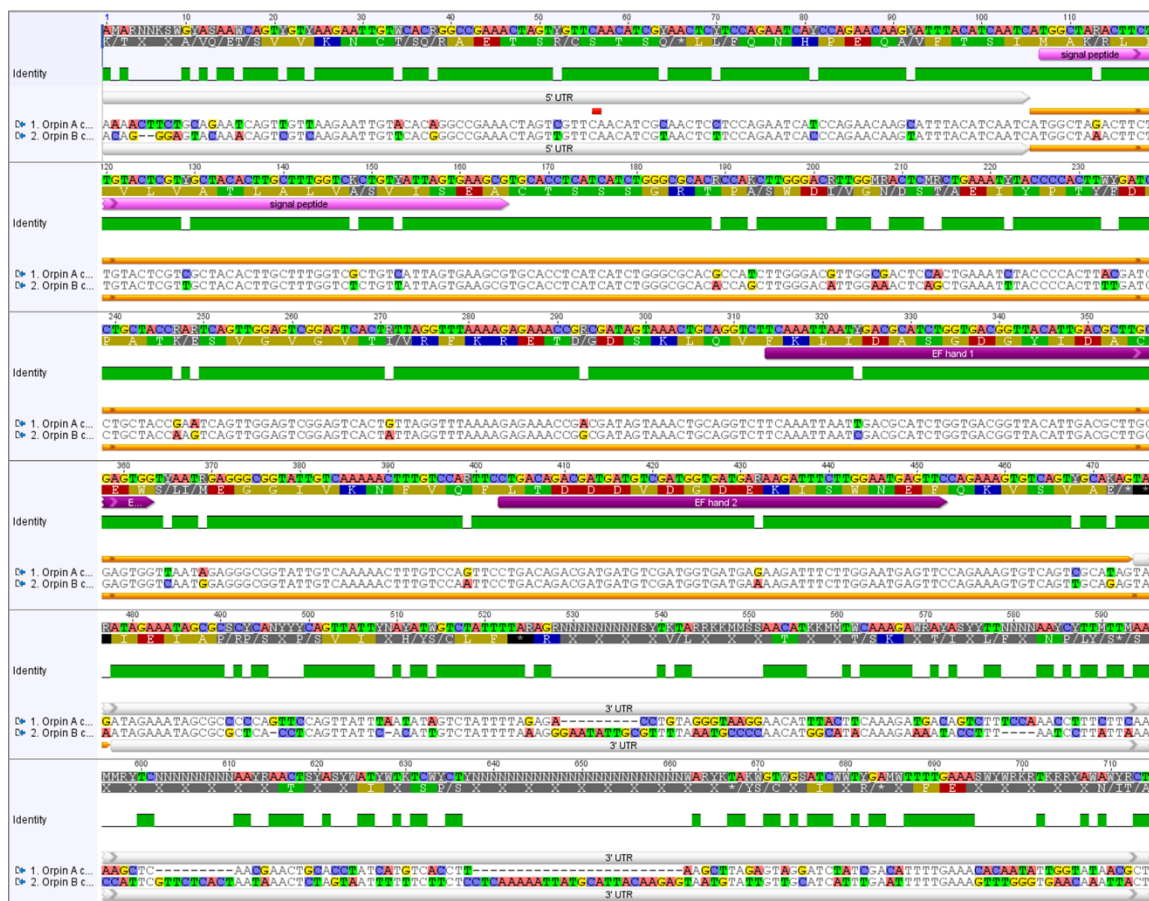
Figure 2.1 *Orpin A* and *Orpin B* are isoforms

Differences between sequences are highlighted. White bars: 5' UTR and 3' UTR regions of both sequences; pink bar: predicted signal peptides; orange bar: ORF regions; purple bars: predicted EF-hand motifs; green: conservation level; top sequences: nucleotide and amino acid consensus sequences. Differences between nucleotide sequences are highlighted. It is shown a significant difference, especially between both 3' UTR sequences. Analysis was done using the MAFFT plugin in Geneious 11.1.5.

ORPIN A

AAA<mark>ACTTCTGCAGAATCAGTTGTTAAGAAT</mark>TGTACACAGGCCGAAACTAGTCGTTCAACATCGCAA
CTCCTCCAGAATCATCCAGAACAAGCATTTACATCAATC

ATGGCTAGACTTCTTGTACTCGTCGCTACACTTGCTTTGGTCGCTGTCATTAGTGAAGCGTGCACCT
CATCATCTGGGCGCACGCCATCTTGGGAC<mark>GTTGGCGACTCCACTGAAA</mark>TCTACCCCACTTACGATCC
TGCTACCGAATCAGTTGGAGTCGGAGTCACTGTTAGGTTTAAAAGAGAAACCGACGATAGTAAACT
GCAGGTCTTCAAATTAATTGACGCATCTGGTGACGGTTACATTGACGCTTGCGAGTGGTTAATAGAG
GGCGGTATTGTCAAAAACTTTGTCCAGTTCCTGACAGACGATG<mark>ATGTCGATGGTGATGAGAAGATT</mark>
TCTTGGAATGAGTTCCAGAAAGTGTCAGTCGCATAG

TAGATAGAAATAGCGCCCCCAGTTCCAGTTATTTAATATAGTCTATTTTAGAGACCTGTAGGGTAAGG
AACATTTACTTCAAAGATGACAGTCTTTCCAAACCTTTCTTCAAAAGCTCAAC<mark>GAACTGCACCTATC</mark>
<mark>ATGTCACCTT</mark>AAGCTTAGAGTAGGATCTATCGACATTTTGAAACACAATATTGGTATAACGCTCTTTG
AATACCGATAATACCGGATGCATGGGTATATATGCAAGCAGAAAATAAATACATTGTCTCTATGTGAG
CCATTGTGAAAAACCGTGAAAA

Figure 2.2. Primers for sq-RT-PCR of *Orpin A*
Orpin A UTRs regions (blue boxes) and coding region (green box) of the *Orpin A* gene. Primer sequences designed to specifically amplify *Orpin A* (light red letters). Primer sequences used for the identification of the original *Orpin* sequence in previous reports (green letters). These primers were designed prior to identification *Orpin* isoform.

ORPIN B

<mark>ACAGGGAGTACAA</mark>ACAGTCGTCAAGAATTGTTCACGGGCCGAAACTAGTTGTTCAACATCGTAACT
CTTCCAGAATCACCCAGAACAAGTATTTACATCAATC

ATGGCTAAACTTCTTGTACTCGTTGCTACACTTGCTTTGGTCTCTGTTATTAGTGAAGCGTGCACCT
CATCATCTGGGCGCACACCAGCTTGGGACATTGGAAACTCAGCTGAAATTTACCCCACTTTTGATC
CTGCTACCAAGTCAGTTGGAGTCGGAGTCACTATTAGGTTTAAAAGAGAAACCGGCGATAGTAAAC
TGCAGGTCTTCAAATTAATCGACGCATCTGGTGACGGTTACATTGACGCTTGCGAGTGGTCAATGG
AGGGCGGTATTGTCAAAAACTTTGTCCAATTCCTGACAGACGATGATGTCGATGGTGATGAAAAGA
TTTCTTGGAATGAGTTCC<mark>AGAAAGTGTCAGTTGCAGAGTAA</mark>

ATAGAAATAGCGCGCTCACCTCAGTTATTCACATTGTCTATTTTAAAGGGAATATTGCGTTTTAAATG
CCCCAACATGGCATACAAAGAAAATACCTTTAATCCTTATTAAACCATTCGTTCTCACTAATAAACTC
TAGTAATTTTTTCTTCTCCTCAAAAATTATGCATTACAAGAGTAATGTATTGTTGCATCATTTGAATTT
TTGAAAGTTTGGGTGAACAAATTACTTTTGTGCAACAAATAATCGAATTACTTCCTGTGAGAGCCTT
ATTTCTTGAAAAGAAGCTATCGTACTGCCTTATTTTGAATTGTCTATCATACTGTATCTAACATATTG
GTTTAAAATATATGAAAATTTTCT

Figure 2.3. Primers for sq-RT-PCR of *Orpin B*
Orpin B UTR sequences (blue boxes) and coding region (green box). Primer sequences designed to specifically amplify *Orpin B* (blue letters).

Sequence comparisons among the two *Orpins* from *H. glaberrima* and the two *Orpin*-like sequences from *S. kowalevskii* show that the latter shared 46–50% identity and 76–77% similarity with the *Orpin A* (Fig. 2.4). Similarly, *Orpin B*

translated amino acid sequence shared 46–50% identity and 66–67% similarity with the sequences from *S. kowalevskii* (Fig. 2.4). Furthermore, we identified three additional putative *Orpin* homologs from another sea cucumber species, *Apostichopus japonicus*, one from the starfish *Acanthaster planci*, and two from the sea urchin *Strongylocentrotus purpuratus* with expected values (*E*-value < 0.001 and total scores > 47.8. All *Orpin*-like sequences contain one domain that is predicted to be a calcium-binding domain composed of two EF-hand motifs at their carboxy-terminal (Figs. 2.5 and 2.6).

A

	ORPIN A OR...	ORPIN B OR...	S. kowalevsk...	S. kowalevsk...	A. japonicus ...	A. japonicus ...	A. japonicus ...	A. planci LO...	S. purpuratu...	S. purpuratu...
ORPIN A ORF translation		90.16%	43.08%	45.38%	46.81%	46.81%	35.00%	40.48%	34.35%	25.52%
ORPIN B ORF translation	90.16%		43.51%	42.75%	47.52%	47.52%	33.89%	37.80%	31.82%	22.80%
S. kowalevskii LOC100375...	43.08%	43.51%		86.23%	30.87%	32.21%	23.40%	25.74%	28.06%	20.00%
S. kowalevskii LOC100369...	45.38%	42.75%	86.23%		29.53%	30.87%	23.40%	25.00%	31.65%	22.50%
A. japonicus ARI48335.1	46.81%	47.52%	30.87%	29.53%		81.07%	49.72%	37.32%	30.20%	22.86%
A. japonicus PIK49419.1	46.81%	47.52%	32.21%	30.87%	81.07%		49.72%	36.62%	30.20%	22.86%
A. japonicus PIK49418.1	35.00%	33.89%	23.40%	23.40%	49.72%	49.72%		25.27%	20.74%	16.47%
A. planci LOC110976159	40.48%	37.80%	25.74%	25.00%	37.32%	36.62%	25.27%		26.09%	18.09%
S. purpuratus LOC105438...	34.35%	31.82%	28.06%	31.65%	30.20%	30.20%	20.74%	26.09%		47.09%
S. purpuratus LOC100890...	25.52%	22.80%	20.00%	22.50%	22.86%	22.86%	16.47%	18.09%	47.09%	

B

	ORPIN A OR...	ORPIN B OR...	S. kowalevsk...	S. kowalevsk...	A. japonicus ...	A. japonicus ...	A. japonicus ...	A. planci LO...	S. purpuratu...	S. purpuratu...
ORPIN A ORF translation		98.36%	76.15%	76.92%	68.79%	67.38%	54.44%	67.46%	59.54%	45.31%
ORPIN B ORF translation	98.36%		77.10%	77.86%	69.50%	68.09%	52.78%	66.14%	57.58%	44.04%
S. kowalevskii LOC100375...	76.15%	77.10%		97.10%	56.38%	55.03%	45.21%	59.56%	56.12%	40.00%
S. kowalevskii LOC100369...	76.92%	77.86%	97.10%		58.39%	57.05%	46.28%	60.29%	57.55%	41.50%
A. japonicus ARI48335.1	68.79%	69.50%	56.38%	58.39%		89.94%	69.06%	63.38%	51.68%	40.95%
A. japonicus PIK49419.1	67.38%	68.09%	55.03%	57.05%	89.94%		69.61%	64.08%	51.68%	40.95%
A. japonicus PIK49418.1	54.44%	52.78%	45.21%	46.28%	69.06%	69.61%		47.25%	44.15%	35.74%
A. planci LOC110976159	67.46%	66.14%	59.56%	60.29%	63.38%	64.08%	47.25%		51.45%	38.69%
S. purpuratus LOC105438...	59.54%	57.58%	56.12%	57.55%	51.68%	51.68%	44.15%	51.45%		56.80%
S. purpuratus LOC100890...	45.31%	44.04%	40.00%	41.50%	40.95%	40.95%	35.74%	38.69%	56.80%	

Figure 2.4. Orpin homologs pairwise sequence divergence
Translated amino acid sequences comparison by identity% (A) and similarity% (B). The alignments were done using Muscle with 50 iterations using Geneious 11.1.5.
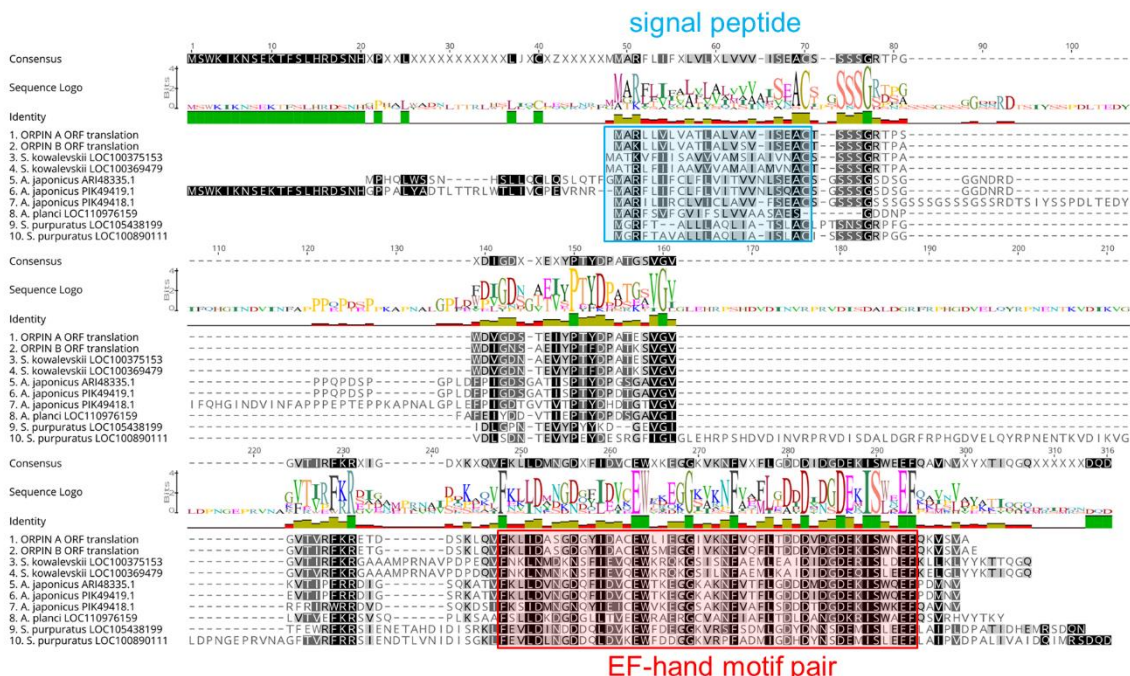
Figure 2.5. Orpin homologs alignment
The most conserved residues are indicated by letters in black boxes, green identity regions, and large cartoon letters at the sequence Logo. The exception is PIK49419.1 because 20 amino acid residues from the N-terminal portion are not compared to other sequences. We can see the additional N-terminal regions from *A. japonicus* sequences ARI48335.1 and PIK49419.1 that did not match to the other homologs. Blue box: signal peptide prediction; red box: EF-Hand motif pair prediction. This alignment was done by Muscle plugin with 50 iterations using Geneious 11.1.5.

2.4.2 Domain analyses

Orpin A and Orpin B amino acid sequences were analyzed using different bioinformatics tools (refer to methodology) for evidence that could point towards a possible function. After evaluating these sequences for domain composition using InterProScan and NCBI's CDD, and Phobos (P. Jones et al., 2014; Marchler-bauer et al., 2017; Quevillon et al., 2005; Yao et al., 2016), we identified that both sequences contain putative calcium-binding domain regions. Both Orpin isoforms shared identical calcium-binding loops residue composition. The key residue

positions that participate in calcium chelation within these loops are conserved when compared with other known EF-hand proteins. The X, Y, Z, –X, –Z positions from each loop of the EF-hands are Asp, Asp, Asp, Asp, Glu ("odd loop") and Asp, Asp, Asp, Ser, Glu ("even loop"), respectively (Fig. 2.6). The only difference is located downstream to the "odd" loop. There are two consecutive amino acids, Leu87 and Ile88, immediately after the Trp86 (–Z+1) of the first calcium-binding loop from Orpin A which are changed to Ser87 and Met88 in Orpin B. Interestingly, Orpin A and Orpin B included a Cys residue at –Z–1 position which is particular to both sea cucumber sequence homologs and is an unusual feature in EF-hand proteins.

When we compared *H. glaberrima* Orpin EF-hand sequences to the other identified putative homologs from *S. kowalevskii*, *A. japonicus, A. planci,* and *S. purpuratus*, we found that additional positions are highly conserved as well. All Orpin sequences share conserved positions at X–4 (Phe), –Z (Glu), –Z+1 (Trp) and –Z+6 (Gly) positions from the "odd" EF-hands, and X–8 (Phe), X (Asp), Z (Asp), –X–1(Ile), –X (Ser), –Z (Glu) and –Z+1 (Phe) positions from the "even" EF-hands. Alternatively, there are residues particular to the EF-hands from *H. glaberrima* Orpin isoforms, such as Ala at X+1, Ser at Y, Ala at –X+1 from the odd EF-hand, and Val at X+1 and Asn at –X+2 from the even EF-hand. Moreover, there are residues that are particular to the holothurians such as Lys at X–3, Cys at –Z–1, Lys at –Z+9 from the odd EF-hand, and Lys at –Y from the even EF-hand (Fig. 2.6).
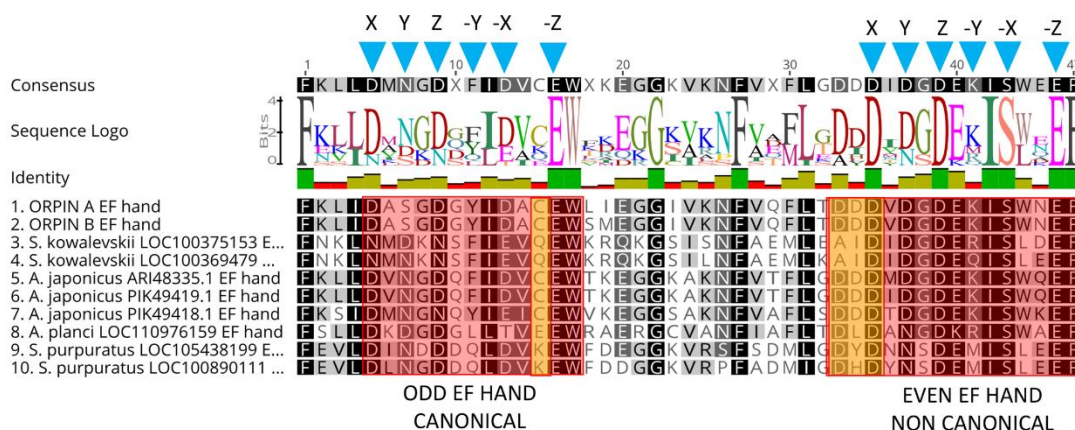
Figure 2.6. Orpin homologs EF-hand motifs alignment
The predicted odd EF-hands match with the canonical EF-hand pattern and the predicted even EF-hands were identified as non-canonical motifs (14 residues vs 12 residues) (red boxes). The non-canonical motifs are similar to vertebrates S100s. Predicted calcium coordinating residues from the identified EF-hands patterns are indicated by blue triangles. Holothurian Orpins contain a Cys residue at the −Z−1 position of the predicted calcium-binding loop (left orange box). The characteristic residues from Orpin residues are highlighted by orange boxes. Alignment was done using the MAFFT plugin in Geneious 11.1.5.

Additional bioinformatics analyses revealed the presence of a signal peptide in the N-terminal of both isoforms (Fig. 2.7). These signal peptides are 20 amino acids long each and are mainly composed of hydrophobic residues. The predicted signal peptides of Orpin A and Orpin B are nearly identical, with the exception of two residues at positions 3 (Arg/Lys) and 15 (Ala/Ser). Furthermore, InterProScan and Phobius identified the same region of 20 residues as a possible transmembrane region. In both cases, a high probability of a cleavage site was identified at the Cys21 residue of each isoform.

If the signal peptide is eliminated, the remaining sequence is predicted to be localized outside the cytoplasm (Fig. 2.7). This strongly suggests that these

peptides could be secreted to the extracellular space and not targeted to the membrane of other cell organelles.

The average length for the predicted signal peptides of the Orpin homologs is 20–22 residues based on SignalP and Phobius predictions (Favio et al., 2019; Juan et al., 2019; Petersen, Brunak, Heijne, & Nielsen, 2011), from the initial Met residue. The predicted signal peptides from the two *S. kowalevskii* sequences are longer (22 residues) than the other Orpins. The predicted signal sequence of a sea urchin homolog (XP_011664021.1) is the shortest (18) of the Orpins (Fig. 2.5).
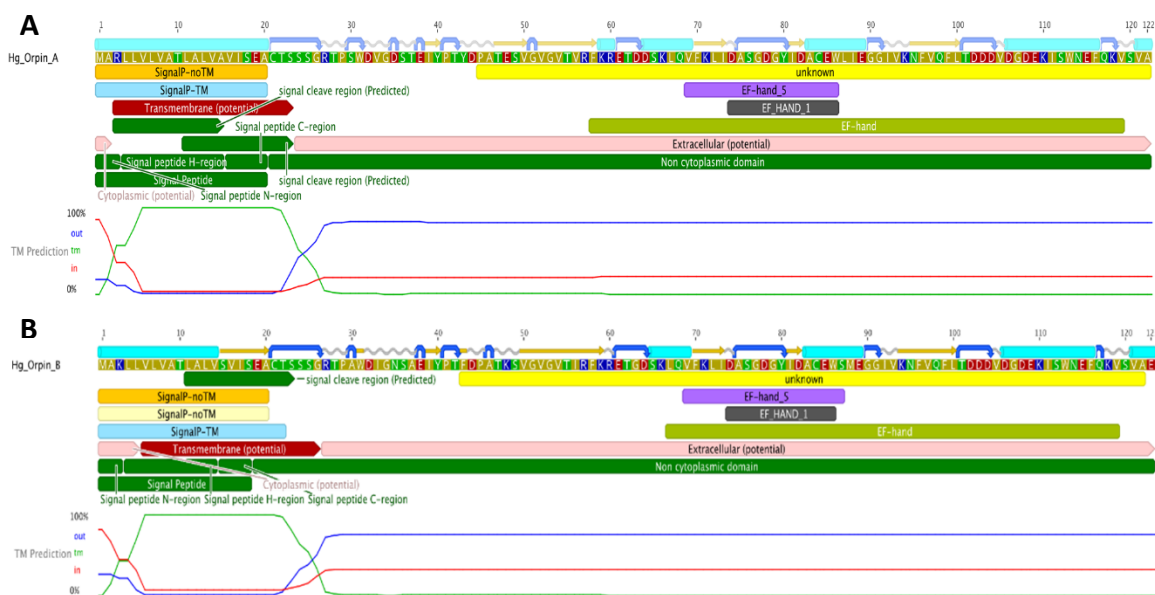
Figure 2.7. Orpin A and Orpin B bioinformatics characterization
Orpin A (A) and Orpin B (B) have predicted signal peptides at their transmembrane N-terminal regions including cleavage sites. Also, the two isoforms have predicted EF-hand motifs in their non-cytoplasmic regions. These were predicted by various bioinformatics plugin tools using Geneious 11.1.5.

The UTR's of the *H. glaberrima Orpin* sequences were also analyzed. Even though the 5' UTR's from *Orpin A* and *Orpin B* are 80.2% identical, there are 20 nucleotide differences between them, mainly SNPs. A ribosome binding site within the 5' UTR of each isoform sequence was identified. In contrast, the retrieved 3' UTR's of both *Orpin* isoforms were completely different. Polyadenylation sites were identified in both *Orpin A* (Ortiz-Pineda et al., 2009) and *Orpin B* downstream to their corresponding stop codons. Interestingly, these analyses revealed the presence of two putative Musashi binding elements (MBEs) within the 3'UTR of *Orpin A*. Even though the available retrieved 3' UTR from *Orpin B* is longer than its paralog, no MBEs were identified within this sequence. Surprisingly, two putative MBEs were also found within the coding sequence of each *Orpin* isotype.

2.4.3 Orpin phylogenetic analysis

In order to determine the relationship of the different Orpin homologs among themselves and with other EF-hand proteins, a phylogenetic tree was constructed with the PhyML program using a MAFFT alignment as input (Guindon et al., 2010; Katoh & Standley, 2013). Orpin A and Orpin B amino acid sequences were used as probes to identify the closest sequences through BLAST searches against the public databases. In addition, representative sequences from different EF-Hand subfamilies of various organisms were obtained from the scientific literature and available databanks. These sequences included members from the following protein families: S100s, calcineurin, recoverin, calbindin, parvalbumin,

oncomodulin, osteonectin, SPARC, troponin C, calmodulin, centrin, Spec, and recoverin (Supp. Table 2.1 and Supp. Table 2.2). Thus, these sequences were used for the final alignment to generate the phylogenetic tree.

The results from this analysis cluster Orpin and the identified hypothetical homologs from *S. kowalevskii* (acorn worm)*, A. japonicus* (sea cucumber)*, A. planci* (sea star)*,* and *S. purpuratus* (sea urchin) together with a bootstrap value of 92, separately from other subfamilies of EF-Hand proteins (Fig. 2.8). The other EF-Hand protein sequences cluster together as individual groups. The Orpin-like cluster was the most distant group after the outgroup sequences of mouse and frog calcineurin A, which do not contain EF-Hand motifs, suggesting that Orpins have evolved separately and are not direct homologs of EF-Hand proteins from other species. As expected, *H. glaberrima* Orpins were close to the other sea cucumber *A. japonicus* Orpin-like sequences. The most distant Orpin homologs were those from sea urchin *S. purpuratus.* The closest protein cluster was the osteonectins, BM-40, or SPARC proteins, which comprise a group of secreted CaBP modulators with a single pair of EF-Hand motifs. After these, the other group of proteins that appeared close by were the S100s, which also are small secreted proteins with two EF-hand motifs. The tree also showed the other outgroup EF-hand lacking protein, calcineurin A from humans, was placed separately from the other EF-Hand proteins of a high number of motifs (3 to 6 EF-Hands).
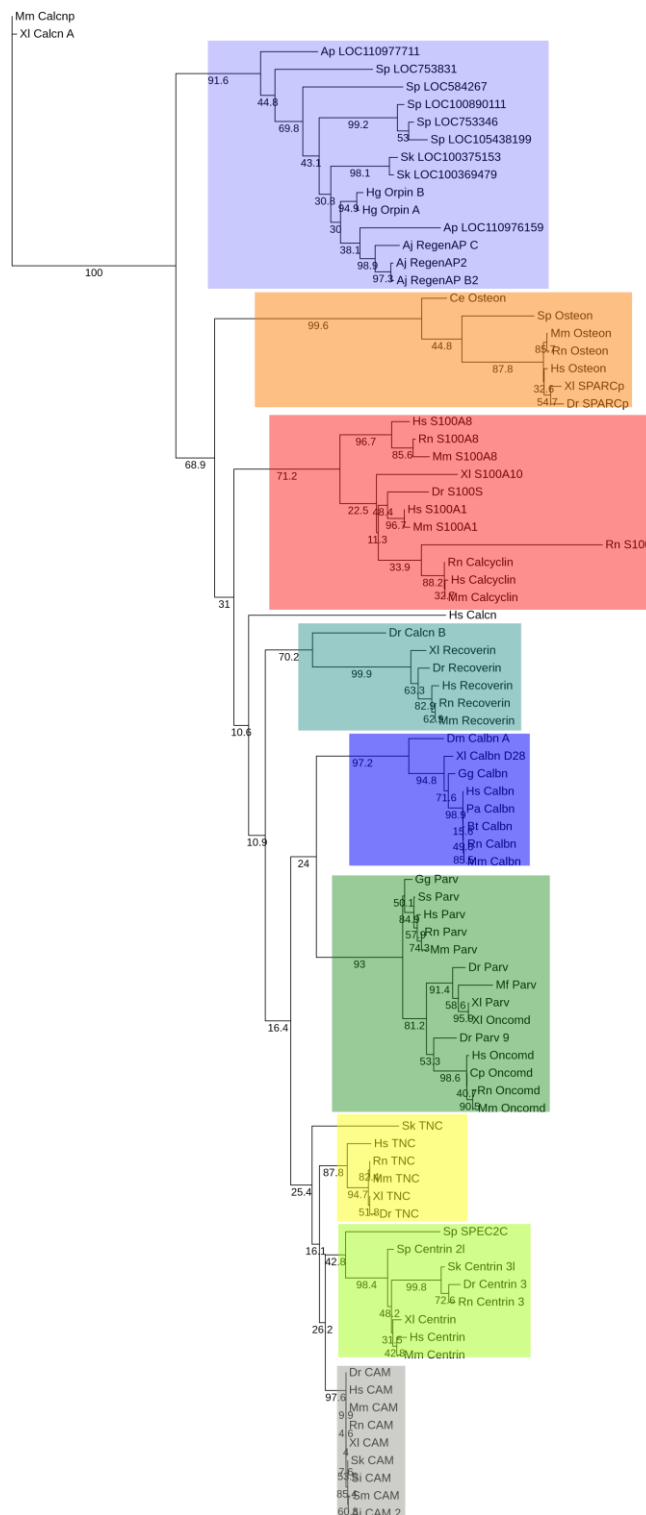
Figure 2.8. Orpin isoforms are specific to the Ambulacraria clade
EF-hand protein representative sequences from different subfamilies were aligned to build a phylogenetic tree. Orpin homologs were clustered together as a group, separated to the other EF-hand proteins. The tree was made using the PhyML plugin ran through Geneious 11.1.5. The parameters used for this analysis were JTT model of amino acid substitution and 1000 bootstraps. Scale bar: 1. Protein sequences accession numbers are included in Supp. Tables 2.1 and 2.2.

2.4.4 *Orpin* gene is expressed in several tissues of *H. glaberrima*

In order to determine the distribution of *Orpin* expression, mRNA was obtained from different tissues or organs of normal (non-regenerating) *H. glaberrima* specimens and processed for PCR analysis. The tissues and organs selected were: small intestine, large intestine, mesentery, radial nerve complex, longitudinal body wall muscle, gonads, and respiratory tree. Primers were designed for the specific detection of *Orpin A* and *Orpin B* mRNA sequences (Figs. 2.2 and 2.3). Transcript levels were evaluated relative to the expression of *NADH subunit 5*, a constitutively expressed housekeeping gene. The results showed that *Orpin A* and *Orpin B* shared similar tissue specificity (Figs. 2.9 and 2.10). Transcripts were detected in the gonads, muscle, mesentery, and haemal system but not in the respiratory tree nor in the nerve. Tissue expression varies significantly, with higher expression levels in the mesentery followed by the expression in muscle and gonads where it is slightly higher than in other tissues. Interestingly, a faint second lighter band was detected from *Orpin B* samples from gonads and muscle tissues.
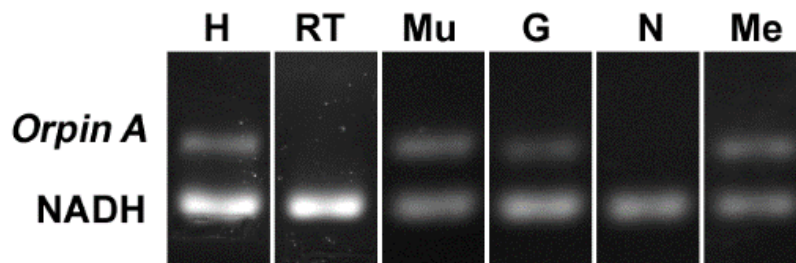
Figure 2.9. *Orpin A* expression in different tissues
Composite image from RT PCR amplification of *Orpin A* from *H. glaberrima* tissues. *Orpin A* expression (top band) was detected in haemal system (H), muscle (Mu), gonads (G), and mesentery (Me) relative to the expression of *NADH*. *Orpin A* was detected neither in the nerve (N) nor in the respiratory tree (RT). The image is a composite from different gels and is divided by a white line.
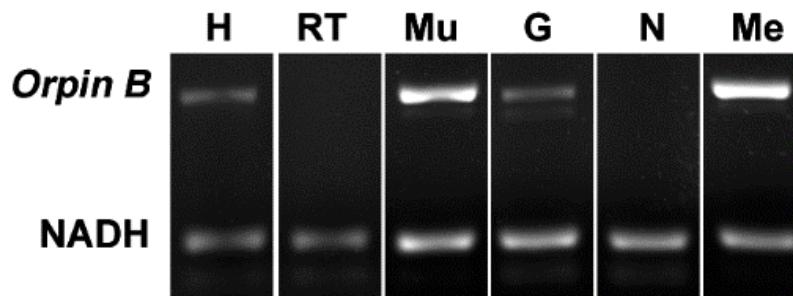
Figure 2.10. *Orpin B* **expression in different tissues**
Composite image from RT PCR amplification of *Orpin B* from *H. glaberrima* tissues. *Orpin B* expression (top band) was detected in haemal system (H), muscle (Mu), gonads (G), and mesentery (Me) relative to expression of *NADH*. *Orpin B* was detected neither in the nerve (N) nor in the respiratory tree (RT). The faint band below the *NADH* band corresponded to primer dimers. The image is a composite from different gels and is divided by a white line.

2.4.5 *Orpin* expression during intestinal regeneration in the sea cucumber

H. glaberrima

Previous results from our laboratory have shown that *Orpin* was differentially expressed in regenerating intestinal tissues when compared to normal intestinal tissues (Ortiz-Pineda et al., 2009; Rojas-Cartagena et al., 2007). In order to validate the upregulation of this novel sequence during regenerative processes, *Orpin* transcript levels were measured during different stages of intestine regeneration. In contrast to previous experiments where no particular effort was made to separate the intestine of normal animals from the attached mesenteries, in the present experiments we measured separately the intestine (a mixed portion from the small intestine and from the large intestine) and the mesentery that attaches the intestine to the body wall, for the normal (non-regenerating) samples. Orpin transcript levels were measured relative to the housekeeping gene *NADH subunit 5*. The gene expression levels were monitored using semi-quantitative RT-

PCR of tissue extracts from 3 days post evisceration (dpe), 5 dpe, 7 dpe, and 10 dpe along with tissues from normal intestine and normal mesentery.

Previously it was found that the expression levels of *Orpin A* increased after 3 days of intestine regeneration and then gradually returned to the basal levels at 14 days of regeneration. In contrast, to those findings (Ortiz-Pineda et al., 2009; Rojas-Cartagena et al., 2007), there was no statistically detected difference found between the transcript expression of *Orpin A* from normal intestine samples and those from any of the studied regenerative days (Data not shown). This was also true for *Orpin B* (Data not shown). However a high differential expression was detected between tissues from *Orpin A* from normal mesentery and tissues from 7–10 dpe sample group, with a $P < .05$ ($P = .002$) (Fig. 2.11). *Orpin B* exhibited a high differential expression between tissues from normal mesentery and tissues from 3–5 dpe with a $P < 0.05$ ($P = .02$), and 7–10 dpe with a $P < .001$ (Fig. 2.12).

Interestingly, we found a different expression profile between *Orpin A* and *Orpin B* transcript levels. While both *Orpin* forms show subsequently decreases in their expression to similar levels at 7-dpe to 10-dpe, the decrease of *Orpin B* seems to occur much faster than that of *Orpin A*.
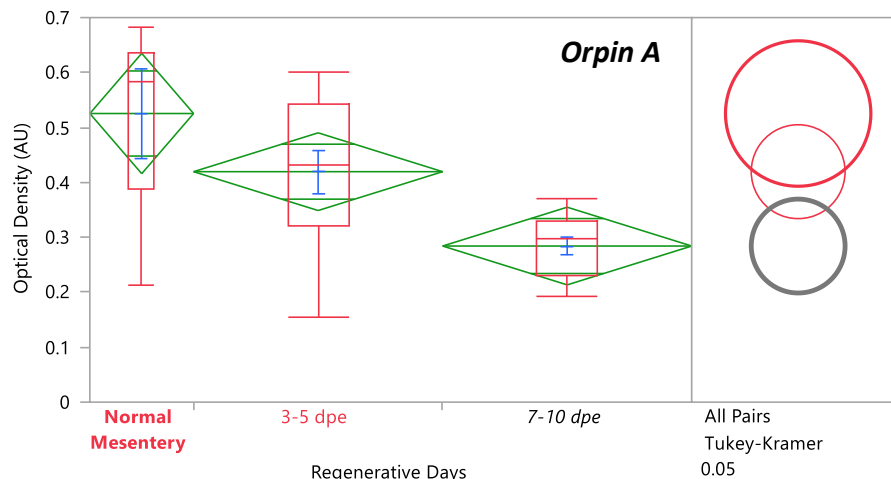
Figure 2.11. *Orpin A* **expression during intestine regeneration grouped tissues**
Semi-quantitative RT-PCR amplification of *Orpin A* transcripts from mRNA samples from different intestine regenerative days compared to the corresponding expression in samples from normal intestine (NI) and normal mesentery (NM). A statistical high differential expression was found between NM and 7–10 days post evisceration (dpe) ($P < .05$; $P = .002$); and between 3–5 dpe and 7–10 dpe ($P < .05$; $P = .03$) as indicated in the all pairs Tukey-Kramer HSD test. The large red circle (low number of data points) displays the significant difference between the small grey circles (high number of data points) group means. Red boxes: outlier box plots summarizing the distribution of points at each factor level from the quantiles report. Green diamonds: sample mean and confidence interval ([1 − alpha] x 100). Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.
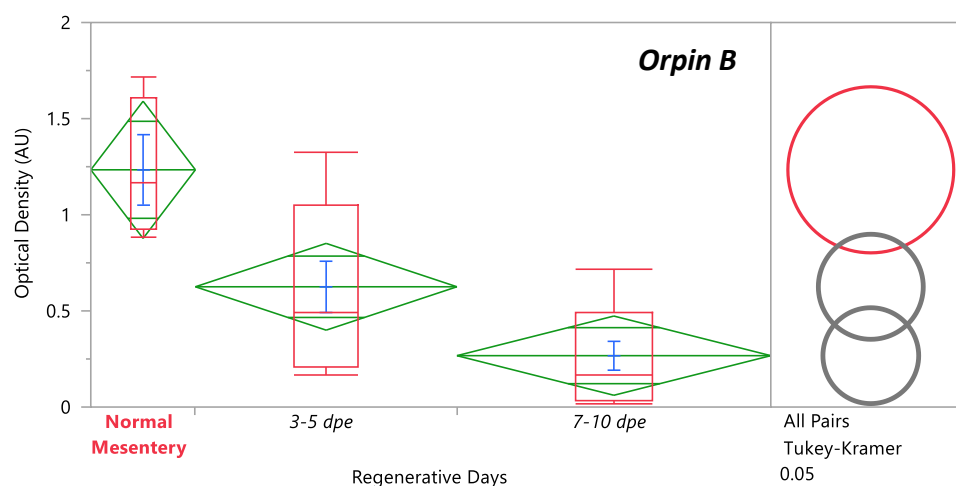
Figure 2.12. *Orpin B* **expression during intestine regeneration grouped tissues**
Semi-quantitative RT-PCR amplification of *Orpin B* transcripts from mRNA samples from different intestine regenerative days compared to the corresponding expression in samples from normal intestine (NI) and normal mesentery (NM). A statistical high differential expression was found between NM and 3–5 dpe ($P < .05$; $P = .02$), and to 7–10 dpe ($P < .001$) as indicated in the all pairs Tukey-Kramer HSD test. The large red circle (low number of data points) displays the significant difference between the small grey circles (high number of data points) group means. Red boxes: outlier box plots summarizing the distribution of points at each factor level from the quantiles report. Green diamonds: sample mean and confidence interval ([1 − alpha] x 100). Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.

2.5 Discussion

We have now described the presence of two predicted EF-hand domain-containing proteins from the sea cucumber *H. glaberrima*. These putative proteins apparently belong to a unique group that is present in echinoderms and hemichordates. According to the mRNA distribution in the sea cucumber, the translated proteins are expressed in multiple organs. Moreover, they are highly represented within the mesentery of the normal and regenerating intestine. The possibility that these are Ca^{2+}-binding proteins is discussed below.

2.5.1 *Orpins* are novel genes

When the first *Orpin* sequence from *H. glaberrima* was identified, no other sequence that showed significant similarity to it could be found within databases (Ortiz-Pineda et al., 2009; Rojas-Cartagena et al., 2007). A few months later, two highly similar sequences (and possible homologs) were identified in the hemichordate *Saccoglossus kowalevskii* (acorn worm) and were added to the databases. These sequences were annotated with accession numbers XM_006824918.1, XP_006824981 (*E*-value: 9E-31) and XM_002736690.2, XP_002736736 (*E*-value: 2E-29). Later on, several homologs from closely related organisms of the Echinodermata phylum were added to the public databases: two from the sea urchin *Strongylocentrotus purpuratus*, three from other sea cucumber *Apostichopus japonicus*, and one from the starfish *Acanthaster planci*. The finding of an additional *Orpin* sequence in *H. glaberrima* increased to ten the known sequences and suggested that these sequences belong to a novel family of

proteins within a group of metazoans. In all cases, the sequences have been annotated with little or no descriptive information other than their tissue/organism from where they originated. At present, *Orpins* appear to be restricted to the Ambulacraria clade (the group that encompasses echinoderms and hemichordates), however, it remains to be seen if, with the sequencing of other animal genomes, the specificity of *Orpins* to the Ambulacraria still stands.

Although the *H. glaberrima* A and B *Orpin* variants share a high percentage of similarity at the nucleotide and protein levels, our data suggest that they correspond to distinct genes. First, the nucleotide and putative amino acid differences are distributed throughout the complete sequence of both variants. Second, even though the 5' UTR sequences of both *Orpin* sequences share a large similarity, they are not identical, and similar to the coding region, have multiple nucleotide differences distributed along the nucleotide sequences. Third, their 3' UTR sequences are very different, sharing minimal similarities. Finally, other species also have more than one *Orpin*-like gene. For example, sequence information from *Orpin* homologs from *S. kowalevskii* and *S. purpuratus* were annotated as located in different loci. These differences are characteristic of different genes rather than allele variants or products from differential splicing. Nonetheless, in spite of these results that suggest two different isoforms originating from two different genes, it is necessary to have the genome information as conclusive evidence. Moreover, it should be emphasized that while *Orpin A* and *B* mRNAs have been identified and sequenced from various tissues (see below), the

S. kowalevskii, *A. japonicus*, *A. planci*, and *S. purpuratus* sequences are hypothetical mRNA/protein-coding sequences that remain to be characterized. Even though it was expected that *H. glaberrima Orpin* sequences would be more similar to those from the other sea cucumber *A. japonicus,* the results showed that they shared higher similarity to the hemichordate *S. kowalevskii* sequences. This can be attributed to longer N-terminal regions from two of the *A. japonicus* sequences (ARI48335.1 and PIK49419.1) that did not match with any of the other homologs. These additional regions were annotated as part of the corresponding ORFs due to an identified methionine upstream to the one that matched the other homologs. Given the fact that these sequences were not validated, it has to be considered the possibility that these first encoding methionine residues could be the result of a PCR artifact, and could be in fact part of their corresponding 5' UTR regions. Hopefully, the characterization of *Orpin* isoforms will provide essential insights that eventually would make feasible the characterization of these homolog sequences.

2.5.2 Are Orpins calcium-binding proteins?

Rigorous analysis of the residue composition of the EF-hand domains coupled to structural and functional experimentation is the mainframe of the study of uncharacterized CaBPs. Thus, understanding the architecture of EF-hand domains provides a hint of the role that a particular EF-hand protein might have. One of the most prominent characteristics of the putative protein Orpin is the presence of EF-hand motifs that are a distinctive signature of calcium-binding

proteins. The EF-hand motif has been used as a standard of reference for the description of the calcium-binding loops from the corresponding domain by the residues at key positions for the chelation of each calcium ion. Even though the EF-hand motifs are a characteristic feature of many CaBPs, few of them consist of less than four EF-hands. As mentioned before, Orpin A and Orpin B paralogs comprise a single putative calcium-binding domain composed of two EF-hands. The canonical EF-hand motif topology is a helix-loop-helix conformation, which regularly binds calcium ions (Kretsinger et al., 1991; Persechini, Moncrief, & Kretsinger, 1989). Usually, this conformation is composed of a highly conserved 12 residues calcium-binding loop flanked on both sides by alpha-helices (Nelson & Chazin, 1998). The residues that participate in calcium coordination were labeled as X, Y, Z and −X, −Y, −Z (Kretsinger et al., 1991; Malmendal et al., 1998), and those conserved positions are conventionally used as a reference frame to analyze the calcium-binding potential and dynamics of the EF-hand CaBPs. A typical EF-hand domain is composed of two calcium-binding loops containing motifs flanked by alpha-helices. The adjacent alpha helices are named incoming and exiting helices from the odd (N-terminal) and even (C-terminal) EF-hand motifs (Gifford, Walsh, & Vogel, 2007; Grabarek, 2006; Kawasaki, Nakayama, & Kretsinger, 1998; Moncrief, Kretsinger, & Goodman, 1990). *H. glaberrima* Orpin isoforms odd calcium-binding loops are composed of highly conserved key amino acids of the canonical domain structure. We showed that these domains shared a high level of conservation in the residues that participate in the chelation of calcium ions. However, the "even" calcium-binding loop from sea cucumber Orpins slightly

deviates from the canonical pattern. The conserved Gly6 residue that provides for loop flexibility is substituted by a Glu6. This substitution is well conserved throughout all the available Orpin homologs with the exception of the starfish sequence. Also, they share a conserved Trp13 at position −Z+1 of the "odd" EF-hand. Furthermore, this conserved residue seemed to be particular to Orpins after comparison to the other 84 EF-Hand sequences from this study. Interestingly, a Cys11 residue is particular to the holothurian Orpins.

In view of these facts, Orpin homologs can be classified as novel EF-hand proteins. Although the bioinformatics analysis strongly suggests that they might be a new type of CaBPs, in order to assure this, experimental confirmation of the actual binding of calcium ions will be required along with the phylogenetic evidence provided in this study.

2.5.3 Orpin relationship with other Ca^{2+} binding proteins

Whether they are indeed CaBPs or not, the sequence comparisons show that Orpins share several characteristics with CaBP subfamilies and that in fact, CaBPs account for the most similar proteins in the database. Several patterns have been developed to accurately classify newly discovered EF-hand proteins. The main pattern is the calmodulin canonical EF motif mostly known as the DXDXDG pattern (Denessiouk, Permyakov, Denesyuk, Permyakov, & Johnson, 2014; Rigden & Galperin, 2004), which contrasts with the pseudo-EF-hand binding loops from the most recent vertebrate S100 family of proteins.

One of the main EF-hand protein subfamilies are the S100 proteins. These proteins are small CaBPs containing only two EF-hand motifs. Nevertheless, their N-terminal motifs are considered pseudo-EF-hands, which is the main characteristic of this protein family. At the moment of this study, this subfamily has only been found in vertebrates (Morgan, Martin-Almedina, Garcia, Jhoncon-Kooyip, & Fernandez, 2006). Although there were no pseudo-EF-hand predicted from both Orpin isoforms sequences, they share several characteristics with S100 proteins, such as the small size, acidic composition, secretion to extracellular location, and EF-hand motif number. Thus, Orpin sequences share structural features with the main CaBP subfamilies, making it difficult to classify them as any of them. Moreover, we have shown that Orpin and Orpin-like sequences clustered together more closely to osteonectins (BM-40/SPARC) proteins, a group of secreted CaBPs with only two EF-Hand motifs, suggesting that these two groups shared a common ancestral origin. In addition, the best-known CaBPs that grouped closely to Orpin-like sequences were mouse (*M. musculus*, AAA37432.1) and frog calcineurin A isoforms (*X. laevis*, AAC23449.1). These two sequences and the human calcineurin A (*H. sapiens*, AAC37581.1) do not contain EF-hands and were included as outgroups for this analysis. We emphasize the fact that the calcineurin isoform B from zebrafish does contain an EF-hand domain, thus it is separated from the outgroup calcineurin A sequences and is included with the other aligned CaBPs. These results suggest that Orpin sequences comprise a specific EF-hand protein group that is different from the other known subfamilies

of calcium-binding proteins. These data suggest that we can be dealing with a new subfamily of EF-hand proteins that is specific to the Ambulacraria clade.

2.5.4 Orpins as secreted proteins

In addition to their EF-hand motif, an additional feature of Orpins is the presence of a signal peptide. In this respect, Orpin isoforms strongly resemble the groups of EF-hand proteins that are secreted, namely the osteonectins, oncomodulins, and S100s. Similar to Orpin sequences, osteonectins (BM-40/SPARC), oncomodulin, and S100 proteins are small peptides containing two EF-hand motifs each and are secreted to the extracellular milieu. Interestingly, these proteins display calcium-mediated dimerization either as heterodimers as in the case of S100A8/S100A9 (Edgeworth, Gorman, Bennett, Freemont, & Hogg, 1991) or homodimers as in the case of S100P (Vivo, Koltzscher, & Gerke, 2000), osteonectins (Maurer et al., 1995), and oncomodulin (Mutus, Palmer, & MacManus, 1988). Usually, EF-hand proteins containing signal peptides are targeted to the outer plasma membrane. There, these CaBPs can act as growth factors recognizing binding targets located on other cell surfaces, thus activating different signaling pathways. Such is the case of osteonectin (BM-40/SPARC), which promotes changes in cell morphology, disrupt cell adhesion, inhibit cell cycle, regulate extracellular matrix, and modulate cell proliferation and migration (Phan, Ahluwalia, & Tarnawski, 2007), thus underlying the process of wound repair. Similarly, the secreted (although lacking a signal peptide) oncomodulin and S100 proteins, are involved in a variety of biological processes including: cell

proliferation, differentiation, survival, nerve regeneration, interaction with transcription factors, and calcium homeostasis (R Donato et al., 2013; Rosario Donato, 1999, 2001; Rosario Donato et al., 2009; Kurimoto et al., 2013; Schaub & Heizmann, 2008; Yin et al., 2009, 2003, 2006) among other functions.

2.5.5 *Orpins* and regeneration

Previous results from our laboratory had shown differential expression of the *Orpin* transcript when mRNAs levels from 3- and 7- day regenerating intestines were compared to normal intestinal tissues. Thus, we had concluded that *Orpin* was over-expressed during intestinal regeneration. Our new data, where we detect high levels of the transcript in the mesentery region, questions our previous interpretation. Thus, if we consider that the samples containing the 3- and 5- day regenerating intestinal rudiments contain a large proportion of the remaining mesentery (that remains attached to the body wall), then the high expression of *Orpin A* and *B* transcript levels that were detected in the 3- or 5- day regenerating tissues can be interpreted as representing the expression in the mesenteric portion and not necessarily in the rudiment itself. As the rudiment itself grows and encompasses a larger proportion of the dissected tissues (in relation to the mesentery) then the *Orpin* expression would appear to decrease. This is why there is no difference between normal mesentery and early regenerating rudiments. The previously observed difference between regenerating rudiments and "normal" intestine would be merely a reflection of the proportion of mesenterial tissue present in both samples; low in "normal" intestines and high in regenerating ones.

In this respect, the isolated "normal" mesentery is a more appropriate control to compare the relative expression of *Orpin* transcript sequences between regenerating and normal tissues. This is particularly true in early regenerating stages (3–5 days) when the proportion of tissues corresponding to the mesentery is quite high.

While the lack of *Orpin* differential expression argues against a possible role in the intestinal regeneration process, we cannot completely exclude this possibility. The presence of Musashi binding elements within *Orpin* isoform sequences suggests that the post-transcriptional regulation of their mRNAs might be controlled by RNA-binding proteins. This element is present in genes that are post-transcriptionally regulated in a spatial and temporal dependent manner (MacNicol, Cragle, & MacNicol, 2011; Takahashi et al., 2013). Moreover, this type of regulation has been implicated in the self-renewal of epithelial, neural and hematopoietic stem and progenitor cells (Asai, Okano, & Yasugi, 2005; Hope et al., 2010; Kharas et al., 2010; M. Nakamura, Okano, Blendy, & Montell, 1994; Okano et al., 2005; Sakakibara et al., 2002; Wang et al., 2008). Such is the case of the target transcript encoding the transcription factor TTK69 in *Drosophila,* where translational activation is mediated by the neural *Drosophila* Musashi. In this way, the Musashi protein induces the differentiation of *Drosophila* IIb cells as neural precursor cells by repressing the translation of the mRNA of this neural differentiation inhibitory factor (Okano et al., 2005). Furthermore, the expression of mammalian Numb protein (m-Numb) induces the expression of regeneration-

related genes such as prostate stem cell antigen (PSCA) and metallothionein-2 (Mt2) in gastric mucosal regeneration in mice. Musashi protein (Msi1) enhances the expression of m-Numb during this regenerative process through post-transcriptional regulation. Having stated this, we cannot disregard the possibility that *Orpin* isoforms play a role during the initial stages of intestine and nerve regeneration. Nonetheless, to explore this possibility, we need to determine if a Musashi protein is present in the *H. glaberrima* proteome during regenerative processes of the sea cucumber and that it binds to *Orpins* mRNA.

It is of interest that *Orpin* isoforms are found to be differentially expressed in a transcriptomic library of regenerating nerve from *H. glaberrima* (Mashanov et al., 2014), also suggesting a possible regeneration-associated function. The two sequences displayed an increase in expression during the regeneration of the radial nerve complex after induced injury. By day 2 and also by day 20 after nerve injury, *Orpin A* expression was significantly higher than non-regenerating radial nerve ($P < .001$). In addition, *Orpin B* was higher in the same samples after day 2, 12, and 20 after nerve injury ($P < .001$). However, in view of our findings in the intestinal system, it remains to be determined whether this differential expression is also the product of the gene is expressed preferentially in the remaining tissues following injury, and not necessarily of increasing its expression.

In summary, we have identified and characterized a group of Orpin-like proteins from a particular group of invertebrate deuterostomes and shown they all share similarities in size, domain composition, and little significant similarities to other

known EF-hand protein sequences. We provide bioinformatics evidence for the presence of signal peptides and cleavage sites in these proteins that suggest secretion of the putative proteins to the extracellular environment. Together, with the identification of predicted EF-hand domains with unique features, we can suggest that these might comprise a novel subfamily of EF-hand containing proteins specific to the Ambulacraria clade. Finally, we studied their expression in normal and regenerating tissues, with the surprise finding that they are highly expressed in the intestinal mesentery.

2.6 Supplementary Figures
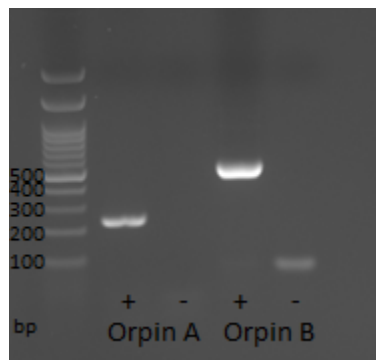
Supplemental Figure 2.1. RT-PCR primers test
Orpin A and *Orpin B* RT-PCR amplification using primers for semi-quantitative analysis. cDNA used was from a pool of cDNAs normal intestine, mesentery, regenerating days tissues. Run cycles: 35. *Orpin B* primers show amplification of dimers, also shown in cDNA (-) control.

Supplemental Table 2.1. EF-hand proteins sequences for the phylogenetic analyses

Sequence Name	Short Name	Accession No.
gi|270313555| regeneration associated protein [*Holothuria glaberrima*]	*Hg*_Orpin_A	ACZ73832.1
ORPIN_B [*Holothuria glaberrima*]	*Hg*_Orpin_B	MT364419
gi|4826655| calbindin [*Homo sapiens*]	*Hs*_Calbn	NP_004920.1
gi|197102844| calbindin [*Pongo abelii*]	*Pa*_Calbn	NP_001126892.1
gi|45382893| calbindin [*Gallus gallus*]	*Gg*_Calbn	NP_990844.1
gi|14010887| calbindin [*Rattus norvegicus*]	*Rn*_Calbn	NP_114190.1
gi|6753242| calbindin [*Mus musculus*]	*Mm*_Calbn	NP_033918.1
gi|7302852| calbindin 53E, isoform A [*Drosophila melanogaster*]	*Dm*_Calbn_A	AAF57925.1
gi|115495899| calbindin [*Bos taurus*]	*Bt*_Calbn	NP_001069663.1
gi|4098291| calbindin D28k [*Xenopus laevis*]	*Xl*_Calbn_D28	AAD00259.1
gi|35290| parvalbumin [*Homo sapiens*]	*Hs*_Parv	CAA44792.1
gi|297747319| parvalbumin [*Sus scrofa*]	*Ss*_Parv	NP_001177086.1
gi|235323| parvalbumin [chickens, muscle, Peptide, 109 aa] [*Gallus gallus*]	*Gg*_Parv	AAB19756.1
gi|537462538| parvalbumin [*Micrurus fulvius*]	*Mf*_Parv	JAB53742.1
gi|214644| parvalbumin, partial [*Xenopus laevis*]	*Xl*_Parv	AAA49925.1
gi|33636707| NP_891983.1 parvalbumin 9 [*Danio rerio*]	*Dr*_Parv9	NP_891983.1
gi|8698675| AF180888_1 parvalbumin [*Danio rerio*]	*Dr*_Parv	AAF78471.1
gi|205973| parvalbumin [*Rattus norvegicus*]	*Rn*_Parv	AAA41799.1
gi|53819| parvalbumin [*Mus musculus*]	*Mm*_Parv	CAA38434.1
gi|145279226| oncomodulin [*Mus musculus*]	*Mm*_Oncomd	NP_149028.2
gi|290491124| oncomodulin [*Cavia porcellus*]	*Cp*_Oncomd	NP_001166459.1
gi|47480974| Oncomodulin [*Homo sapiens*]	*Hs*_Oncomd	AAH69550.1
gi|291290895| oncomodulin [*Xenopus laevis*]	*Xl*_Oncomd	NP_001167478.1
gi|1289279| oncomodulin [*Rattus norvegicus*]	*Rn*_Oncomd	CAA33840.1
gi|5542035| AC006536_1 calmodulin [*Homo sapiens*]	*Hs*_CAM	AAD45181.1
gi|148225823| calmodulin [*Xenopus laevis*]	*Xl*_CAM	NP_001080864.1
gi|6753244| calmodulin [*Mus musculus*]	*Mm*_CAM	NP_033920.1
gi|14010863| calmodulin [*Rattus norvegicus*]	*Rn*_CAM	NP_114175.1
gi|47550761| calmodulin [*Danio rerio*]	*Dr*_CAM	NP_999901.1
gi|20152223| calmodulin [*Strongylocentrotus intermedius*]	*Si*_CAM	BAB89361.1
gi|656991664| calmodulin [*Stichopus monotuberculatus*]	*Sm*_CAM	AID21732.1
gi|64446704| calmodulin 2 [*Apostichopus japonicus*]	*Aj*_CAM_2	AAY41437.1
gi|296434295| calmodulin [*Saccoglossus kowalevskii*]	*Sk*_CAM	NP_001171813.1
gi|291240933| PREDICTED: troponin C, slow skeletal and cardiac muscles-like [*Saccogloss*	*Sk*_TNC	XP_002740370.1
gi|1220372| troponin C [*Homo sapiens*]	*Hs*_TNC	AAA91854.1
gi|6678369| troponin C, slow skeletal and cardiac muscles [*Mus musculus*]	*Mm*_TNC	NP_033419.1
gi|77627992| troponin C, slow skeletal and cardiac muscles [*Rattus norvegicus*]	*Rn*_TNC	NP_001029277.1
gi|56118753| troponin C, slow skeletal and cardiac muscles [*Danio rerio*]	*Dr*_TNC	NP_852475.2
gi|148223295| cardiac troponin C [*Xenopus laevis*]	*Xl*_TNC	NP_001083764.1
gi|50979313| calcium-binding protein SPEC 2C [*Strongylocentrotus purpuratus*]	*Sp*_SPEC2C	NP_999769.1
gi|338325| osteonectin [*Homo sapiens*]	*Hs*_Osteon	AAA60993.1
gi|304334| osteonectin [*Caenorhabditis elegans*]	*Ce*_Osteon	AAA16827.1

Supplemental Table 2.2. (continued) EF-hand proteins sequences for the phylogenetic analyses

Sequence Name	Short Name	Accession No.
gi|201022| osteonectin [*Mus musculus*]	*Mm*_Osteon	AAA40125.1
gi|600381| osteonectin [*Rattus norvegicus*]	*Rn*_Osteon	BAA06029.1
gi|50080160| SPARC precursor [*Danio rerio*]	*Dr*_SPARCp	NP_001001942.1
gi|148222476| SPARC precursor [*Xenopus laevis*]	*Xl*_SPARCp	NP_001079752.1
gi|117371494| osteonectin [*Strongylocentrotus purpuratus*]	*Sp*_Osteon	ABK33665.1
gi|15680116| S100 calcium binding protein A1 [*Homo sapiens*]	*Hs*_S100A1	AAH14392.1
gi|13542773| S100 calcium binding protein A1 [*Mus musculus*]	*Mm*_S100A1	AAH05590.1
gi|28570182| S100 calcium-binding protein, ventral prostate [*Rattus norvegicus*]	*Rn*_S100	NP_788265.1
gi|51591887| S100 calcium binding protein S [*Danio rerio*]	*Dr*_S100S	NP_001003989.1
gi|213626526| S100 calcium binding protein A10 (calpactin) [*Xenopus laevis*]	*Xl*_S100A10	AAI69571.1
gi|965468| calcineurin [*Homo sapiens*]	*Hs*_Calcn	AAC37581.1
gi|192657| calcineurin, partial [*Mus musculus*]	*Mm*_Calcnp	AAA37432.1
gi|41054369| calcineurin B homologous protein 1 [*Danio rerio*]	*Dr*_Calcn_B	NP_956009.1
gi|3213204| calcineurin A [*Xenopus laevis*]	*Xl*_Calcn_A	AAC23449.1
gi|192346| calcyclin [*Mus musculus*]	*Mm*_Calcyclin	AAA37358.1
gi|179766| calcyclin [*Homo sapiens*]	*Hs*_Calcyclin	AAA51905.1
gi|4753894| calcyclin [*Rattus norvegicus*]	*Rn*_Calcyclin	CAB42002.1
gi|21614544| protein S100-A8 [*Homo sapiens*]	*Hs*_S100A8	NP_002955.2
gi|281485600| protein S100-A8 [*Rattus norvegicus*]	*Rn*_S100A8	NP_446274.2
gi|7305453| protein S100-A8 [*Mus musculus*]	*Mm*_S100A8	NP_038678.1
gi|147906697| centrin [*Xenopus laevis*]	*Xl*_Centrin	NP_001081398.1
gi|63101779| Centrin 3 [*Danio rerio*]	*Dr*_Centrin_3	AAH95095.1
gi|269785117| centrin 3-like protein [*Saccoglossus kowalevskii*]	*Sk*_Centrin_3l	NP_001161514.1
gi|115898527| PREDICTED: centrin-2-like [*Strongylocentrotus purpuratus*]	*Sp*_Centrin_2l	XP_001177362.1
gi|5669593| AF080592_1 centrin [*Mus musculus*]	*Mm*_Centrin	AAD46391.1
gi|149058943| centrin 3 [*Rattus norvegicus*]	*Rn*_Centrin_3	EDM09950.1
gi|414993| centrin [*Homo sapiens*]	*Hs*_Centrin	AAC27343.1
gi|71834632| recoverin [*Danio rerio*]	*Dr*_Recoverin	NP_001025419.1
gi|148236627| recoverin, gene 2 [*Xenopus laevis*]	*Xl*_Recoverin	NP_001087534.1
gi|4506459| recoverin [*Homo sapiens*]	*Hs*_Recoverin	NP_002894.1
gi|6677693| recoverin [*Mus musculus*]	*Mm*_Recoverin	NP_033064.1
gi|18266710| recoverin [*Rattus norvegicus*]	*Rn*_Recoverin	NP_543177.1
gi|585712788| PREDICTED: uncharacterized protein LOC100375153 [*Saccoglossus kowale*	*Sk*_LOC100375	XP_006824981.1
gi|115653113| PREDICTED: uncharacterized protein LOC753346 [*Strongylocentrotus purpu*	*Sp*_LOC753346	XP_001175931.1
gi|72157804|r PREDICTED: uncharacterized protein LOC584267 [*Strongylocentrotus purpu*	*Sp*_LOC584267	XP_800280.1
gi|115738363|PREDICTED: uncharacterized protein LOC753831 [*Strongylocentrotus purpu*	*Sp*_LOC753831	XP_001183518.1
gi|291233601|PREDICTED: uncharacterized protein LOC100369479 [*Saccoglossus kowale*	*Sk*_LOC100369	XP_002736736.1
gi|390353994|PREDICTED: uncharacterized protein LOC100890111 [*Strongylocentrotus pu*	*Sp*_LOC100890	XP_003728236.1
ARI48335.1 Regeneration Associated Protein [*Apostichopus japonicus*]	*Aj*_RegenAP2	ARI48335.1
PIK49419.1 regeneration associated protein [*Apostichopus japonicus*]	*Aj*_RegenAP_B	PIK49419.1
PIK49418.1 regeneration associated protein [*Apostichopus japonicus*]	*Aj*_RegenAP_C	PIK49418.1
XP_022084912.1 uncharacterized protein LOC110976159 [*Acanthaster planci*]	*Ap*_LOC110976	XP_022084912.1
XP_011664021.1 PREDICTED: uncharacterized protein LOC105438199 [*Strongylocentrotu*	*Sp*_LOC105438	XP_011664021.1
uncharacterized protein LOC110977711 [*Acanthaster planci*]	*Ap*_LOC110977	XP_022087750.1

The sequences were found through BLASTp and individually representations for EF-hand families from the NCBI database. Sequences Hs_Calcn (AAC37581.1), Mm_Calcnp (AAA37432.1), and Dr_Calcn_B (NP_956009.1) does not contain predicted EF-hand motifs.

2.7 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Asai, R., Okano, H., & Yasugi, S. (2005). Correlation between Musashi-1 and c-hairy-1 expression and cell proliferation activity in the developing intestine and stomach of both chicken and mouse. *Development Growth and Differentiation*, *47*(8), 501–510. https://doi.org/10.1111/j.1440-169X.2005.00825.x

Bengert, P., & Dandekar, T. (2003). A software tool-box for analysis of regulatory RNA elements, *31*(13), 3441–3445. https://doi.org/10.1093/nar/gkg568

Bias, C., & Gish, W. (1994). Combined Use of Sequence Similarity and Codon Bias for Coding Region Identification. *Journal of Computational Biology*, *1*(1), 39–50.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*, *47*(April), 402–407. https://doi.org/10.1093/nar/gkz297

Castro, E. De, Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-genevaux, P. S., Gasteiger, E., … Hulo, N. (2006). ScanProsite : detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, *34*, 362–365. https://doi.org/10.1093/nar/gkl124

Denessiouk, K., Permyakov, S., Denesyuk, A., Permyakov, E., & Johnson, M. S. (2014). Two Structural Motifs within Canonical EF-Hand Calcium- Binding Domains Identify Five Different Classes of Calcium Buffers and Sensors. *PLoS ONE*, *9*(10), 1–14. https://doi.org/10.1371/journal.pone.0109287

Donato, R, Cannon, B. R., Sorci, G., Riuzzi, F., Hsu, K., Weber, D. J., & Geczy, C. L. (2013). Functions of S100 proteins. *Current Molecular Medicine*, *13*(1), 24–57. https://doi.org/10.2174/1566524011307010024

Donato, Rosario. (1999). Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1450*(3), 191–231. https://doi.org/10.1016/S0167-4889(99)00058-0

Donato, Rosario. (2001). S100: A multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *International Journal of Biochemistry and Cell Biology*, *33*(7), 637–668. https://doi.org/10.1016/S1357-2725(01)00046-2

Donato, Rosario, Sorci, G., Riuzzi, F., Arcuri, C., Bianchi, R., Brozzi, F., … Giambanco, I. (2009). S100B's double life: Intracellular regulator and extracellular signal. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1793*(6), 1008–1022. https://doi.org/10.1016/j.bbamcr.2008.11.009

Edgar, R. C., Drive, R. M., & Valley, M. (2004). MUSCLE : multiple sequence alignment with high accuracy and high throughput, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edgeworth, J., Gorman, M., Bennett, R., Freemont, P., & Hogg, N. (1991). Identification of p8,14 as a highly abundant heterodimeric calcium binding protein complex of myeloid cells. *Journal of Biological Chemistry*, *266*(12), 7706–7713.

Favio, M., Park, Y., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., … Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019, *47*(April), 636–641. https://doi.org/10.1093/nar/gkz268

García-Arrarás, J. E., Estrada-Rodgers, L., Santiago, R., Torres, I. I., Díaz-Miranda, L., & Torres-Avillán, I. (1998). Cellular mechanisms of intestine regeneration in the sea cucumber, Holothuria glaberrima Selenka (Holothuroidea:Echinodermata). *The Journal of Experimental Zoology*, *281*(4), 288–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9658592

García-Arrarás, J. E., & Greenberg, M. J. (2001). Visceral Regeneration in Holothurians. *Microscopy Research and Technique*, *55*(March), 438–451. https://doi.org/10.1002/jemt.1189

Gifford, J. L., Walsh, M. P., & Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. *The Biochemical Journal*, *405*(2), 199–221. https://doi.org/10.1042/BJ20070255

Grabarek, Z. (2006). Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. *Journal of Molecular Biology*, *359*(3), 509–525. https://doi.org/10.1016/j.jmb.2006.03.066

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies : assessing the performance of PhyML 3 . 0 New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hope, K. J., Cellot, S., Ting, S. B., MacRae, T., Mayotte, N., Iscove, N. N., & Sauvageau, G. (2010). An RNAi screen identifies msi2 and prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. *Cell Stem Cell*, *7*(1), 101–113. https://doi.org/10.1016/j.stem.2010.06.007

Huang, H.-Y., Chien, C.-H., Jen, K.-H., & Huang, H.-D. (2006). RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Research*, *34*(Web Server issue), W429-34. https://doi.org/10.1093/nar/gkl333

Jones, D. T. (1999). Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices, *292*, 195–202.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., Mcanulla, C., … Hunter, S. (2014). Sequence analysis InterProScan 5 : genome-scale protein function classification, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Juan, J., Armenteros, A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., … Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kawasaki, H., Nakayama, S., & Kretsinger, R. H. (1998). Classification and evolution of EF-hand proteins. *BioMetals*, *11*(4), 277–295. https://doi.org/10.1023/A:1009282307967

Kharas, M. G., Lengner, C. J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., … Daley, G. Q. (2010). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. *Nature Medicine*, *16*(8), 903–908. https://doi.org/10.1038/nm.2187

Kretsinger, R. H., Tolbert, D., Nakayama, S., & Pearson, W. (1991). The EF-Hand, Homologs and Analogs. *Novel Calcium-Binding Proteins*, 17–37. https://doi.org/10.1007/978-3-642-76150-8_3

Kurimoto, T., Yin, Y., Habboub, G., Gilbert, H.-Y., Li, Y., Nakao, S., … Benowitz, L. I. (2013). Neutrophils express oncomodulin and promote optic nerve regeneration. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(37), 14816–14824. https://doi.org/10.1523/JNEUROSCI.5511-12.2013

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

MacNicol, M. C., Cragle, C. E., & MacNicol, A. M. (2011). Context-dependent regulation of Musashi-mediated mRNA translation and cell cycle regulation. *Cell Cycle*, *10*(1), 39–44. https://doi.org/10.4161/cc.10.1.14388

Malmendal, A., Carlström, G., Hambraeus, C., Drakenberg, T., Forsén, S., & Akke, M. (1998). Sequence and Context Dependence of EF-Hand Loop Dynamics. An 15 N Relaxation Study of a Calcium-Binding Site Mutant of Calbindin D9k. *Biochemistry*, (37), 2586–2595. https://doi.org/10.1021/bi971798a

Marchler-bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., … Bryant, S. H. (2017). CDD / SPARCLE : functional classification of proteins via subfamily domain architectures, *45*(November 2016), 200–203. https://doi.org/10.1093/nar/gkw1129

Mashanov, V. S., Zueva, O. R., & Garcia-Arraras, J. E. (2010). Organization of glial cells in the adult sea cucumber central nervous system. *Glia*, *58*(13), 1581–1593. https://doi.org/10.1002/glia.21031

Mashanov, V. S., Zueva, O. R., & García-Arrarás, J. E. (2013). Radial glial cells play a key role in echinoderm neural regeneration. *BMC Biology*, *11*, 49. https://doi.org/10.1186/1741-7007-11-49

Mashanov, V. S., Zueva, O. R., & García-Arrarás, J. E. (2014). Transcriptomic changes during regeneration of the central nervous system in an echinoderm. *BMC Genomics*, *15*(1), 1–21. https://doi.org/10.1186/1471-2164-15-357

Maurer, P., Hohenadl, C., Hohenester, E., Göhring, W., Timpl, R., & Engel, J. (1995). The C-terminal Portion of BM-40 (SPARC/Osteonectin) is an Autonomously Folding and Crystallisable Domain that Binds Calcium and Collagen IV. *Journal of Molecular Biology*, *253*(2), 347–357. https://doi.org/10.1006/jmbi.1995.0557

Moncrief, N. D., Kretsinger, R. H., & Goodman, M. (1990). Evolution of EF-hand Calcium-modulated Proteins. I. Relationships Based on Amino Acid Sequences. *Journal of Molecular Evolution*, *30*(6), 522–562. https://doi.org/Doi 10.1007/Bf02101108

Morgan, R. O., Martin-Almedina, S., Garcia, M., Jhoncon-Kooyip, J., & Fernandez, M. P. (2006). Deciphering function and mechanism of calcium-binding proteins from their evolutionary imprints. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1763*(11), 1238–1249. https://doi.org/10.1016/j.bbamcr.2006.09.028

Mutus, B., Palmer, E. J., & MacManus, J. P. (1988). Disulfide-Linked Dimer of Oncomodulin: Comparison to Calmodulin. *Biochemistry*, *27*(15), 5615–5622. https://doi.org/10.1021/bi00415a033

Nakamura, M., Okano, H., Blendy, J. A., & Montell, C. (1994). Musashi, a neural RNA-binding protein required for Drosophila adult external sensory organ development. *Neuron*, *13*(1), 67–81. https://doi.org/10.1016/0896-6273(94)90460-X

Nelson, M. R., & Chazin, W. J. (1998). Structures of EF-hand Ca2+-binding proteins: Diversity in the organization, packing and response to Ca2+ binding. *BioMetals*, *11*(4), 297–318. https://doi.org/10.1023/A:1009253808876

Okano, H., Kawahara, H., Toriya, M., Nakao, K., Shibata, S., & Imai, T. (2005). Function of RNA-binding protein Musashi-1 in stem cells. *Experimental Cell Research*, *306*(2), 349–356. https://doi.org/10.1016/j.yexcr.2005.02.021

Ortiz-Pineda, P. A. (2010). Analysis and characterization of genes associated with intestinal regeneration in the sea cucumber (Echinodermata: Holothuroidea) (UMI No. 3412547) (Doctoral dissertation, University of Puerto Rico Río Piedras Campus, San Juan, Puerto Rico). Available from ProQuest Dissertations Publishing. https://search.proquest.com/docview/743816791

Persechini, A., Moncrief, N. D., & Kretsinger, R. H. (1989). The EF-hand family of calcium-modulated proteins. *Trends in Neurosciences*, *12*(11), 462–467. https://doi.org/10.1016/0166-2236(89)90097-0

Petersen, T. N., Brunak, S., Heijne, G. Von, & Nielsen, H. (2011). Correspondence SignalP 4.0 : discriminating signal peptides from transmembrane regions. *Nature Publishing Group*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Phan, E., Ahluwalia, A., & Tarnawski, A. S. (2007). Role of SPARC-matricellular protein in pathophysiology and tissue injury healing: Implications for gastritis and gastric ulcers. *Med. Sci. Monit.*, *13*(2), RA25-30.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., … Res, A. (2005). InterProScan : protein domains identifier, *33*, 116–120. https://doi.org/10.1093/nar/gki442

Rigden, D. J., & Galperin, M. Y. (2004). The DxDxDG motif for calcium binding: Multiple structural contexts and implications for evolution. *Journal of Molecular Biology*, *343*(4), 971–984. https://doi.org/10.1016/j.jmb.2004.08.077

Rojas-Cartagena, C., Ortíz-Pineda, P., Ramírez-Gómez, F., Suárez-Castillo, E. C., Matos-Cruz, V., Rodríguez, C., … García-Arrarás, J. E. (2007). Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber Holothuria glaberrima. *Physiological Genomics*, *31*(2), 203–215. https://doi.org/10.1152/physiolgenomics.00228.2006

Sakakibara, S., Nakamura, Y., Yoshida, T., Shibata, S., Koike, M., Takano, H., … Okano, H. (2002). RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15194–15199. https://doi.org/10.1073/pnas.232087499

San Miguel-Ruiz, J. E., & García-Arrarás, J. E. (2007). Common cellular events occur during wound healing and organ regeneration in the sea cucumber Holothuria glaberrima. *BMC Developmental Biology*, *7*(1), 115. https://doi.org/10.1186/1471-213X-7-115

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., … Ostell, J. (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *47*(D1), D23–D28. https://doi.org/10.1093/nar/gky1069

Schaub, M. C., & Heizmann, C. W. (2008). Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies. *Biochemical and Biophysical Research Communications*, *369*(1), 247–264. https://doi.org/10.1016/j.bbrc.2007.10.082

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675.

Suárez-Castillo, E. C., & García-Arrarás, J. E. (2007). Molecular evolution of the ependymin protein family: a necessary update. *BMC Evolutionary Biology*, *7*(1), 23. https://doi.org/10.1186/1471-2148-7-23

Takahashi, T., Suzuki, H., Imai, T., Shibata, S., Tabuchi, Y., Tsuchimoto, K., … Hibi, T. (2013). Musashi-1 post-transcriptionally enhances phosphotyrosine-binding domain-containing m-Numb protein expression in regenerating gastric mucosa. *PloS One*, *8*(1), e53540. https://doi.org/10.1371/journal.pone.0053540

Vivo, S. P. H., Koltzscher, M., & Gerke, V. (2000). Identification of Hydrophobic Amino Acid Residues Involved in the Formation of, 9533–9539.

Wang, X.-Y., Yin, Y., Yuan, H., Sakamaki, T., Okano, H., & Glazer, R. I. (2008). Musashi1 modulates mammary progenitor cell expansion through proliferin-mediated activation of the Wnt and Notch pathways. *Molecular and Cellular Biology*, *28*(11), 3589–3599. https://doi.org/10.1128/MCB.00040-08

Yao, X., Liu, Y., Tan, Y., & Song, Y. (2016). The complete chloroplast genome sequence of Helwingia himalaica (Helwingiaceae , Aquifoliales) and a chloroplast phylogenomic analysis of the Campanulidae. https://doi.org/10.7717/peerj.2734

Yin, Y., Cui, Q., Gilbert, H.-Y., Yang, Y., Yang, Z., Berlinicke, C., … Benowitz, L. I. (2009). Oncomodulin links inflammation to optic nerve regeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19587–19592. https://doi.org/10.1073/pnas.0907085106

Yin, Y., Cui, Q., Li, Y., Irwin, N., Fischer, D., Harvey, A. R., & Benowitz, L. I. (2003). Macrophage-derived factors stimulate optic nerve regeneration. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(6), 2284–2293. https://doi.org/23/6/2284 [pii]

Yin, Y., Henzl, M. T., Lorber, B., Nakazawa, T., Thomas, T. T., Jiang, F., … Benowitz, L. I. (2006). Oncomodulin is a macrophage-derived signal for axon regeneration in retinal ganglion cells. *Nature Neuroscience*, *9*(May), 843–852. https://doi.org/10.1038/nn1701

Chapter 3 Expression and Purification of Recombinant Orpin Proteins

3.1 Abstract

Orpin A and Orpin B are putative EF-hand protein isoforms from *Holothuria glaberrima* that are highly similar. Previous bioinformatics analyses predicted N-terminal signal peptides along with a pair of EF-hand motifs. To further characterize these putative proteins, their expression and purification is required. However, Orpins are difficult to express in soluble form by standard methods alone. A genetically modified version of Orpin B was expressed and purified combining previous methods with predicted calcium-binding activity. In this study, we present two protocols that were developed to avoid the solubility problems.

The first protocol consisted in the production of a modified version of the Orpin B open reading frame (ORF) in order to be expressed and purified in a soluble form called His-OBSm. His-OBSm is His tagged and contains two additional genetic modifications: the deletion of the signal peptide encoding sequence and the addition of a 25 residues peptide from the pET200 plasmid vector. The addition of this peptide at the C-terminal, made possible the purification of a recombinant version of Orpin B in a soluble form.

The second protocol was developed in order to produce a soluble recombinant Orpin B that best resembles the original protein, without the addition of the C-terminal peptide as in His-OBSm. This protocol consisted of cloning the Orpin B coding sequence lacking the signal peptide region, into a pGEX plasmid as a fusion partner of GST.

For this protocol, we tested different combinations of expression conditions, additive combinations, and protein extraction methods such as lowering the expression induction temperature, OD_{600}, IPTG concentrations, host cells, growth media, oxygenation conditions, supplementation with D-glucose and GlyGly, and cell lysis and protein extraction with BugBuster® and B-PER™ reagents. We identified that the best parameters to increase GST-OBS expression and solubility, were to express the plasmid into a BL21 (DE3)pLysS cell line grown in LB pre-culture overnight, and then diluted in TB media supplemented with 1% D-glucose and growth at 37°C until OD_{600} of 0.5–0.7 at low oxygenation / low mechanical stress conditions. The cells were then induced with 0.5 mM IPTG at 22°C for 16 h. The harvested cell pellets were lysed for the protein extraction with B-PER™ reagent protocol to minimize the harsh sonication conditions.

Acceptable levels of expression and solubility were also achieved with BL21 (DE3)-RIL cell lines and with the media supplementation of GlyGly. However, glucose supplementation was more readily available and cost-effective. Thus, these materials can be considered as an alternative for the production of GST-OBS. These strategies can be used to increase the soluble expression and purification of other EF-hand proteins that are difficult to express by standard methods.

3.2 Introduction

In the previous chapter, we described Orpin as an EF-hand containing putative protein identified from the sea cucumber *Holothuria glaberrima.* This protein has two highly similar isoforms that we have named Orpin A and Orpin B. These putative proteins contain N-terminal signal peptide and two EF-hand motifs that suggest they might be Ca^{2+}-binding proteins. Nonetheless, our characterization was limited to sequence and bioinformatics analyses and measurements of the expression of *Orpin* mRNA.

In order to further characterize Orpin, we needed to develop new tools. On the one hand, it was necessary to obtain the expressed protein to be able to perform functional assays. On the other hand, the expressed protein could also be used to produce antibodies that will serve as tools to further explore Orpin organ distribution and/or cellular localization. Specifically, one of our goals as part of the characterization of EF-hand containing Orpin isoforms was to determine the cells that expressed said proteins. A requirement to achieve this goal was the availability of Orpin-specific antibodies. Being Orpin a novel protein, no commercial antibodies were available. Thus, we focused on the production of antibodies that recognize this protein. An initial step toward this goal was to develop the methodology to produce a recombinant version of Orpin.

A first step toward antibody production is the need of the target protein in a highly purified form. This led us to attempt to express and purify a recombinant

version of Orpin B in bacteria to be used as an immunogen. For this, it was necessary to produce the protein in a soluble form to obtain the required amounts for the antigen to induce the specific immunological response. This chapter presents the multiple attempts to express the Orpin protein and the hurdles that had to be overcome in order to obtain a product. We show here different strategies that can be performed in order to produce proteins that are difficult to express, and specific strategies that can be applied to EF-hand calcium-binding proteins (CaBPs).

3.3 Materials and Methods

3.3.1 Preparation of recombinant constructs

The open reading frame (ORF) of *Orpin B* was amplified by PCR from the cDNA product of RNA extraction from a somatic muscle tissue of *Holothuria glaberrima.* Oligodeoxyribonucleotide primers were chemically synthesized (IDT, Integrated DNA Technologies), to amplified from the initial codon as well (His-Orpin B) or immediately upstream the signal peptide encoding region to delete this portion (His-OBS), and another primer was designed to avoid the ORF stop codon in order to use the stop from the plasmid, downstream the insertion site (His-OBSm). In addition, we followed the appropriate indications for the forward primer design suitable for in-frame cloning into the expression vectors, with N-terminal 6X-histidine tags based on the champion pET200 Directional TOPO Expression Kit (Invitrogen, Cat. No. K2001) manufacturer's instructions. The primers used for the

amplification are shown in Supplemental Table 3.1. The cloning into pET200/D-TOPO plasmids was performed following the manufacturer's protocol. The recombinant proteins were expressed as N-terminal His-tagged fusions. The positive clone sequences were confirmed by Sanger sequencing excised electrophoresed sample bands at the Sequencing and Genotyping Facility (UPR-RP) analyzed through Geneious software 11.1.5 (https://www.geneious.com).

3.3.2 GST-OBS cloning

The sequences were cloned by restriction-free (RF) and Exponential Mega Priming PCR (EMP) cloning protocol (Lund, Leiros, Elin, & Bjerga, 2014). For this purpose, specific oligonucleotides that shared regions with the *Orpin B* insert and with the pGEX-6P1 plasmid were synthesized to be inserted by means of RT-PCR. These "mega primers" were used to amplify an oligonucleotide that incorporated the *Orpin B* ORF encoding region, using a sample of somatic muscle cDNA as a template. After verifying the presence of the insert by means of agarose electrophoresis, the insertion was carried out into the plasmid. The synthesis (PCR insertion) was carried out using a high fidelity Taq polymerase, PfuUltra II Fusion HS DNA Polymerase (Agilent) in 1x PfuUltra II reaction buffer. After the sequence was inserted, the original plasmids, that did not contain the insert, were degraded. These "parental" plasmids were methylated and therefore, they were degraded by using the *Dpn*I enzyme. Given the fact that BL21 strains had the *hsd*SB mutation that disrupts methylation and degradation of the DNA (Rosano & Ceccarelli, 2014), the pGEX-6P1 plasmid was transformed into an *E. coli* Star™ strain that lacked

this mutation, followed by a plasmid extraction that yielded methylated versions of the plasmid. This plasmid was used for the RT-PCR insertion. The cloned plasmid samples were incubated with the *Dpn*I enzyme according to the manufacturer's protocol at 37°C for 3 h. The enzyme was then inactivated by incubation for 20 min at 80°C. The presence of the generated vector was monitored using agarose electrophoresis. The GST-OBS sequence was confirmed by Sanger sequencing excised electrophoresed sample bands at the Sequencing and Genotyping Facility (UPR-RP) analyzed through Geneious software 11.1.5 (https://www.geneious.com)

3.3.3 *Orpin B* ORF without signal peptide region PCR product

PCR amplification was prepared combining 5 µL of 10x Taq polymerase buffer, 1.0 µL of deoxyribonucleoside triphosphate (dNTP), 3.0 µL of $MgCl_2$, 2.5 µL of each primer (Supp. Figure 3.1), 2.5 µL of DMSO, 3 µL (637.8 ng) of cDNA, 0.5 µL of Taq polymerase, and completed with MilliQ water to a final volume of 50 µL. The samples were denatured by heating to 95°C for 4 min, followed by 40 cycles of 30 s denaturation at 94°C, 30 s annealing at 60°C and 50 s elongation at 72°C, and a final elongation step of 5 min at 72°C. PCR products were analyzed by 1% agarose gel electrophoresis.

3.3.4 PCR insertion of *OBS* product into pGEX-6P1 plasmid

For the PCR insertion, 200 ng of the PCR product were combined with 100 ng of the plasmid vector (pGEX-6P1), 0.5 µL of dNTP, 2.5 µL of 10x PfuUltra II buffer,

0.5 µL of PfuUltra II polymerase, and completed with MilliQ water to a final volume of 25 µL. The samples were denatured by heating to 95°C for 30 s, followed by 18 cycles of 30 s denaturation at 95°C, 30 s annealing at 55°C, and 12 min elongation at 68°C. PCR products were analyzed by 1% agarose gel electrophoresis. After incubation with *DpnI* enzyme, the cloned plasmids were verified for the correct inserted sequence with two additional PCR reactions following the previous *OBS* PCR product amplification program. One amplification included GST-OBS primers (annealing: 65°C) and another one included pGEX-6P1 primers (annealing: 55°C). These samples were sequenced in duplicates at the Sequencing & Genotyping Facility (UPR-RP) and the results were evaluated on Geneious 11.1.5 software (https://www.geneious.com).

3.3.5 Cell culture and recombinant proteins expression

3.3.5.1 His tag fusion proteins

Competent *E. coli* One-Shot™ BL21 Star™ (DE3) and DH5α (Invitrogen) competent cell lines were transformed with the pET200/D-TOPO plasmids by heat shock for the plasmid propagation during cloning. After positive clones were confirmed, the plasmids were extracted as minipreps and transformed into competent BL21 (DE3)-RIL cells (Invitrogen) as the expression vector. Colonies were grown overnight on Luria Bertani (LB) plates (Sambrook, W., Russell, & Coldspring Harbor Laboratory (2001), n.d.) containing kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL). Fresh colonies were first inoculated into 10 mL media containing the appropriate antibiotics and grown overnight at 37°C with shaking of

225 rpm. These pre-cultures were diluted into 1 L of LB medium with the appropriate antibiotics and incubated at 37°C with shaking until OD_{600} 0.5–0.7. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactosidase (IPTG) for 16 h at 37°C (His-Orpin B) or 22°C (His-OBS and His-OBSm) with shaking. A similar protocol was followed for the His-β-galactosidase control protein.

3.3.5.2 GST-OBS and GST Proteins

A seeding pre-culture of 30 mL of LB media containing ampicillin (100 μg/mL) and chloramphenicol (50 μg/mL) was inoculated with transformed BL21 (DE3)pLysS (or BL21 (DE3)-RIL) cells with either GST-OBS or GST plasmid, from a 50% glycerol stock and was incubated at 37°C overnight. Later the pre-culture was diluted in (1 L to 3 L) TB media (Sambrook et al., n.d.) (or LB as specified in the corresponding experimental designs) supplemented with 1% D-glucose (some experiments did not include this additive as specified by the corresponding experimental design) and incubated at 37°C until OD_{600} 0.5–0.7 (or OD_{600} 0.7–0.8; OD_{600} 1.0; OD_{600} 1.3 in specified experiments), then samples were induced with 0.5 mM (or 1.0 mM) IPTG at 22°C (37°C or RT in specified experiments) for 16 h.

3.3.6 Cell lysis and protein extraction

3.3.6.3 His tag fusion proteins

The induced cells were harvested by centrifugation at 11,000 x g and 4°C for 35 min on a Sorvall Lynx 4000 Centrifuge and the cell pellets were stored at −80°C

overnight. Later, these cell pellets were resuspended in extraction buffer (50 mM Tris mM, 150 mM NaCl, 1 mM DTT, 20% glycerol, pH 8.0) with lysozyme (10 mg/mL), DNase (0.1 mg/mL), and 10% of 1x proteases inhibitors cocktail (Pierce) on ice for 30 min, followed by sonication 3 times with a burst duration of 30 s each. The sonicated lysates were centrifuged at the previous conditions, and the supernatant containing the soluble proteins were collected into 15 mL tubes.

3.3.6.4 GST-OBS and GST proteins

The cell lysis carried out with B-PER™ Bacterial Protein Extraction Reagent (ThermoFisher Scientific) or BugBuster® Protein Extraction Reagent (EMD Millipore, Cat. No. 70921-4) were performed following the corresponding manufacturers' protocols using 1 µL Benzonase® Nuclease (Novagen) for the DNA degradation (for each 1 mL of supernatant) and 10% of 1x Proteases Inhibitors Cocktail (Pierce™).

3.3.7 Protein purification

3.3.7.5 Affinity purification of His-Orpin B, His-OBS, His-OBSm, and His-β-galactosidase

The soluble lysates were incubated with nickel and nitrilotriacetic acid coupled to a cross-linked 6% agarose resin (Ni-NTA) (Sigma-Aldrich, Cat No. P6611-100ML) that was previously equilibrated with the extraction buffer without the lysozyme, DNase, and protease inhibitor cocktail, at 2–8°C for 30 min shaking on

a rocker. Then, the flow-through was collected by gravity and the resin was washed 5 times (15 mL each) with a buffer of 50 mM Tris, 150 mM NaCl, 1 mM DTT, 20% glycerol, pH 8.0 (10 mM imidazole were also included to the wash buffer for His-OBSm only and no imidazole was added to the wash buffer before purification optimization). The resin was incubated with elution buffer of 50 mM Tris, 150 mM NaCl, 1 mM DTT, 20% glycerol, pH 8.0, including 200 mM imidazole and 10% of 1x proteases inhibitors cocktail at 2–8°C for 30 min shaking on a rocker. The eluted samples (10 mL each) were collected by gravity on 15 mL tubes.

3.3.7.6 Glutathione-coupled resin affinity purification

Supernatants from lysed cell pellets were resuspended with glutathione coupled resin beads (1 mL = 1 CV) and incubated in a cold room at 2–8°C in a test tube rocker for ~ 1 h. Then, the flow-through material was collected by gravity and the resin was washed with 30 CV of wash buffer. The elution was done by incubation of the resin beads with the elution buffer ~ 1 h (For the double incubation experiments, the resin was washed again at this step, incubated again with the initial flow-through, and washed again as before). Later, the elution was done by incubating the elution buffer for 30 min each time, rocking at 2–8°C and collected by gravity. Samples of 50 µL were collected at each step for SDS-PAGE analysis and stored at –20°C. The eluted samples concentration determination was done by NanoDrop Spectrophotometer (Thermo).

3.3.7.7 GST-OBS and GST 1 - 3 L scale-up purification

For the experiment of optimal purification wash buffer determination, prior to cell harvesting at 4°C and 11,000 x g for 35 min, two samples of 10 mL each, were collected from the induced cell culture with either GST-OBS or GST plasmid (The remaining cell pellet stored at −20°C from the 1 L cell culture was sonicated and purified as described above). For the scale-up experiments, the harvested cell pellet from the corresponding cell cultures (1 L to 3 L) were used completely for the corresponding purification procedures. These samples were centrifuged as described in the other methodologies and stored at −80°C overnight. One of the wash buffers was composed of 50 mM Trizma, 150 mM NaCl, and 10% glycerol pH 8.0, and the other was composed of Phosphate-buffered saline (PBS) buffer at pH 7.4. The affinity purification was performed using glutathione-coupled resin following the method described previously. The washing steps were done using either of the mentioned buffers. For the elution steps, 20 mM of reduced GSH was added to the corresponding buffer. A similar protocol was followed for the GST control protein.

3.3.8 GST-OBS and GST purification for $CaCl_2$ solubilization experiments

[100 mM] $CaCl_2$ stock solution

0.147 g of $CaCl_2$ x $2H_2O$ were dissolved in 10 mL of Chelex® 100 (Sigma-Aldrich) treated dH_2O (Sambrook et al., n.d.).

[100 mM] EGTA stock solution

0.380 g of EGTA were dissolved in 10 mL of Chelex® 100 (Sigma-Aldrich) treated dH_2O and adjusted to pH 8.0 (Sambrook et al., n.d.).

Transformed BL21 (DE3)pLysS cells from a 50% glycerol stock were grown at 37°C in 30 mL of LB media with kanamycin (50 µg/mL) and ampicillin (100 µg/mL), overnight as a pre-culture. This material was used to inoculate 2 L of TB media supplemented with 1% D-glucose, ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL), until OD_{600} ~ 0.5, then induction was done with 0.5 mM IPTG at 22°C for 16 h. Cells were harvested by centrifugation and stored overnight at −80°C. Pellet samples of 0.514 g were used for the extraction and purification. Cell pellets were lysed in lysis buffer: 576 µL of B-PER™ (ThermoFisher, Cat. No. 78248), 128 µL of 1x proteases inhibitors cocktail, 2 µL of Benzonase® Nuclease, 281.6 µL of 100 mM $CaCl_2$ stock solution (or 100 mM EGTA stock solution), and 422.4 µL of Chelex® 100 (Sigma-Aldrich) treated dH_2O (This solution was done for the 20 mM $CaCl_2$ buffer, and corresponding adjustments were done for 0, 2.5, 5.0, 10.0, and 50 mM $CaCl_2$ or EGTA experiments); with a metal spatula and by pipetting repeatedly. Then the samples were kept rocking at RT for 55 min and centrifuged at 15,000 x g and 4°C for 10 min. The supernatant was removed and mixed with the glutathione-coupled resin for 45 min at RT in a rocker. A volume of 52 µL from 50% slurry from glutathione coupled resin was used for the purification of each sample. The resin was centrifuged and the flow-through material was stored at 2–

8°C for a second incubation. The corresponding resin mixture was centrifuged and washed 3 times with 200 µL of cold PBS at pH 7.4.

The initial flow-through was incubated again with the washed resin for 45 min at RT in a rocker. The resin mixture was centrifuged and the flow-through discarded. The resin was washed again with 200 µL of cold PBS at pH 7.4 and the elution was done with 60 µL of 20 mM reduced glutathione diluted in the wash buffer. Samples from the eluted samples were diluted in loading buffer and analyzed through a 15% SDS-PAGE.

3.3.9 SDS-PAGE

Sodium dodecyl sulfate (SDS) Resolving gels were prepared at 10% or 15% of polyacrylamide with 1 mm of thickness. The samples (30 µL) were diluted 1:1 with SDS loading dye with the exception of pellet samples that were diluted in a 3:1 ratio, boiled at 90°C for 5 min, cooled down on ice for 1 min, and centrifuged before adding 30 µL into the gel wells. The molecular weight marker used was Precision Plus Protein™ Standards Dual Color (Bio-Rad, Cat. No. 161-0374). The electrophoresis was run 30 min at 50 V (during migration through the stacking gel portion) and 35–45 min at 150 V (during migration through the resolving gel portion) to increase band sharpness. Then, the gels were cleaned with dH_2O to remove the SDS residues from the running buffer and stained with Coomassie Brilliant Blue R250 Dye (Thermo Scientific, Cat. No. 20278) solution (heated for 30 s in a microwave) for 30 min. The gels were cleaned with dH_2O and destained with

a 10% acetic acid, 45% methanol, 45% dH$_2$O solution until the bands were visible. Kim wipe papers were submerged into the destaining solution with the gels to accelerate the destaining time by absorption of the Coomassie excess. After that, the gels were cleaned with dH$_2$O and rehydrated. This was seen to increase the band stain intensity and resolution.

3.3.10 Western Blots

3.3.10.8 Buffer preparation

TBS 10x buffer

   30.3 g of Tris and 43.8 g of NaCl were diluted in 500 mL of dH$_2$O and adjusted to pH 7.6.

TBS-T buffer

   50 mL of TBS 10x buffer and 500 µL of Tween 20 were diluted in 500 mL of dH$_2$O.

SDS running buffer 10x

   144 g of glycine and 30 g of Tris Base were diluted in 100 mL of 10% SDS solution.

Transfer buffer

   100 mL of SDS Running Buffer 10x and 200 mL of methanol were diluted in 700 mL of dH$_2$O.

3.3.10.9 Protein band transference to nitrocellulose membranes

The different purified protein samples were electrophoresed on 15% polyacrylamide gels of 1 mm thickness following the described SDS-PAGE methodology. The molecular weight marker used was Precision Plus Protein™ Standards Dual Color (Bio-Rad, Cat. No. 161-0374). The gels were transferred to Hybond-C Extra nitrocellulose membranes (GE, Cat. No. RPN303E) using an Owl™ VEP-2 Electroblotting System (ThermoFisher) for the Western Blot (ThermoFisher Scientific) for 3 h at 30 V. The corresponding gels were stained with Coomassie Brilliant Blue R250 Dye (Thermo Scientific, Cat. No. 20278) solution and then destained as in the SDS-PAGE methodology, to confirm that the protein bands were transferred to the membranes. The nitrocellulose membranes were blocked with TBS-T buffer with 0.5% dry milk for 1 h at RT in motion with a rocker. The membranes were washed three times with TBS-T for 15 min each time. Then the membranes were incubated with the His-OBSm anti-serum diluted 1:1000 in TBS-T buffer and 5% dry milk (10 µL/10 mL) at 2–8°C overnight in a rocker. Membranes were washed three times again as before and incubated with the secondary antibody, HRP-conjugated, anti-mouse IgG from sheep (GE, Cat. No. NXA931) diluted 1:5000 in TBS-T buffer and 5% dry milk (2 µL/10 mL). The membranes were washed again as before, including an additional wash of TBS for 10 min to remove the Tween 20. The antibody cross-reactivity was detected by a chemiluminescent method using SuperSignal™ West Dura Extended Duration Substrate (ThermoFisher Scientific, Cat. No. 37071) following the manufacturer's protocol and visualized in a ChemiDoc XRS+ (BioRad).

3.3.11 Size exclusion chromatography purification

The recombinant GST-OBS was purified through a size exclusion chromatography column, Superdex 200 10/300 GL using an AKTA Explorer system FPLC. The sample was buffer exchanged after the Ni-NTA purification to 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.8 (including DTT and proteases inhibitors and phosphatase inhibitors concentrations) using an Amicon Ultra 10 K (500 µL to 50 µL each centrifugation, 6 times) (Millipore). The sample was filtered through a 0.22 µm filter and injected with a glass syringe to the AKTA system. The sample was run into the column at 0.5 mL/min flowrate. The separation was monitored by UV and fractions for each peak (13.0 mL, 14.5 mL, 18.5 mL, 20.5 mL, and 21.5 mL) were collected. Each fraction sample was analyzed through SDS-PAGE following the procedure described before.

3.4 Results

Initial attempts to express the Orpin protein included:

1. Producing the recombinant protein fused to a poly-His tag by cloning the coding sequence into a pET200/D-TOPO plasmid vector.

2. Deleting the hydrophobic transmembrane encoding region (signal peptide amino acid sequence described in the previous chapter).

3. Lowering the induction temperature.

4. Inducing protein expression at low IPTG concentrations, and at different growth states (OD_{600}).

In all these expression and purification attempts, the different recombinant versions of Orpin were difficult to express and were produced in an insoluble form. The proteins could not be solubilized from the cell pellet after expression and cell lysis as no proteins of the corresponding molecular weight were detected in the eluted samples after purification. The band corresponding to the molecular weight of the overexpressed protein was only detected in the cell pellets after treatment with SDS and electrophoresis. Thus, we proceeded to develop different strategies to be able to produce a soluble form of Orpin. The two main methodologies consisted in:

1. Producing a His-tagged Orpin B fused to a short peptide at the C-terminal that increased its solubility. This sequence version was inserted into a PET200/D-TOPO plasmid and expressed into BL21 (DE3)-RIL cells (See Section A. Expression Protocol 1).

2. Producing a Glutathione *S*-transferase (GST) fused version cloned into a different plasmid (pGEX-6P1) and expressed into different cell lines (BL21 (DE3)pLysS or BL21 (DE3)-RIL) through different conditions, growth mediums and including the presence of additives such as D-glucose and/or glycylglycine (GlyGly). Using this expression protocol, we also showed that the addition of $CaCl_2$ during the extraction procedure increased the amount of purified protein that could be recovered in soluble form (See Section B. Expression Protocol 2).

Here we detailed the expression/purification procedures and provide the results obtained:

3.4.1 A. EXPRESSION PROTOCOL 1- Production of a His-tagged Orpin B with a small peptide sequence in the C-terminal (His-OBSm).

3.4.1.10 Increasing the soluble extraction of recombinant Orpin B by fusing to a poly-His tag and attaching a peptide at the C-terminal (His-OBSm).

Initially, we cloned the *Orpin B* sequence into a vector and expressed it as a polyhistidine (6XHis) tagged protein into *E. coli* BL21 (DE3) cells. However, the protein could not be recovered in a soluble form as the expected protein band was observed only in the insoluble pellet. Thus, an evaluation of the primary translated amino acid sequence was performed in order to generate a version of the peptide that could be produced in a soluble stable form. As discussed in the previous chapter, the amino-terminal portion of the sequence has a predicted signal peptide in its N-terminal portion. Signal peptides of secreted proteins are known to have a high composition of hydrophobic amino acids as this segment is targeted to a transmembrane region of cells that express these types of proteins (von Heijner, 1990; Zheng & Gierasch, 1996). These hydrophobic stretches are able to interact with similar regions when expressed at high concentrations and can lead to protein instability and aggregation (Carrió & Villaverde, 2002; Hartley & Kane, 1988).

The signal peptides of secreted proteins are eventually cleaved from the functional, final native form (von Heijner, 1990; Zheng & Gierasch, 1996). As mentioned previously, Orpin A and Orpin B sequences have a predicted cleavage site between residues 20 and 21 that is located between the predicted signal

peptide and the non-cytoplasmic portion of each sequence (Ortiz-Pineda et al., 2009).

The elimination of the signal peptide encoding nucleotides from the sequence inserted into the plasmid is an alternative modification to decrease the overall hydrophobicity of the expressed peptide and can help to improve its solubilization following cell lysis. Thus, we proceeded to delete the nucleotide sequence that codes for this hydrophobic portion from the insert and cloned the new (shorter) nucleic acid sequence into a pET200/D-TOPO vector to be able to increase the solubility of the recombinant peptide and make possible its purification. This modification alone was not enough for the soluble expression of this version of the peptide, as the band of the expected weight (37.8 kDa) was not detected in any of the elution samples (Data not shown).

We then evaluated the sequence hydrophobicity potential using bioinformatics analyses and found that if the original stop codon were deleted, translation would continue until encountering the next stop codon from the plasmid vector. This would produce a recombinant peptide that included the complete amino acid sequence for the expected final native Orpin B but it would also include 25 additional residues from the corresponding portion of the vector. By doing this, we could achieve a reduction in the GRAVY (Grand Average of Hydropathicity) number, increasing its solubility potential (Kyte & Doolittle, 1982) (Table 3.1). This

sequence would be a suitable candidate for the production of a soluble protein that

would include the important epitopes for antibody production.

Table 3.1. Solubility prediction of the different protein sequences

Protein Sequence	GRAVY
Orpin B	0.025
OBS	−0.355
His-Orpin B	−0.328
His-OBS	−0.660
His-OBSm	−0.712
GST	−0.410
GST- Orpin B	−0.210
GST- OBS	−0.344

The solubility potential was based on the comparison to the value associated with the GST protein sequence, which is considered highly soluble. Hence, a low negative GRAVY value correlates to high solubility potential. The low negative number of the GRAVY value associated as compared to the GST sequence value shows an increase in the predicted solubility of the His's-OBSm peptide. Thus, the addition of the 25 residues from the pET200/D-TOPO plasmid insertion site to the C-terminal of His-OBS coupled to the deletion of the N-terminal signal peptide sequence, were expected to increase the solubility potential of the recombinant peptide. Note that even though GST is considered highly soluble, the fusion protein GST-OBS associated values displayed a low solubility potential for the expressed sequence. GRAVY: "Grand Average of Hydropathicity". Orpin B: Orpin A isoform sequence, OBS: Orpin B with the signal peptide sequence deletion, His-Orpin B: Orpin B sequence fused to a 6XHis N-terminal tag, His-OBS: 6XHis tag fused to OBS sequence, His-OBSm: His-OBS sequence with the additional 25 residues from the insertion site of the pET200/D-TOPO plasmid to the C-terminal, GST: Glutathione S-transferase sequence, GST-Orpin B: GST tag fused to Orpin B sequence, and GST-OBS: GST tag fused to OBS sequence.

The expression methodology consisted in the transformation of BL21 (DE3)-RIL cells and their pre-culture of 10 mL of medium with the corresponding antibiotics (kanamycin and chloramphenicol) at 37°C overnight. The pre-culture was then diluted in 1 L of medium with the same antibiotics and grown to an OD_{600} of 0.7-0.8. At this point, the expression of the recombinant protein was induced with 1 mM IPTG for 16 h, incubating in motion until the cells were harvested. These

cells were lysed in a buffer compatible with Ni-NTA (Nickel bound to nitriloacetic acid coupled to agarose beads) resin using sonication, in the presence of lysozyme, DNase and protease inhibitors. We collected 10 mL fractions corresponding to two elution steps from the Ni-NTA resin (Fig. 3.1).
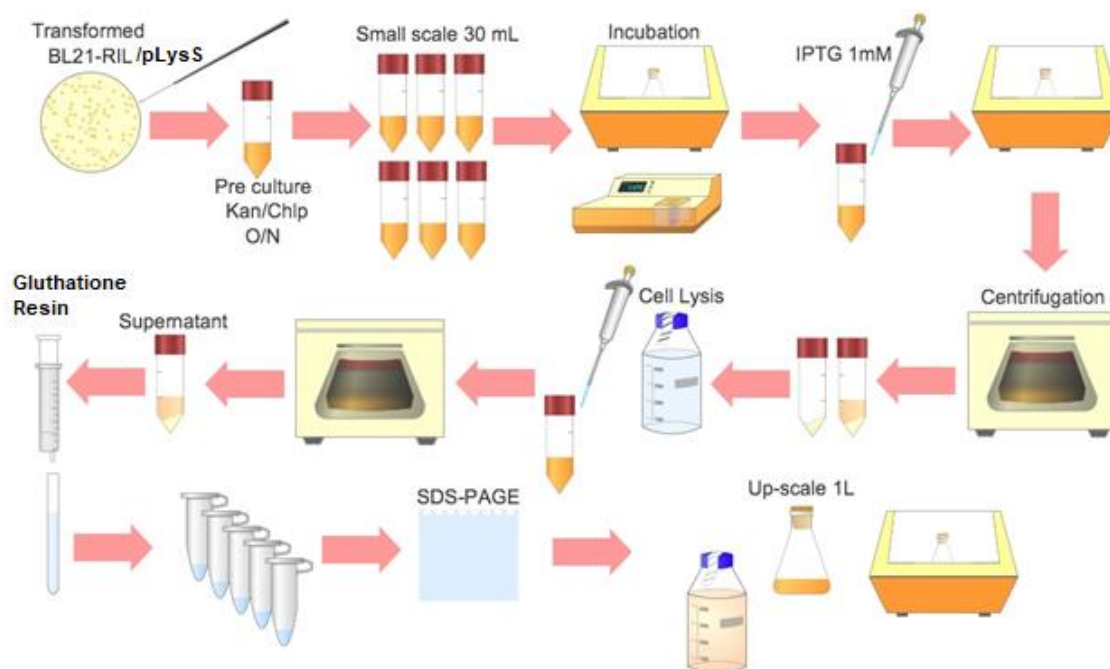
Figure 3.1. GST-OBS and GST expression and purification methodology diagram
As mentioned previously, BL21 (DE3)-RIL and BL21 (DE3)pLysS cell lines were used for the experimental design. The small scale was also done preliminary in 15 mL conical tubes. In addition, the eluted material from the scale-up expression was collected in 15 mL conical tubes. Cell lysis was performed through sonication in the presence of lysis buffer in ice or through BugBuster® or B-PER™ at RT.

When assayed, this modified version of the Orpin B peptide could be expressed as a soluble form and extracted from the hydrophobic pellet after cell lysis (Fig. 3.2). We refer to this modified Orpin B (OB) peptide as His-OBSm because it is fused to a poly-histidine (6XHis) tag in the N-terminal portion (His), it lacks the signal peptide encoding residues (S) and includes 25 amino acids from the expression vector (m) at its C-terminal region (Fig. 3.3).
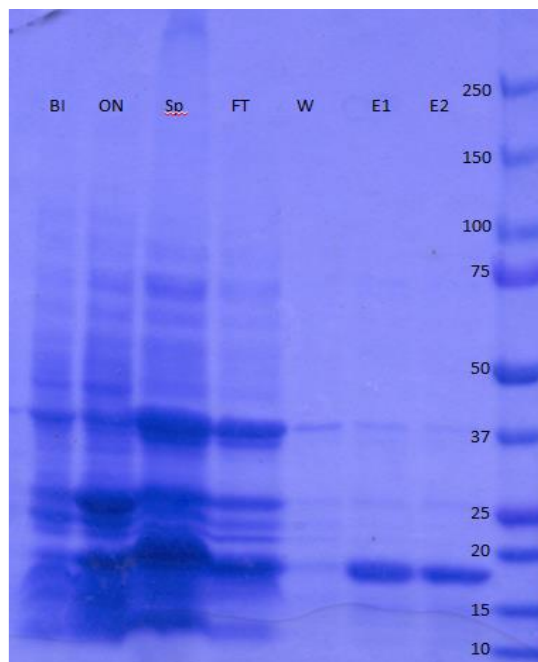
Figure 3.2. Expression and purification of His-OBSm
The purification was done with Ni-NTA resin and eluted with imidazole. The samples collected from expression and purification procedure were separated in SDS-PAGE 10% gels and stained with Coomassie Brilliant Blue R-250. The band corresponding to ~ 17 kDa was eluted with high purity in elution steps E1 and E2. BI: Before Induction; ON: overnight induced sample; Sp: soluble supernatant from the protein extraction; FT: flow-through; W: wash sample; E1: first eluted sample; E2: second eluted sample.

Figure 3.3. Additional 25 residues attached to Orpin B C-terminal region for His-OBSm cloning
The reverse primer was designed without including the stop codon and the encoding additional 25 amino acid residues are contained in the pET200/D-TOPO plasmid cloning site (red box) downstream to the final glutamate residue from OBS (inside red box).

The expression of His-OBSm was done in parallel to a control sequence (His-β-galactosidase) that was treated exactly under the same conditions and using the same vector plasmid. The control sequence also contained the 6XHis tag and corresponded to a weight of 121 kDa (Fig. 3.4).
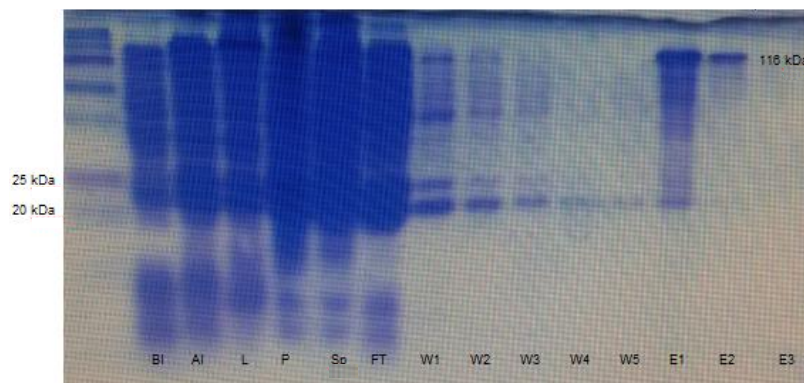
Figure 3.4. Expression and purification of the control pET200 product His-β-galactosidase
The protein band corresponding to 121 kDa was purified as an expression control using the pET200/D/lacZ. The purification was done with Ni-NTA resin and eluted with imidazole. Higher purity was achieved in the second elution step, E2. BI, Before Induction; AI, After Induction; L, protein extraction lysate; P, insoluble pellet from the extraction; Sp, soluble supernatant from the extraction; FT, flow-through; W, washes; E, imidazole eluted samples. It was not observed protein recovery in the third eluted sample, E3. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.2 B. EXPRESSION PROTOCOL 2. Production of Orpin B fused to GST (GST-OBS)

3.4.2.11 Production of a soluble version of Orpin B peptide that resembles the original native form.

Even though we were able to produce a soluble version of Orpin B, it was far from the ideal protein for characterization studies. Therefore, we aimed at obtaining an expressed protein that did not include the foreign 25 residues attached to the C-terminal to provide for a more accurate characterization of the protein. At this point, a second strategy was used to produce a recombinant version of Orpin that best resembled the original Orpin sequence. A strategy was pursued to enhance the solubility of this protein while maintaining a means to easily purify it. Techniques, where chimeric proteins are made by fusing the protein

of interest with protein partners such as maltose-binding protein (MBA), N-utilization substance protein A (NusA), ubiquitin, SUMO and glutathione *S*-transferase (GST), are widely used for this purpose. However, their large size may lead to unpredictable assessment of the fused protein solubility. Even though GST displays lesser solubility enhancement attributes in comparison to the other partners, it still has an additional advantage over the other alternatives. Specifically, this protein tag while being smaller than MBA, can, not only increase its solubility but also be used to purify the target proteins by affinity chromatography. Therefore, a GST plasmid was used to generate a soluble version of Orpin without modifications in its C-terminal. This recombinant version was also difficult to express initially and needed to be optimized for the soluble expression and purification of higher quantities that make it suitable for structural characterization. Here we show several strategies that made possible the production of a soluble version of this protein. These strategies included the use of different cell lines, growth media, additives, expression conditions, and extraction techniques.

3.4.2.12 Preliminary Expression Screening of Orpin B fused to GST (GST-OBS)

The *Orpin B* sequence, without the signal peptide region, was cloned and expressed using the pGEX-6P1 plasmid (GST-OBS). The inserted sequence was verified by sequencing of the extracted plasmids after cell transformation. This recombinant protein was called GST-OBS as it is the Orpin B sequence (OB) fused

to a GST tag at the N-terminal (GST), without the signal peptide coding sequence (S).

The expression of GST-OBS was performed at a small scale and was carried out in conjunction with bacteria transformed with the non-modified vector, containing only the GST sequence. The GST expression was performed under the same conditions as an expression control. To express these proteins, the corresponding transformed cells were incubated with fresh Luria Bertani (LB) medium (30 mL) with ampicillin (100 µg/mL) overnight at 37ºC and 225 rpm. These seeding culture samples were then diluted to ~ OD_{600} 0.3. After growing to ~ OD_{600} 0.5, cell cultures were induced with 1 mM IPTG to activate the recombinant protein over expression. The induced culture samples were incubated for 16 h at the same previous parameters. Samples (500 µL) from induced cultures were collected after 4 and 16 h and compared to uninduced samples through SDS-PAGE analysis (Fig. 3.5). As expected, the 4 h induced samples displayed the protein overexpression and the 16 h samples showed even higher expression based on band thickness and intensity (Fig. 3.5). Even though the band corresponding to GST was detected at the expected molecular weight of ~ 26 kDa, the band corresponding to the over expressed protein in the GST-OBS samples migrated faster than the molecular weight marker band corresponding to 37 kDa, thus appearing to have a smaller molecular weight (Fig. 3.5). This was surprising since, as we mentioned before, the plasmid used for the expression was verified to contain the complete and correct sequence in frame with the GST tag. Although SDS-PAGE displays an

estimate for molecular weight, it is not considered to provide high precision to validate this, as other factors such as the ionic composition of the sample buffer could affect the migration velocity of the bands. Other methods were used to determine that this over-expressed protein was GST-OBS. One of these was using affinity purification to assess if the GST tag was present. In this respect, binding to glutathione coupled resin showed the expected weight band in the elution samples on SDS-PAGE gels (Figs. 3.9 and 3.11–3.18). In later experiments, we also used Western blotting to determine if there was cross-reaction with the serum (Fig. 3.21).
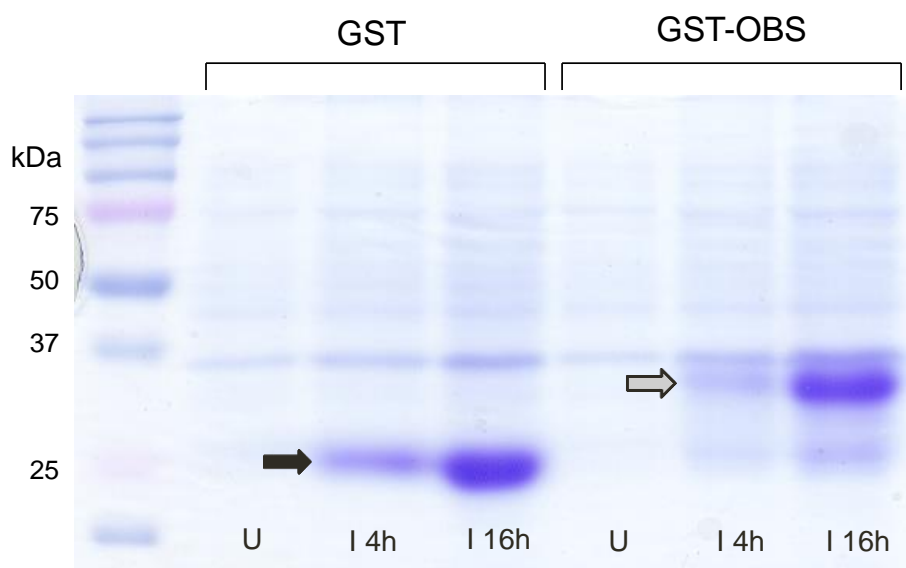
Figure 3.5. Expression of GST-OBS and GST in BL21 Star™
Plasmids were inserted into BL21 Star™ host cells and were induced for 4 and 16 h with 1.0 mM IPTG. Although protein expression was detected at 4 h of induction, samples induced for 16 h show higher protein bands thickness and intensity at the corresponding molecular weights. Black arrow: GST (~ 26 kDa), white arrow: GST-OBS (< 37 kDa). Lanes: U, uninduced cell cultures; I 4h, induced cell cultures for 4 h; I 16h, induced cell cultures for 16 h. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.2.13 GST-OBS and GST Expression Scale-up (1 L)

After confirming the overexpression of the expected bands, we performed a scale-up expression of GST-OBS and GST (control) in 1 L culture medium and purified the expressed protein using a glutathione-coupled resin. This was done using the same standard parameters that were used to express the corresponding peptides; inducing the cell cultures with 1 mM IPTG at 37°C for 16 h. Cells were harvested by centrifugation at 4,000 rcf and 4°C for 15 m. The cell pellet was stored at −20°C overnight. Then, cell lysis was done by sonication in presence of lysis buffer (50 mM Trizma, 150 mM NaCl, 10% glycerol pH 8.0 and lysozyme (1 mg/mL), DNase (0.1 mg/mL) and 10% of 1x proteases inhibitors cocktail). Later, soluble supernatants were incubated with glutathione resin beads (Pierce™, Cat No. 16100) at 4°C for 1 hour. After 18 column volumes (CV = 5 mL) of washes three elution samples (1 CV each) were collected using 10 mM reduced glutathione buffer. After SDS-PAGE analysis, none of the expected bands (corresponding to the expected molecular weight) were detected in any of the three elution samples of the GST-OBS purification (Fig. 3.6A). However, the GST control was successfully purified and detected in the three elution samples at high concentrations as based on band thickness, intensity, and purity (Fig. 3.6B).

The hydrophobic pellet generated after cell lysis was used to determine if the target protein was expressed as an insoluble form. After SDS-PAGE analysis, the expected bands were detected in both GST-OBS and GST pellet samples (Fig. 3.6C) based on electrophoretic mobility and band intensity. Thus, in order to

extract GST-OBS from the insoluble pellet, optimal expression and extraction conditions had to be determined to increase the solubility of GST-OBS.
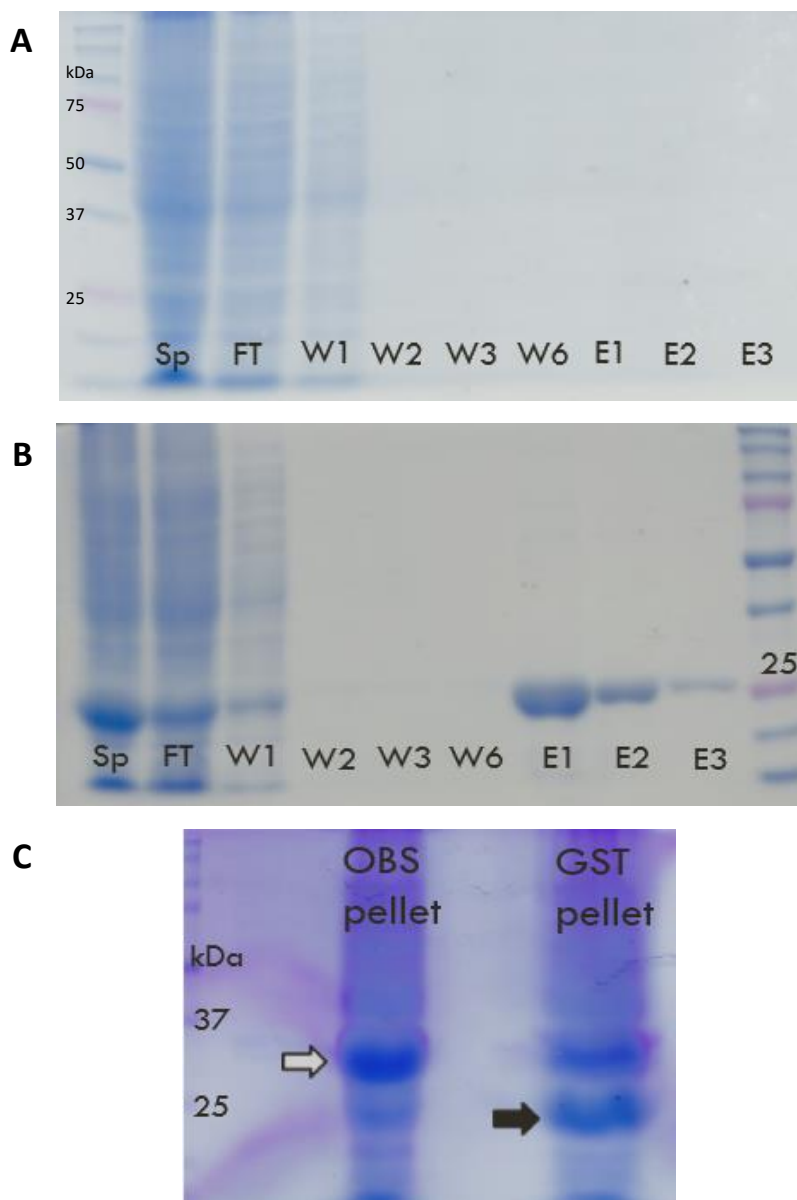
Figure 3.6. Scale-up preliminary expression and purification of GST-OBS and GST
BL21 Star™ strain was used for the expression of GST-OBS and GST. GST-OBS was not recovered at any of the three eluates (A), whereas GST (~ 26 kDa) was recovered with high purity and high concentration (B). GST-OBS band remains mainly in the insoluble pellet fraction, as well as some of the expressed GST (C). Sp: supernatant, FT: flow-through, W#: number of wash step (3 CV each one), E#: number of eluate with 10 mM reduced glutathione. Black arrow: GST band, white arrow: GST-OBS band. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.2.14 Determination of optimal conditions to increase GST-OBS solubility.

When a foreign gene is inserted in *E. coli,* the spatio-temporal control of its expression is altered. The encoding protein will be expressed in a microenvironment that may be different from the original source and the rate of protein synthesis can be faster than required to fold properly. These factors can lead to protein instability and aggregation (Carrió & Villaverde, 2002; Hartley & Kane, 1988). The buildup of these aggregated species is known as inclusion bodies (IBs). As the target recombinant protein band was being observed mainly in the insoluble pellet after cell lysis and protein extraction, we considered the possibility that the protein could be synthesized as aggregates in IBs. In order to minimize the formation of IBs, several strategies can be implemented to slow the rate of protein synthesis that favors the proper folding of newly synthesized proteins. One way to achieve this is to reduce the incubation temperature and to carry out the protein synthesis in the range 15-25°C. (Schein, C. H., Noteborn, 1988; Vasina & Baneyx, 1997; Vera, A., Gonzalez-Montalban, N., Aris, A., Villaverde, 2006). Another way to minimize IBs formation is to reduce the cellular protein concentration from basal synthesis prior to the induction with IPTG. This can be achieved through the addition of 0.2-1% w/v D-glucose in the growth medium as recommended by Studier (2005). Therefore, an experiment was designed to test various expression parameters (Table 3.2) in order to increase GST-OBS solubility. The parameters that were tested were: (1) variations in IPTG concentration (0.5 mM vs 1.0 mM), (2) induction temperature (37°C vs 22°C), (3)

induction OD_{600} (0.5 vs 1.3), and (4) supplementation of media with 1% D-glucose. These parameters serve to analyze the effect of slowing down bacterial host metabolism rate to provide the time for the recombinant target peptide to properly express and fold. As previously, the GST peptide was expressed as a control for these experiments. Small scale preparations (10 mL) were used for this experiment with BL21 Star™ cells as expression host incubated in LB medium supplemented with proper concentration of ampicillin (100 µg/mL).

Table 3.2. Experimental design for the determination of the optimal expression conditions for GST-OBS

Sample	Induction T (°C)	IPTG (mM)	Induction OD_{600}	Glucose
1	37	0.5	0.5–0.7	-
2	37	1.0	0.5–0.7	-
3	37	0.5	0.5–0.7	1%
4	37	1.0	1.3	-
5	RT	1.0	0.5–0.7	-
6	RT	0.5	0.5–0.7	-
7	RT	0.5	0.5–0.7	1%
8	RT	1.0	1.3	-
9	37	1.0	0.5–0.7	1%
10	RT	1.0	0.5–0.7	1%
11	37	1.0	1.3	1%
12	RT	1.0	1.3	1%
13	37	1.0	0.5–0.7	-

RT: room temperature, 22°C ± 2°C.

First, we analyzed the expression in induced versus uninduced samples qualitatively based on the intensity and thickness of the expected band of < 37 kDa. We found that induction at OD_{600} 0.5–0.7 at 37°C and increasing the IPTG concentration to 1 mM increased the expression of the corresponding band (Fig.

3.7A: 1 and 2). In contrast, inducing with 1 mM IPTG and 37°C at higher OD_{600} of 1.3 resulted in a decreased band thickness and intensity (Fig. 3.6A: 4). In addition, another band with high expression was observed in each of these samples, migrating slightly slower than 25 kDa, based on the corresponding band of the molecular weight marker. The sample expressed at the same temperature and OD_{600} 0.5–0.7 with 0.5 mM IPTG, and supplemented with 1% glucose resulted in a similar lower expression as the sample without glucose (Fig. 3.7A: 1 and 3). However, the > 25 kDa band was not as evident in the sample with 1% glucose. As in the case of the first two samples induced without 1% glucose supplementation, the thickness of the GST-OBS band increased with 1 mM IPTG and decreased after induction at OD_{600} 1.3 with 1% glucose (Fig. 3.7A: 3 and Fig. 3.7C: 9 and 11). The samples with expression induced at a lower temperature (RT), resulted in a higher overall increase in band thickness with similar patterns (Figs. 3.7B: 5–8 and 3.7C: 10 and 12). Although, at RT, induction at OD_{600} 0.5–0.7 and supplementation with 1% glucose, a higher band thickness was observed with 0.5 mM IPTG versus 1.0 mM IPTG (Figs. 3.7B: 7 and 3.7C: 10). In summary, the bands expected to be GST-OBS (< 37 kDa) displayed what appeared to be their highest intensity in the samples with the conditions of induction at OD_{600} 0.5–0.7 with 1.0 mM IPTG either at 37°C or RT, without glucose, or 0.5 mM IPTG at RT, with supplementation of 1% glucose (Figs. 3.7A: 2 and 3.7B: 5 and 7). The additional > 25 kDa band was observed in the samples proportionally with the expression level. Given that the GST tag is expected to weight > 26 kDa (including linker portion), this suggests that these are fragments from GST-OBS generated

by proteolysis during expression. If so, they would co-elute after purification with glutathione-coupled beads.
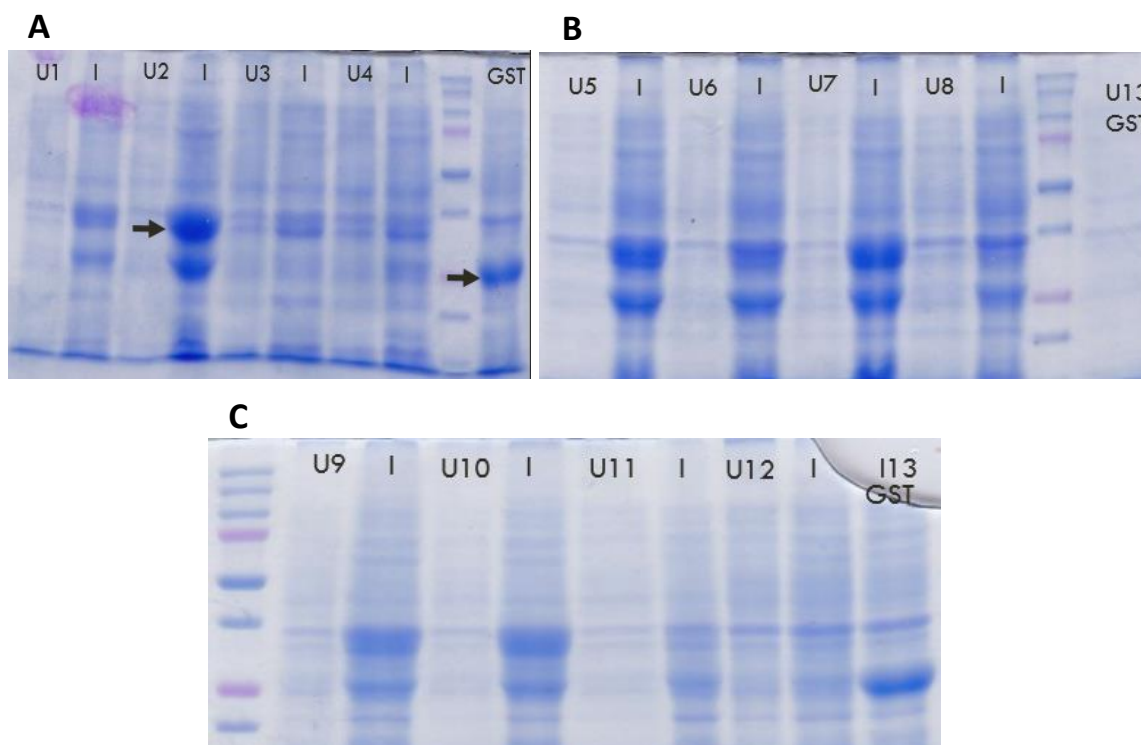

Figure 3.7. Expression optimization of GST-OBS in BL21 Star™
There is some level of expression of GST-OBS at every tested conditions (A–C). Although, additional bands (> 25 kDa) of similar intensity and thickness, are present in each case at different levels. Numbers correspond to expression parameters from the previous table. U#: number of uninduced cell cultures, I: IPTG induced cell cultures. Arrows: GST-OBS band (< 37 kDa) and GST (~ 26 kDa) control induced samples. The samples of GST are shown in the last lane of each gel image: induced (A, GST), uninduced (B, U13 GST), and induced (C, I12 GST) for comparison of band expression. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

The effect of cell lysis and protein extraction protocols was also evaluated in terms of the expression of GST-OBS and GST as determined by electrophoresis. After considering that the sonication procedure could be affecting the stability of the extracted recombinant protein, we proceeded to assess the use of mild conditions for the cell lysis and protein extraction. For this, we used detergent based reagents BugBuster® and B-PER™ protocols. These reagents are used for

gentler lysis of the cell pellets at RT, in contrast to sonication procedures (which are deemed harsher), in presence of 10% of 1x proteases inhibitors cocktail. After cell harvest, cell pellets were stored at −20°C overnight. Each pellet was weighted and split into two samples. One half of the pellet was used for the BugBuster® protocol. The protein concentration recovered by this method was low. Thus, we proceeded to try another similar reagent protocol, B-PER™. This reagent is recommended (by the manufacturer) as being more suitable for GST tagged proteins than BugBuster®. The samples processed using the B-PER™ reagent resulted in an overall slightly higher protein recovery, and more pronounced in the GST control samples (Fig. 3.8C: 13BB and 13P). With the exception of three sample conditions, the majority of the samples using the same reagent had slightly the same overall protein recovery. The sample conditions with the lowest soluble protein recovery were induced at OD_{600}: 0.5–0.7, 37°C with 1.0 mM IPTG and supplemented with 1% glucose (Fig. 3.8B: 9); and two samples induced at OD_{600}: 1.3 with 1.0 mM IPTG and supplemented with 1% glucose, either at 37°C or RT (Fig. 3.8C: 11 and 12). The recovery from the latter two sample was expected, as the expression level was low (Fig. 3.7C). An additional band corresponding to a smaller protein than the band of interest was present in several samples but more evident on: sample induced at OD_{600}: 0.5–0.7, 37°C, 1.0 mM IPTG, no glucose (Fig. 3.8A: 2); and four samples supplemented with 1% glucose: induced at OD_{600}: 0.5–0.7, RT with either 0.5 mM or 1.0 mM IPTG (Figs. 3.8B: 7 and 3.8C: 10); OD_{600}: 1.3, with 1.0 mM IPTG, either at 37°C or RT (Fig. 3.8C: 11 and 12). There was no visible presence of the > 25 kDa additional band in the samples induced at OD_{600}:

0.5–0.7, 37°C, no glucose with either 0.5 mM or 1.0 mM IPTG (Fig. 3.8A: 1P and

2BB); and OD$_{600}$: 1.3, 37°C with 1.0 mM IPTG, and 1% glucose (Fig. 3.8C: 11).

Based on these results, we proceeded to purify the samples extracted with B-

PER™, in a glutathione-coupled resin.
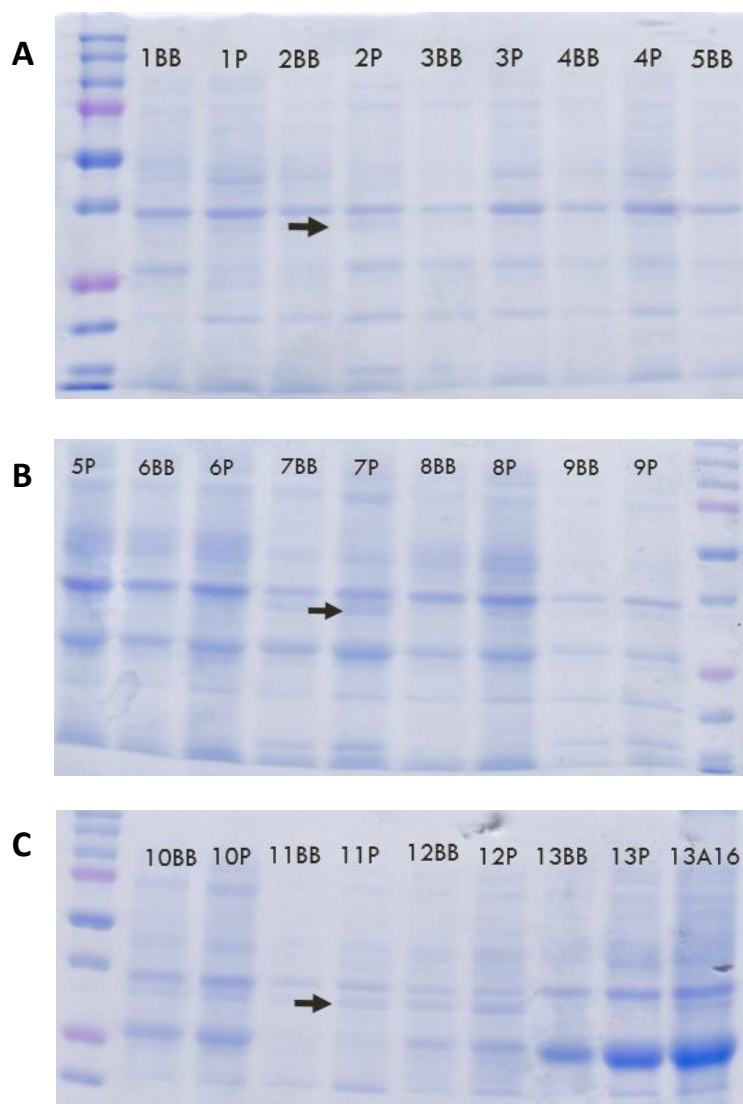
Figure 3.8. BL21 Star™ cell lysis and protein isolation
Numbers correspond to parameters from the previous table. Each sample was supplemented with proteases inhibitors cocktail. B-PER™ reagent extracted supernatants recovered a higher concentration of proteins in each sample. Samples 13: GST control, A16: after 16 h of induction, BB: BugBuster® reagent protocol, P: B-PER™ reagent protocol, arrow: possible GST-OBS band. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

Batch scale purification was used for each of the B-PER™ extracted supernatants. Glutathione resin beads (1 CV = 20 µL) were incubated with the corresponding supernatants following the previous purification protocol. GST was purified side by side as a control and was successfully recovered at high concentrations, based on band thickness and intensity of the two elution samples (Fig. 3.9E: E1, and E2). In contrast, the purified GST-OBS samples resulted in no recovery of the expected band in the majority of the samples with only a slight recovery from the samples induced at OD_{600}: 0.5–0.7, RT, and 1% glucose with either 0.5 mM or 1.0 mM IPTG (Figs. 3.9C: 7 and 3.9D: 10). Hence, a new experimental design was considered that included a different expression host, medium, and the presence of additives.
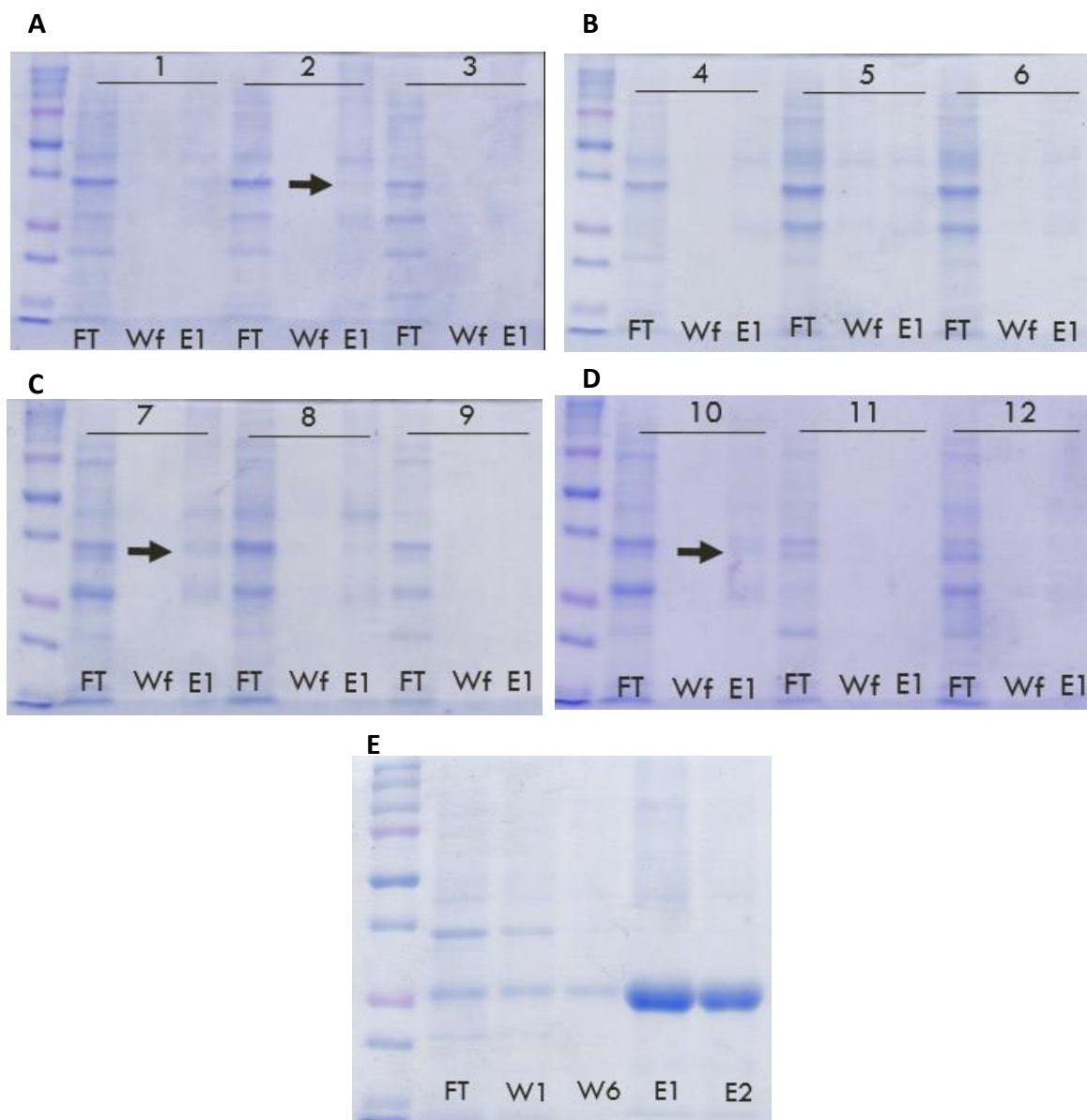
Figure 3.9. GST-OBS and GST purification from BL21 Star™ cells

A-E: Gels from purification of B-PER extractions. GST-OBS was eluted soluble in low concentrations at several parameters at different levels (arrow). GST side-by-side purification as a control (E). Numbers correspond to parameters from the previous table. FT: flow-through, W#: number of wash (3 CV), Wf: final wash after 18 CV, E#: eluted sample, arrow: expected GST-OBS band. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.2.15 Effect of different expression hosts, culture medium and additives on GST-OBS expression.

Different *E. coli* strains were considered to help increase protein production and two strains were chosen; the BL21 (DE3)pLysS and BL21 (DE3)-RIL since they displayed characteristic features that could be suitable to achieve our goal. BL21 (DE3)pLysS cells are recommended for the expression of peptides that could be toxic to host cells. As other BL21 (DE3) strains, these cells have T7 promoter systems under transcriptional control of a *lac*UV5 promoter to achieve recombinant protein expression in presence of glucose (Lanzer, 1988; Silverstone, Arditti, & Magasanik, 1970). These promoter systems consist of the cloning of the target sequence downstream a promoter that is recognized by phage T7 polymerase. This strain contains an additional pLysS plasmid that codes for a T7 lysozyme that binds to the highly active T7 RNA polymerases and hence reduces the basal expression of the recombinant protein that occurs in the absence of the inducer (IPTG) (Moffatt & Studier, 1987; Muller-Hill, B., Crapo, L., Gilbert, 1968). The cells containing the insert were grown in liquid medium with ampicillin at 37°C overnight at 250 rpm. Extractions of plasmids were done from each sample and analyzed through sequencing to confirm the correct (in-frame) sequence insertion.

In addition, the BL21 (DE3)-RIL strain is used for recombinant plasmid expression containing rare codon usage. Sometimes the occurrence of synonymous codons in the coding DNA from the original organism is significantly different from that of the bacterial expression host. At the moment of recombinant protein synthesis induction, a depletion of low-abundance tRNAs occurs, leading

to amino acid misincorporation in the polypeptide. Therefore, the expression levels of the heterologous protein is affected (Gustafsson, Govindarajan, & Minshull, 2004). To avoid this issue, an expression host carrying plasmids containing extra copies of the most common rare tRNAs can be used. BL21 (DE3)-RIL strain provides extra genes for the tRNAs for AGG/AGA (Arg), AUA (Ile), and CUA (Leu). We also incorporated the supplementation of the medium with the dipeptide GlyGly that has been proven to be effective in increasing the solubility of difficult to express proteins while avoiding the loss of their corresponding biochemical function (Sudip Ghosh et al., 2004). Thus, a new experimental design was generated to test the effect of BL21 (DE3)-RIL, BL21 (DE3)pLysS, 1% D-glucose and GlyGly addition to the culture medium following the published protocol (Sudip Ghosh et al., 2004).

This protocol consisted of cell growth at 37°C until OD_{600} 1.0 in Terrific Broth (TB) medium, then inducing protein expression with 0.5 mM IPTG for 16 h at 22°C. Even though LB is the most commonly used medium for growing *E. coli,* cell growth cannot achieve higher cell densities due to its scarce amounts of carbohydrates and divalent cations (Sezonov, Joseleau-Petit, & D'Ari, 2007). Based on this, TB medium has been shown to be superior to LB for achieving higher cell densities (Madurawe, Chase, Tsao, & Bentley, 2000; Studier, 2005). TB medium is used to increase the cell mass in spite of the catabolic repression caused by glucose addition and contains glycerol as a carbon source that is preferred to lactose for further reduction of basal expression. Therefore, this experimental design consisted of the addition of variable concentrations of GlyGly ranging from 0 to 1000 mM (1 M) to TB culture medium (Table 3.3).

Table 3.3. Experimental design of the effect of GlyGly on BL21 (DE3)pLysS and BL21 (DE3)-RIL cells

Sample	Host Cells	GlyGly (mM)	Glucose
1	BL21 (DE3)pLysS	0	-
2	BL21 (DE3)pLysS	50	-
3	BL21 (DE3)pLysS	100	-
4	BL21 (DE3)pLysS	200	-
5	BL21 (DE3)pLysS	500	-
6	BL21 (DE3)pLysS	1000	-
7	BL21 (DE3)-RIL	0	-
8	BL21 (DE3)-RIL	50	-
9	BL21 (DE3)-RIL	100	-
10	BL21 (DE3)-RIL	200	-
11	BL21 (DE3)-RIL	500	-
12	BL21 (DE3)-RIL	1000	-
13	BL21 (DE3)pLysS	0	-
14	BL21 (DE3)pLysS	0	1%
15	BL21 (DE3)-RIL	0	1%
16	BL21 (DE3)pLysS	1000	1%
17	BL21 (DE3)-RIL	1000	1%

Expression samples were grown and induced in Terrific Broth (TB) media.

The results showed that culture samples supplemented with > 100 mM GlyGly had viability problems, thus they were not used for the following analyses (Data not shown). The expression profiles of the remaining samples were evaluated through SDS-PAGE analysis. The samples supplemented with either no GlyGly, 50 mM GlyGly or 1% glucose expressed in BL21 (DE3)pLysS or BL21 (DE3)-RIL exhibited overexpression of GST-OBS and GST as determined by the thickness and intensity of the corresponding protein bands of the expected weights < 37 kDa (for GST-OBS) or ~ 26 kDa (for GST control) in the induced samples of each case (Fig. 3.10). This suggested that the host cells promote higher target peptide

expression. In addition, we compared the expression intensity to samples expressed in LB medium without GlyGly or 1% glucose, included in the same analysis (Fig. 3.10B). The samples grown in LB medium displayed slightly less intensity of the expected weight band including the GST control sample (Fig. 3.10A and 3.10B). Then, these samples were lysed and purified to evaluate protein recovery.
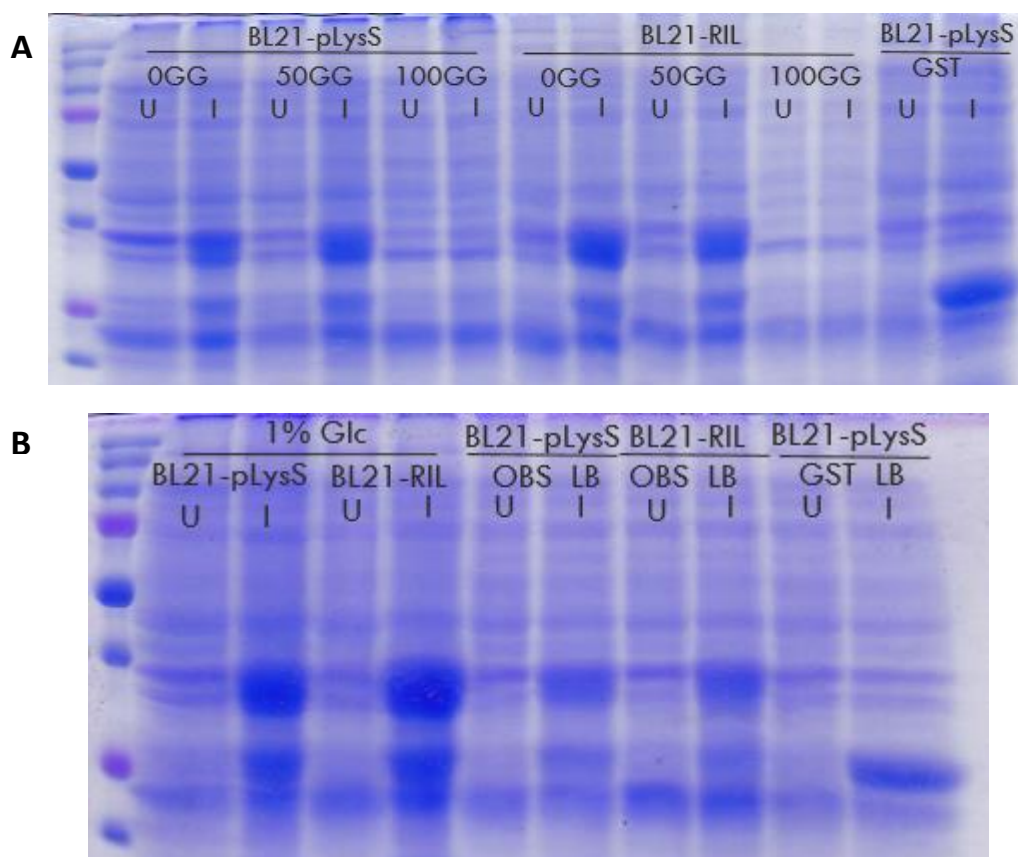

Figure 3.10. Effect of 1% glucose or GlyGly supplemented medium on GST-OBS and GST expression

Expression was done in BL21 (DE3)pLysS and BL21 (DE3)-RIL host cells. Various concentrations of GlyGly were used as additive to Terrific Broth (TB) culture medium except indicated otherwise. Induction was done with 0.5 mM IPTG at 22°C for 16 h. Numbers correspond to the previous parameters table. Bacterial cultures with higher than 100 mM GlyGly were not viable and are not shown. Both host cells displayed an expression of GST-OBS and GST. Higher expression was noted in TB cultures rather than in LB cultures. U: uninduced cell cultures, I: IPTG induced cell cultures. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

The cell pellets were lysed with B-PER™ reagent protocol supplemented with protease inhibitor cocktail. Supernatants were purified with glutathione resin beads as previously described but increasing washing volume from 18 CV to 30 CV. Samples from different steps of the batch scale purification were analyzed through SDS-PAGE (Fig. 3.11). Three of these samples expressed in BL21 (DE3)pLysS with no GlyGly, 50 mM GlyGly, and 1% glucose, respectively (Fig. 3.11A: 1 and 2; Fig. 3.11B: 14), displayed the presence of the soluble GST-OBS band in the corresponding elution samples. Also, the sample expressed in BL21 (DE3)-RIL with 1% glucose also presented the expected band (Fig. 3.11B: 15). However, the samples expressed in BL21 (DE3)pLysS with either 50 mM GlyGly or 1% glucose, showed less proteolysis degradation (band > 25 kDa band) (Figs. 3.11A and 11B).
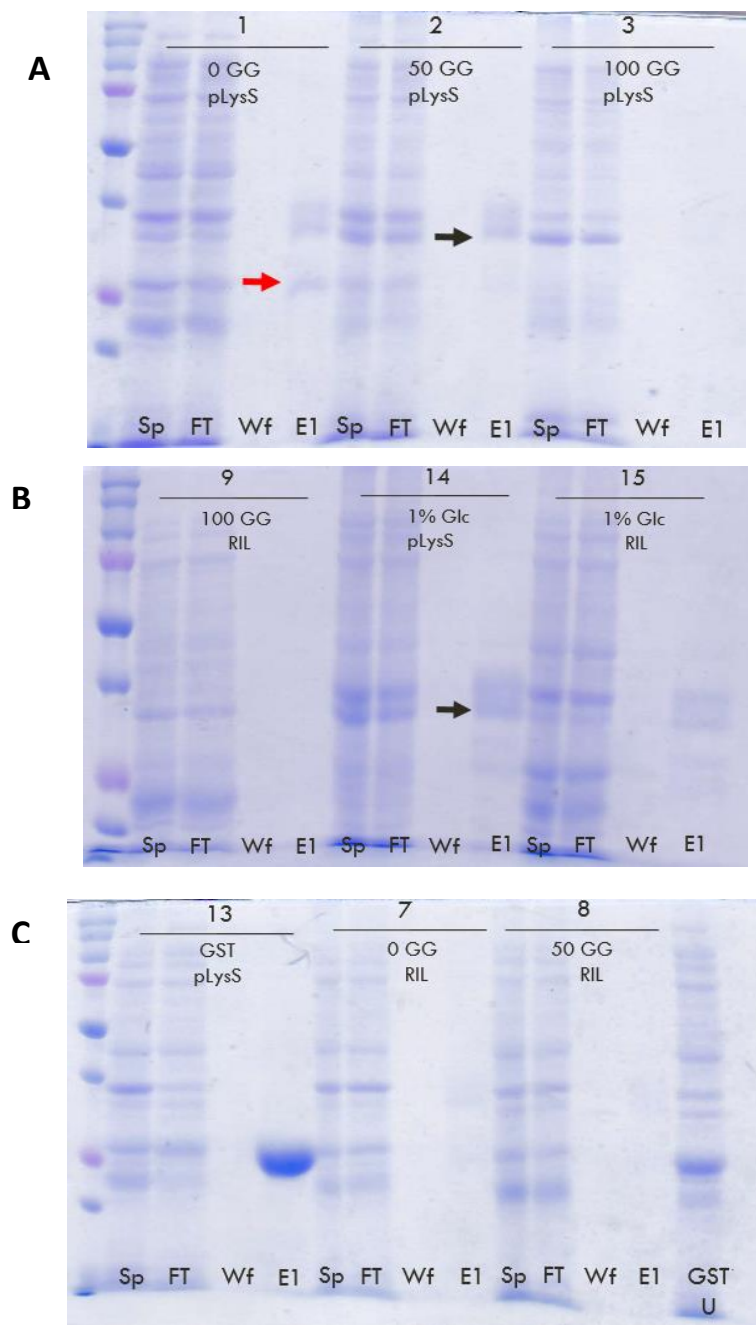
Figure 3.11. Effect of GlyGly and 1% Glucose supplemented medium on GST-OBS purification from BL21 (DE3)pLysS and BL21 (DE3)-RIL cells
Samples supplemented with 1% glucose and 50 mM GlyGly resulted in a higher concentration of the expected band (black arrows) along with lower proteolysis presence (red arrow). GST was purified from BL21 (DE3)pLysS cells side by side as a purification control (C). Numbers correspond to parameters from the previous table. Sp: soluble supernatant, FT: flow-through, Wf: final wash after 30 CV, E1: first elution with 10 mM reduced glutathione, #GG: mM concentration of GlyGly, Glc: D-glucose, RIL: BL21 (DE3)-RIL cells, pLysS: BL21 (DE3)pLysS cells, black arrows: expected GST-OBS band, red arrow: expected proteolysis band. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

Since proteins of interest were detected at low level, the next experimental design included increasing the supernatant volume incubated with resin beads. After washing several times (30 CV) the collected flow-through was incubated again with the resin beads. After this incubation, the resin was washed for a second time (30 CV) before elution with reduced glutathione 10 mM. This was expected to increase the binding concentration of GST-OBS molecules to the resin after removal of non-specific binding of contaminant species, resulting in higher peptide recovery.

An additional experimental design was generated to optimize the expression of soluble GST-OBS peptide using the parameters that resulted in satisfactory protein recovery from the previous experiments. The parameters tested in these experiments included protein expression induction at OD_{600}: 0.5 or 1.0, IPTG concentration of 0.5 mM or 1.0 mM, and medium supplementation with either 1% glucose, 50 mM GlyGly, or the combination of both (1% glucose + 50 mM GlyGly) (Table 3.4). BL21 (DE3)pLysS cells were used as expression host this time.

Table 3.4. Experimental design of the determination of the optimal expression conditions of 1% D-glucose vs 50 mM GlyGly

Sample	OD600	IPTG (mM)	Additive
1	0.5	0.5	1% Glc
2	0.5	1.0	1% Glc
3	1.0	0.5	1% Glc
4	1.0	1.0	1% Glc
5	0.5	0.5	50 mM GG
6	0.5	1.0	50 mM GG
7	1.0	0.5	50 mM GG
8	1.0	1.0	50 mM GG
9	0.5	0.5	1% Glc + 50 mM GG
10	0.5	1.0	1% Glc + 50 mM GG
11	1.0	0.5	1% Glc + 50 mM GG
12	1.0	1.0	1% Glc + 50 mM GG

Glc: glucose; GG: glycylglycine.

After purification, the recovery of the eluted protein was increased by incubating the flow-through with the resin beads a second time after several washes. However, a few contaminants remained attached to the column. Suggesting that these contaminants might be degradation products of GST-OBS.

Although the contamination band of > 25 kDa was not expected to be produced in samples grown in 50 mM GlyGly supplemented medium, it was observed in these samples (Figs. 3.12B: 5 and 6, and 3.12C: 7 and 8). This might be due to alterations in some of the expression conditions. In this experiment, 10 mL of cell culture media were placed in 50 mL tubes to increase oxygenation. Also, seeding cultures were grown in LB medium instead of TB as done in the previous experimental design. These conditions of oxygenation and pre-culture sample

media could have had an adverse effect on the GST-OBS solubility. After taking into consideration these unaccounted parameter differences, an additional experimental design was generated to compare the effect of the level of oxygenation and seeding culture medium to GST-OBS solubility.
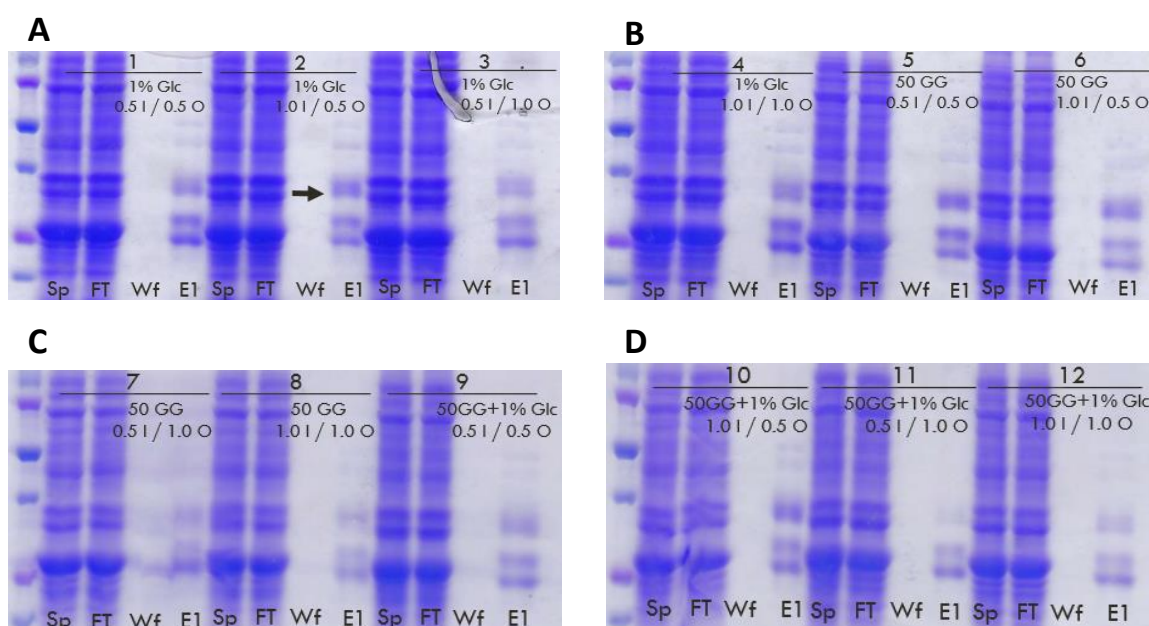

Figure 3.12. Evaluation of 1% D-glucose vs 50 mM GlyGly supplemented mediums for determination of optimal conditions for solubilization of GST-OBS

Gels showing purification of samples from Table 3.4 conditions (A–D). Samples with higher expected band intensity and lower proportion of contaminant bands were detected in samples grown in 1% glucose supplemented medium, induced at OD_{600} 0.5 with 1.0 mM IPTG (A: 2) or 50 mM GlyGly supplemented medium, induced at OD_{600} 0.5 with 0.5 mM IPTG (B: 5). Numbers correspond to the parameters from Table 3.3. 0.5 I: samples induced with 0.5 mM IPTG, 1.0 I: samples induced with 1.0 mM IPTG, 0.5 O: samples induced at OD_{600} 0.5, 1.0 O: samples induced at OD_{600} 1.0, Glc: D-glucose, 50 GG: 50 mM GlyGly supplemented medium, Sp: soluble supernatant, FT: flow-through, Wf: final wash after 30 CV, E1: first elution with 10 mM reduced glutathione, black arrow: expected migration position of GST-OBS. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

This experimental design included the previous conditions (in 50 mL tubes) in addition to the same conditions expressed in low oxygenation (same volume but in 15 mL tubes). Surprisingly, after performing this new experimental design, the results showed that contaminant bands were reduced proportionally to the band of interest in elution samples from low oxygenation suggesting that low oxygenated samples displayed less proteolysis degradation (Fig. 3.13). Samples grown in LB medium with low oxygenation displayed fewer bands above 37 kDa, from byproduct proteins than samples grown at higher oxygenation. All of the 50 mM GlyGly supplemented cultures elution samples exhibited low protein yield at the expected weight (< 37 kDa) band (Fig. 3.13B: 5 and 6; Fig. 3.13C: 7 and 8). These cells grew slower than highly oxygenated ones after comparison of OD_{600} growth monitoring data points (Data not shown). This strengthened the idea that reducing the oxygenation and mechanical stress to cells could promote the stability of the recovered recombinant product. After this evaluation, we proceeded to scaling-up (1 L) of the expression conditions for GST-OBS (Fig. 3.14).
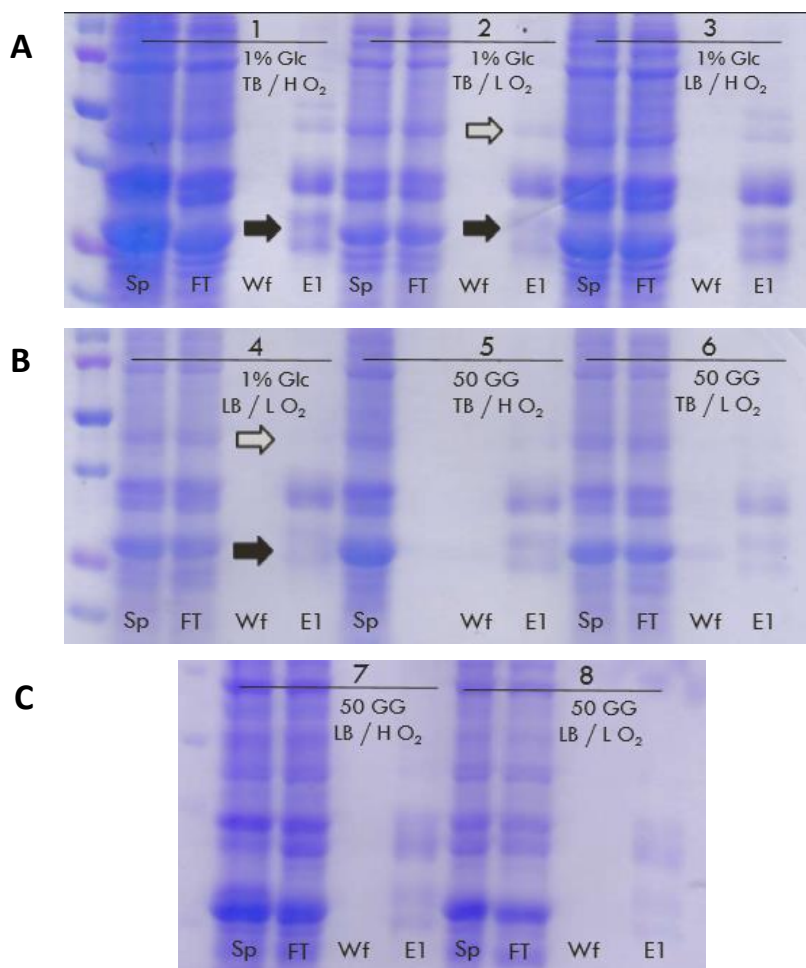
Figure 3.13. Effect of oxygenation and seeding culture medium on GST-OBS solubility
Gels showing purification of samples with high or low oxygenation (A–C). Degradation bands were reduced in eluate samples from low oxygenation (black arrows). Sample with low oxygenation and LB medium appear to have reduced contamination above 37 kDa (white arrows). All 50 mM GlyGly supplemented cultures eluate samples exhibited low protein yield at the < 37 kDa band. H O$_2$: high oxygenation, L O$_2$: low oxygenation, LB: Luria Bertani medium, TB: Terrific Broth medium, Sp: soluble supernatant, FT: flow-through, Wf: after 30 CV of wash buffer, E1: first eluate. Numbers correspond to parameters in the previous table. Note: FT sample condition 5 is missing. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

Several protocols in the scientific literature implemented PBS pH 7.4 as the washing buffer. A difference in buffer pH could remove the non-specific contaminant bands from the resin beads prior to elution. Thus, PBS was evaluated as an alternate buffer for the purification steps and compared side by side with the previously used Tris-NaCl buffer at pH 8.0. For this experiment, we used the expression conditions: (1 L) TB medium supplemented with 1% glucose, induced

at OD_{600} 0.5–0.7, RT (22°C) with 1 mM IPTG. After 16 h of induction, two 10 mL samples were collected prior to final cell harvesting and centrifuged for separate purification procedures using B-PER™ reagent as described. The remaining cell pellet was stored at −80°C for the other batch scale purification experiments. After SDS-PAGE analysis, there was no notable decrease in non-specific contaminant bands after washing the resin with PBS pH 7.4 (Fig. 3.14A).
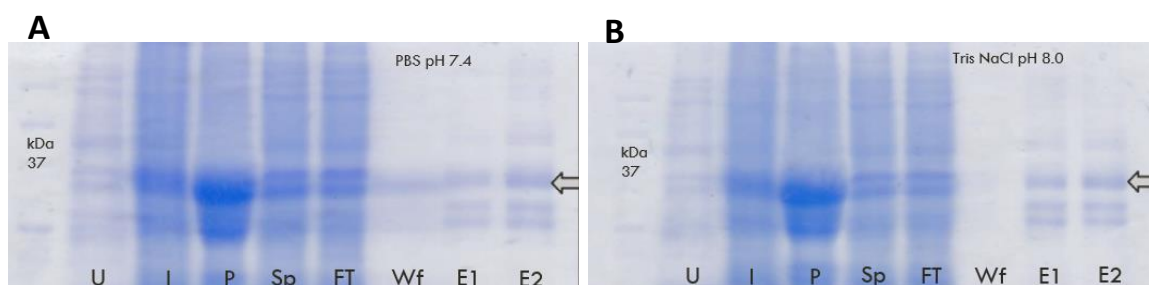

Figure 3.14. Optimization of purification washing step
Lysates of two samples from the 1 L induced culture were purified with glutathione resin beads. Recombinant GST-OBS were eluted with 10 mM reduced glutathione buffer. A total of 30 CV were used after incubation with the flow-through of the corresponding buffer. GST-OBS purification using PBS pH 7.4 (A), 50 mM Tris, 150 mM NaCl, 10% glycerol (B). The second eluate sample from PBS purification has higher protein recovery than the first eluate. Either of the buffers could not remove the degradation bands. Lanes: U: uninduced cell culture, I: IPTG induced cell cultures, P: insoluble pellet, Sp: soluble supernatant, FT: flow-through, Wf: final wash, E1: first eluate and E2: second eluate. White arrow: GST-OBS band of 37.8 kDa. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.2.16 Effect of $CaCl_2$ on GST-OBS solubility

It is widely accepted that the solubility and stability of proteins can be enhanced by the use of additives during the expression of difficult to express proteins. We showed this effect through the addition of 1% D-glucose or GlyGly to the growth media. Also, it is known that this enhancement can be achieved through the supply of essential cofactors needed for the activity of the target protein. Therefore, in order to determine the use of a possible cofactor that could work as an additive to increase the solubility and stability of GST-OBS, we considered the domain

composition of the Orpin B protein. As mentioned previously, the Orpin protein sequences contain two predicted EF-hand motifs. Each EF-hand motif pattern is composed of a putative calcium-binding loop flanked by two alpha helixes and connected by a linker. Thus, we evaluated the possibility that calcium ions could promote an increase in solubility and or stability during cell lysis and protein extraction. Thus, $CaCl_2$ was included during the lysis and extraction procedure to evaluate this possibility. Initially, 5 mM of $CaCl_2$ were included in the lysis buffer during the protein extraction; this was compared to the effect of adding 5 mM of EGTA, as a calcium depletion condition, as well as to using the buffer without additives (Fig. 3.15). The $CaCl_2$ sample exhibited a dramatic increase in the protein band when compared to the EGTA sample and the sample without any additives.
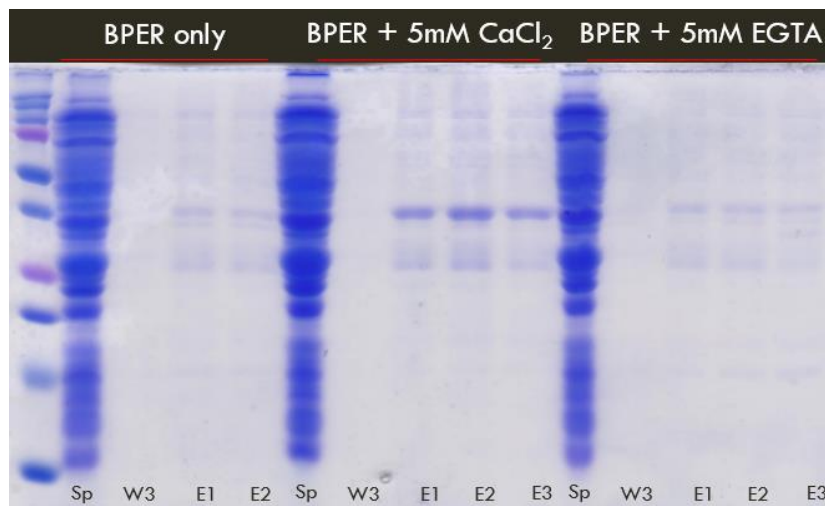

Figure 3.15. Calcium-dependent increased GST-OBS extraction from BL21 (DE3)pLysS pellets

GST-OBS protein was expressed following the methodology described in earlier reports. Protein expression was induced at ~ 0.5 OD_{600} with 1 mM IPTG. Equally weighed cell pellets were lysed using B-PER™ reagent with Benzonase® Nuclease for DNA degradation. B-PER™ only samples did not include any additives, while the other samples were supplemented with either 5 mM $CaCl_2$ or 5 mM EGTA. The soluble portion of the lysate was incubated with glutathione coupled resin beads. The resin was washed extensively with 50 mM Tris, pH 8.0. GST-OBS peptides were eluted from the resin beads using 20 mM reduced glutathione in 50 mM Tris, pH 8.0. The purification was done at RT using a tabletop centrifuge. The GST-OBS expected molecular weight is 37.8 kDa. Sp: soluble lysate supernatant; W3: final 50 mM Tris wash; E1, E2, E3: protein samples eluted with 20 mM reduced glutathione. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.3 Concentration-dependent solubility / stability effect of CaCl₂ on GST-OBS

We evaluated if the solubility / stability enhancement observed by the addition of CaCl₂ during the cell lysis and protein extraction was concentration-dependent. Different additions of CaCl₂ from 0 (zero), to 2.5 mM, 5.0 mM, 10 mM, 20 mM, and 50 mM were used for the assessment (Fig. 3.16A). Additional samples of 5 mM and 50 mM of EGTA were used as calcium depleted conditions. Also, GST protein samples, expressed under the same conditions, were treated using the same concentrations of CaCl₂ or EGTA as a control for this experiment (Fig. 3.16B).
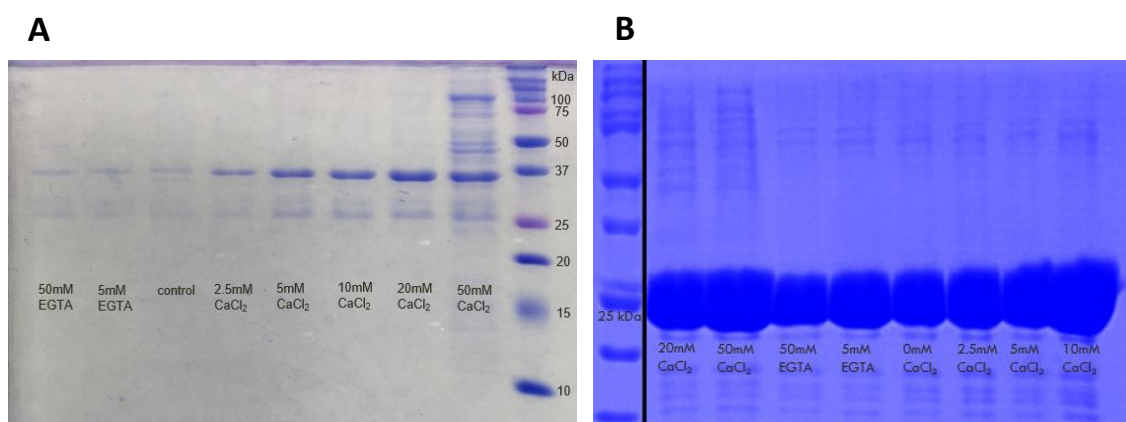
Figure 3.16. Calcium concentration-dependent solubility of GST-OBS
Purified GST-OBS elution samples from glutathione couple resin after the addition of different CaCl₂ concentrations displaying an increase in band thickness on samples with higher CaCl₂ concentration (A). Purified GST control protein samples with the same different CaCl₂ concentrations (B). The GST protein band thickness did not change with the increasing CaCl₂ concentrations during cell lysis and protein extraction. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

After purification, the GST-OBS protein band (< 37 kDa) was detected in all of the samples. However, the band density increased as the $CaCl_2$ concentration was increased from 2.5 mM to 50 mM compared to the calcium depleted samples (5 mM and 50 mM EGTA), as shown in the final eluted samples (Fig. 3.15). Precipitation was observed when 50 mM of $CaCl_2$ was included during the protein extraction and this was reflected in protein bands of higher molecular weight than the target band (< 37 kDa). Thus, 20 mM $CaCl_2$ was identified as the optimal concentration for the protein extraction of GST-OBS. In contrast, the control GST protein band (~ 26 kDa) concentration recovery did not change disregarding the concentration of $CaCl_2$ or EGTA used during the protein extraction. Moreover, the GST band intensity did not exhibit observable differences among the eluted samples.

These results showed that the GST protein could be recovered with high solubility in either case. It is important to highlight the fact that even though the GST tag size (~ 26 kDa + linker) is larger than the OBS (~ 8 kDa) portion of GST-OBS, the majority of the GST-OBS produced remains in the insoluble pellet without the addition of $CaCl_2$ concentrations higher than 2.5 mM.

Interestingly, a migration shift was observed among the purified GST-OBS protein bands extracted in the presence of the different concentrations of $CaCl_2$ after separation in SDS-PAGE (Fig. 3.17). The GST-OBS protein bands were observed to be slightly more separated to the corresponding 37 kDa band from the molecular weight marker as higher concentrations of $CaCl_2$ were used for the

protein extraction (2.5 mM, 5 mM, 10 mM, 20 mM, and 50mM). The protein bands

extracted in calcium depleted conditions with EGTA (5 mM and 50 mM) displayed

a migration closer to the expected weight of 37.8 kDa compared to the

corresponding 37 kDa band from the molecular weight marker. Similar migration

shifts during electrophoresis have been reported to occur to CaBPs after binding

calcium ions (Garrigos et al., 1991; Vogel, 2002; Zhang, Zhang, Zeng, Pi, & Zhu,
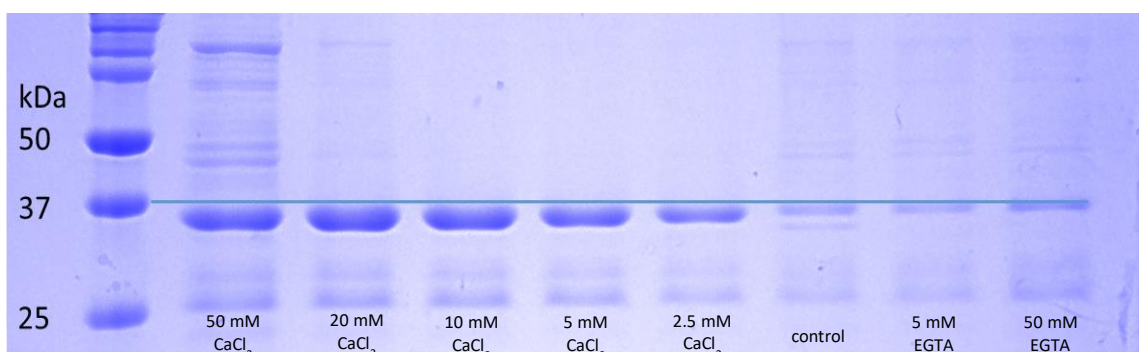
2017).

Figure 3.17. Calcium concentration-dependent band migration shift in SDS-PAGE
A replicate of the experiment from Figure 3.16, displaying the similar faster migration from the
expected molecular weight of 37.8 kDa proportional to the concentration of $CaCl_2$ addition to the
lysis and protein extraction buffer. Blue line: to highlight the expected migration of GST-OBS based
on the molecular marker 37 kDa band. Gels are 10% SDS-PAGE stained with Coomassie Brilliant
Blue R-250.

Based on the results from these strategies, we chose optimal conditions for the

up-scale production of GST-OBS in BL-21 (DE3)-pLysS. This procedure consisted

of growing a pre-culture in LB medium overnight with the appropriate antibiotics,

then using this pre-culture to inoculate a minimum of 1 L of TB medium

supplemented with 1% D-glucose (in a 1 L flask to minimize oxygenation and

mechanical stress) shaking at 225 rpm. The target protein expression was induced

at OD_{600}: 0.5–0.7 with 1 mM IPTG at 22°C and kept shaking for 16 hours. The cells were then harvested by centrifugation and the cell pellet was frozen at −80°C overnight. The cell lysis and protein extraction was done using the B-PER™ reagent protocol including lysozyme, DNase, 1x protease inhibitors cocktail, and 20 mM $CaCl_2$. The soluble supernatant was separated by centrifugation and incubated with glutathione resin at 2–8°C for the purification. The resin was washed with PBS pH 7.4 and incubated a second time with the resin. After additional washes, the recombinant proteins were eluted from the resin by incubation with the wash buffer including 20 mM reduced glutathione and collected by gravity. After the affinity purification, the final production of GST-OBS protein was quantified using the NanoDrop. The corresponding production and recovery of the recombinant protein in each 10 mL eluted sample was as follow: E1: ~ 7.98 mg, E2: ~ 2.97 mg, E3: ~ 0.87 mg, E4: ~ 0.65 mg, and E5: ~ 0.58 mg. The total GST-OBS protein purified and recovered was 13.05 mg (Fig. 3.18).
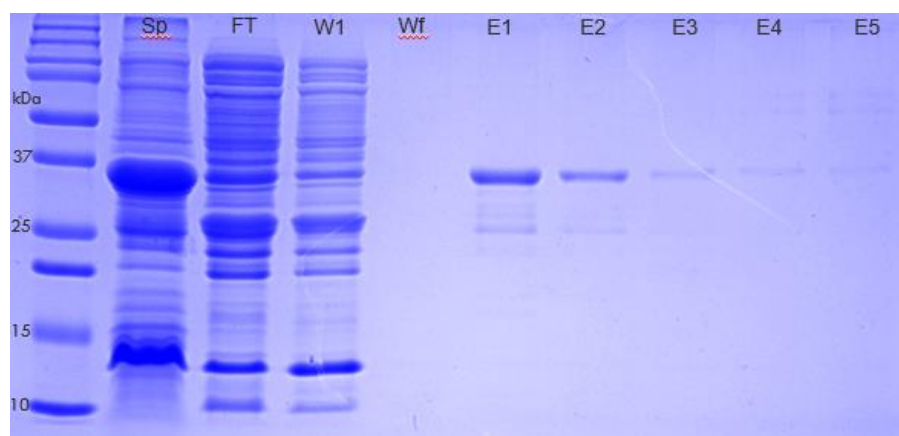

Figure 3.18. GST-OBS purification from 1 L scale-up through GST resin
Sp: soluble supernatant after cell lysis and protein extraction, FT: flow-through, W1: first column wash, Wf: final column wash, E: eluted purified proteins. Gels are 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250.

3.4.4 Size Exclusion Chromatography (SEC) to increase the purity of GST-OBS

Size exclusion chromatography (SEC) was performed in order to further purify the GST-OBS protein. The GST-OBS samples eluted from the glutathione-coupled resin were buffer exchanged to be able to be purified through this gel chromatography. The band corresponding to GST-OBS could not be completely separated from the GST tag band as the two bands were present in the same eluted fraction samples observed in the SDS-PAGE gel (Fig. 3.19: F1, F2, and F3). The majority of the GST tag band of > 25 kDa was recovered in the third eluted fraction sample, F3. An additional protein band was detected at a molecular weight lower than the corresponding 15 kDa band from the molecular weight marker in the sample fractions F1, F2, F5, and F6, with higher concentration in the last two samples. As the expected weight from the OBS sequence alone is 11.3 kDa, we considered the possibility that some amount of the GST tag could be cleaved from the purified GST-OBS, producing two bands corresponding to each portion of the sequence (Fig. 3.19: F1 and F2).
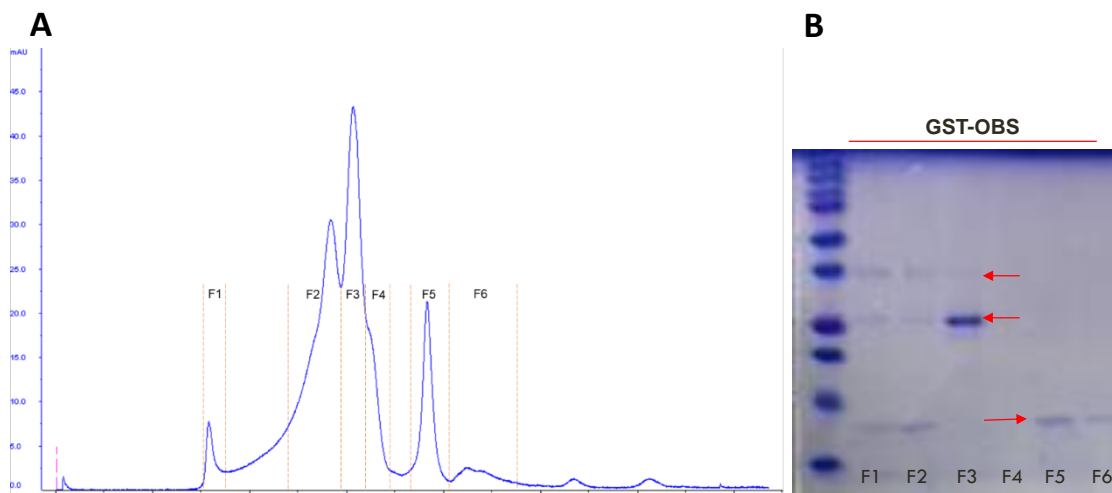
Figure 3.19. Gel filtration chromatography purification of GST-OBS
A Superdex 200 SEC column coupled to an ÄKTA Explorer FPLC. Samples eluted from glutathione resin were concentrated in 0.5 mL Ultracentricons 10 K and further buffer exchanged to 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.8 before sample loading to the column. Fractions from different eluted absorbance peaks were collected and further concentrated. UV profile from the SEC purification (A). Samples were separated on 15% SDS-PAGE gels and stained with Coomassie blue (B). F1 – F6: fraction samples collected from each peak after 13 mL; Red arrows: GST-OBS protein band at ~ 37 kDa, auto-cleaved GST tag at > 25 kDa, auto-cleaved OBS at < 10 kDa.

3.4.5 Oligomerization of GST-OBS

During the extraction and purification of the GST-OBS protein in the presence of 50 mM $CaCl_2$, we observed the presence of protein bands of higher molecular weight with a particular intense band close to the 100 kDa molecular weight marker. As some EF-hand proteins are found to form dimers and even large aggregates in the presence of Ca^{2+} (A. Nakamura et al., 2000), there is a possibility that GST-OBS can oligomerize. Another possibility could be that the GST fragments could be forming oligomers, but this is unlikely as we did not observed evident higher molecular weight bands in the calcium-dependent concentration solubility experiments, that were able to resist the reducing conditions of the SDS electrophoresis (Fig. 3.20).
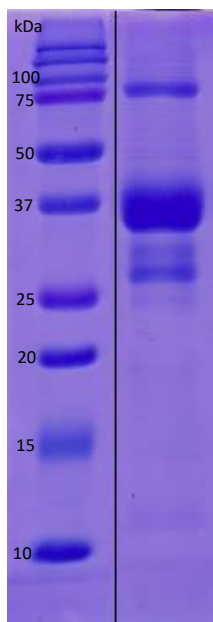
Figure 3.20. Co-elution of a high molecular weight protein after GST-OBS extraction using 50 mM CaCl₂

A protein band > 100 kDa based on the molecular marker, was observed after using 50 mM CaCl₂ during the cell lysis and protein extraction of GST-OBS. Gels are 15% SDS-PAGE stained with Coomassie Brilliant Blue R-250. The gel presented is a composite image from the same gel electrophoresis.

To test the possibility that the protein bands of high molecular weight (> 100 kDa) corresponded to different oligomeric states of the eluted proteins, the protein samples extracted in the presence of 50 mM of CaCl₂ were transferred to a nitrocellulose membrane and incubated with the serum of polyclonal antibodies available (Description of the antibody production can be found in the next chapter). The eluted GST material containing 50 mM CaCl₂ or 50 mM EGTA concentrations and GST-OBS containing 50 mM EGTA, were also included in the assay as control samples. The serum showed positive immune reactivity to two band fragments from the 50 mM EGTA sample and to three band fragments in the membrane

corresponding to the 50 mM CaCl$_2$ samples (Fig. 3.21). No immune reactivity was

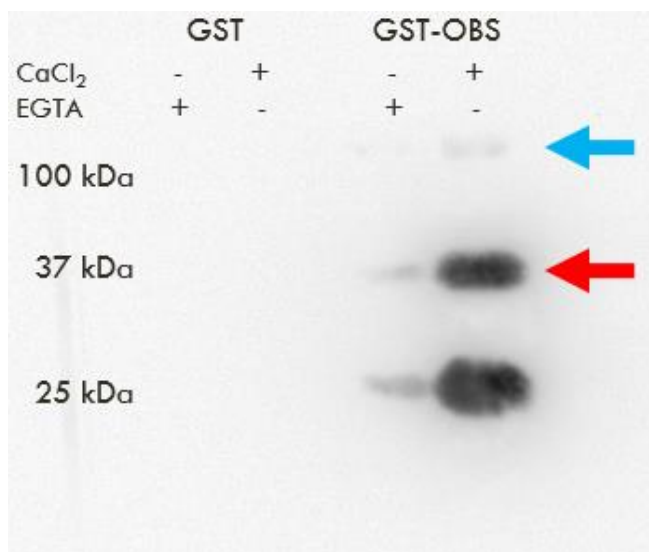detected in either of the GST corresponding lanes of transferred protein samples.

Figure 3.21. His-OBSm anti-serum detection of different peptide bands from proteins extracted with 50 mM CaCl$_2$ or 50 mM EGTA
Western Blot membrane from Figure 3.16 calcium binding assay samples. Orpin-derived antibody specific immunoreactivity to GST-Orpin B. It is shown that in the presence of 50 mM of CaCl$_2$, GST-OBS solubility increased along with the presence of possible oligomerization species. No cross-reactivity was detected in the corresponding GST samples. Red arrow: GST-OBS; blue arrow: aggregates. Protein samples were ran in SDS-PAGE 15% gels and transferred to nitrocellulose membranes.

This can suggest that some of the higher molecular weight band fragment from

the higher CaCl$_2$ concentration samples corresponded to an oligomeric state of the

recombinant protein that was able to resist the reduced conditions of the SDS

electrophoresis. In addition, as the > 25 kDa band was also recognized by the

antibody sera and did bind to the glutathione-coupled resin, this suggests that this

band corresponds to a cleaved portion of GST-OBS, including the GST tag and a

small portion of the OBS sequence. The remaining OBS portion of the cleaved

GST-OBS is expected to be < 12 kDa based on theoretical molecular weight of GST-OBS (37.8 kDa) without the GST tag portion > 25 kDa as is shown in the SEC purification samples (Fig. 3.19B). As this OBS portion is of very low molecular weight, could got through the membrane. This possibility was investigated and the results are discussed in Chapter 4.

3.5 Discussion

This study shows different strategies that were performed to be able to produce two recombinant variants of the Orpin B isoform. The first variant (His-OBSm) corresponds to a His-tagged Orpin with an additional 25 amino acids in its C-terminal. The second variant (GST-OBS) corresponds to a GST-tagged Orpin. This is a more soluble form that better resembles the native Orpin B.

This chapter entails the many attempts and experimental modifications performed to express and purify an Orpin isoform. Thus, the results section includes a large discussion component that serve to analyze a particular experimental procedure and to decide what next step to follow in improving the expression or purification methodology. Nonetheless, other, more general aspects of the work are discussed below.

3.5.1 Protocol modifications that improve the expression and/or purification of Orpin.

A recombinant Orpin B protein was difficult to produce following commonly used expression conditions, and protein extraction protocols. Different initial attempts were performed in order to produce the Orpin B gene product. These included: protein expression induction at 37°C and lower temperatures in LB medium, induction at different cell growing states based OD_{600} using different concentrations of IPTG and using sonication in the presence of lysozyme for the protein extraction. Furthermore, the signal peptide coding sequence was deleted from the coding sequence to minimize the hydrophobic residues composition of the final product but proved futile, as the overexpress protein could not be extracted from the insoluble lysed pellet. Many heterologous proteins produced in *E. coli* are produced as insoluble inclusion bodies (IBs). IBs are insoluble aggregates of misfolded proteins without activity caused during over-expression (Mogk, Mayer, & Deuerling, 2002). One of the hypothesis for this is that the proteins are synthesized at a faster rate than protein folding allows and thus cannot acquire the final stabilized conformation. Proteins that are expressed as IBs have to be isolated and unfolded by detergents and denaturants such as urea, arginine and/or guanidinium ions that must be removed later using dialysis or dilution strategies in order to slowly refold into their native conformations. These procedures are time-consuming and provide no guarantee that the protein will acquire its native conformation without losing activity and stability (Sanderson & Skelly, 2007).

In order to avoid IB formation, several expression strategies could be followed. These consisted of using different host cell lines, genetic modifications, induction temperatures and IPTG concentrations, growth media, additives, methods of cell lysis, and protein extraction explained before in this report. However, as we described in this study, trying different combinations of these modifications alone was not enough to attain a desirable solubility for the different Orpin recombinant variants described before.

Other strategies to avoid the formation of insoluble aggregates of the expressed target protein consisted of using engineered fusion technologies such as grafting the domain sequence of interest and fusing it into another scaffold protein (Sørensen, H. P., & Mortensen, 2005). The other main protein engineering technology is based on subdomain expression that consists of separating different segments of the protein of interest and expressed separately. Even though these approaches were followed in order to obtain sufficient quantities of a purified target protein, they often resulted in insoluble aggregates and could not display the original properties of their native counterpart. However, there are studies of CaBPs that were difficult to express and could only be produced following these approaches (Castiblanco, 2009). This allowed us to understand the factors that contributed to Ca^{2+} binding and Ca^{2+} affinities through the investigation of the relationship between the net charges of key coordination residues (Wilkins, Yang, & Yang, 2003; Yang et al., 2005).

Given that the electrostatic nature of Ca^{2+} is highly affected by charged residues, binding geometry, and protein environment (Maniccia, Yang, Li, Johnson, & Yang, 2006), we wanted to express the recombinant putative CaBP Orpin B to closely resemble the native protein based on expression optimization. Given that Orpin contains two putative EF-hand motifs with the potential of calcium-binding, we proceeded to express the sequence including the original putative calcium-binding motifs architecture to be able to take advantage of this feature in order to increase the stability and/or solubility of this recombinant protein. As we will discuss further, the protocol developed using $CaCl_2$ as an additive during the cell lysis and protein extraction made it feasible to increase soluble protein recovery. Furthermore, these results provided evidence to support the calcium-binding activity of the isoform Orpin B.

3.5.2 What can we learn about Orpin from these experimental results?

3.5.2.17 Solubility / stability enhancement in the presence of $CaCl_2$

As mentioned before, the Orpin B sequence contains two EF-hand motif residues that have a calcium-binding loop within each one. These motifs are also known as helix-loop-helix motifs and the corresponding calcium-binding loop pattern has the potential to coordinate calcium ions. Examples of EF-hand proteins that bind calcium ions are S100A2, S100A4, S100A6, and S100B, which are present in the nucleus where they control gene expression (Boye & Mælandsmo, 2010; Ikura & Ames, 2006; Leclerc, E., Heizmann, 2011; Wolf, Haase-Kohn, & Pietzsch, 2011). However, there are cases where EF-hand protein do not exhibit

calcium-binding activity as in the case of the 10 residue loop, Kv4K+ channel interacting protein, 20 residue loop of EF-hand1 of CIB1, a protein related to neuronal Ca^{2+} sensor (Grabarek, 2006). As we have shown in this study, there was a dramatic increase in final GST-OBS recovery correlated to increasing amounts of $CaCl_2$ during the cell lysis and protein extraction in contrast to the absence of $CaCl_2$. This suggested that the presence of $CaCl_2$ could increase GST-OBS stability and or solubility. However, additional experiments would be required to confirm that this enhancement is attributed to an affinity for calcium and establish GST-OBS as a calcium-binding protein.

3.5.2.18 Structural changes in the presence of $CaCl_2$

The Orpin antisera exhibited cross-reactivity to the recombinant protein expressed from another host and modified plasmid, GST-OBS. This heterologous protein had an expected molecular weight of 37.8 kDa and displayed a faster migration than the 37 kDa protein band from the molecular weight marker. There are numerous reports showing aberrant electrophoresis migration after binding Ca^{2+} ions as in the case of longistatin (Anisuzzaman et al., 2010). In fact, this feature was used to identify CaBPs through gel mobility shift assays (GMSA), as the calcium-binding activity is evident by the changes in movement of these proteins in the Ca^{2+}-bound or -unbound states (Anisuzzaman et al., 2010; Heizmann, 2013; Klee, Crouch, & Krinks, 1979). Thus, the changes in migration that GST-OBS exhibited in the presence or absence of $CaCl_2$ (5 mM and 50 mM EGTA), suggested that GST-OBS displays calcium-binding activity. It is important

to remember that Orpin isoforms had only two predicted Ca^{2+} binding sites; thus, the migration shift is expected to be less than other CaBPs of higher number of Ca^{2+} binding sites. Notwithstanding this indirect evidence, in order to fully confirm Orpin B calcium-binding activity, additional characterization studies, such as UV and CD spectra experiments, radioactive calcium assays, and calcium-binding loop dye assays such as Ruthenium Red (Heizmann, 2013) need to be performed.

3.5.2.19 GST-OBS Aggregation in the Presence of Ca^{2+}

Lastly, it is important to discuss the intense band of high molecular weight (> 100 kDa) recovered from the purification of GST-OBS extracted in the presence of 50 mM $CaCl_2$. As this band was not observed during the purification of the control protein GST with the same amount of $CaCl_2$, we evaluated the possibility that it could be an oligomeric state of the GST-OBS protein as seen for other CaBPs. Previous studies revealed that CaBPs such as CBP40 precipitate forming large oligomers in the presence of calcium ions and that this oligomerization is reversible upon calcium chelation (A. Nakamura et al., 2000). Another example is the protein STIM1 that undergoes the formation of high-order oligomers also in the presence of Ca^{2+} ions (Convington, E. D., Wu, M. M., Lewis, 2010; Liou, Fivaz, Inoue, & Meyer, 2007; Stathopulos, Peter B., Li, Guang-Yao, Plevin, Ames, & Ikura, 2006). STIM1 oligomerization is a critical role in the activation of the CRAC channel (Liou et al., 2007; Luik, Wang, Prakriya, Wu, & Lewis, 2008).

The > 100 kDa band found in the 50 mM $CaCl_2$ GST-OBS samples, exhibited cross-reactivity to the Orpin antibodies sera in Western blots, suggesting that this might be an oligomerized version of GST-OBS that is able to resist the reducing conditions of SDS-PAGE. However, additional experimentation needs to be performed in order to confirm this assumption.

In summary, we have found that with a combination of different materials and expression conditions coupled to genetic engineering and the addition of $CaCl_2$, we were able to increase the production of two pure Orpin variants. We also found that a recombinant Orpin B displays the faster migration during electrophoresis characteristic of other CaBPs due to structural changes upon Ca^{+2} binding. The final purified recovery of this protein was increased gradually by using $CaCl_2$ concentrations up to 50 mM. Moreover, this protein appears to be able to oligomerize in the presence of $CaCl_2$ higher than 50 mM. Therefore, we propose that Orpins (and in particular Orpin B) are EF-hand motif containing proteins that have calcium-binding activity. These protocols can help to develop new strategies for the expression and production of other difficult to express EF-hand proteins. Thus, enabling the further characterization and understanding of the calcium-binding activity and affinity of other CaBPs.

3.6 Supplementary Data

Supplemental Table 3.1. Oligonucleotide primer sequences for Orpin B proteins cloning

Plasmid	Name	Sequence	Tm (°C)
pET200/D-TOPO	His-OBSm fwd	5'-CACCTGCACCTCATCATCTGGG-3'	59.9
pET200/D-TOPO	His-OBSm rev	5'-CTCTGCAACTGACACTTTCTGGAACTC-3'	60.3
pGEX-6P1	GST-OBS fwd	5'-CTGGAAGTTCTGTTCCAGGGGCCCTGCACCTCATCATCTGGGCG-3'	72.3
pGEX-6P1	GST-OBS rev	5'-ACGCGCGAGGCAGATCGTCAGTCATTACTCTGCAACTGACACTT-3'	69.8
pGEX-6P1	pGEX-6P1 fwd	5'-ATCCTCCAAAATCGGATCTGGAAGTTCTG-3'	60.0
pGEX-6P1	pGEX-6P1 rev	5'-GCATGTGTCAGAGGTTTTCACCGTC-3'	60.1

His-OBSm: N-terminal His-tagged Orpin B sequence without signal peptide region. The corresponding primers were used for the cloning into pET200/D-TOPO plasmid. GST-OBS: N-terminal GST-tagged Orpin B sequence without signal peptide region. The corresponding primer sequences were used for the cloning into pGEX-6P1 plasmid. pGEX-6P1 primers were used for sequence insertion verification of GST-OBS. Tm: melting temperature, fwd: forward primer and rev: reverse primer.

3.7 References

Anisuzzaman, Islam, M. K., Miyoshi, T., Alim, M. A., Hatta, T., Yamaji, K., … Tsuji, N. (2010). Longistatin, a novel EF-hand protein from the ixodid tick Haemaphysalis longicornis, is required for acquisition of host blood-meals. *International Journal for Parasitology*, *40*(6), 721–729. https://doi.org/10.1016/j.ijpara.2009.11.004

Boye, K., & Mælandsmo, G. M. (2010). S100A4 and metastasis: A small actor playing many roles. *American Journal of Pathology*, *176*(2), 528–535. https://doi.org/10.2353/ajpath.2010.090526

Carrió, M. M., & Villaverde, A. (2002). Construction and Deconstruction of Inclusion Bodies. *Journal of Biotechnology*, *96*, 1–10. Retrieved from www.elsevier.com/locate/jbiotec%0Apapers2://publication/uuid/56943B93-1DD2-4DEC-A85C-B8D144F85B66

Castiblanco, A. (2009). Expression and Purification of Engineered Calcium Binding Proteins. *Chemistry Theses*. Retrieved from http://scholarworks.gsu.edu/chemistry_theses/20

Convington, E. D., Wu, M. M., Lewis, R. S. (2010). Essential Role for the CRAC Activation Domain in Store-dependent Oligomerization of STIM1. *Molecular Biology of the Cell*, *21*(24), 4325–4337. https://doi.org/10.1091/mbc.E10

Garrigos, M., Deschamps, S., Viel, A., Lund, S., Champeil, P., Møller, J. V., & Maire, M. le. (1991). Detection of Ca2+-binding proteins by electrophoretic migration in the presence of Ca2+ combined with 45Ca2+ overlay of protein blots. *Analytical Biochemistry*, *194*(1), 82–88. https://doi.org/10.1016/0003-2697(91)90154-L

Ghosh, S., Rasheedi, S., Rahim, S. S., Banerjee, S., Choudhary, R. K., Chakhaiyar, P., … Hasnain, S. E. (2004). Method for enhancing solubility of the expressed recombinant proteins in Escherichia coli. *BioTechniques*, *37*(3), 418–423.

Grabarek, Z. (2006). Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. *Journal of Molecular Biology*, *359*(3), 509–525. https://doi.org/10.1016/j.jmb.2006.03.066

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology*, *22*(7), 346–353. https://doi.org/10.1016/j.tibtech.2004.04.006

Hartley, D. L., & Kane, J. F. (1988). Properties of inclusion bodies from recombinant Escherichia coli. *Biochemical Society Transactions*, *16*(2), 101–102. https://doi.org/10.1042/bst0160101

Heizmann, C. W. (2013). *Calcium-Binding Proteins and RAGE*. (C. W. Heizmann, Ed.), *Methods in Molecular Biology* (Vol. 963). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-230-8

Ikura, M., & Ames, J. B. (2006). Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: Two ways to promote multifunctionality. *Proceedings of the National Academy of Sciences*, *103*(5), 1159–1164. https://doi.org/10.1073/pnas.0508640103

Klee, C. B., Crouch, T. H., & Krinks, M. H. (1979). Calcineurin: a calcium- and calmodulin-binding protein of the nervous system. *Proceedings of the National Academy of Sciences*, *76*(12), 6270–6273. https://doi.org/10.1073/pnas.76.12.6270

Kyte, J., & Doolittle, R. F. (1982). A Simple Method for Displaying the Hydropathic Character of a Protein, *157*, 105–132.

Lanzer, M. (1988). Promoters Largely Determine the Efficiency of Repressor Action. *Proceedings of the National Academy of Sciences*, *85*(23), 8973–8977. https://doi.org/10.1073/pnas.85.23.8973

Leclerc, E., Heizmann, C. W. (2011). The importance of Ca2+/Zn2+ signaling S100 proteins and RAGE in translational medicine, (5), 1232–1262.

Liou, J., Fivaz, M., Inoue, T., & Meyer, T. (2007). Live-cell imaging reveals sequential oligomerization and local plasma membrane targeting of stromal interaction molecule 1 after Ca2+ store depletion. *Proceedings of the National Academy of Sciences*, *104*(22), 9301–9306. https://doi.org/10.1073/pnas.0702866104

Luik, R. M., Wang, B., Prakriya, M., Wu, M. M., & Lewis, R. S. (2008). Oligomerization of STIM1 couples ER calcium depletion to CRAC channel activation. *Nature*, *454*(7203), 538–542. https://doi.org/10.1038/nature07065

Lund, B. A., Leiros, H. S., Elin, G., & Bjerga, K. (2014). A high-throughput , restriction-free cloning and screening strategy based on ccd B-gene replacement, 1–7.

Madurawe, R. D., Chase, T. E., Tsao, E. I., & Bentley, W. E. (2000). A recombinant lipoprotein antigen against Lyme disease expressed in E. coli: Fermentor operating strategies for improved yield. *Biotechnology Progress*, *16*(4), 571–576. https://doi.org/10.1021/bp0000555

Maniccia, A. W., Yang, W., Li, S. Y., Johnson, J. A., & Yang, J. J. (2006). Using protein design to dissect the effect of charged residues on metal binding and protein stability. *Biochemistry*, *45*(18), 5848–5856. https://doi.org/10.1021/bi052508q

Moffatt, B. A., & Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, *49*(2), 221–227. https://doi.org/10.1016/0092-8674(87)90563-0

Mogk, A., Mayer, M. P., & Deuerling, E. (2002). Mechanisms of protein folding: molecular chaperones and their application in biotechnology. *Chembiochem*, *3*(9), 807–814.

Muller-Hill, B., Crapo, L., Gilbert, W. (1968). Mutants That Make More Lac Repressor, 1259–1264.

Nakamura, A., Okagaki, T., Takagi, T., Nakashima, K. I., Yazawa, M., & Kohama, K. (2000). Calcium binding properties of recombinant calcium binding protein 40, a major calcium binding protein of lower eukaryote Physarum polycephalum. *Biochemistry, 39*(13), 3827–3834. https://doi.org/10.1021/bi991855v

Ortiz-Pineda, P. A., Ramírez-Gómez, F., Pérez-Ortiz, J., González-Díaz, S., Santiago-De Jesús, F., Hernández-Pasos, J., … García-Arrarás, J. E. (2009). Gene expression profiling of intestinal regeneration in the sea cucumber. *BMC Genomics*, *10*(1), 262. https://doi.org/10.1186/1471-2164-10-262

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*(APR), 1–17. https://doi.org/10.3389/fmicb.2014.00172

Sambrook, J., W., Russell, D. W. (1957), & Coldspring Harbor Laboratory (2001). (n.d.). *Molecular cloning: a laboratory manual* (3rd ed.). Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Sanderson, M., & Skelly, J. (2007). *Macromolecular Crystallography: Conventional and High-Throughput Methods*. (M. Sanderson & J. Skelly, Eds.), *Oxford biosciences. The Oxford University Press*. Oxford University Press.

Schein, C. H., Noteborn, M. H. M. (1988). Formation of soluble recombinant proteins in Escherichia coli is favored by lower growth temperature. *Nature Biotechnology*, *6*, 709–712.

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23), 8746–8749. https://doi.org/10.1128/JB.01368-07

Silverstone, A. E., Arditti, R. R., & Magasanik, B. (1970). Catabolite-insensitive revertants of lac promoter mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *66*(3), 773–779. https://doi.org/10.1073/pnas.66.3.773

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*(1), 1475–2859. https://doi.org/10.1186/1475-2859-4-1

Stathopulos, Peter B., Li, Guang-Yao, Plevin, M. J., Ames, J. B., & Ikura, M. (2006). Stored Ca 2+ Depletion-induced Oligomerization of Stromal Interaction Molecule 1 (STIM1) via the EF-SAM Region. *Journal of Biological Chemistry*, *281*(47), 35855–35862. https://doi.org/10.1074/jbc.m608247200

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016

Vasina, J. A., & Baneyx, F. (1997). Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the *Escherichia coli* cspA and tac promoter. *Protein Expression and Purification*, *9*(2), 211–218.

Vera, A., Gonzalez-Montalban, N., Aris, A., Villaverde, A. (2006). Anaerobic Fermentation of Glycerol by Escherichia coli: A New Platform for Metabolic Engineering. *Methods*, *96*(6), 1101–1106. https://doi.org/10.1002/bit

Vogel, H. J. (2002). *Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques.* (H. J. Vogel, Ed.) (Vol. 173). Calgary, AB, Canada: Humana Press.

von Heijner, G. (1990). The Signal Peptide. *The Journal of Membrane Biology*. *115*, 195–201.

Wilkins, A. L., Yang, W., & Yang, J. J. (2003). Structural Biology of the Cell Adhesion Protein CD2: From Molecular Recognition to Protein Folding and Design. *Current Protein & Peptide Science*, *4*(5), 367–73.

Wolf, S., Haase-Kohn, C., & Pietzsch, J. (2011). S100A2 in cancerogenesis: A friend or a foe? *Amino Acids*, *41*(4), 849–861. https://doi.org/10.1007/s00726-010-0623-2

Yang, W., Wilkins, A. L., Ye, Y., Liu, Z. R., Li, S. Y., Urbauer, J. L., … Yang, J. J. (2005). Design of a calcium-binding protein with desired structure in a cell adhesion molecule. *Journal of the American Chemical Society*, *127*(7), 2085–2093. https://doi.org/10.1021/ja0431307

Zhang, Y., Zhang, X., Zeng, H., Pi, E., & Zhu, Y. (2017). Analysis of EF-Hand Proteins in Soybean Genome Suggests Their Potential Roles in Environmental and Nutritional Stress Signaling. *Frontiers in Plant Science*, *8*(May), 1–15. https://doi.org/10.3389/fpls.2017.00877

Zheng, N., & Gierasch, L. M. (1996). Signal Sequences : The Same Yet Different. *Cell, 86*(6), 849–852.

Chapter 4 Insights to Orpin Physiological Role

4.1 Abstract

The identification of Orpin expressing cells in tissues from the sea cucumber *Holothuria glaberrima* is required to characterize Orpin protein expression profiles and gather insights into their physiological role. The availability of a recombinant version of Orpin B made feasible the production of an antibody for the identification of Orpins using Western Blot and immunohistochemistry.

We use a his-tagged recombinant version of Orpin B (His-OBSm), (previously described in Chapter 3), as an antigen to immunize two mice and produce an antiserum. We showed that the antiserum cross-reacted with proteins that shared similar sequence regions with the antigen His-OBSm, and not to a recombinant protein control, GST. Western Blot results showed antibody cross-reactivity to His-OBSm, GST-OBS, and His-β-galactosidase. The antiserum was also used to identify Orpin expressing cells in the mesentery, intestine, and longitudinal muscle from the sea cucumber *Holothuria glaberrima*, through immunohistochemistry. The antiserum can be used as a tool to develop an assay for the immunoprecipitation of Orpin protein partners in future experiments to gather more insights into Orpin functional role.

In other aspects, we investigated the effect of Wnt signaling pathway activation on *Orpin* mRNA expression in regenerating gut explants. The Wnt/B-catenin pathway was activated by incubation with 20 mM LiCl for 48–72 h. We showed an apparent decrease of *Orpin B* transcript levels after treatment with LiCl compared

to NaCl and non-supplemented media controls using semi quantitative RT-PCR analysis. The decrease was not detected in *Orpin A*. These results provide preliminary hints into the possible association of *Orpin* with this known regenerative pathway in an isoform specific manner.

4.2 Introduction

Further studies on Orpin were focused on two different but related issues. First, to describe the cells that express Orpin and in what tissues are these cells localized. Second, to determine what factors might have an effect on *Orpin* expression. Both of these studies are expected to provide clues to Orpins role. The recombinant His-OBSm was used as an antigen to produce an antibody that could be used in immunohistochemistry experiments. The obtained antibodies were tested in various tissues, including visceral muscle, mesentery, and intestine from *Holothuria glaberrima*. The antibody specificity was evaluated through western blot using different recombinant peptides that shared portions of the antigen sequence and compared to GST and His-β-galactosidase as controls.

In another aspect, *Orpin A* and *Orpin B* gene expression was evaluated *ex vivo*. For this, we took advantage of ongoing experiments in the lab where explants of regenerating intestine were exposed to LiCl, an activator of the Wnt signaling pathway to determine if Wnt might have a role in *Orpin* expression. These results provided insights for the possible role of Orpins in *Holothuria glaberrima*.

4.3 Materials and Methods

4.3.1 Antibodies anti-serum production

Purified His-OBSm peptide was used for the immunization of two mice. The protein expression and purification was performed as previously described in Chapter 3. The samples were desalinated and ultrafiltered with Amicon® Ultra-0.5 centrifugal filters (Millipore) as described below. Approximately 300 µg of purified protein were diluted with PBS and Freund's Complete Adjuvant (Sigma) in a 1:1 solution to a final volume of 300 µL. Each mouse was injected intraperitoneally with 150 µL of the sample solution. After a month, both animals were injected again in the same way, this time using Freund's Incomplete Adjuvant (Sigma) for the solution preparation. Blood samples of approximately 500 µL from retro-orbital plexus bleeding of each animal were collected at 7, 16, and 21 days after. The blood samples were incubated with 1 mM PBS at 37°C for 30 min and then centrifuged for 2 min at 13,200 rpm to retrieve sera. The corresponding supernatants were then collected and diluted with 50% glycerol to be stored at −20°C.

This procedure was carried out once again after one month. However, this time, a second bleed of each animal was done two weeks after the last immunization. In this way, it was possible to obtain enough polyclonal antibodies for the detection of Orpin during its characterization. The supernatant was subsequently used for immunohistochemistry and western blots.

4.3.2 SDS-PAGE

Sodium dodecyl sulfate (SDS) Resolving gels were prepared at 10% or 15% of polyacrylamide with 1 mm of thickness. Tissue extract samples (30 µL) were diluted 1:1 with SDS loading dye with the exception of pellet samples that were diluted in a 3:1 ratio, boiled at 90°C for 5 min, cooled down on ice for 1 min, and centrifuged before adding 30 µL into the gel wells. The molecular weight marker used was Precision Plus Protein™ Standards Dual Color (Bio-Rad, Cat. No. 161-0374). The electrophoresis was run 30 min at 50 V (during migration through the stacking gel portion) and 35–45 min at 150 V (during migration through the resolving gel portion) to increase band sharpness. Then, the gels were cleaned with dH_2O to remove the SDS residues from the running buffer. The gels were stained with Coomassie Brilliant Blue R250 Dye (Thermo Scientific, Cat. No. 20278) for 30 min with a solution heated for 30 s in a microwave, immediately before use. The gels were cleaned with dH_2O and destained with a 10% acetic acid, 45% methanol, 45% dH_2O solution until the bands were visible. Kim wipe papers were submerged into the distaining solution with the gels to accelerate the distaining time by absorption of the Coomassie excess. After that, the gels were cleaned with dH_2O and rehydrated. This was seen to increase the band stain intensity and resolution.

4.3.3 Western Blots

4.3.3.20 Buffers preparation

TBS 10x buffer

    30.3 g of Tris and 43.8 g of NaCl were diluted into 500 mL of dH_2O and adjusted to pH 7.6.

TBS-T buffer

    50 mL of TBS 10x buffer and 500 µL of Tween 20 were diluted into 500 mL of dH_2O.

SDS running buffer 10x

    144 g of glycine and 30 g of Tris Base were diluted into 100 mL of 10% SDS solution.

Transfer buffer

    100 mL of SDS Running Buffer 10x and 200 mL of methanol were diluted into 700 mL of dH_2O.

4.3.3.21 Protein band transference to nitrocellulose membranes

    The different purified protein samples were electrophoresed on 10% polyacrylamide gels of 1 mm thickness following the described SDS-PAGE methodology. The molecular weight marker used was Precision Plus Protein™

Standards Dual Color (Bio-Rad, Cat. No. 161-0374). The gels were transferred to Hybond-C Extra nitrocellulose membranes (GE, Cat. No. RPN303E) using an Owl™ VEP-2 Electroblotting System (ThermoFisher) for the Western Blot (ThermoFisher Scientific) for 3 h at 30 V. The corresponding gels were stained with Coomassie Brilliant Blue R250 Dye (Thermo Scientific, Cat. No. 20278) solution and then destained as in the SDS-PAGE methodology, to confirm that the protein bands were transferred to the membranes. The nitrocellulose membranes were blocked with TBS-T buffer with 0.5% dry milk, for 1 h at RT in motion with a rocker. The membranes were washed three times with TBS-T for 15 min each time. Then the membranes were incubated with the His-OBSm anti-serum diluted 1:1000 in TBS-T buffer and 5% dry milk (10 µL/10 mL) at 2–8°C overnight in a rocker. Membranes were washed three times again as before and incubated with the secondary antibody, HRP-conjugated, anti-mouse IgG from sheep (GE, Cat. No. NXA931) diluted 1:5000 in TBS-T buffer and 5% dry milk (2 µL/10 mL). The membranes were washed again as before, including an additional wash of TBS for 10 min to remove the Tween 20. The antibody cross-reactivity was detected by a chemiluminescent method using SuperSignal™ West Dura Extended Duration Substrate (ThermoFisher Scientific, Cat. No. 37071) following the manufacturer's protocol and visualized in a ChemiDoc XRS+ (BioRad).

4.3.4 Protein ultrafiltration and buffer exchange

Aliquots of 500 µL were filtered through Amicon® Ultra-0.5 centrifugal filters (Millipore) of 3 K cutoff for GST-OBS produced before the optimization and His-

OBSm, 10K for GST, and 50 K for His-β-galactosidase, and centrifuged at 14,000 x g for 10 min at RT and the filtrates were discarded. This procedure was repeated 5 times each. To recover the concentrated solute, the Amicon Ultra filter device was placed upside down in a clean microcentrifuge tube and centrifuged for 2 min at 1,000 × g to transfer the concentrated sample from the device to the tube.

4.3.5 Animals

Adult specimens (10–15 cm in length) of the sea cucumber *H. glaberrima* were collected in coastal areas of northeastern Puerto Rico and kept in indoor in aerated seawater aquaria at RT (22°C ± 2°C). Evisceration was induced by 0.35 M KCl injections (3–5 mL) into the coelomic cavity (García-Arrarás et al., 1998). Eviscerated animals were let to regenerate for 4 days before the dissection and tissue extraction. For the dissection, organisms were anesthetized by placement in ice-cold water for 1 h. (García-Arrarás et al., 1998; Rojas-Cartagena et al., 2007; Suárez-Castillo & García-Arrarás, 2007).

4.3.6 Immunohistochemistry

Gut rudiments were fixed in 4% paraformaldehyde (PFA) in 0.1M phosphate buffered saline (PBS) overnight at 4°C. After fixation, tissues were rinsed thrice with PBS (15 min each) and cryoprotected in 30% sucrose/PBS at 4°C. Gut rudiments were mounted in OCT compound (Sakura, Finetek, USA) and cryosectioned (20 μm) using a Leica CM1850 cryostat. Briefly, tissue sections

were permeabilized with 1% Triton X-100 for 15 min, followed by two washes with PBS. Then, slides were treated with 0.05M HCl for 1h. After an additional wash with PBS, the tissues were blocked with normal goat serum. Then the slides were incubated with the primary antibody OBSm antisera diluted in (1/200 or 1/500) RIA buffer (0.05 M PBS- pH 7.4, 0.15 M NaCl, 0.5% BSA, and 1.5 mM NaN_3) overnight. Next day, the slides were washed three times with PBS for 15 min each wash. Then, the tissue sections were incubated with the secondary antibody goat anti-mouse labeled with Cy3 (GAM-Cy3, 1:1,000, Jackson Immuno Research Laboratories) for 1 h at RT. After three additional washes with PBS, slides were mounted in buffered glycerol solution containing 1 µg/ml of 4',6-diamidino-2-phenylindole (DAPI, Sigma). Finally, tissue sections were visualized and analyzed using a Nikon Eclipse Ni fluorescence microscope.

4.3.7 *Ex vivo* studies

4.3.7.22 Cell culture

Gut rudiments with their associated mesenteries were dissected out from sea cucumbers at 4 dpe. Gut rudiment explants were surface disinfected as described previously (Bello, Abreu-Irizarry, & García-Arrarás, 2015) and carefully placed in 24-well plates, one explant per well. Each explant was covered with 1 mL of Leibovitz-15 cell culture medium (L-15) conditioned for marine species (Schacher & Proshansky, 1983) and supplemented with antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin, 50 µg/mL gentamicin), antifungal (2.5 µg/mL amphotericin B),

1x MEM nonessential amino acids, 1 mM sodium pyruvate, α-tocopherol acetate (1.75 µg/mL), and 1x ITS (Insulin, Transferrin, and Sodium Selenite). All cell culture reagents were purchased from Sigma-Aldrich. Explants were incubated in a modular incubator chamber (Billups-Rothenberg Inc.) at RT (20–24°C).

4.3.7.23 Pharmacological treatment of explants

Intestine explants dissected at 4 dpe were treated with LiCl (agonist). LiCl (Sigma) was added to the culture medium at a final concentration of 20 mM. NaCl was added to increase the concentration by 20 mM to treat control explants. Explants were grown in 700 µL of culture medium with LiCl, NaCl or without supplementation for 48–72 h. The medium was changed after 24 and 48 h.

4.3.7.24 RNA extraction and cDNA synthesis

RNA extraction was performed on tissue extracts of regenerating rudiments. After tissue removal from 24-well plates, tissue explants were placed in 1 mL of TRIzol reagent (Invitrogen), homogenized with a PowerGen Model 125 Homogenizer (Thermo Scientific) and incubated 30 min on ice. These samples were mixed vigorously with 200 µL of chloroform and incubated 10 min at RT. After centrifuged at 12,000 rpm at 4°C, the aqueous RNA phase was separated, mixed with 70% ethanol, and transferred to an RNeasy Mini Kit column (QIAGEN) for deoxyribonuclease (DNase) treatment (QIAGEN). Total RNA was extracted following the manufacturer's protocol. The concentration and purity of

the total RNA was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The cDNA was synthesized from 1 µg of the total RNA using the ImProm-II Reverse Transcription System (Promega) and oligo $(dT)_{23}$ primers.

4.3.7.25 Semi-quantitative RT-PCR

RT-PCR reactions were performed using cDNAs prepared from extracted RNA. These reactions were set up in a reaction volume of 25 µL with the final concentration of the PCR primers of 100 nM. Specific primers for the most variable regions between *Orpin A* and *Orpin B* sequences were designed. The primers used were: *Orpin B* forward: 5'-ACAGGGAGTACAAACAGTCGTCAA-3' and *Orpin B* reverse: 5'-CTATTTACTCTGCAACTGACACTTTCT-3'; *Orpin A* forward: 5'-ACTTCTGCAGAATCAGTTGTTAAGA-3' and *Orpin A* reverse: 5'-TTCAGTGGAGTCGCCAAC-3'. RT-PCR reactions were performed on three independent RNA samples purified from each of the regeneration stages (previously mentioned) as well as from the normal intestines. The PCR amplification was done by an initial denaturation step of 94°C (45 s), a primer annealing step of 50.2°C (45 s), and an extension step of 72°C (45 s) with a final additional 72°C (10 min) for 33 cycles for *Orpin A*, 28 cycles for *Orpin B*, and 26 cycles for *NADH*, as the amplification parameters for each pair of primers. All samples were analyzed in triplicate. The relative expression of *Orpin A* and *Orpin B* was normalized relative to the expression of the housekeeping gene *NADH dehydrogenase subunit 5* using ImageJ software (Schneider et al., 2012) from the optical density values from electrophoresed sample bands on agarose gels, using

a Molecular Imager ChemiDoc XRS+ (BioRad). The primers used for the *NADH* sequence amplification were: forward: 5'-CGGCTACTTCTGCGTTCTTC-3' and reverse: 5'-ATAGGCGCTGTCTCACTGGT-3'. The *Orpin A* and *Orpin B* sequences were confirmed by Sanger sequencing excised electrophoresed sample bands at the Sequencing and Genotyping Facility (UPR-RP) analyzed through Geneious software 11.1.5 (https://www.geneious.com).

4.3.7.26 Statistical analyses:

Statistical significance of the resulting data was evaluated through one-way ANOVA using the JMP®, Version 12. SAS Institute Inc., Cary, NC, 1989-2019. The multiple comparison procedure and statistical test Tukey-Kramer HSD (honestly significant difference) was used to determine significant differences between means from optical densities determined by ImageJ software as mentioned before (Schneider et al., 2012). The Tukey-Kramer results are displayed as circles that represents number of data points. Since all the data point groups have the same numbe, all the circles are of the same size. The red circles shows that no statistically detected difference was found between the group means. All values were reported as the mean ± standard mean error, including mean diamond with confidence interval ([1 − alpha] x 100). While a $P < .05$ and $P < .001$, were use as a criteria to determine statistical significance difference between groups.

4.4 Results

4.4.1 Production of Orpin anti-serum

To determine the cellular expression of Orpin, we immunized mice to obtain antibodies against Orpin. The expressed His-OBSm recombinant protein was used as an immunogen, as it does not includes the larger portion of the GST tag even though it does contains the additional 25 amino acid peptide in its C-terminal. The methodology for antibody production is described in the Material and Methods section. Two mice were injected with approximately 150 µg of the peptide diluted in 150 µL of 1 mM of PBS and Freund's Complete Adjuvant buffer, each. A boost was given one month following the initial injection and several blood samples were collected from the mice, 7, 16, and 21 days after. The serum was extracted from the collected blood and used for immunoblots and immunohistochemistry. Since this is a mixture of polyclonal antibodies, it is likely that antibodies contained in this serum can recognize proteins other than Orpin B. Among these are poly-histidine containing proteins (His tag), and proteins that might contain sequences corresponding to the 25 amino acids peptide added to the C-terminal of His-OBSm.

4.4.2 OBSm anti-serum recognized the Orpin B recombinant peptide and other poly histidine-tagged peptides

The antibody sera with the higher apparent affinity were chosen based on the intensity of labeling and decreased background interaction in immunohistochemistry studies. To determine the specificity of the His-OBSm

antisera, different protein samples were expressed and assayed for cross-reactivity in a Western Blot. These proteins were His-OBSm (antigen used as a positive control, 17 kDa), His-tagged fused protein β-galactosidase (His-β-galactosidase, 121 kDa), GST-OBS (37.8 kDa), and GST (negative control, 26 kDa) (Figs. 4.1A–C). These recombinant proteins were electrophoresed and transferred to a nitrocellulose (NC) membrane for the immunoblot assay. The bands of the corresponding molecular weights were analyzed (Fig. 4.1). The immunoblots showed that the antisera cross-reacted with the original antigen (His-OBSm), and with the other His-tagged fused protein, β-galactosidase. In addition, the antisera exhibited positive immune recognition to GST-OBS. However, the GST protein expressed as a control from pGEX-6P1, did not exhibit cross-reaction, showing the specificity of the antibody sera. The antisera also identified a > 26 kDa band in the GST-OBS sample. Previous discussed results suggested that this band corresponds to a proteolysis product that might contain the GST tag retaining a small part of the Orpin B sequence (See Chapter 3).

The labeling of this band (> 26 kDa) required further exploration and additional experiments were performed to understand its origin. In previous results (Chapter 3), we showed that after purification by gel filtration, three peptides eluted, GST-OBS, a > 26 kDa peptide, and a < 15 kDa peptide. In addition, the > 26 kDa peptide was shown to cross-react with the OBSm antisera. However, the < 15 kDa peptide was not detected. Given the low molecular weight of that peptide and low abundance in the GST-OBS sample, there was a possibility that it was lost,

passing through the NC membrane during the transfer. To clarify the meaning of these results and ensure retaining the < 15 kDa band peptide in the membrane, we transferred an ultrafiltered GST-OBS sample to a membrane, and processed it for Western Blot analysis. As we previously showed, GST-OBS expressed before optimization showed a high content of the two peptide bands in question. To ensure a high amount of the < 15 kDa peptide, a sample of this GST-OBS was ultrafiltered with an Amicon Ultra device (3 kDa molecular weight cutoff), electrophoresed and transferred to a NC membrane. A sample obtained before ultrafiltration was included for comparison. Cross-reactivity was detected with the GST-OBS band (37.8 kDa), the > 26 kDa and the < 15 kDa bands, along with a band migrating below the GST-OBS and two bands close to the 100 kDa based on the molecular weight marker (Fig. 4.1D). The sample obtained before ultrafiltration exhibited cross-reactivity to the latter two bands only. These results suggest that the > 26 kDa and < 15 kDa band peptides are proteolysis products from degradation of GST-OBS.
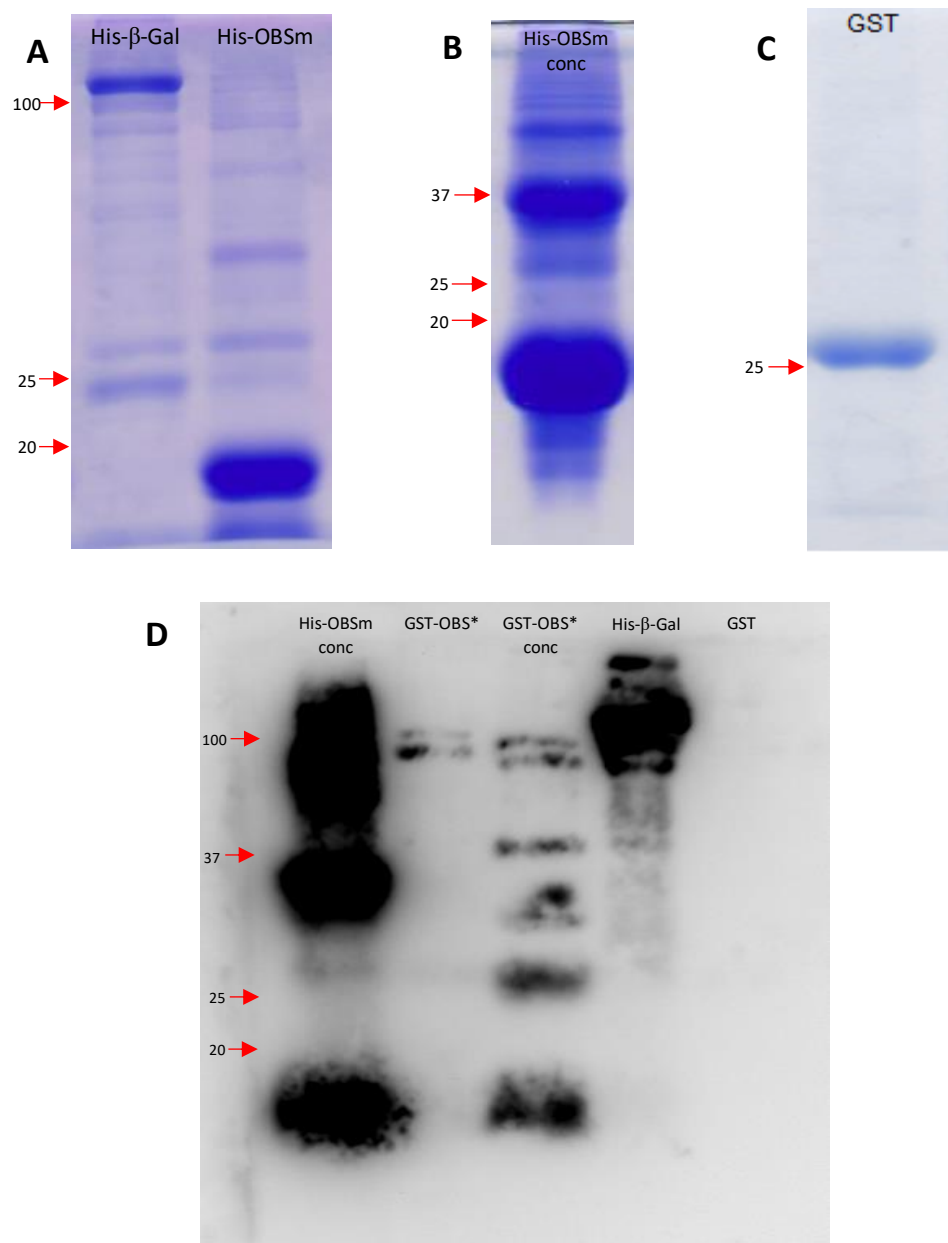
Figure 4.1. Anti-OBSm immunoreactivity in purified peptides

Protein samples were concentrated using Amicon Ultra filtration devices before separation through SDS-PAGE gel (A–C). Coomassie-stained SDS-PAGE gel 10% containing His-β-galactosidase and His-OBSm (A). Coomassie-stained SDS-PAGE gel 15% containing concentrated His-OBSm (B). Coomassie-stained SDS-PAGE gel 10% containing GST (C). 10%. Western Blot image from peptides separated in SDS-PAGE gel 10% (D). His-OBSm lane show several bands reacting to the antibody. The GST-OBS samples used for this immunoblot were expressed with conditions before the production optimization (*, asterisk). GST did not positively reacted with anti-OBSm. His-β-Gal: His-β-galactosidase, conc: concentrated sample, numbers and arrows: molecular weight in kDa based on the molecular marker.

As expected, the antibodies in the serum were able to recognize the proteins that share sequence regions to the His-OBSm antigen, showing its specificity. Nonetheless, antibodies in the sera also reacted to the His-tag as shown from the labeling of the His-β-galactosidase (Fig. 4.1D). The cross-reactivity detection specificity for the antiserum to the different recombinant proteins produced in this study is summarized in Figure 4.2.

Figure 4.2. Solubility and Immunoreactivity of the different recombinant peptides to the His-OBSm anti-serum
Different recombinant peptides expressed and purified for characterization or as control. Orpin B corresponds to the sequence including signal peptide. His-OBSm was the recombinant version of Orpin B fused to an N-terminal poly His tag and used for the production of the antibody sera. GST-OBS is the recombinant version of Orpin B fused to an N-terminal GST tag. GST and His-β-Gal (β-galactosidase fused to an N-terminal poly His tag) was used for experimental controls. Solubility (red) indicates degree of solubility based on GRAVY index prediction; (+): high solubility and (-): less solubility. GST-OBS (-) less solubility without Ca++ presence and (+) high solubility with Ca++ presence. Immunoreactivity, green checkmark: cross-reactivity with antibody sera; red X: no detected cross-reactivity with antibody sera.

4.4.3 Immunohistological localization of Orpin-expressing cells

4.4.3.27 Characterization of OBSm antibodies

The anti-OBSm serum was tested by immunohistochemistry in several holothurian tissues. Various cell types exhibited immunofluorescent labeling to anti-OBSm, including cells within the coelomic epithelia of the intestine, longitudinal muscle, and the mesothelium from the mesentery (composed of peritoneocytes and muscle cells) (Figs. 4.3–4.6). In some cells, labeling was observed in what appeared to be the plasma membrane while in muscle fibers of the longitudinal muscle the labeling appeared to be in the nuclear membrane (Fig. 4.6).
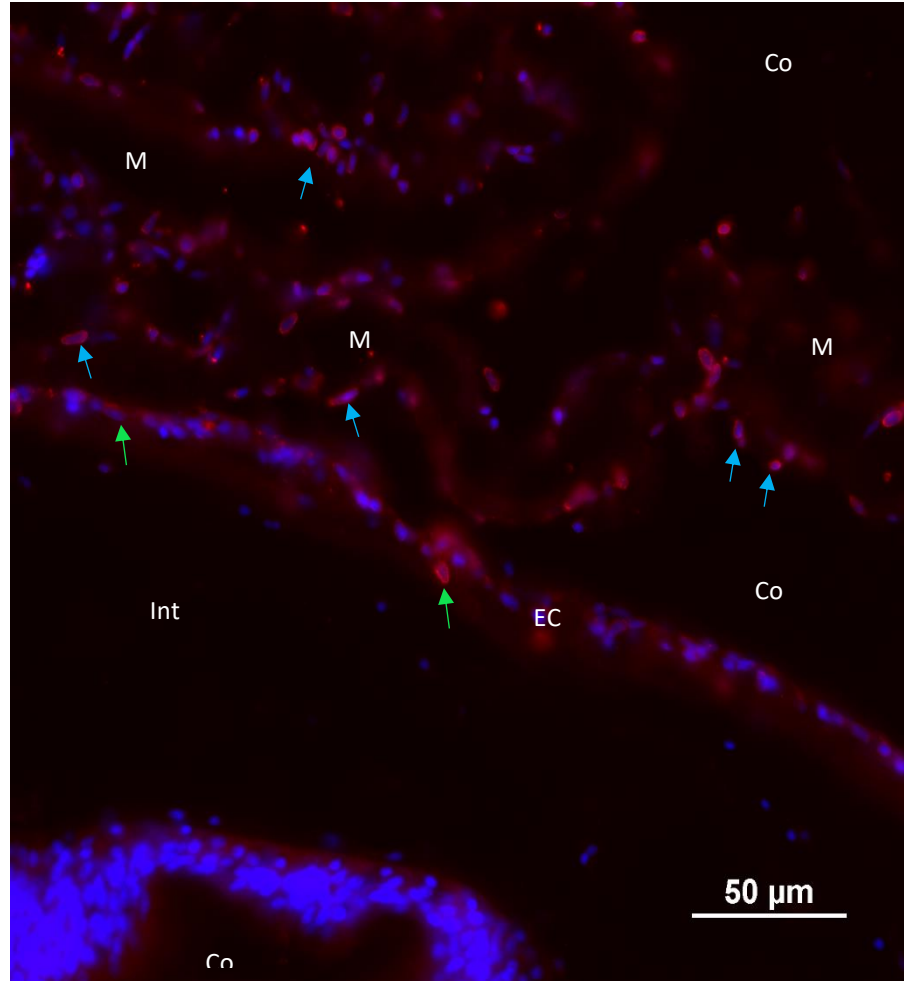
Figure 4.3. OBSm labeling in intestine coelomic epithelium and mesentery
Cells in the mesentery tissue and several coelomic epithelium cells in the intestine display labeling of the OBSm antibody. Red: anti-OBSm, blue: DAPI (nucleus). Magnification: 40 X, primary antibody dilution: 1/500, mouse 1.
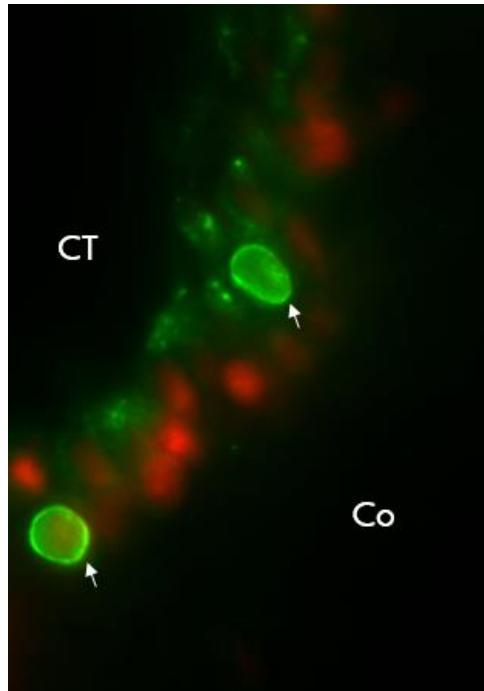
Figure 4.4. OBS immuno-positive cells in the normal intestine epithelia
Cells showing the labeling of OBSm antibodies are part of the coelomic epithelium of the normal non-regenerating intestine. The image represents some of the few immune-positive cells present in the intestine. Green: anti-OBSm, red: DAPI (nucleus), Co: coelom, CT: connective tissue, white arrows: OBSm immune-positive cells. Magnification: 100 X, primary antibody dilution: 1/500, mouse 1.
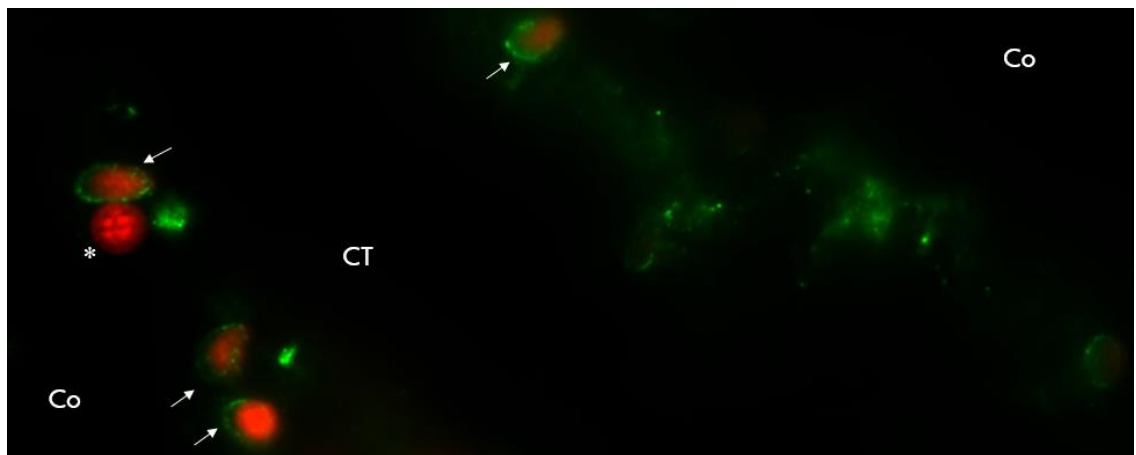

Figure 4.5. Normal non-regenerating mesentery IHC
OBSm immuno-positive cells display an irregular morphology compared to immune-negative cells (*). Green: anti-OBSm, red: DAPI (nucleus), Co: coelom, CT: connective tissue, white arrows: OBSm immuno-positive cells, *: immuno-negative cell nucleus. Magnification: 100 X, primary antibody dilution: 1/500, mouse 1.
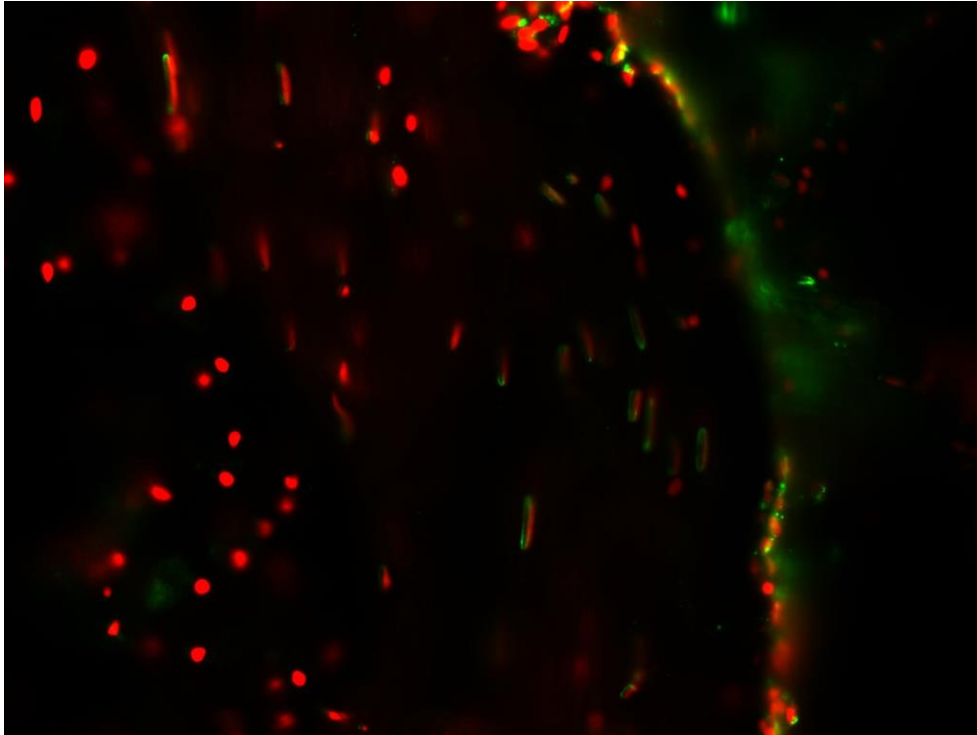
Figure 4.6. OBSm labeling in the longitudinal muscle
Longitudinal muscle tissue show labeling of the OBSm antibody surrounding several muscle fibers nucleus membrane. Green: anti-OBSm, red: DAPI (nucleus). Magnification: 40 X, primary antibody dilution: 1/200, mouse 1.

The cells labeled in the coelomic epithelium have a large irregular nucleus surrounded by a small amount of cytoplasm in comparison to other nearby non-labeled cells (Figs. 4.3 and 4.4).

4.4.4 Modulation of *Orpin* expression by the Wnt signaling system

4.4.4.28 *Ex vivo* experiments

A second set of preliminary experiments was aimed at determining if *Orpin* expression responded to cell signaling pathways. These experiments were part of

an ongoing effort in the lab to determine the role of signaling pathways on the initial steps of intestinal regeneration. We had shown that the Wnt signaling pathway was involved in the intestinal regeneration process (Bello et al., 2019). There is evidence in the scientific literature that EF-Hand calcium-binding proteins such as Naked Cuticle (Nkd) (Wharton, Zimmermann, Rousset, & Scott, 2001; D. Yan et al., 2001) and tescalcin (Kobayashi, Nakamura, & Wakabayashi, 2015) have been shown to interact with proteins directly involved in Wnt/β-catenin signaling pathway such as Dishevelled (Dvl) and GSK3, respectively. Since *Orpin* had been thought to be associated with the regenerative process we were therefore interested in exploring the possibility that *Orpin* expression itself was being modulated by the signaling pathways activated during the regenerative process. Thus, we proceeded to determine if the treatment with the classic activator of the Wnt signaling pathway, lithium chloride (LiCl), had an effect on *Orpin* expression.

For these *ex vivo* experiments the intestines of sea cucumbers that had been regenerating (gut rudiments) for 4 days were dissected out and placed in culture. These tissues were then treated with 20 mM LiCl and cultured for 7 days. Tissues were then processed for semi-quantitative RT-PCR measurement of *Orpin* isoforms mRNA expression. Semi-quantitative RT-PCR analyses were performed using isoform-specific primers to measure *Orpin A* and *Orpin B* levels from tissue extracts processed immediately after dissection of animals (Refer to Chapter 2 for primers information). The measurements were normalized relative to the expression of *NADH subunit 5*.

The controls used for these experiments were NaCl supplemented medium and non-treated medium. There was no statistically detected difference between *Orpin A* transcript expression from LiCl and the control samples relative to individual *NADH* expression ($P = .42$) (Fig. 4.8). Although, no statistically detected difference was found between experimental and control Orpin B levels ($P = .13$), a clear trend is visible that suggests that there is indeed an effect but that it appears to be non-significant due to a small number of experiments. The effect of LiCl was shown to be specific to *Orpin B*. The amplification of *Orpin B* bands from LiCl treatment samples was irregular in comparison to the NaCl samples and to the non-treated medium (Figs. 4.7 and 4.9).
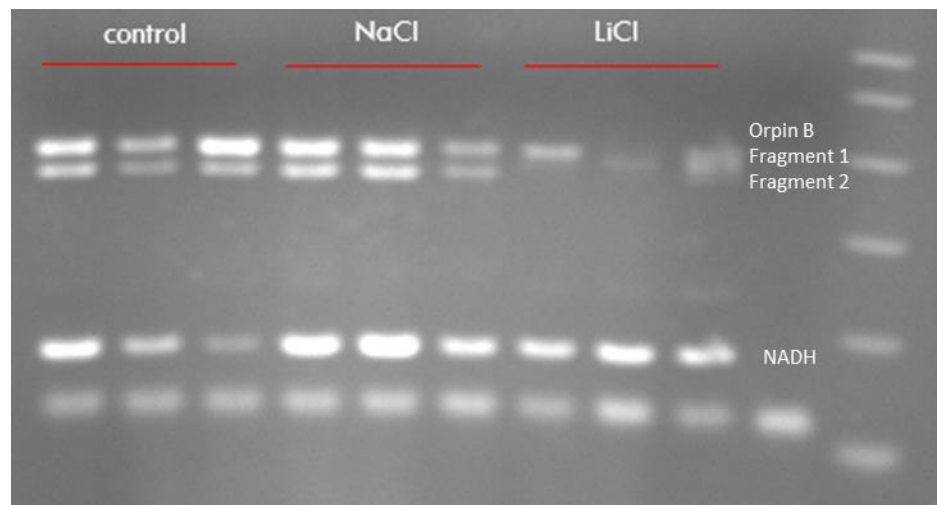

Figure 4.7. *Orpin B* semi-quantitative RT-PCR from gut rudiment explants treated with LiCl NaCl was used as a negative control and additional control samples included only the corresponding dilution volume of culture medium (non-treated medium). Two bands from *Orpin B* were amplified in the control samples. The transcription expression was compared to *NADH* expression. The lower bands corresponded to primer dimers. Three blastema-like rudiment explants from different animals at 4 days of regeneration (dpe), were used for RNA extraction and cDNA synthesis. Lanes 1–3: *Orpin A* transcript band from non-treated medium, lanes 4–6: NaCl treated explant samples, lanes 7–9: LiCl treated explant samples, lane 10: negative control with no cDNA, showing only the primer dimers. Electrophoresis was run in a 1.0% agarose gel.
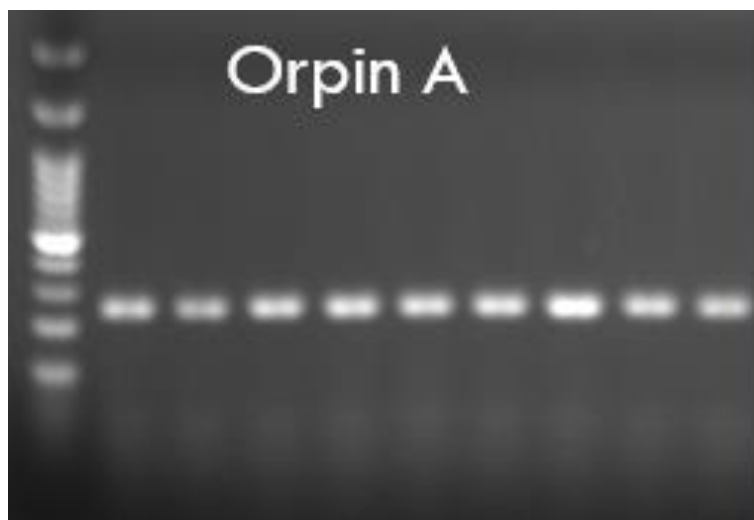
Figure 4.8. *Orpin A* semi-quantitative RT-PCR from gut rudiment explants treated with LiCl
No significant changes were detected in *Orpin A* amplicon expression levels during LiCl experiments relative to *NADH* expression (shown in Figure 4.7). Lanes 2–4: *Orpin A* transcript band from non-treated medium, lanes 5–7: NaCl treated explant samples, lanes 8–10: LiCl treated explant samples. Electrophoresis was run in a 1.0% agarose gel.
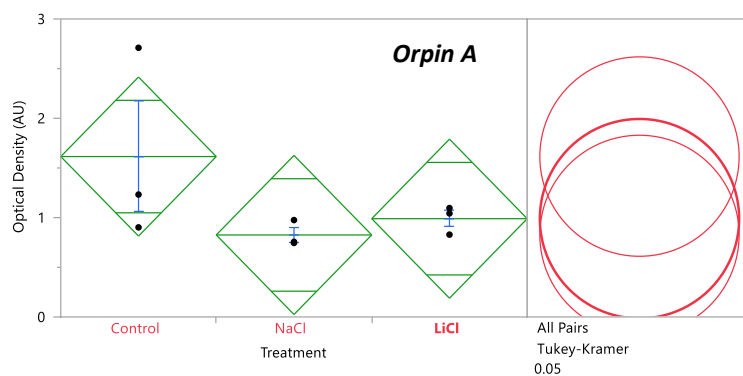

Figure 4.9. *Orpin A* transcript levels from explants treated with LiCl
Semi-quantitative RT-PCR amplification of *Orpin B* transcripts from mRNA samples from LiCl compared to the corresponding expression in samples from NaCl and non-treated growth media (control). No statistically detected difference was found between the group means as indicated in the all pairs Tukey-Kramer HSD test. Red circles shows that no statistically detected difference group means. Green diamonds: mean diamonds with long mid line representing the group mean, while the top and bottom of each diamond represents the (1 − alpha) x 100 confidence interval of each group. Green lines above and below the group mean: overlapping marks indicate that the two group means are not significantly different at the given confidence level. Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.

Figure 4.10. *Orpin B* transcript levels (top band) from explants treated with LiCl
Semi-quantitative RT-PCR amplification of *Orpin B* transcripts from mRNA samples from LiCl compared to the corresponding expression in samples from NaCl and non-treated growth media (control). No statistically detected difference was found between the group means as indicated in the all pairs Tukey-Kramer HSD test. Red circles shows that no statistically detected difference group means. Green diamonds: mean diamonds with long mid line representing the group mean, while the top and bottom of each diamond represents the (1 − alpha) x 100 confidence interval of each group. Green lines above and below the group mean: overlapping marks indicate that the two group means are not significantly different at the given confidence level. Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.
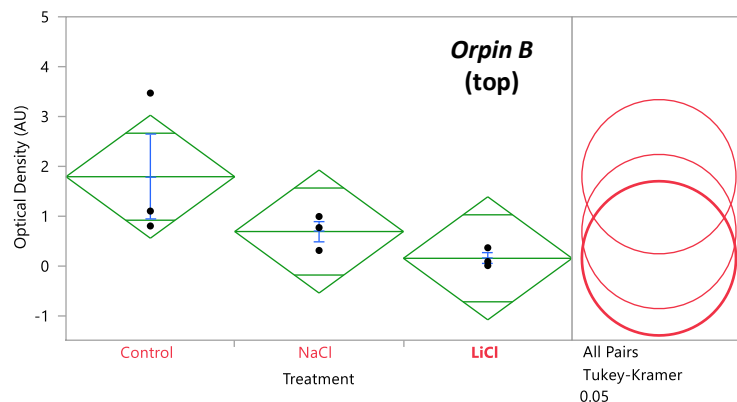

Figure 4.11. *Orpin B* transcript levels (bottom band) from explants treated with LiCl
Semi-quantitative RT-PCR amplification of *Orpin B* transcripts from mRNA samples from LiCl compared to the corresponding expression in samples from NaCl and non-treated growth media (control). No statistically detected difference was found between the group means (LiCl and non-treated media (control), *P* = .05) as indicated in the all pairs Tukey-Kramer HSD test. Red circles shows that no statistically detected difference group means. Green diamonds: mean diamonds with long mid line representing the group mean, while the top and bottom of each diamond represents the (1 − alpha) x 100 confidence interval of each group. Green lines above and below the group mean: overlapping marks indicate that the two group means are not significantly different at the given confidence level. Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.
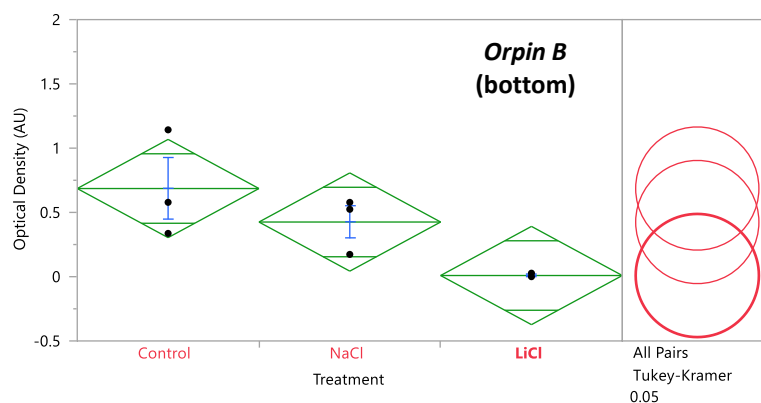
Moreover, the LiCl experiments resulted in another discovery. Although *Orpin A* and *NADH* amplicons resulted in the amplification of the corresponding bands each, *Orpin B* primers amplified the expected band of 475 bp plus an additional smaller band of 415 bp (Fig. 4.7).

To evaluate the possibility that the additional band is a splice variant, gel bands were further separated using 1.5% agarose gels and excised for the sequencing of the products (Fig. 4.10). The results quality was not great, even though, careful analyses of the sequences revealed two details. First, additional amplicons did not correspond to *Orpin A* sequence because the sequencing results matched only with the *Orpin B* sequence. Second, the longer band corresponded to the expected region encompassing from the 5' UTR to the second EF-hand motif encoding-segment of *Orpin B*. However, the shorter band sequence matched with a region of the 5' UTR and with a region of the coding sequence but lacked the portion encoding the signal peptide. This suggests the presence of a splicing variant from *Orpin B* with a deletion of the first ~ 50 bp of the coding sequence corresponding to the predicted signal peptide encoding-region (Fig. 4.11).
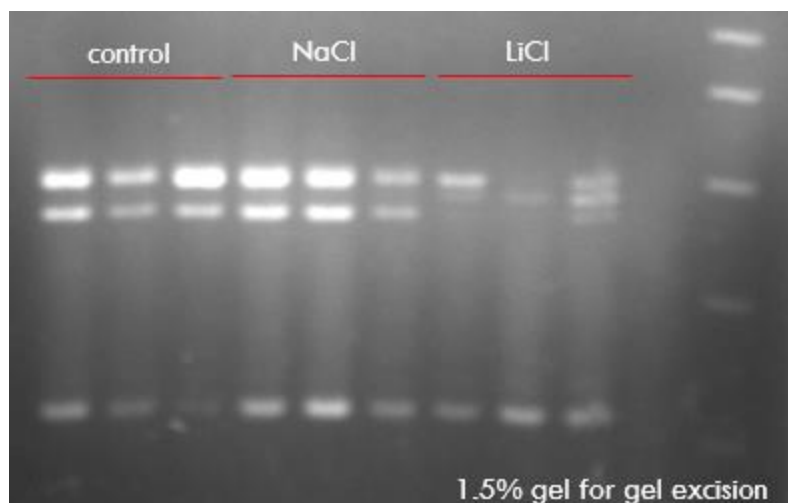
Figure 4.12. Separation of samples from the Figure 4.7 on 1.5% agarose gel for band excision and sequencing analysis
The bands from the same lane were further separated to facilitate gel band excision and further sequencing analysis. Lanes 1–3: *Orpin A* transcript band from non-treated medium, lanes 4–6: NaCl treated explant samples, lanes 7–9: LiCl treated explant samples. The electrophoresis was run for 60 min at 150 V.
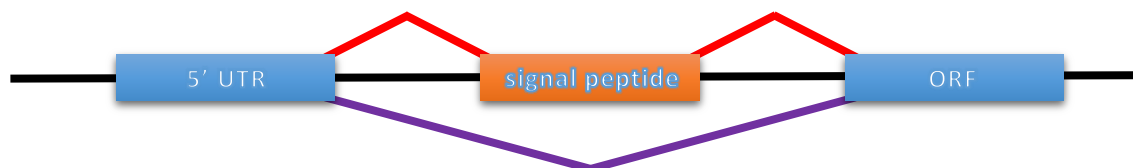
Figure 4.13. Schematic of alternative splicing found between *Orpin B* RT-PCR amplicons from *ex vivo* gut rudiment explants cultures
The diagram shows the two transcript variants found in gut rudiment explant cultures. Blue boxes represent the conserved region between both transcript variants. The red line represents the transcript sequence of *Orpin B* including the signal peptide encoding-region, whereas the purple line represents the transcript sequence lacking the signal peptide encoding-region.

4.5 Discussion

These studies were performed to provide insights into the possible role of Orpin

in the sea cucumber *Holothuria glaberrima*. First, we produced an antibody serum

to help characterize the Orpin-expressing cells. The antibody sera were produced

with the expression of a recombinant version of Orpin B (His-OBSm) with genetic modifications such as deletion of the N-terminal signal peptide encoding-region, N-terminal polyhistidine tag to facilitate the purification of the peptide, and the insertion of 25 residues from the pET200 plasmid vector to the C-terminal. Although these modifications were done to increase the final production of the expressed recombinant product in a soluble and stable form, they also complicate the analysis of the obtained results. In particular, because antibodies might be produced against some of the additional amino acid sequences, and these antibodies could cross-react with some holothurian proteins. For example, our group had reported the interaction of antibodies raised against different proteins from other organisms exhibited cross-reaction with a homolog protein that shares a similar domain sequence. In that report they showed that antibodies made against rat EF-Hand proteins calbindin 1 (calbindin-D28k), parvalbumin, and calbindin 2 (calretinin), displayed labeling of cells and fibers within the neural system of *H. glaberrima* (Díaz-Balzac et al., 2012). We also demonstrated that the first two antibodies worked for Western Blot analyses as well as for the detection of a recombinant calbindin-D32K (calbindin 2) and other proteins in extracted tissues from this sea cucumber and sea urchin. Even though each antibody was produced against a different EF-Hand protein (calbindin 1 and parvalbumin), the two different antibodies cross-reacted with the same protein band from the tissues extracted. Thus, we cannot disregard that anti-OBSm could interact with other proteins with similar topology.

The antibody sera was tested for specificity through Western Blot using different recombinant products that some shared sequence portions with the original Orpin B and the others were used as controls. As expected, the antisera exhibited cross-reactivity with the original antigen, His-OBSm (positive control), GST-OBS, and His-β-galactosidase and none to GST (negative control). As we described early, GST-OBS is another recombinant version of Orpin B that also lacks the signal peptide encoding region and was expressed fused to an N-terminal GST tag. The detection of GST-OBS suggested that the antisera is able to cross-react with Orpin B. In addition, as the antisera did not cross-react with the recombinant GST, the detection is considered specific to the Orpin B sequence directly. As previously described (see chapter 2), Orpin B is an isoform of Orpin A and shares high sequence similarity and identity. Thus, it is expected that the serum is able to recognize both proteins in cells that express either or both of the isoforms in tissues. Furthermore, as the recombinant His-β-galactosidase also exhibited cross-reaction, this suggested that the antisera also detected the N-terminal poly-histidine tag as His-OBSm antigen shares this tag sequence. In the future, this feature can be exploited by binding the antisera to polyhistidine-coupled resin, to identify protein partners of Orpin isoforms through co-immunoprecipitation procedures.

4.5.1 Antibody sera labeling in tissues from *H. glaberrima*

After determining that the antibody cross-reacted with the Orpin sequence, we proceeded to investigate different tissues from the sea cucumber to identify cells that express this protein isoforms.

Positively labeled cells were found in tissues such as normal non-regenerating intestine, mesentery, and longitudinal muscle. Moreover, the labeling of immune-positive cells was distinctive in a tissue-specific manner. The positive-labeled cells from the longitudinal muscle tissue exhibited localized cross-reactivity in the nuclear membrane on muscle fibers. Various EF-Hand proteins had been identified in different muscle tissues and intracellular localization. For example, members of the EF-Hand subfamily of S100s were identified in cells from the human vascular smooth muscle and from intestinal smooth muscle, with a differential localization pattern with use of polyclonal antibodies (Mandinova et al., 1998). S100A1 and S100A4 were found only in the cytoplasm and strongly associated with the sarcoplasmic reticulum and with actin stress fibers. S100A2 was localized only in the cell nucleus. S100A6 was located in in both the sarcoplasmic reticulum and in the cell nucleus. In another study, osteonectin or SPARC (Secreted Protein, Acidic and Rich in Cysteine), was detected in the skeletal muscle and vascular smooth muscle as in a wide range of other normal human tissues and tumors (Porter, Sage, Lane, Funk, & Gown, 1995). A monoclonal antibody was produced to a Ca^{+2}-binding region of SPARC and the immunostaining in human tissues was almost exclusively intracellular.

We cannot disregard the possibility that positively labeled cells can express molecules that interact indirectly with the antisera. Although, based on sequence similarity to other protein sequences from the sea cucumber encoding transcriptome, this possibility is unlikely. As previously discussed, the recombinant version of Orpin B used as an antigen, contained an N-terminal polyhistidine tag and a 25 residues long peptide attached to its C-terminal, and thus it is expected that the sera includes antibodies to either of these two regions. We searched the sea cucumber encoding transcriptome for similar sequences using the complete recombinant protein (His-OBSm), the His tag portion, or the additional C-terminal 25 residues portion sequence. Although, no additional sequences than Orpin A and Orpin B were found based in sequence similarity.

In summary, the evidence suggests that Orpin proteins are expressed in the cells from the coelomic epithelium of non-regenerating intestine although in a notable high number in the associated mesentery tissue, with cytoplasmic localization. Furthermore, it is suggested that Orpin proteins are also expressed in cells from the longitudinal muscle, localized in the nuclear membrane. However, additional antibodies produced against specific portions of each isoform will be needed to confirm if the protein expression is spatio-temporal and cell-specific. In order to confirm the specificity of the antisera labeling, pre-absorption experiments will be required.

4.5.2 *Orpin B* specific transcription alteration during activation of Wnt pathway activation

In another aspect, we performed experiments to investigate the effect of *Orpin* isoform transcription during the activation of the Wnt pathway, a known signaling pathway that is activated during sea cucumber regeneration. Upon activation of the canonical or Wnt/β-catenin signaling, Disheveled (Dvl) inhibits the constitutively destruction complex of β-catenin, consisting of Axin, GSK3β and APC. Consequently, β-catenin accumulates in the cytoplasm, and then translocates to the nucleus to modulate gene transcription by interacting with LEF1/TCF transcription factors. Conversely, Wnt signaling pathways can be inhibited by the EF-Hand protein, Nkd, upon binding to Dvl and β-catenin (Rousset et al., 2001; Van Raay, Coffey, & Solnica-Krezel, 2007; Van Raay et al., 2011). Also, other studies showed that tescalcin, another EF-Hand protein, inhibits GSK3 (Kobayashi et al., 2015). Previous work from our laboratory group and that of others has provided evidence of the involvement of the Wnt pathway during the intestinal regeneration in sea cucumbers (Bello et al., 2019; Girich, Isaeva, & Dolmatov, 2017; Mashanov, Zueva, & Garcia-Arraras, 2012; Ortiz-Pineda et al., 2009; Sun, Yang, Chen, & Xu, 2013; Yuan et al., 2019). It was shown that the activation of this pathway had an effect on cell differentiation and cell proliferation. LiCl has been used to activate the Wnt pathway both *in vivo* and *in vitro* (Jope, 2003; Klein & Melton, 1996; Stambolic, Ruel, & Woodgett, 1996). LiCl activation of this pathway is known to promote activation of cell proliferation via an inhibition of the enzyme GSK3, which then results in an increase in nuclear β-catenin levels.

Bello et al. observed that LiCl enhanced the formation of the gut rudiment, increasing its size. In our experiments, we observed a trend where *Orpin B* transcript expression was decreased in LiCl treated explants when compared to the experimental controls (NaCl or non-treated culture media). Although the decrease was notable, the difference was not statistically detected ($P = .05$) with the 3 replicates used in this experiment. Additional experiments with higher sample number will be required for confirmation of the differential expression. The decrease in transcript expression was not observed in *Orpin A* transcript expression, thus LiCl effect was isoform-specific. LiCl induced mesenchymal-epithelial differentiation during human kidney development by activation of the Wnt signal pathway (Price, Kolatsi-Joannou, Mari, Long, & Winyard, 2018). Following LiCl exposure, LEF1 mRNA levels showed a significant decrease in mesenchymal cells. Furthermore, led to the downregulation of renal progenitor (*SIX2*, *EYA1*, and *CD133*) and mesenchymal markers (*HGF, CD24*). Previous work showed that the EF-Hand protein, naked cuticle homolog 1 (NKD1) was downregulated in human osteosarcoma, suggesting that this may be involved in the proliferation and migration of osteosarcoma cells through the activation of the canonical Wnt signaling pathway (Chen, Xu, Zhu, & Liu, 2018).

In addition, we found another interesting effect upon treatment with LiCl that was also isoform specific. The corresponding band pattern was also affected by LiCl presence as additional shorter bands appeared in some samples. This transcription alteration was not observed in *Orpin A* transcript expression.

Although, the obtained results in this study did not provided additional evidence to understand this surprising finding.

4.5.3 Alternate splicing on *Orpin B* transcription during cultured of regenerating intestine explants culture

Interestingly, we found the presence of an additional *Orpin B* amplicon band in the control samples incubated with NaCl or with non-treated medium. This was not observed in *Orpin A* nor in *NADH* amplicon bands. We can exclude some of the possibilities regarding the cause of the second *Orpin B* amplicon band. First, RNA samples' integrity was not compromised because the other gene amplicons amplified as expected. Second, cDNA samples were not degraded because there were only amplified two bands instead of a smear. Third, the cause of the additional band did not affect the expression of *Orpin A* nor *NADH* amplicons. After sequencing of the excised gel bands from *Orpin B*, it was found that the longer band corresponded to a region of the 5' UTR to the C-terminal EF-hand motif encoding sequence of the *Orpin B* (See Chapter 2 for sequence reference). The shorter band sequence started with the same 5' UTR region and ended with the same EF-hand motif encoding sequence with a deletion of an intermediate sequence region immediately downstream to the 5' UTR. Interestingly, the sequence portion missing in this amplicon band corresponded to the signal peptide encoding-region from the original identified *Orpin B*. This suggested that an alternative splicing event was caused by the culture procedure of the explants. Splice variants lacking the signal peptide region from the extended SPARC family

of protein have been previously identified. The detection of multiple cell-type specific isoforms arise due to a combination of alternative splicing and post-translational modification revealed the complexity of this family proteins (Viloria et al., 2016). For four of the seven currently known members of the extended SPARC family of proteins, at least one alternative variant lacking the signal peptide was identified, such as SPARC-like 1 or Hevin, SPOCK-1, SPOCK-3 and SMOC-2. This family of matricellular proteins are multifunctional regulators of cellular and matrix interactions. A detailed summary of the diverse functions that have been identified for each of the members of this family as in extracellular matrix activity, growth factor activity, and cell adhesion, is discussed by Bradshaw, 2012 (Bradshaw, 2012). Viloria et al., 2016 reported expression of each member of the extended SPARC family in pancreatic β cells, stromal cells and in ducts. The report shows the identification *in vivo* in pancreatic sections from normal mouse through immunohistochemistry and *in vitro* from numerous pancreas derived cell lines through western blot. Even though all of the main representative members of this family contain a signal peptide and thus expected to be found extracellularly, they observed extensive cytoplasmic staining and in some cases in the nucleus. Moreover, they identified multiple isoforms and splicing variants, which included versions lacking the signal peptide encoding-region, from ENSEMBL database. The alternative splice variants lacking the signal peptide suggested that both intracellular and extracellular isoforms of these proteins exist. These findings share similarities with Orpin such as having EF-Hand calcium binding domain, signal peptide, and alternative splice variants lacking this signal peptide. These also

provide a possible explanation for the intracellular identification of Orpin by immunohistochemistry, although it contains a predicted signal peptide encoding-region.

These results suggested that Orpin isoforms might have specific differential roles in the cellular and molecular events from the sea cucumber. The experiments described in this study provided insights that will help in future experimental designs for the elucidation of the role of Orpin isoforms in *H. glaberrima*.

4.6 References

Bello, S. A., Abreu-Irizarry, R. J., & García-Arrarás, J. E. (2015). Primary cell cultures of regenerating holothurian tissues. *Methods in Molecular Biology (Clifton, N.J.), 1189*, 283–297. https://doi.org/10.1007/978-1-4939-1164-6_19

Bello, S. A., Torres-Gutiérrez, V., Rodríguez-Flores, E. J., Toledo-Román, E. J., Rodríguez, N., Díaz-Díaz, L. M., … García-Arrarás, J. E. (2019). Insights into intestinal regeneration signaling mechanisms. *Developmental Biology*. https://doi.org/10.1016/j.ydbio.2019.10.005

Bradshaw, A. D. (2012). Diverse Biological Functions of the SPARC Family of Proteins. *Int J Biochem Cell Biol., 44*(3), 480–488. https://doi.org/10.1016/j.biocel.2011.12.021.Diverse

Chen, X., Xu, P., Zhu, J., & Liu, F. (2018). Downregulation of NKD1 in human osteosarcoma and its clinical significance. *Molecular Medicine Reports, 17*(1), 1111–1117. https://doi.org/10.3892/mmr.2017.7968

Díaz-Balzac, C. A., Lázaro-Peña, M. I., García-Rivera, E. M., González, C. I., & García-Arrarás, J. E. (2012). Calbindin-D32k is localized to a subpopulation of neurons in the nervous system of the sea cucumber Holothuria glaberrima (Echinodermata). *PloS One, 7*(3), e32689. https://doi.org/10.1371/journal.pone.0032689

García-Arrarás, J. E., Estrada-Rodgers, L., Santiago, R., Torres, I. I., Díaz-Miranda, L., & Torres-Avillán, I. (1998). Cellular mechanisms of intestine regeneration in the sea cucumber, Holothuria glaberrima Selenka (Holothuroidea:Echinodermata). *The Journal of Experimental Zoology, 281*(4), 288–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9658592

Girich, A. S., Isaeva, M. P., & Dolmatov, I. Y. (2017). Wnt and frizzled expression during regeneration of internal organs in the holothurian Eupentacta fraudatrix. *Wound Repair and Regeneration, 25*(5), 828–835. https://doi.org/10.1111/wrr.12591

Jope, R. S. (2003). Lithium and GSK-3: one inhibitor, two inhibitory actions, multiple outcomes. *Trends in Pharmacological Sciences, 24*(9), 441–443. https://doi.org/10.1016/S0165-6147(03)00206-2

Klein, P. S., & Melton, D. A. (1996). A molecular mechanism for the effect of lithium on development. *Proceedings of the National Academy of Sciences, 93*(16), 8455–8459. https://doi.org/10.1073/pnas.93.16.8455

Kobayashi, S., Nakamura, T. Y., & Wakabayashi, S. (2015). Calcineurin B homologous protein 3 negatively regulates cardiomyocyte hypertrophy via inhibition of glycogen synthase kinase 3 phosphorylation. *Journal of Molecular and Cellular Cardiology*, *84*, 133–142. https://doi.org/10.1016/j.yjmcc.2015.04.018

Mandinova, A., Atar, D., Schäfer, B. W., Spiess, M., Aebi, U., & Heizmann, C. W. (1998). Distinct subcellular localization of calcium binding S100 proteins in human smooth muscle cells and their relocation in response to rises in intracellular calcium. *Journal of Cell Science*, *111*(14), 2043–2054.

Mashanov, V. S., Zueva, O. R., & Garcia-Arraras, J. E. (2012). Expression of Wnt9, TCTP, and Bmp1/Tll in sea cucumber visceral regeneration. *Gene Expression Patterns*, *12*(1–2), 24–35. https://doi.org/10.1016/j.gep.2011.10.003

Ortiz-Pineda, P. A., Ramírez-Gómez, F., Pérez-Ortiz, J., González-Díaz, S., Santiago-De Jesús, F., Hernández-Pasos, J., … García-Arrarás, J. E. (2009). Gene expression profiling of intestinal regeneration in the sea cucumber. *BMC Genomics*, *10*, 262. https://doi.org/10.1186/1471-2164-10-262

Porter, P. L., Sage, E. H., Lane, T. F., Funk, S. E., & Gown, A. M. (1995). Distribution of SPARC in normal and neoplastic human tissue. *Journal of Histochemistry and Cytochemistry*, *43*(8), 791–800. https://doi.org/10.1177/43.8.7622842

Price, K. L., Kolatsi-Joannou, M., Mari, C., Long, D. A., & Winyard, P. J. D. (2018). Lithium induces mesenchymal-epithelial differentiation during human kidney development by activation of the Wnt signalling system. *Cell Death Discovery*, *4*(1). https://doi.org/10.1038/s41420-017-0021-6

Rojas-Cartagena, C., Ortíz-Pineda, P., Ramírez-Gómez, F., Suárez-Castillo, E. C., Matos-Cruz, V., Rodríguez, C., … García-Arrarás, J. E. (2007). Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber Holothuria glaberrima. *Physiological Genomics*, *31*(2), 203–215. https://doi.org/10.1152/physiolgenomics.00228.2006

Rousset, R., Mack, J. A., Wharton, K. A., Axelrod, J. D., Cadigan, K. M., Fish, M. P., … Scott, M. P. (2001). Naked cuticle targets dishevelled to antagonize Wnt signal transduction. *Genes and Development*, *15*(6), 658–671. https://doi.org/10.1101/gad.869201

Schacher, S., & Proshansky, E. (1983). Neurite regeneration by Aplysia neurons in dissociated cell culture: modulation by Aplysia hemolymph and the presence of the initial axonal segment. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *3*(12), 2403–2413.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675.

Stambolic, V., Ruel, L., & Woodgett, J. R. (1996). Lithium inhibits glycogen synthase kinase-3 activity and mimics Wingless signalling in intact cells. *Current Biology*, *6*(12), 1664–1669. https://doi.org/10.1016/S0960-9822(02)70790-2

Suárez-Castillo, E. C., & García-Arrarás, J. E. (2007). Molecular evolution of the ependymin protein family: a necessary update. *BMC Evolutionary Biology*, *7*(1), 23. https://doi.org/10.1186/1471-2148-7-23

Sun, L. N., Yang, H. S., Chen, M. Y., & Xu, D. X. (2013). Cloning and expression analysis of Wnt6 and Hox6 during intestinal regeneration in the sea cucumber Apostichopus japonicus. *Genetics and Molecular Research*, *12*(4), 5321–5334. https://doi.org/10.4238/2013.November.7.7

Van Raay, T. J., Coffey, R. J., & Solnica-Krezel, L. (2007). Zebrafish Naked1 and Naked2 antagonize both canonical and non-canonical Wnt signaling. *Developmental Biology*, *309*(2), 151–168. https://doi.org/10.1016/j.ydbio.2007.04.018

Van Raay, T. J., Fortino, N. J., Miller, B. W., Ma, H., Lau, G., Li, C., … Coffey, R. J. (2011). Naked1 Antagonizes Wnt Signaling by Preventing Nuclear Accumulation of β-Catenin. *PLoS ONE*, *6*(4), e18650. https://doi.org/10.1371/journal.pone.0018650

Viloria, K., Munasinghe, A., Asher, S., Bogyere, R., Jones, L., & Hill, N. J. (2016). A holistic approach to dissecting SPARC family protein complexity reveals FSTL-1 as an inhibitor of pancreatic cancer cell growth. *Scientific Reports*, *6*(August), 1–19. https://doi.org/10.1038/srep37839

Wharton, K. A., Zimmermann, G., Rousset, R., & Scott, M. P. (2001). Vertebrate proteins related to Drosophila naked cuticle bind dishevelled and antagonize Wnt signaling. *Developmental Biology*, *234*(1), 93–106. https://doi.org/10.1006/dbio.2001.0238

Yan, D., Wallingford, J. B., Sun, T. Q., Nelson, A. M., Sakanaka, C., Reinhard, C., … Williams, L. T. (2001). Cell autonomous regulation of multiple Dishevelled-dependent pathways by mammalian Nkd. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 3802–3807. https://doi.org/10.1073/pnas.071041898

Yuan, J., Gao, Y., Sun, L., Jin, S., Zhang, X., Liu, C., … Xiang, J. (2019). Wnt Signaling Pathway Linked to Intestinal Regeneration via Evolutionary Patterns and Gene Expression in the Sea Cucumber Apostichopus japonicus. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00112