



Università degli Studi di Padova  
Facoltà di Scienze MM.FF.NN.

Laurea Specialistica in Biologia Sanitaria

# **Molecular Studies of Piscine Hatching Enzymes**

## **Choriolysin gene structure and expression**

Relatore: **Prof. Francesco Argenton**  
Dipartimento di Biologia, Università degli Studi di Padova

Correlatore: **Prof. Bernt T. Walther**  
Molekylærbiologisk Institutt, Universitetet i Bergen

Laureanda: **Valeria Orsetti**

Anno Accademico: anno 2006/2007



## *Acknowledgements*

*First of all I wish to thank my family, for the loving support, morale as well as economic. I am much grateful to my parents, Mirka and Giuliano, for enabling me to know one of the most amazing country in the world, Norway, and to choose it as destination in my last year of university. I also want to thank my grandparents, Lisa and Sergio, even if they fed me a bit too much.*

*Many thanks to the Fantoft students. Thank you Pie, for making my Erasmus, rolling down the lawns and swimming in the fjords, Bergen wouldn't have been the same without you. Thanks also to Christina Benedetta and Petra, for sharing joy and sorrow of our distance from home; Matteo Rok Claudia Steffie and Mathias for parting every night, because we don't care about the rain!*

*I need to thank my Norwegian supervisor, Bernt Walther, for welcoming me and making me part and parcel of his team. Thanks to the laboratory staff, Javad and Mirushe who taught me all the lab-tricks I know and looked after me for saving our equipment. Thanks to salmons and zebrafishes for their embryos.*

*For all this... Thank you, Bergen!!*

*Thanks to Francesco Argenton, my Italian supervisor, for rushing a correction of this thesis in few days only, anybody else probably fell asleep long before the end.*

*Last but not least my Italian friends, without whom I could not live, maybe too many for listing them all but not for taking into my heart. Special thanks to Christian Nicola Marta Elena Betty Alessandro Francesco Anna M. Simon Grazia and Anna Z. for supporting me here and from afar. Thanks for the cheerful days I spent with you and for your patience, I know it was not easy, but with you I improved myself.*



## *Abstract*

*Hatching* refers to the embryonic exit from the eggshell (Chorion). It is classified in two types, mechanical and enzymatic hatching.

The enzymes involved in this phenomenon are named Hatching Enzymes and they are known and characterized in many species. They are produced by hatching gland cells, stored in granules and released into perivitelline space at the time of hatching.

Chorion consists of two layers: a thick inner part called *zona radiata*, composed of 3 main protein monomers, specific target of the hatching proteases, and a thin outer layer, *zona pellucida*. During hatching, protein aggregates are depolymerised and become water soluble.

Choriolysin (salmon hatching enzyme) is a Zn-protease which belongs to the Astacin family, a group of metallo-endopeptidases with more than 20 members now identified. Some proposed functions include activation of growth factors, degradation of polypeptides and processing of extracellular proteins.

The family members are characterized by a unique 18-amino acids signature sequence, which encompassed the Zn binding motif “HExxH”.

Atlantic salmon *Salmo salar* enzymatic hatching is the subject of the thesis. The present work shows the last 558 nucleotides of the total 786, which compose Choriolysin sequence and its gene expression in embryo development.



## *Riassunto in italiano*

Il termine “Schiusa” fa riferimento alla fuoriuscita dell’embrione dal guscio dell’uovo (il Chorion). Questa può essere classificata in due tipi, schiusa meccanica o enzimatica.

Gli enzimi coinvolti in nel fenomeno di rottura del Chorion sono detti enzimi della Schiusa e sono noti e caratterizzati in molte specie. Questi enzimi sono prodotti da specifiche cellule ghiandolari, dette cellule della ghiandola della schiusa, successivamente immagazzinati in granuli e rilasciati nello spazio perivitellinico al momento in cui l’embrione è pronto ad uscire dal guscio.

Il Chorion è formato da due strati: uno interno più spesso chiamato *zona radiata*, composto da 3 principali proteine monometriche (target specifico degli hatching enzymes), e uno strato esterno più fino detto *zona pellucida*. Durante la rottura del Chorion, gli aggregati proteici sono depolimerizzati e diventano idrosolubili.

La Choriolisina (l’enzima della schiusa di salmone) è una zinco-proteasi che appartiene alla famiglia delle Astacine. Questo gruppo di metallo-endopeptidasi comprende più di 20 membri e sono caratterizzati da un’unica sequenza di 18 aminoacidi, che comprende il motivo di legame allo zinco HExxH.

La schiusa enzimatic del salmone atlantico *Salmo salar* è oggetto di questa tesi. Il mio lavoro aveva due scopi principali, determinare la sequenza nucleotidica di Choriolisina di salmone e studiarne la sua espressione genica in diversi stadi di sviluppo embrionale.

L’elaborato mostra gli ultimi 558 nucleotidi dei 786 totali che compongono la sequenza genica (cDNA), trovati con le tecniche di Polymerase Chain Reaction (PCR) e Rapid Amplification of cDNA ends (RACE) -PCR. L’espressione genica durante lo sviluppo embrionale è stato analizzato dallo stadio di uova non fertilizzate alla schiusa, utilizzando la Real-Time PCR.

Choriolysin dovrebbe essere un gene con normali caratteristiche, quali la presenza di una copia che trascrive per una proteina e l’eredità mendeliana. Studi che sostengono questa ipotesi possono essere utilizzati nel confronto con il gene e la proteina di Zonase, altro enzima della schiusa trovato nel fluido perivitellinico di salmone.



# *Contents*

<b>ACKNOWLEDGEMENTS.....</b>	<b>I</b>
<b>ABSTRACT .....</b>	<b>III</b>
<b>RIASSUNTO IN ITALIANO.....</b>	<b>V</b>
<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 HATCHING PHENOMENON.....	1
1.1.1 The teleostean hatching.....	1
1.1.2 The egg envelope: Chorion.....	2
1.1.3 The hatching-gland cells .....	3
1.1.3.1 Origin and location.....	3
1.1.3.2 Morphology.....	4
1.1.3.3 Secretion.....	5
1.1.4 Digestion of the eggshell: Choriolysis .....	6
1.1.5 Regulation of the hatching process .....	6
1.1.5.1 Oxygen.....	7
1.1.5.2 Temperature.....	7
1.1.5.3 Light.....	7
1.1.5.4 Other factors.....	7
1.1.6 Embryonic development of Zebrafish <i>Danio rerio</i> .....	8
1.2 THE ASTACIN FAMILY.....	9
1.2.1 Proposed functions .....	9
1.2.2 Primary structure.....	11
1.2.3 Tertiary structure and domains.....	11
1.2.4 Effects of non-catalytic domains: Activation.....	14
1.3 THE HATCHING PROTEASES .....	14
1.3.1 Medaka <i>Oryzias latipes</i> Choriolysin .....	14
1.3.1.1 Proteolytic action .....	15
1.3.1.2 The genomic sequence.....	15
1.3.1.3 HCE, High Choriolytic Enzyme .....	16
1.3.1.4 LCE, Low Choriolytic Enzyme.....	17
1.3.1.5 Differences in chemical characteristics and enzymological properties.....	17
1.3.2 Zebrafish <i>Danio rerio</i> Choriolysin .....	18
1.3.3 Phylogenetic evolution.....	18
1.4 ATLANTIC SALMON <i>SALMO SALAR</i> HATCHING ENZYMES.....	19
1.4.1 Choriolysin.....	19
1.4.2 Zonase .....	20
1.5 AIM OF THE STUDY .....	21
<b>2. MATERIALS AND METHODS .....</b>	<b>23</b>
2.1 MATERIALS .....	23
2.1.1 Starting material .....	23
2.1.2 Kits .....	24
2.1.3 Electric tools.....	25
2.1.4 Solutions.....	26

2.1.5 Primers.....	27
<b>2.2 METHODS .....</b>	<b>28</b>
2.2.1 Total RNA extraction .....	28
2.2.2 cDNA syntesis by RT-PCR.....	28
2.2.3 Primer design.....	29
2.2.3.1 Degenerated primers .....	29
2.2.3.2 Primers based on Medaka .....	29
2.2.3.3 Sequence-specific primers.....	29
2.2.4 Gene amplification by PCR.....	30
2.2.5 Rapid Amplification of cDNA Ends (RACE) -PCR .....	31
2.2.6 Real time-PCR.....	32
2.2.7 Agarose gel electrophoresis.....	34
2.2.8 Gel extraction .....	35
2.2.9 Sequencing .....	35
2.2.10 Bioinformatics research.....	36
<b>3. RESULTS.....</b>	<b>39</b>
3.1 PREAMBLE .....	39
3.2 CHORIOLYSIN NUCLEOTIDE SEQUENCE .....	41
3.2.1 RNA extraction.....	41
3.2.2 Amplification with Ch(F/R)1 and Ch(F/R)2 .....	41
3.2.3 Primers Ch(F/R)3 .....	43
3.2.4 Amplification with Ch(F/R)3 .....	45
3.2.4.1 531bp sequence .....	46
3.2.4.2 620bp unknown band .....	47
3.2.5 Primers Ch(F/R)4 and ChR5 .....	47
3.2.6 RACE-PCR.....	48
3.3 CHORIOLYSIN GENE EXPRESSION .....	50
3.3.1 Amplification of Choriolysin by PCR.....	50
3.3.2 Amplification of Choriolysin by Real time-PCR .....	51
<b>4. DISCUSSION.....</b>	<b>55</b>
4.1 CHORIOLYSIN NUCLEOTIDE SEQUENCE .....	55
4.1.1 Amplification with Ch(F/R)1 and Ch(F/R)2 .....	55
4.1.2 Primers Ch(F/R)3 .....	56
4.1.3 Amplification with Ch(F/R)3 .....	57
4.1.4 RACE-PCR.....	59
4.2 CHORIOLYSIN GENE EXPRESSION .....	60
4.2.1 Amplification by PCR .....	61
4.2.2 Amplification by Real time PCR.....	62
4.3 FURTHER WORKS .....	64
<b>5. CONCLUSIONS.....</b>	<b>65</b>
<b>REFERENCES .....</b>	<b>67</b>
<b>APPENDIX .....</b>	<b>71</b>

# ***1. Introduction***

## **1.1 Hatching phenomenon**

The term *hatching* refers to the embryonic exit from an egg and is a process by which animals change their life from an intracapsular to a free-living type. This transition in development occurs both in oviparous and in viviparous animals such mammals [42]. Thus, hatching is a universal phenomena during the development of animals of all vertebrate phyla and most invertebrate phyla (except unicellular organisms).

Hatching can be classified in two types, mechanical hatching and enzymatic hatching [34]. In the former, the egg envelope (or eggshell, which protects the embryo in the early stage of development) is broken as a result of increasing pressure inside the egg either from the embryo or from extraembryonic origin, or due to pecking, mastication or emaciating by the embryo [4]. In the latter case, the escape of the embryo from the egg is preceded by softening the egg envelope by an enzyme (or enzymes) secreted by the embryo [4] [34] [35]. The teleostean hatching mechanism considered in this work is the enzymatic process.

### **1.1.1 The teleostean hatching**

In 1900 Kerr studied the lungfish *Lepidosiren paradoxa* and showed that the horny eggshell became quite soft so that the embryo could break it by a violent movement. He attributed the softening of the egg envelope to a digestion by some “ferments” secreted by the embryo. The first experimental study of egg envelope-digestion came with Moriwaki (1910) and Wintrebert (1912), who studied hatching in salmon. Moriwaki found that the inner layer of chum salmon *Oncorhynchus keta* eggshell was dissolved by the contents of perivitelline fluid collected during hatching. He also found that the outer layer remained undigested, although it was very fragile and later torn open by the embryo. It was possible conclude that the egg envelope-dissolving substance seemed to be a kind of ferment [10].

Furthermore, a large amount of unicellular glands were found to become differentiated on the surface of embryonic body about 10 days before hatching, and it was considered that the dissolving enzyme must have been secreted from mature glands only at the time of hatching, as the perivitelline fluid obtained before the time of hatching was inactive in dissolving egg envelope.

In 1931 Needham coined the term “hatching enzyme”, based on long studies on several teleosts like rainbow trout, goldfish and perch [17].

### 1.1.2 The egg envelope: Chorion

In general, the teleostean egg envelope consists in two distinct layers. The thick inner part called *zona radiata* [8] [38] causes the mechanical rigidity of the egg [23]. In this layer one may further distinguish a lamellate internal part, and an external part which is rather homogeneous. The whole *zona radiata* is externally covered by a thin layer termed *zona pellucida*, which probably protects the embryo from bacterial infection and sets the osmotic barrier of the perivitelline space. An additional jelly covering bearing attaching filaments is often observed in demersal eggs.

*Zona pellucida* is probably produced by the surrounding follicle cells [33]; recent studies have shown that protein components of *zona radiata* appear to be produced in the liver and transported in the blood to the ovarian follicle in a manner resembling vitellogenesis [29]. *Zona radiata* seems to be composed of three main protein monomers which are covalently aggregated during the hardening process [19].

Following activation or fertilization, the weak and fragile egg envelope of the unfertilised fish egg is transformed into a tough structure through a process called “hardening”. The egg envelope (chorion) of the fertilized egg consists of a thin outer layer and a thick inner layer. The salmon eggs chorion is composed of a scleroprotein, classified as pseudokeratin and was later named ichthulokeratin.

Mechanisms for the hardening process is not fully understood: it's known that the process is  $\text{Ca}^{2+}$ -dependent and generates a 10-fold increase in the mechanical strength of the eggshell [5] [16]. It was recently postulated that the hardening process is due to the action of a transglutaminase enzyme which crosslink the protein chain [9] [19].

The crosslinking of the three protein chain and the similarity in amino acid composition between *zona radiata* proteins and fibrin, possibly suggest that the hardening process in teleostean eggs may be analogous to the blood coagulation process [20].

The result of the hardening process is an eggshell with rigid mechanical properties, which necessitates enzymatic digestion prior to the larval breaking-out.

### 1.1.3 The hatching-gland cells

The hatching enzyme-producing cells are present early in the *intra ovo* development as specialized cells containing a small number of eosinophilic granules. At this stage the hatching glands (HG) cells may be distinguished morphologically from other cells by their larger size, abundance of cisternae in the endoplasmic reticulum and a larger electron-dense nucleus with a large nucleolus [36]. In zebrafish and rainbow trout the HG cells develop at the stage of eye pigmentation [9] [31] and in medaka at the stage of 10-12 somites [20], which is before the eye pigmentation stage. The stage of HG development is probably species-dependent.

#### 1.1.3.1 Origin and location.

Differentiation of hatching-gland cells has been studied first in medaka, *Oryzias latipes*, where the HG cells are located in the pharyngeal wall (Ishida in 1944 reported an endodermal origin [12]). At the same time the HG cells are supposed to be of ectodermal origin in fishes which have the HG cells distributed on the outer surface of the embryonic body/yolk sac [41]. The hatching gland cells of zebrafish are in the epidermis of the yolk sac and those of Masu salmon are both in the epithelium of the pharyngeal cavity and the lateral epidermis of the head. However, it was found that the hatching gland cells of zebrafish and Masu salmon originated from the anterior end of the hypoblast, the Polster, as did those of medaka. It was clarified, therefore, that such difference in the final location of hatching gland cells among these species resulted from the difference in the migratory route of the hatching gland cells after the Polster region. In salmonoid fishes the gland cells are distributed in the anterior surface of the embryonic body and the yolk sac, and on the inner surface of the larynx and gill.

Many unicellular hatching glands appeared on the surface of embryos as they reached the hatching stage. A few secretory granules found in the trans face of the Golgi apparatus were less electro-dense than most other granules, probably representing the immature state.

### **1.1.3.2 Morphology.**

The development and maturation of the HG cells have been studied by electron microscopy in some teleostean species, such as medaka [36], salmonids [42] and some tropical fishes [31]. The ultrastructure of the HG cells in these species is similar, even if they are located in different areas.

The developing immature HG cells are characterized by electron-dense nuclear material and amount of rough endoplasmic reticulum. As the HG cells differentiate, secretory granules appear near the Golgi apparatus and increase in numbers as the production of hatching enzyme continues. The hatching enzyme is stored in the granules in a crystalline state. Before hatching the mature HG cells are fused with zymogen granules. The nucleus has no conspicuous nucleolus, the cisternae of endoplasmic reticulum are fragmentary and the Golgi complex is not prominent [37].

There are three types of secretory granules in the hatching-gland cells of medaka embryos. Type 1 granules are homogeneously electron-dense and were predominant at earlier developmental stage. Type 2 granules are as electron-dense as type 1 but contain a crescent-shaped shell of higher electron-density. Type 3 granules contain somewhat granular contents with as low an electron density as the cytoplasmatic matrix. This kind of granules are predominant in the embryo at latter developmental stages. Just before secretion, a small hole appears at the apical end of the cell and type 3 granules seem to be disintegrated within the cell [35].

Bourdin (1926) regarded the hatching glands as being morphologically merocrine but functionally holocrine.

### 1.1.3.3 Secretion.

After being packaged in the secretory granules, the hatching enzyme (HE) is secreted into the perivitelline space, where it has access to the eggshell [35].

During the release of the HE, the swelling of the glandular cells is associated with separation of the covering epithelial cells. Inside the gland cell, the granules become angular in shape, their membranes are dissolved partly and their contents are mixed with cytoplasm before they are secreted from the cell.

In salmonid the HE granules fuse with each other shortly before the release of the HE [42]. Yamamoto (1979) [37] observed no coalescence of the zymogen granules during natural secretion of the HE.

In pike [24] was reported a release of the granules by exocytosis. HG secretion involves release of granules of unequal sizes which at least temporarily leaves larger pits in the cell surface and then are associated with cell death.

Before hatching, the zymogen granules become less dense [37], which may imply that the HE undergoes some sort of hydration prior to release. Secretory granules isolated from unhatched medaka embryos show that the HE in the granule is active and has the same molecular weight as secreted enzyme [13]; this suggests that the hatching enzyme does not undergo any catalytic activation at hatching, but is stored in an active form.

After hatching, the HG cells die after an apoptotic degeneration mechanism [24]. The remaining cells are condensed and phagocytosed by the surrounding epidermal cells and further digested by these cells.

### 1.1.4 Digestion of the eggshell: Choriolysis

There are two factors involved in the breakout of the eggshell: the lashing movement of the tail and the secreted hatching enzyme. These are both necessary and the tail movement is known to be effectively only after the digestion of inner layer by the enzyme. Usually the hatching enzyme is secreted shortly before an actual hatching occurs.

During hatching, the eggshell changes: the solid protein aggregate is depolymerised and becomes water soluble. Digestion of the eggshell starts in *zona radiata*, adjacent to the perivitelline space, and continues out towards the *zona pellucida*. The hatching enzyme does not digest the *zona pellucida* and the larvae break this covering by muscular forces. The structural differences between these two layers which limit the hatching enzyme to dissolving only the *zona radiata*, are unknown. In the *zona pellucida*, a higher carbohydrate content has been reported [9], which may prevent digestion by the hatching enzyme.

Electron microscopic observations on the digestion of *zona radiata* [38] indicate that the digestion products possess high molecular weights. The hatching enzyme selectively hydrolyses only very few peptide-bonds in lamellar proteins and gives rise to seven soluble glycoproteins. These glycoproteins seem to be resistant to further digestion by the hatching enzyme. The ability of HE to cleave only at restricted sites in the eggshell allows dissolution of the rigid eggshell without a need for complete digestion of the eggshell proteins.

### 1.1.5 Regulation of the hatching process

Hatching in fishes is a developmental stage-specific phenomenon. The release of the proteolytic hatching enzyme is possible only after a specific stage of development. Embryo must have attained this particular developmental stage and have fully matured hatching-gland cells before hatching occurs.

Attained a specific developmental stage is however not sufficient to cause actual hatching. The beginning of this process in Teleosts is a complex phenomena: there are many factors that stimulate or suppress the hatching, influencing secretion of hatching enzyme. The exact hatching point seems to be affected by environmental stimuli like oxygen, temperature and light [35].

#### **1.1.5.1 Oxygen.**

In medaka, an oxygen shortage stimulates the respiratory activity of the embryo and the hatching is seen markedly accelerated [35].

Hatching in salmon embryos can be accelerated if they are subjected to asphyxia by bubbling hydrogen through the hatchery water.

#### **1.1.5.2 Temperature.**

A previous cooling of the embryos followed by an elevation of the temperature, facilitates synchronization of hatching and controls the hatching time. Once the enzyme is secreted, it solubilises the egg envelope faster at higher temperature than at lower [10].

#### **1.1.5.3 Light.**

Studies in medaka and zebrafish embryos [25] reported that hatching rate was significantly higher in the light period than in the dark period. Other studies showed that the developmental rate from fertilization to hatching was not affected by any change in light/dark time, but hatching was significantly suppressed in constant dark condition. Moreover, embryos developed before hatching in a 14h light/10h dark cycle [35].

Therefore, hatching enzyme secretion seems to be controlled by stimulation of photoreceptors such as eyes, probably via central nervous system. A similar effect was also observed in Atlantic salmon: embryos kept in continuous darkness hatched random. Stimulation may take place via photoreceptor in eyes and/or pineal gland by way of a neuroendocrine or a neural mechanism.

#### **1.1.5.4 Other factors.**

Preliminary studies determined a possible role of hormones in the regulation of hatching [22]. The agents may be classified into two categories: one acting directly on the gland cells, and the other acting indirectly and probably via nervous system.

### 1.1.6 Embryonic development of Zebrafish *Danio rerio*

Zebrafish is commonly used in biology as model organism, due to its short life cycle, 12 weeks, and its transparent embryo, which allows to follow every single cell during development [32]. A series of stages for development of Zebrafish embryos (Fig. 1.1) has been described and all of them are characterized by morphological criteria, based on the number of somites and the organs present. After fertilization, the zygote begins the segmentation that give rise to a blastoderm. This blastoderm expands in vegetative direction with an epiboly movement. At 5½ hours post fertilization starts the gastrula stage: cells from ectoderm and mesoderm spin inside, around the blastoderm border (involution). 9 hours post fertilization the notocord starts to be traced, and at 10 hours gastrulation is complete. Follow neurula stage and somites shaping. Next 12 hours are important for organs and systems formation. At 18 hours there are 18 somites. Nervous system has built up very fast and optic vesicles give rise to the eyes. After 52 hours post fertilization, the embryo goes out of the chorion (hatching) and the fry start swimming and feeding.

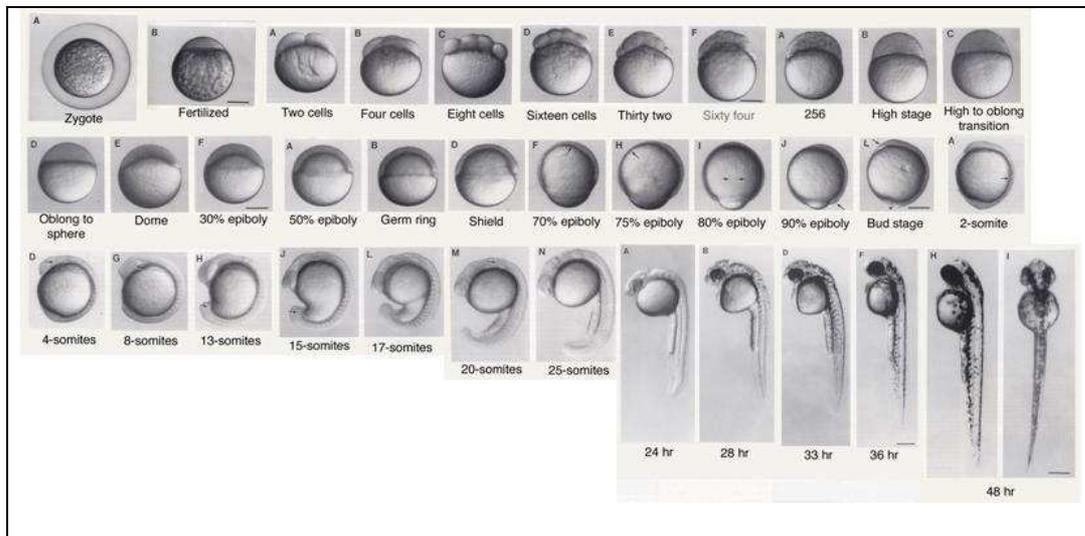


Fig. 1.1: Zebrafish *Danio rerio* embryo development

## 1.2 The Astacin Family

By the end of the 1980s four mechanistic classes of proteases and six families were identified. Each family had characteristic active site residues and was supposed to descend from a common ancestor by divergent evolution [18].

Thanks to the information explosion that occurred along with the advances in molecular and cell biology, more than 60 families of peptidases have now been proposed on the basis of similarities in primary structure [21].

The Astacin family of metalloendopeptidases was recognized as a novel family of proteases in the 1990s [6]. The family was named “Astacin family” because of its origin: the first to be sequenced and biochemically characterized was astacin, a protease from the crayfish *Astacus astacus* L. [27]. The crayfish enzyme astacin is one of the smallest members of the family and its gene, the first described, spans a region of 2616bp, with 5 exons and 4 introns.

More than 20 Astacin family members have now been identified. Some proposed functions of these proteases include activation of growth factors, degradation of polypeptides and processing of extracellular proteins.

Astacin family members are characterized by a unique 18-amino acid signature sequence, HExxHxxGFxHExxRxDR, which begins with the pentacoordinated HExxH zinc-binding motif found in most metalloendopeptidases. The signature sequence is part of an approximately 200-amino acid sequence, which is the entire mature crayfish astacin (molecular mass 22,614Da) and the catalytic or protease domain of all the members of the family [3].

### 1.2.1 Proposed functions

Astacin family enzymes have roles in mature and developmental systems. They are expressed in a tissue specific manner in mature organisms and are temporally and spatially expressed in developmental systems.

Astacin is synthesized in the crayfish hepatopancreas, an organ that has intestinal, hepatic, and pancreatic functions [28]. It is stored extracellularly as an active proteinase in the stomach-like cardia and is thought to be a digestive enzyme also in many other crustaceans.

Several of the Astacin family members are implicated in the hatching process of embryos and in skeleton formation. The fish enzymes HCE 1 and 2 and LCE work in concert to degrade the egg shell (chorion) of embryos. In *Oryzias latipes*, these enzymes are stored in zymogen granules and secreted/activated at the appropriate times to serve this function [39]. HCEs act directly on the “hard” chorion to digest protein into Pro-rich pieces; LCE further degrades the remaining polypeptides [15]. CAM-1 of quail embryos is also implicated in egg shell proteins degradation [7]. The hydra HMP1 enzyme seems to be involved in pattern formation and morphogenesis: immunohistochemical studies showed a role in head regeneration, tentacle battery cells formation and degradation of extracellular matrix proteins of the developing tentacles. Meprin subunits are the only Astacin family members known to form homo- or hetero-oligomers and to contain putative COOH-terminal transmembrane domains. Other examples are shown in Table 1.1.

Member	Species/Tissue	Natural substrates	Functions
Astacin	Crayfish stomach	Ingested proteins	Digestion of food
Chiriolysin H (HCE 1 and 2)	Fish hatching gland cells	Hard chorion, egg envelope	Embryo hatching
Choriolysin L (LCE)	Fish hatching gland cells	HCE-digested chorion peptides	Embryo hatching
CAM-1	Quail embryos	Extraembryonic choriollantoic membrane	Release Ca <sup>2+</sup> from eggshell protein; hatching
HMP1	Hydra tentacles	Extracellular matrix protein; fibronectin, type 1 gelatin	Pattern formation; formation of tentacle cells
SpAN and BP10	Seq urchin; early stages of embryogenesis	Latent <i>univir</i>	Early developmental decisions
BMP-1	Animal embryos; mouse and human tissues	Latent TGF- $\beta$ growth factors	Pattern formation; biomineralization; bone/cartilage formation
<i>Tolloid</i>	<i>Drosophila</i> embryos; mouse and human embryos	Latent decapentaplegic (DPP)	Dorsal/ventral patterning; early embryo development
<i>Tolloin-related-1</i>	<i>Drosophila</i> embryos	Latent growth factors	Larval and pupal development
Meprin A ( $\alpha$ - $\alpha$ and $\alpha$ - $\beta$ )	Mouse kidney; rat and human kidney and intestine; embryos	Urinary peptides; PTH; MSH; bradykinin; extracellular matrix proteins	Urinary peptideprocessing; extracellular protein processing and digestion
Meprin B ( $\beta$ - $\beta$ )	Mouse kidney and intestine; embryos	Extracellular proteins	Extracellular protein interactions; remodeling

Tab. 1.1. Tissue distribution and proposed functions of Astacin family members.

### 1.2.2 Primary structure

The signature sequence for the family, **HE<sub>xx</sub>H<sub>xx</sub>GF<sub>x</sub>HE<sub>xx</sub>R<sub>x</sub>DR**, spans from astacin His 92 to Asp 109 (Fig. 1.4, positions 101-118); this sequence contains the three imidazole-zinc ligands. There are four conserved cysteine residues that are known to form intradomain sulfhydryl bridges in astacin: Cys42/Cys198 and Cys64/Cys84. Because of the conservation of these cysteine residues, equivalent disulfide bridges were supposed to exist also in the proteases domain of the other family members [3].

Another conserved region, **SxMHY** (astacin Ser145 to Tyr149; Fig. 1.4, positions 156-160), contains the tyrosine residue involved in zinc binding and the methionine involved in a “Met-turn” of the peptide chain [2] [14] [26].

Other conserved specific residues were identified and are probably crucial to the overall structure of the protease domain. Comparison of the 22 family members indicates that there are 28 residues that are totally conserved. 17 of those are accounted for within the 18-residue family signature sequence, the sequence containing the zinc-binding Tyr, and by the conserved cysteines [3]. The other 11 conserved residues are primarily involved in internal bonds in astacin.

The amino acid sequences of the Astacin family member proteases domains are 29 to 99% identical. The fish enzymes HCE1 and HCE2 are products of two different genes that are highly conserved, 95% identities.

### 1.2.3 Tertiary structure and domains

The X-ray crystal structure of astacin shows a compact bilobal structure with a long, deep active-site cleft [1] that divides it into two parts, NH<sub>2</sub>- and COOH-terminals.

Astacin family proteases are synthesized with NH<sub>2</sub>-terminal signal and proenzyme sequences, and astacin itself is now known to contain a prepro segment of 49 amino acidic residues. This signal peptide leads the protein into the endoplasmatic reticulum during biosynthesis. The prosequences vary greatly in size and are considered to be important for regulating activity and perhaps expression of the proteases. The NH<sub>2</sub>-terminal 100 amino acids are organized into five-stranded pleated  $\beta$ -sheets and two long  $\alpha$ -helices. One helix forms the top part of the active

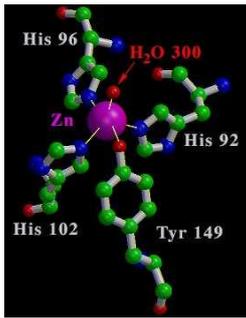


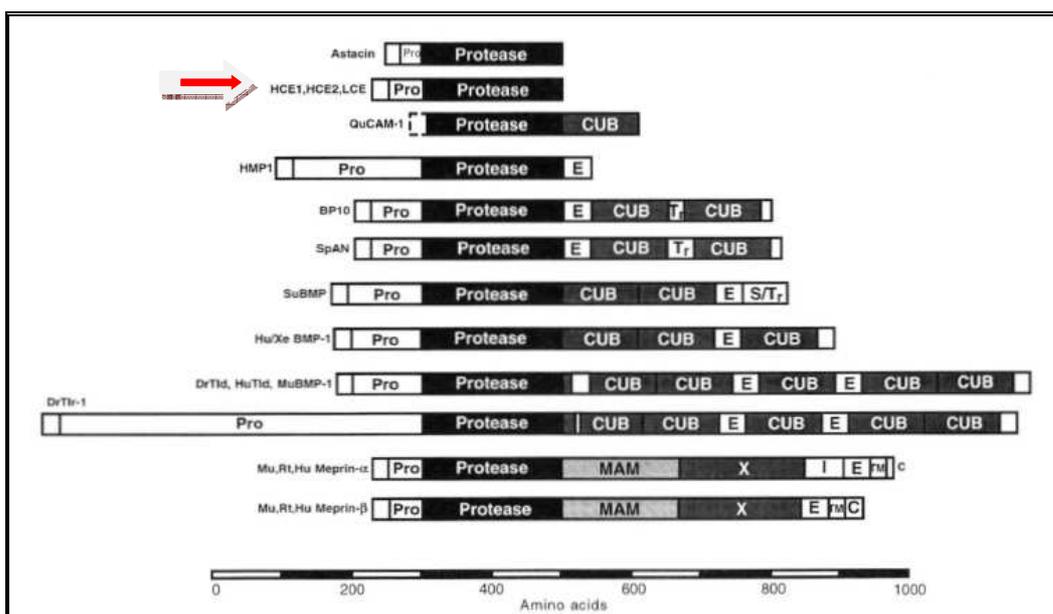
Fig. 1.2: Zn-binding motif.

site of astacin and includes the zinc ligands His92, His96, and His102. Zinc atom is complexed by three histidine residues (His 92,96, and 102 of astacin; positions 101, 105, and 111 of Fig. 1.4), a glutamate residue (Glu 93; position 102) through a water molecule, and a tyrosine residue (Tyr 149; position 160) from the “Met turn”. Glu 103 (position 112) is oriented by a salt bridge to the NH<sub>2</sub>-terminal residue (Fig. 1.2).

The COOH-terminal half of astacin has little defined secondary structure except for a three-turn  $\alpha$ -helix before a disulfide bond, which ends the domain and connects it to the NH<sub>2</sub>-terminal half.

The smallest members of the family (crayfish astacin and the teleostean choriolysins L and H, LCE and HCE) have no domains COOH-terminal to the protease domain. Many others indeed (such as meprins, BMP-1, tolloid) contain multiple domains COOH-terminal to the protease domain, organized in several turns and irregular structures. They can be either secreted from cells or plasma membrane-associated enzymes.

Most of the known Astacin family members contain one or more non-catalytic “interaction” domains, like E (epidermal-growth-factor-like), and CUB (Clr/s complement-like). These domains could promote protein-protein and substrate interactions.

Fig. 1.3. Domains structure of Astacin family members. Red arrow: medaka *Oryzias latipes* Choriolysin.

Several other Astacin family members have been identified but not included in Fig. 1.3. Fig. 1.4 shows the primary sequence of the protease domain of many of these members.

The domain structure of the Astacin family represents a pattern seen in several protease families, such as the coagulation proteases, complement proteases, proprotein convertases and matrixins.

Several members of these families result from gene duplication, evolution, gene fusion and exon shuffling. All of these families present a conserved protease domain is associated with a variety of non-catalytic domains that yield different proteins, with somewhat different enzymatic activities, regulation, expression and interactions with small ligands and other proteins.

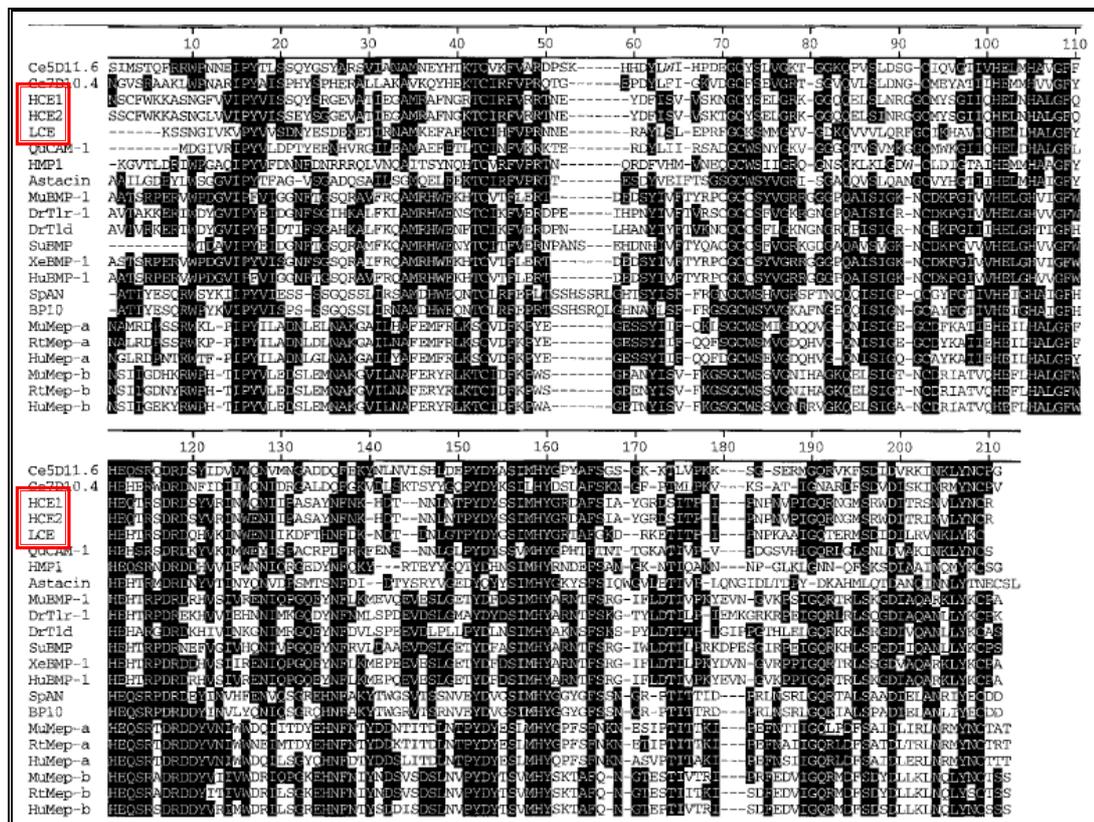


Fig. 1.4. Sequence comparison of the protease domain of 22 members of the Astacin family. Red panel: medaka *Oryzias latipes* Choriolysin.

### 1.2.4 Effects of non-catalytic domains: Activation

Astacin family enzymes are synthesized as inactive proenzymes. Removal of the prosequences constitutes a major mechanism for activation. The crystal structure of astacin indicates that there is a salt bridge between the NH<sub>2</sub>-terminal amino acid of the mature protease domain, Ala1 and Glu103 (position 112 of Fig. 1.4), which is next to the third zinc ligand (astacin residue His102, position 111) and near the active site. This bridge may inhibit the enzymatic activity because the prosequence would prevent the formation of the salt bridge.

Removal of the prosequences is considered a form of post-translational regulation in Astacin family proteinase activity.

## 1.3 The Hatching Proteases

Dissolution of tough egg envelope by secreted hatching enzymes is the major feature of hatching in fishes. Hatching enzyme is a general term for an enzyme secreted by the embryo in order to break down the egg envelope and release the embryo from its intracapsular life. Recently this enzyme was also found in mammalian embryos.

Hatching enzyme has been purified from several species like: *Gobius joso* (Denucé 76), medaka *Oryzias latipes* (Yasumasu et al 88), rainbow trout *Onkoryncus mykiss* (Hagenmaier 1974) and *Fundulus heteroclitus* (DiMichele et al 81).

### 1.3.1 Medaka *Oryzias latipes* Choriolysin

Medaka *Oryzias latipes* is the closest to salmon in terms of hatching enzymes and for this reason taken as example to explain the system. It consists of two zinc metalloproteases called Choriolysin H and L: HCE, high choriolytic enzyme, and LCE, low choriolytic enzyme [15]. Observations suggested that they are essentially the same enzyme, but they behave differently because of their different

states of association with some heterologous substances such as hydrolysed chorion. In medaka they have 24kDa and 25,5kDa respectively.

Characterization of their cDNAs showed that both the enzymes consisted of 200 amino acids containing the common His-Glu-x-x-His motif for Zn protease and belonged to the Astacin family. Similarity between HCE and LCE is 55%.

The consensus sequence motif which let them belong to the astacin proteases family is **HELLHALGFYHEHTRSDR**.

### 1.3.1.1 Proteolytic action

The HCE is the zinc-proteases which have a binding site for the eggshell in addition to its catalytic site. The enzyme binds tightly to the inner layer of the hardened chorion and it swells the substrate by its partial proteolytic action, remaining bound to the swollen eggshell. The LCE is efficient in digestion of previously swollen eggshell but can't digest intact chorion. Proteolytic action of HCE to the chorion is prerequisite for the initiation of the chorion-digesting process by the hatching enzyme system. The two enzymes differ both in biochemical characteristics and catalytic properties.

### 1.3.1.2 The genomic sequence

In spite of the similarity of the amino acid primary structure of medaka hatching enzymes, HCE and LCE genomic structures are quite different from each other. LCE is composed of 8 exons and 7 introns, while HCE is an intron-less gene.

Medaka has eight copies of HCE genes (all of them intron-less): five for HCE1 and three for HCE2. Six (three of each) of the genes form a cluster within approximately 25kb of the genomic DNA; the two others are located separately. The 5'-flanking regions of all the HCE genes are 80-95% similar for 200-400bp [3].

In contrast, there is only one copy of the LCE gene: it is within 3,6kb and contains eight exons and seven introns. For LCE, the TATA box consensus sequence is 28bp upstream from the transcription start site.

One explanation for the evolutionary occurrence of the "intron-less" medaka HCE gene is the participation of a retroposable element such as long interspersed

elements (LINE). However, this intron-less pseudogenes are rarely found in non-mammalian vertebrates and most of them have no promoter, unlike the functional medaka HCE gene.

Another explanation is that the HCE gene lost its introns step-by-step during evolution, but it is difficult to understand why “stepwise intron-loss” occurred in HCE and not in LCE [11].

There are other studies on the genomic sequence of these proteases. Exon-intron structure of *Fundulus heteroclitus* HCE and LCE genes was found to conserve those of medaka. Analysis of their proteolytic action on the egg envelope confirmed the similarity in the hatching system. cDNA and genomic DNA isolated from Japanese eel show that intron-exon structure of both the genes is similar to medaka LCE (it has 9 exons and 8 introns) but not to HCE.

### 1.3.1.3 HCE, High Choriolytic Enzyme

HCE is a unique enzyme with regard to its proteolytic action on the Chorion (amino acid sequence in Tab. 1.2). It tightly binds to the chorion prior to partial digestion and swelling of it.

It was easy to think about the possibility for HCE to recognize and bind specific site(s) on the hardened chorion and cleave and release specific polypeptides or peptides. Investigations in this way showed that the various kind of released polypeptides had unique characteristics, such as the high contents of proline, but also glutamic acid, glutamine and  $\gamma$ -glutamyl  $\epsilon$ -lysine isopeptide.

HCE itself can't cleave the  $\gamma$ -glutamyl  $\epsilon$ -lysine isopeptide bond responsible for the chorion hardening. It seems to digest the surroundings of them and bring about the chorion swelling probably because of the exposure of hydrophilic regions. In addition, the chorion-swelling action of HCE results in making the substrate susceptible to Low Choriolytic Enzyme attack [15]. LCE, in fact, can easily digest the soft chorion: experiments came up successfully with unfertilized eggs, which contain a lower quantity of  $\gamma$ -glutamyl and  $\epsilon$ -lysine isopeptide.

```
MNLAPSTCLL LLLLLGIAQA LPIQNEEGHE EGEDDFVDI TTRILTSNNN TDQLLLEGDL
VAPTNRNAMK CWSSSCFWKK ASNGLVVIPY VISSEYSGGE VATIEGAMRA FNGKTCIRFV
RRTNEYDFIS VVSKTGCYSE LGRKGGQLQEL SINRGGCMYS GIIQHELNHA LGFQHEQTRS
DRDSYVRINW ENIIPASAYN FNKQDTNNLN TPYDYSSIMH YGKDAFSIAY GRDSITPIP
PNVPIGQRNG MSRWDITRIN VLYNCR
```

Tab. 1.2: HCE, High Choriolytic Enzyme protein sequence. Consensus motif red highlighted; Met-turn green highlighted.

### 1.3.1.4 LCE, Low Choriolytic Enzyme

LCE is a basic protein with a molecular weight of about 25,5kDa (amino acid sequence in Tab. 1.3) which differs from HCE because its incapability of digesting the intact inner layer of chorion, but only the one that has been swollen previously by HCE. Many studies demonstrated this essential role of choriolysis in cooperation with HCE [40].

This enzyme is characterized by high content of aspartic acid, glutamic acid and the most abundant lysine.

MDLLAKASVL	LLLLLSLSNA	QTDNMEEAEN	GSSKEEIVES	ELEDVSSIIF	RMNNSMEEL
LEGDLVLPKT	RNAMKCFGAP	DSCRWPKSSN	GIVKVPYVVS	DNYESDEKET	IRNAMKEFAE
KTCIHFVPRN	NERAYLSLEP	RFGCKSMMGY	VVDKQVVVLQ	RFGCIKHAVI	QHELLHALGF
YHEHTRSDRD	QHVKINWENI	IKDFTHNFDK	NDTDNLGTPY	DYGSIMHYGR	TAFGKDRKET
ITPIPNPKAA	IGQTERMSDI	DILRVNKLYK	C		

Tab. 1.3: LCE, Low Choriolytic Enzyme protein sequence. Consensus motif red highlighted; Met-turn green highlighted.

### 1.3.1.5 Differences in chemical characteristics and enzymological properties

The molecular weights of LCE and HCE are 25,5 and 24kDa respectively. LCE is more thermostable than that of HCE.

HCE partially digests the chorion with swelling and LCE effectively digest the swollen chorion. Thus, HCE changes the tough structure of the intact chorion into the looser structure, producing a specific substrate for LCE.

The swelling of the inner layer of chorion represents an intermediate stage in the process of breakdown.

The chorion swelling was attributed to a partial proteolysis by HCE and the chorion-digesting activity was caused by the cooperative action of two kind of proteases, HCE and LCE.

### 1.3.2 Zebrafish *Danio rerio* Choriolysin

In zebrafish two types of hatching enzymes have been identified. They are called *Hatching enzyme 1* (Accession number NP\_998800.1) and *Hatching enzyme 2* (Accession number NP\_998711.1) and have the same function of HCE and LCE in medaka.

```
MDIRASLSIL LLLFGLSQAS PLREFEAVFV SEPETVDITA QILETNKGSS EVLFEGDVVL
PKNRNAFICE DKSCFWKNA NNIVEVPYV SGEFSINDKS VIANAISIFH AQT CIRFVPR
SIQADYLSIE NKDGCYSAIG RTGGKQVVS NRKGCVYSGI AQHELNHALG FYHEQSRSDR
DQYVRINWNN ISPGMAYNFL KQKTNNQNTP YDYGSLMHYG KTAFAIQPGL ETITPIP DEN
VQIGQRQGLS KIDILRINKL YGC
```

Tab. 1.4: HE1, Hatching Enzyme 1 protein sequence. Consensus motif red highlighted; Met-turn green highlighted.

```
MDPKISLSIQ LLLVGISLAA PVGEYDNSNG IETPQNVDIT TLEETNKGSS RRLIEGDMLY
PQTRNALVCG NNNCFWKNS SNFVEVPYIV SSEYSATEIS VIQKAMSGIH NKTCIRFVPR
ISQTDYISIE NQDGCFAFIG KKGKQLVSL RKKGCVYHSI VQHELNHALG FYHEHVRSDR
DSYITIHWEY IATNEIRNFM KKNTNSQNTT YDYGSIMHYG KTAFTTVK GK ETMTYPDET
VPIGKAKEMS DIDILRINMM YSCNISDDLK I
```

Tab. 1.5: HE2, Hatching Enzyme 2 protein sequence. Consensus motif red highlighted; Met-turn green highlighted.

The consensus sequence motif which let them belong to the Astacin proteases family is **HELNHALGFYHEQSRSDR** (amino acid sequence of HE1 and HE2 respectively in Tab. 1.4 and 1.5).

### 1.3.3 Phylogenetic evolution

According to the phylogenetic tree of Teleost fishes, Japanese eel *Anguilla japonica* is one of the early diverged fishes, followed by zebrafish and the acanthopterygian fishes (medaka and *Fundulus* [11]).

The strong similarity of HCE and LCE proteins indicates a common ancestral gene and it is interesting that their gene structures have evolved to differ so markedly.

The ancestral gene of medaka HCE and LCE (paralogous to each other) was supposed to be Japanese eel hatching enzyme gene-like exon-intron structure, and that the HCE genes had lost all their introns during the evolutionary process of Teleostea, while LCE genes conserved the exon-intron structure of the ancestral gene.

Analysis of gene synteny and cluster structure showed that the syntenic genes around the HCE and LCE genes were highly conserved between medaka and *Tetraodon*, but such synteny was not found around zebrafish hatching enzyme genes. It was hypothesized that the zebrafish hatching enzyme genes were translocated from chromosome to chromosome, and lost some of their introns during evolution.

## 1.4 Atlantic salmon *Salmo salar* Hatching Enzymes

The work for this thesis has been based on Atlantic salmon *Salmo salar* as first organism, which as well as other fishes, it has an hatching system composed of proteases. Previously hatching enzymes were known as Metallo-proteases called Choriolysins but there are also Serine-proteases called Zonases (named by their work on the Chorion or Zona, *zonalysis*). These have been sequenced and patented [30].

### 1.4.1 Choriolysin

This enzyme belongs to the Zn-proteases like in the other species, even if in salmon its gene sequence has not been characterized yet. Only the protein sequence is known and the start point of my work was based on that one.

```
MDHRPTLSLL LLLLLLGLSQ ASGNEFHDEP DHVSITSVIL KSNNGTNELL LDGDILAPRT
RNAMKCFSSQ YSCLWKKSSD GLVYVPYILS AVYSSLEVET IETAMKYFQG KTCIRFIPRK
TQTAYLDIQS SGGCFGTVGT VGDRQTLSLA QFGCVQHGII QHELLHALGF YHEHNRSDR
QYIRINWQYI YDYAVGNFQK EDTNNLHTAY DYSVMHYDR TAYTNDYGKE TITPIPDPSV
AIGQRLGMSD IDVLKVNKLY QC
```

Tab. 1.6: Salmon Choriolysin protein sequence. Consensus motif red highlighted; Met-turn green highlighted.

The sequence presents all the characteristics of the ones in the other species. The consensus sequence comprehensive of Zinc-binding motif, **HELLHALGFYHEHNRSDR**, is highly conserved and also the Met-turn, **SVMHY**.

### 1.4.2 Zonase

Hatching enzyme of Atlantic salmon was at first purified and characterized as a Serine protease, Zonase. This enzyme was found to be a dimer and present in three iso-forms (Zonase I, II and III), all of them with a molecular weight about 26kDa. The complete amino acid sequence was determined from cDNA and the genomic sequence(s) characterization is in progress.

The protease possesses two sites that interact with the substrate (one binding site, by hydrophobic interactions, and one active site). This enzyme cleaves peptide bonds C-terminal to lysine and arginine but only when enzyme binding takes place on separate hydrophobic region of the attacked protein.

Zonase is considered quite special: it is stable and highly specific for hydrophobic substrates, with a remarkable catalytic resilience under varied reaction conditions (wide range of temperature, pH, salt and detergents). Peculiarity of this enzyme is the non-selfdegradation, it does not digest itself, which is a property unknown to many other proteases. As a consequence, the enzyme can be kept at room temperature and in solution without losing its activity for months; it can also retain its activity in extremely hydrophobic, acidic or alkaline environs.

Structural proteins of the eggshell are shown to share many similarities to human fibrinogen such as number of monomers, molecular weight, amino acid contents and cross links between each other. Hardening of fish eggshell proteins after fertilization is essentially the same process as blood coagulation and the enzymatic digestion by Zonase provides skin regeneration. Due to its unusual behaviour and its effects on human skin this protease is highly studied.

## 1.5 Aim of the study

My studies for this thesis work focus on Choriolysin gene. The two main purposes were:

1. Determine the nucleotide sequence of Atlantic salmon *Salmo salar* Choriolysin
2. Study the gene expression in embryos at several developmental stages, through its mRNA

Choriolysin is supposed to be a simple gene, present in only one copy for one protein, and with mendelian inheritance. Studies that support this belief can be used to compare its features with the ones of Zonase gene and protease.



## 2. Materials and Methods

### 2.1 Materials

#### 2.1.1 Starting material

Atlantic salmon (*Salmo salar*) embryos were taken at seven stages.

In this study also Zebrafish (*Danio rerio*) embryos were used: 100 embryos per stage were collected at different hours, from hour 0 to hour 120, in order to cover the most important stages of development.

<i>Salmo salar</i>	<i>Danio rerio</i> embryos
<ul style="list-style-type: none"><li>• Sperm</li><li>• Non-fertilized embryos</li><li>• 1 day/degree (dd) after fertilization embryos</li><li>• 200dd embryos</li><li>• 370dd embryos</li><li>• 550dd embryos (close to hatch)</li><li>• Hatched embryos</li></ul>	<ul style="list-style-type: none"><li>• 0 hours</li><li>• 4 hours</li><li>• 6 hours</li><li>• 9 hours</li><li>• 12 hours</li><li>• 18 hours</li><li>• 24 hours</li><li>• 36 hours</li><li>• 48 hours</li><li>• 72 hours</li><li>• 120 hours</li></ul>

Both Salmon and Zebrafish living embryos were stored at their favorite temperature (4°C in 5mM NaCl ddH<sub>2</sub>O, bubbling hydrogen, for Salmon and 28°C for Zebrafish) in order to collect them when they reached the right stage.

## 2.1.2 Kits

### cDNA synthesis kit

#### ThermoScript RT-PCR System, Invitrogen

- 15U/μl ThermoScript RT
- 5X cDNA Synthesis Buffer (250 mM Tris acetate pH 8.4, 375 mM potassium acetate, 40 mM magnesium acetate, stabilizer)
- 0,1M DTT
- 10mM dNTP Mix
- 40U/μl RNaseOUT
- 50μM Oligo (dT)<sub>20</sub>
- 50ng/μl Random Hexamers
- DEPC-Treated Water
- 2U/μl E. coli RNase H

### Real time PCR kit

#### LightCycler 480 SYBR Green I Master, Roche Applied Science

- 2X Ready-to-use hot-start PCR reaction mix (FastStart Taq DNA Polymerase, reaction buffer, dNTP mix with dUTP instead of dTTP, SYBR Green I dye, and MgCl<sub>2</sub>)
- H<sub>2</sub>O, PCR-grade, to adjust the final reaction volume

#### LightCycler 480 Multiwell Plate 96, Roche Applied Science

- 5 x 10 plates
- Sealing foils

### Gel extraction – DNA purification kit

#### Wizard SV gel and PCR clean-up system, Promega

- Membrane Binding Solution
- Membrane Wash Solution (concentrated)
- Nuclease-Free Water
- Wizard SV Minicolumns
- Collection Tubes (2ml)

### Sequencing staff

- Big Dye version 3.1
- Sequencing buffer

## Taq DNA polymerase kit

### TaKaRa Ex Taq, TaKaRa Bio Inc.

5U/μl TaKaRa Ex Taq  
 10X Ex Taq Buffer  
 2,5mM (each) dNTP Mixture

## RACE PCR kit

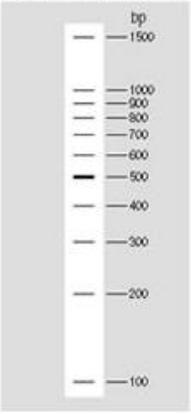
### 5'/3' RACE kit, 2<sup>nd</sup> generation, Roche Applied Science

5X cDNA synthesis buffer (250 mM Tris-HCl, 40 mM MgCl<sub>2</sub>, 150 mM KCl, 5 mM dithiothreitol, pH 8.5)  
 25 U/μl Transcriptor Reverse Transcriptase (in 200 mM potassium phosphate, 2mM dithiothreitol, 0.2% (v/v) Triton X-100, 50% glycerol (v/v), pH 7.2)  
 10mM (each) Deoxynucleotide mixture (mixture of dATP, dCTP, dGTP, dTTP)  
 2mM dATP (in Tris-HCl, pH 7.5)  
 10X Reaction buffer (100mM Tris-HCl, 15mM MgCl<sub>2</sub>, 500mM KCl, pH 8.3)  
 80U/μl Terminal Transferase, recombinant (in 60mM K-phosphate (pH 7.2 at 4°C), 150mM KCl, 1mM 2-Mercaptoethanol, 0.5% Triton X-100, 50% glycerol)  
 1ng/μl Control neo-RNA  
 37.5μM Oligo dT-anchor primer  
 12.5μM PCR anchor primer  
 12.5μM Control primer neo1/rev  
 12.5μM Control primer neo2/rev  
 12.5μM Control primer neo3/for

## 2.1.3 Electric tools

Name	Commercial name	Company
Homogenizer	Polytron PT 1200 E	Kinematica
Thermocycler (PCR)	GeneAmp PCR System 2400	Perkin Elmer
Thermocycler (Real-time PCR)	LightCycler 480	Roche Applied Science
Power supply	Microcomputer electrophoresis power supply	Consort E455
Centrifuge (Eppendorf tubes)	BioFuge Fresco	Heraeus
Centrifuge (Real-time PCR plate)	Allegra X-12R	Beckman Coulter
Spectrophotometer	NanoDrop ND 1000	Saveen Warner

## 2.1.4 Solutions

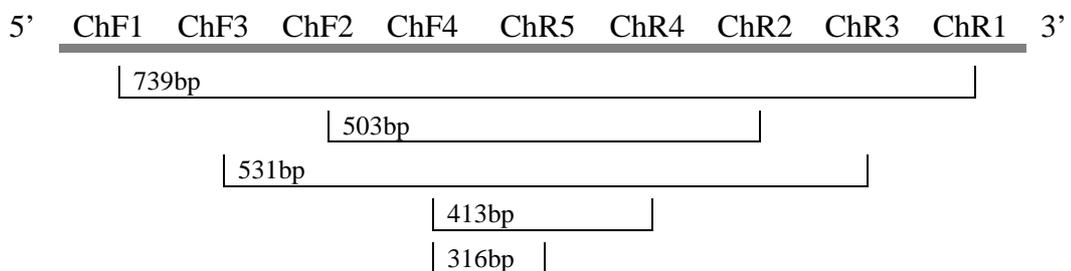
Name	Components	Company
Trizol		Invitrogen
Chloroform	CHCl <sub>2</sub>	Merck
Isopropanol		
DEPC-H <sub>2</sub> O	1% DECP in double distilled H <sub>2</sub> O	
0,5X TEA buffer	5,4gr Tris base 2,75gr Boric acid 0,465gr Na <sub>4</sub> EDTA	
Agarose (standard low-m <sub>r</sub> )		BioRad Laboratories
100bp DNA ladder		TaKaRa Ex Taq, TaKaRa Bio Inc.
6X loading buffer	9,09% bromphenol blue 0,09% xylene cyanol FF 60% glycerol	TaKaRa Ex Taq, TaKaRa Bio Inc.

## 2.1.5 Primers

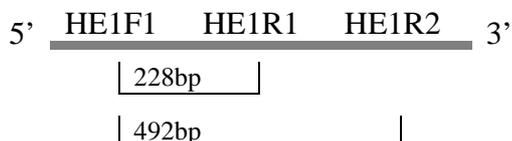
Salmon	ChF1	5' – TGG ATC ATC GTC CTA CTC TTT C – 3' length 22bp (nt2 to 23), CG content 45%, $T_m$ 54°C	Hand made in this thesis
	ChF2	5' – CGT AAT GCT ATG AAA TGT TTT TCT TC – 3' length 26bp (nt181 to 206), CG content 31%, $T_m$ 52°C	
	ChF3	5' – CCA GGA ATG CCA TGA AGT GCT TTG – 3' length 24bp (nt178 to 201), CG content 50%, $T_m$ 52°C	
	ChF4	5' – AAG TCA TCT GAC GGC TTG GTG TAC G – 3' length 25bp (nt229 to 253), CG content 52%, $T_m$ 61°C	
	ChR1	5' – CCA AGA CGT TGA CCA ATA GC – 3' length 20bp (nt721 to 740), CG content 50%, $T_m$ 54°C	
	ChR2	5' – CCA TAA TCA TTA GTA TAA GCA GTA CGA – 3' length 27bp (nt657 to 683), CG content 33%, $T_m$ 53°C	
	ChR3	5' – AGG GAT GGG GGT TAT GGT CTC CT – 3' length 23bp (nt686 to 708), CG content 57%, $T_m$ 53°C	
	ChR4	5' – GAG GAG TAG TCG TAT GCA GTG TGC A – 3' length 25bp (nt617 to 641), CG content 52%, $T_m$ 60°C	
	ChR5	5' – ACT GTT CAC GGT CAC TCC TGT TGT G – 3' length 25bp (nt520 to 544), CG content 52%, $T_m$ 61°C	
	Zebrafish	HE1F1	
HE1R1		5' – GCC ACT GTA CAC ACA GCC TTT CC – 3' length 23bp (nt455 to 477), CG content 57%, $T_m$ 60°C	
HE1R2		5' – CTC CCT CTG TCC GAT TTG CAC G – 3' length 22bp (nt720 to 741), CG content 59%, $T_m$ 60°C	

Ch = salmon Choriolysin; HE1 = zebrafish Hatching Enzyme 1;  
F = Forward; R = Reverse.

Product sizes (salmon):



Product sizes (zebrafish):



## 2.2 Methods

### 2.2.1 Total RNA extraction

Total RNA was isolated from 3 embryos (and about 4gm of sperm) of Salmon and about 100 embryos of Zebrafish at each available stage.

The extraction took place by homogenization with 1 ml Trizol (TRIzol reagent, Invitrogen), followed by Chloroform extraction (0,2ml) and Isopropanol precipitation (0,5ml).

Obtained RNA was resuspended in 50  $\mu$ l DEPC-treated H<sub>2</sub>O and stored in -20°C or -80°C until use.

### 2.2.2 cDNA synthesis by RT-PCR

Reverse transcription polymerase chain reaction (RT-PCR) is used for amplifying RNA molecules. The RNA strand is reverse transcribed into its DNA complement or cDNA (complementary DNA), using dNTPs and an RNA-dependent DNA polymerase, also known as reverse transcriptase. This enzyme (ThermoScript RT-PCR System, Invitrogen) is a DNA polymerase enzyme able to use RNA as template and synthesize DNA.

The whole reaction transcribes single-stranded RNA into double-stranded DNA. The reaction was performed with 1  $\mu$ l 50 $\mu$ M Oligo d(T) as primer, 2 $\mu$ l 10mM dNTPs, 1 $\mu$ l 15U/ $\mu$ l reverse transcriptase and, according to the RNA concentrations found, 1-2 $\mu$ l RNA.

Samples were incubated at  
50°C → 40 min  
85°C → 5 min, to deactivate the reverse transcriptase

1 $\mu$ l 2 U/ $\mu$ l RNase H was added in order to remove the original RNA template and then it was incubated again

37°C → 20 min

RNase H specifically degrades the RNA in RNA:DNA hybrids and will not degrade DNA or unhybridized RNA, it is commonly used to destroy the RNA template after cDNA synthesis.

cDNA was obtained and stored at -20°C until use.

## 2.2.3 Primer design

### 2.2.3.1 Degenerated primers

The nucleotide sequence of salmon Choriolysin was not known, for this reason the first step was to use degenerated primers.

These oligos are drawn based on the amino acid sequence; those fragments are retrotranscribed manually but since the genetic code is degenerated (more than one codon transcribes for the same amino acid), some bases need to be guessed.

These primers are ChF1, ChF2, ChR1, ChR2 (Cap. 2.1.5).

### 2.2.3.2 Primers based on Medaka

The sequence of another organism was used to design other primers. I took Choriolysin sequence of medaka, *Oryzias latipes*, which amino acid sequence had the highest similarity with the salmon protease. The fragments found without mismatches were taken as template for the primers, supposing that where they have the same amino acid sequence, possibly they could have a very similar nucleotide sequence.

These primers are ChF3, ChR3 (Cap. 2.1.5).

### 2.2.3.3 Sequence-specific primers

Other primers were sequence-specific, designed based on the sequence found with those first oligos.

These primers are ChF4, ChR4, ChR5 (Cap. 2.1.5).

Prior to order the best primers, conditions of each one were performed by an oligo analyzer ([www.idtdna.com/analyzer/Applications/OligoAnalyzer](http://www.idtdna.com/analyzer/Applications/OligoAnalyzer)), which was useful to make a selection of the sequences.

### 2.2.4 Gene amplification by PCR

Polymerase Chain Reaction (PCR) is the process used to amplify specific parts of a DNA molecule, typically short DNA fragments (up to 10kb), through the thermal stability of *Taq* (*Thermus aquaticus*) DNA polymerase.

The technique allows small amounts of DNA (low to 5pmol) to be amplified exponentially *in vitro*, without using a living organism. PCR requires several basic components, such *DNA template*, which contains the region of the DNA fragment to be amplified, 2 *primers*, which determine the beginning and end of the region to be amplified, *Taq DNA polymerase*, a DNA polymerase which copies the region to be amplified, *Deoxynucleotide triphosphates*, (dNTPs) from which the DNA polymerase builds the new DNA, *Buffer solution*, which provides a suitable chemical environment for the DNA Polymerase (*Divalent cations*, magnesium or manganese ions, and *Monovalent cations*, potassium ions).

Target gene of Choriolysin was amplified from cDNA obtained from all the stages of embryo development, by PCR (TaKaRa Ex Taq, TaKaRa Bio Inc.). The oligos were combined in every possible way for the 370dd stage which was the most abundant in total RNA, and the couple that gave a better yield was used to perform PCR amplification in all the stages.

For the reaction, 1µl 10µM primers, 0,25µl 5U/µl *Taq* DNA polymerase and 4µl 2,5mM dNTPs were used. The samples were transferred in a thermocycler which heats and cools the reaction tubes within it to the precise temperature required for each step of the reaction. The PCR process consists of a series of 30 cycles (they can be increased or decrease between 25 and 35 cycles). Each of them consists of three steps:

- 1.*denaturation*: the double-stranded DNA has to be heated to 95°C (30 seconds step) in order to separate the strands. This step breaks apart the hydrogen bonds that connect the two DNA strands. Prior to the first cycle, the DNA was denatured for an extended time (5 minutes) to ensure that both the template DNA and the primers have completely separated and are now single-strand only.
- 2.*annealing*: the temperature is lowered for 25 seconds, so the primers can attach themselves to the single DNA strands. It depends on the primers and is usually 5°C below their melting temperature.
- 3.*elongation*: DNA polymerase has to copy the DNA strands. It starts at the annealed primer and works its way along the DNA strand. Taq polymerase elongates optimally at a temperature of 72°C. Elongation step needs 45

seconds and a final longer step (10 minutes) is used after the last cycle to ensure that any remaining single stranded DNA was completely copied.

The resulting products were supposed to be DNAs at about 739bp (ChF1/ChR1), 503bp (ChF2/ChR2), 531bp (ChF3/ChR3), 413bp (ChF4/ChR4) and 316bp (ChF4/ChR5).

### **2.2.5 Rapid Amplification of cDNA Ends (RACE) -PCR**

Rapid Amplification of cDNA Ends (RACE) is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and unknown sequences at either the 3' or the 5' -end of the mRNA (5'/3' RACE kit, 2<sup>nd</sup> generation, Roche Applied Science was used).

3' RACE takes advantage of the natural poly(A) tail in mRNA as a generic priming site for PCR amplification. In this procedure, mRNAs are converted into cDNA using reverse transcriptase (RT) and an oligo-d(T) adapter primer. Specific cDNA is then directly amplified by PCR using a gene-specific primer (GSP) that anneals to a region of known exon sequences and an adapter primer that targets the poly(A) tail region. This permits the capture of unknown 3'end mRNA sequences that lie between the exon and the poly(A) tail.

5' RACE, or anchored PCR, is a technique that facilitates the isolation and characterization of 5'ends from low-copy messages. First, strand cDNA synthesis is primed using a gene-specific antisense oligonucleotide GSP1. Following cDNA synthesis, the first strand product is purified from unincorporated dNTPs and GSP1. TdT (Terminal deoxynucleotidyl transferase) is used to add homopolymeric tails to the 3'ends of the cDNA. Tailed cDNA is then amplified by PCR using GSP2 and homopolymer-containing anchor primer which permit amplification from the homopolymeric tail. This allows amplification of unknown sequences between the GSP2 and the 5'-end of the mRNA.

First strand cDNA synthesis for 3' RACE was performed with 2µl 10mM dNTPs mix, 1µl 37,5µM oligo dT-anchor primer, 1µl total RNA and 1µl 25U/µl transcriptor reverse transcriptase. Samples were incubated 1 hour at 55°C and 5 minutes at 85°C. The PCR amplification of resulting cDNA was performed with 1µl cDNA product, 1µl 12,5µM PCR anchor primer, 1,25µl 10µM gene-specific primer (ChF4), 1µl 10µM dNTPs mix and 0,5µl 5U/µl Taq DNA polymerase.

Samples were placed in a thermocycler as follow:

95°C, 5min

30 cycles: 95°C, 30sec

58°C, 25sec

72°C, 45sec

72°C, 10 min

For first strand cDNA synthesis of 5' RACE, 2µl 10mM dNTPs mix, 1,25µl 10µM ChR4, 1µl total RNA and 1µl 25U/µl transcriptor reverse transcriptase were used. Samples were incubated 1 hour at 55°C and 5 minutes at 85°C. The cDNA product was purified by subsequent centrifugations (Wizard SV gel and PCR clean-up system, Promega) and applied to for the poly(A) tailing of the 3'-end of the first strand cDNA. The reaction was performed with 19µl cDNA and 2,5µl 2mM dATP and incubated 3 minutes at 94°C. After adding 1µl 80U/µl terminal transferase, the sample was incubated 20 (or 30) minutes at 37°C and 10 minutes at 70°C, to inactivate the terminal transferase. The PCR of tailed DNA was performed with 5µl dA-tailed DNA, 1µl 12,5µM oligo dT-anchor primer, 1,25µl 10µM gene-specific primer (ChR5), 1µl 10µM dNTPs mix and 0,5µl 5U/µl Taq DNA polymerase. Samples were placed in a thermocycler as follow:

95°C, 5min

30 cycles: 95°C, 30sec

58°C, 25sec

72°C, 45sec

72°C, 10 min

## 2.2.6 Real time-PCR

Real time Polymerase Chain Reaction (PCR) has become in recent years a robust and widely used methodology for biological investigation because it can detect and quantify very small amounts of specific nucleic acid sequences. It has many applications but as aim of this study it was used to quantify gene expression during development. In order to reach this point, the cDNAs used were synthesized from the RNAs of each stage of development (*Salmo salar*: non fertilized eggs, 1dd embryos, 200dd, 370dd, 550dd and hatched. *Danio rerio*: 0hours post fertilization embryos, 4hours, 6hours, 9hours, 12hours, 18hours, 24hours, 36hours, 48hours and 72 hours).

The term “real time” means that it can monitor PCR product while the amplification is occurring, by including a common fluorescent dye (such as

Ethidium Bromide, SYBR Green I, hydrolysis probes, hybridation probes, molecular beacons, sunrise and scorpion primers and peptide nucleic acid light-up probes) and running the PCR under ultraviolet light. This technique allows to better determine the amount of starting DNA in the sample, before the PCR amplification. In this project, Real-time PCRs were performed with SYBR Green I, which binds to the minor groove of the dsDNA, emitting 1000-fold greater fluorescence than when it is free in solution (absorbing light of 480 nm wavelength and emitting light of 520 nm wavelength). Generation of PCR products can be detected by measurement of the SYBR Green I fluorescence signal. In solution, the unbound dye exhibits very little fluorescence which is greatly enhanced upon DNA-binding. During PCR, the increase in SYBR Green I fluorescence is directly proportional to the amount of dsDNA generated. In addition, the final product can be further characterized by subjecting it to increasing temperatures to determine when the double-strand product melts. This melting point is a unique property dependent on product length and nucleotide composition. In melting curve analysis the reaction mixture is slowly heated up to 97°C, which causes melting of dsDNA and corresponding decrease of SYBR Green I fluorescence. The instrument (LightCycler® 480) continuously detects the decrease and displays it as a melting peak.

The running contains the following programs:

1. *Pre-Incubation* for activation of FastStart Taq DNA polymerase and denaturation of the DNA
2. *Amplification* of the target DNA
3. *Melting Curve* for PCR product identification
4. *Cooling* the multiwell plate

Table 2.1 shows the PCR parameters that were programmed for a LightCycler 480 System PCR run with the LightCycler 480 SYBR Green I Master using a LightCycler 480 Multiwell Plate 96.

The reactions were performed with 3µl water (PCR grade), 2µl primers (1µl each, ChF4/ChR5 for salmon and HE1F1/HE1R1 for zebrafish), 10µl Master Mix and 5µl cDNA containing the target gene.

Detection Format	Block Type		Reaction Volume	
SYBR Green	96		10-100 $\mu$ l	
Programs				
Program Name	Cycles		Analysis Mode	
Pre-Incubation	1		None	
Amplification	45		Quantification	
Melting Curve	1		Melting Curves	
Cooling	1		None	
Temperature Targets				
Target (°C)	Acquisition Mode	Hold (hh:mm:ss)	Ramp Rate (°C/s)	Acquisitions (per °C)
Pre-Incubation				
95	None	00:05:00	4.4	–
Amplification				
95	None	00:00:10	4.4	–
58 (primer dependent)	None	00:00:05	2.2	–
72	Single	00:00:25	4.4	–
Melting Curve				
95	None	00:00:05	4.4	–
65	None	00:01:00	2.2	–
97	Continuous	–	–	5-10
Cooling				
40	None	00:00:10	1.5	–

Tab. 2.1: Real time PCR parameters.

## 2.2.7 Agarose gel electrophoresis

Agarose gel electrophoresis is a method used to separate DNA strands by size and to estimate the size of the separated strands by comparison to known fragments (DNA ladder). This is achieved by pulling negatively charged DNA molecules through an agarose matrix with an electric field. Shorter molecules move faster than longer ones.

The gel was made with 1% agarose in 0,5X TEA buffer (50ml) and 1,7 $\mu$ l Ethidium Bromide. It run 2 hours at 70mA.

At the end, those bands can be seen under UV light thanks to the ethidium bromide that inserts into the DNA.

### 2.2.8 Gel extraction

DNA is extracted from an agarose gel following agarose gel electrophoresis in order to isolate a specific band. To perform this step, UV light is shone on the gel to illuminate all the ethidium bromide-stained DNA. The desired band is identified and physically removed. The removed gel should contain the desired DNA inside. It is placed in a 1,5ml microcentrifuge tube. The gel is melted, placed in a spin-column and used in several following centrifugations (Wizard SV gel and PCR clean-up system, Promega). The DNA was eluted in a total volume of 15-50 $\mu$ l nuclease free-H<sub>2</sub>O.

### 2.2.9 Sequencing

The samples extracted from the agarose gel were immediately prepared for the sequencing. The reaction was performed with 1 $\mu$ l Big-Dye 3.1, 1 $\mu$ l sequencing buffer (both provided by SARS sequencing laboratory), 5-20ng DNA template, 5pmol primer and ddH<sub>2</sub>O up to 10 $\mu$ l.

According to the concentrations we found, 1-2 $\mu$ l of DNA were used. The samples were placed in a thermal cycler and heated with a specific program

1 cycle: 96°C, 5min  
25 cycles: 96°C, 10sec  
          T<sub>annealing</sub>, 5sec  
          60°C, 4min

10 $\mu$ l of ddH<sub>2</sub>O were added after the reaction, to have the samples in a 20 $\mu$ l final volume.

### 2.2.10 Bioinformatics research

- BLAST

<http://www.ncbi.nlm.nih.gov/BLAST>

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help to identify members of gene families.

- Blast2sequences

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>

It's a new BLAST-based tool for aligning two given sequences using BLAST engine for local alignment. Blast2Sequences utilizes the BLAST algorithm for pairwise DNA-DNA or protein-protein sequence comparison. A World Wide Web version of the program can be used interactively at the NCBI.

- ClustalW

<http://www.ch.embnet.org/software/ClustalW-XXL.html>

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

- Molecular Toolkit

<http://www.vivo.colostate.edu/molkit/index.html>

The Molecular Toolkit is a group of programs for analysis and manipulation of nucleic acid and protein sequence data.

Nucleic Acid Analysis and Manipulation Programs:

- Dot Plots: examine the similarity of two DNA (or RNA) sequences by production of a similarity matrix displayed as a dot plot.
- Manipulate and Display Sequences: perform simple manipulations on a DNA sequence (inverse, complement, inverse-complement, double-stranded etc).
- Restriction Maps: generate graphical and text-based maps for restriction endonuclease cleavage of DNA.
- Translate: translate a DNA or RNA sequence and obtain graphical and text depictions of the resulting protein sequences.

Protein Analysis Programs:

- Reverse Translate: reverse translate a protein sequence into DNA.
- Protein Composition: obtain the amino acid composition of a protein.
- Hydrophobicity Plots: plot hydrophobic and hydrophilic domains of a protein.

- Integrated DNA Technologies

<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer>

This is also a group of programs but it was used especially for its Oligo Analyzer tool.

- Salmon Genome Project

<http://www.salmongenome.no/cgi-bin/sgp.cgi>

The Salmon Genome Project (SGP) is working to increase the knowledge of the biology of Atlantic salmon (*Salmo salar*) through generation and analysis of genomic data. SGP is a collaboration between seven Norwegian and Canadian research groups. The project focuses on gene function and genome organization through the development of genetic and physical maps, gene sequencing and data interpretation using bioinformatics approaches.



## 3. Results

### 3.1 Preamble

A previous work on *Salmo salar* Hatching fluid has been done before this study on Choriolysin gene started. As I already reported in the introduction, the term *hatching* refers to an embryo breaking out of the eggshell and it is a crucial event in the life cycle of all vertebrates and in fact of almost all sexual animals. After fertilization, the early development of the embryo always takes place inside the eggshell. This means that the embryo has to hatch by breaking through the zona. During hatching, proteases secreted by the embryo into PVF (PeriVitelline Fluid, which protects the embryo during development intra ovo) serve to sunder the zona, to allow the embryo to hatch. After hatching, the embryos are suspended in hatching fluid, which is extra-ovine water plus hatching enzymes and the remnants of the sundered eggshell (zona). After removing embryos, this liquid is called the hatching fluid (HF).

Studies on Salmon HF has revealed proteases that have been patented, and some other new components. The HF is the result of extensive proteolysis by hatching enzymes (Choriolysin and Zonase) on the eggshell and some of the products can confound and complicate purification of HF components. Investigations have revealed that structural zona-proteins are highly hydrophobic and are cross linked by a transglutaminase type enzyme upon fertilization. The cross links formed are isopeptide bonds between parallel protein chains rendering them totally insoluble without breaking covalent bonds. HF presents a difficult material in terms of conventional protein characterization, as many of the components interact in many ways to disguise their identity during protein purification. For this reason, the purification method itself has been patented.

The aims of my previous work were to characterize the composition of the hatching fluid (HF) in order to understand the identity and possible interactions between its proteins. In particular, zonase and choriolysin have been investigated. The work primarily used alternative purification methods. In particular, it was helpful if ion exchange chromatography could be used which is presently difficult due to interactions of PVF-components with highly variably charged remnants of the zona. Also, the variable size of such components impacts adversely on the use of size exclusion chromatography for studying HF/PVF components. The variability of the zona-fragments is due to the action of hatching enzymes on an

enormous substrate with a high molecular weight, where the products of the reaction are necessarily variable.

My work therefore was to investigate conditions under which these conventional methods could be used. The success of such endeavours depends to an essential degree on the condition of the starting material. Most HF preparations are unsuitable for chromatography, but other works in Bergen have developed methods to produce a starting material that is highly refined by proprietary technology to give a modified HF suitable for chromatographic methods. This was my starting material.

In order to overcome interactions of HF components with zona-remnants of variable sizes, I tried to use agents to break such interactions (e.g. urea). However, in order to evaluate the effects of such agents, I first established results when such agents are not used. This allowed a rational approach to attune the conditions needed to break up macromolecular interactions in the HF and thus use conventional purification methods.

In sum, such work could perhaps allow better characterization of components in HF/PVF. This study was performed by using several chromatographic methods, in particular size exclusion and ion exchange (column pretreated with urea). With the first one I could place the proteases together in a huge complex, at a molecular weight around 650kDa. The ion exchange chromatography showed that the complex was eluted twice, first with urea (2M) and then with 200mM NaCl. SDS-PAGE showed two bands around 37kDa, probably choriolysin, two bands at about 25kDa, supposed to be zonases, and other peptides (eggshell fragments) below 17kDa. 2D-PAGE placed choriolysin and zonases at two different pHs respectively at about 5 and 6.

One way for further works on this biochemistry study of the Hatching fluid could be to start purifications with a material taken from inside the eggs (i.e. PVF), not breaking the egg shells, in order to avoid small fragments that can disturb the background of some experiment. However, this is very hard because we can have few  $\mu$ l of material from each egg instead of the several ml we would need. Moreover, it is possible to characterize this preparation by cutting each spot from the 2D-PAGE gel and analyse them separately by mass spectrometry coupled to data bank identification of peptides.

For my thesis work I decided to focus the attention on the genetic part of these proteases, in particular on Choriolysin, which gene structure and mRNA expression during embryo development is quite well known in many species but not in Salmon.

## 3.2 Choriolysin nucleotide sequence

### 3.2.1 RNA extraction

RNA extraction of salmon and zebrafish embryos gave good concentrations of total RNA:

Salmon	Zebrafish eggs
non fertilized eggs: 0,2µg/µl	0 hours: 1µg/µl
1dd embryos: 0,1µg/µl	4 hours: 1,3µg/µl
200dd: 1,4µg/µl	6 hours: 1,5µg/µl
370dd: 3,2µg/µl	9 hours: 1µg/µl
550dd: 2,2µg/µl	12 hours: 1,2µg/µl
hatched: 1µg/µl	18 hours: 1,5µg/µl
	24 hours: 1,8µg/µl
	36 hours: 2,5µg/µl
	48 hours: 2µg/µl
	72 hours: 0,4µg/µl
	120 hours: 1,6µg/µl

### 3.2.2 Amplification with Ch(F/R)1 and Ch(F/R)2

My study on *Salmo salar* Choriolysin gene started from its protein sequence.

In order to find the gene, RNA extracted from stages mentioned above was retrotranscribed in cDNA and amplified by PCR.

PCR was as first performed with degenerated primers, which means that the inverse translation of the protein (DNA manipulation tool kit, sequence in Appendix) was used to pick up some fragments of that sequence as oligos in the PCR reaction.

The chosen primers were:

ChF1: 5' – TGG ATC ATC GTC CTA CTC TTT C – 3'  
Length 22bp (from nt 2 to nt 23)

ChR1: 5' – CCA AGA CGT TGA CCA ATA GC – 3'  
Length 20bp (from nt 659 to nt 678)

ChF2: 5' – CGT AAT GCT ATG AAA TGT TTT TCT TC – 3'  
Length 26bp (from nt 181 to nt 205)

ChR2: 5' – CCA TAA TCA TTA GTA TAA GCA GTA CGA – 3'  
Length 27bp (from nt 657 to nt 683)

Expected product size ChF1/R1: 739bp

Expected product size ChF2/R2: 503bp

Many PCR conditions were tried (several annealing temperature and different concentrations of the reaction components) but none of them worked properly to give an amplicon (Fig. 3.1).

Every time a positive control was included, sometimes a different PCR-kit or cDNA and primers (Zonase primers) were used as positive control to check the equipment and in those cases we could obtain a product.

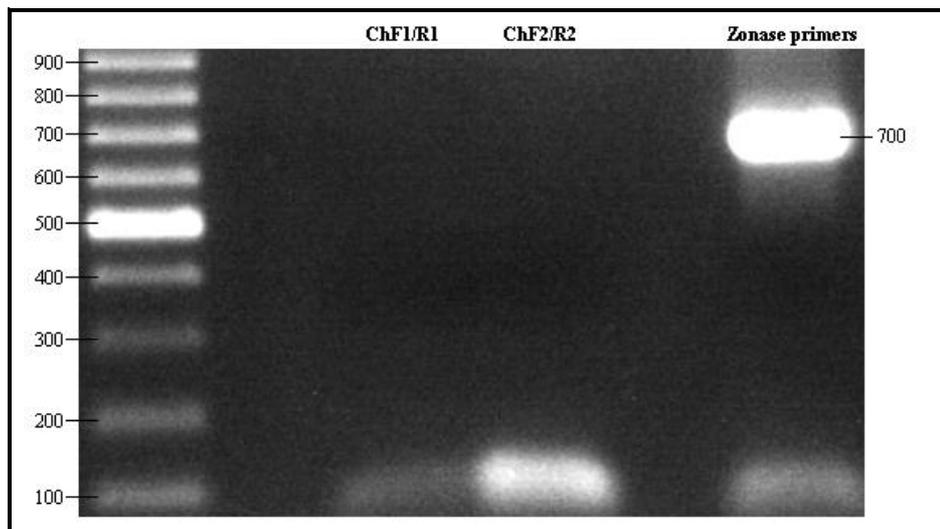


Fig. 3.1: PCR amplification of Choriolysin with both the couples of generated primers [ChF/R1] and [ChF/R2] and of control gene Zonase.

This led us to think about choriolysin primers we used. The bands appeared at about 1kb can be assigned to the annealing of primers with each other.

### 3.2.3 Primers Ch(F/R)3

Since the degenerated primers did not work, another approach has been tried in order to find oligos capable to amplify the gene.

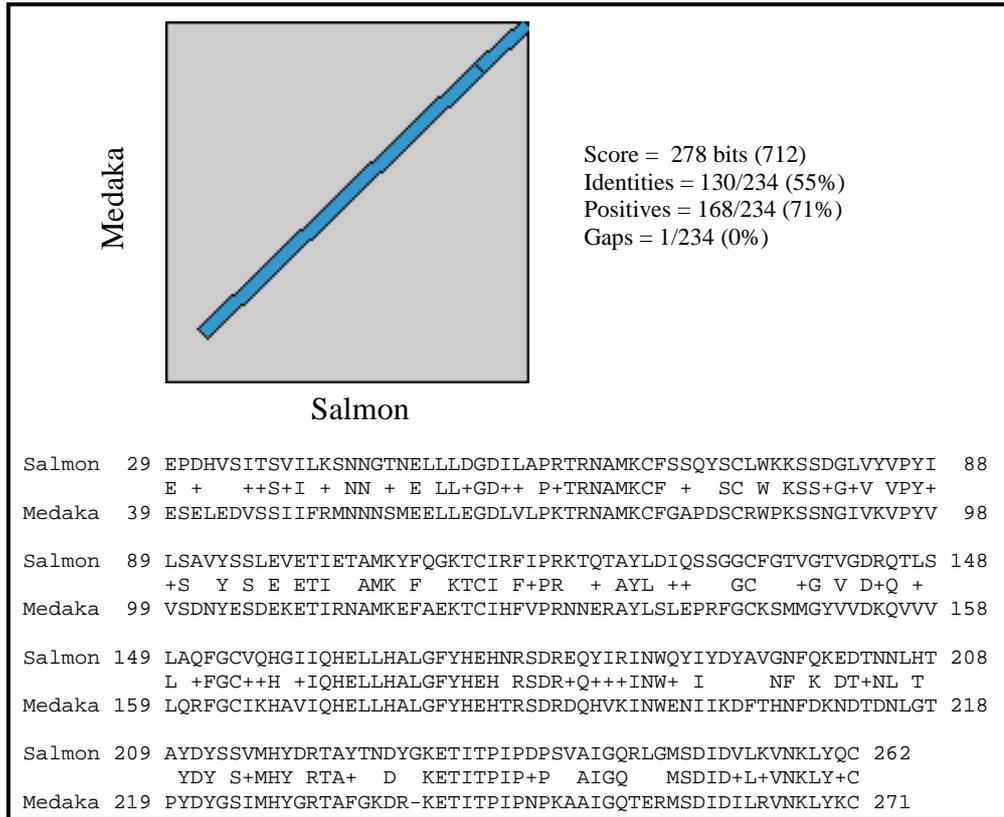
The protein sequence of *Salmo salar* Choriolysin, compared by Blast with protein databases of all the other organisms, gave alignments with many Hatching enzymes. The most similar sequence was medaka *Oryzias latipes* Choriolysin L with 55% identity.

*Oryzias latipes* Choriolysin L protein (Accession number BAA20403.1) and cDNA sequences (Accession number D83949) are shown in Tab. 3.1.

a)	MDLLAKASVL LLLLLSLSNA QTDNMEEAEN GSSKEEIVES ELEDVSSIIF RMNNSMEEL LEGDLVLPKT RNAMKCFGAP DSCRWPKSSN GIVKVPYVVS DNYESEKET IRNAMKEFAE KTCIHVFP RN ERAYLSLEP RFGCKSMGY VVDKQVVVLQ RFGCIKHAVI QHELLHALGF YHEHTRSDRD QHVKINWENI IKDFTHNFDK NDTDNLGTPY DYGSIMHYGR TAFGKDRKET ITPIPNPKAA IGQTERMSDI DILRVNKLYK C
b)	atggacctgc tggccaaagc atctgtgctg ctggtgctgc tcctgagcct cagcaacgct caaactgaca atatggaaga agcagaaaac ggttcatcaa aggaggaaat agttgagtct gaactggagg acgtgtcctc catcatcttc agaatagaca acaactctat ggaggaactg ttggaaggag atcttgttct tcccaaaaacc aggaatgcca tgaagtgtt tggcgctcca gatagctgcc gctggccaaa gtcttccaat ggcacgtga aggttcctta tgtggttagc gacaactatg aaagtgacga gaaggaaacc attcggaacg ccatgaagga gtttgacagaa aaaacctgca ttcactttgt tcctcgcaac aatgagaggg cctacctgag ccttgaaccc agatttggct gcaagtctat gatgggctat gtggttgaca aacaagtggg ggtgctgcag cggtttggct gcataaagca cgccgtcatc cagcatgagc tcctgcatgc tctgggttct taccacgagc acactcggag cgaccgcgac cagcatgtaa aaatcaactg ggaaaacatc attaagatt tcacacacaa ctttgataag aatgataaccg acaatctggg caccctgat gactatggct ccatcatgca ctatggaaga actgcctttg gaaaagacag aaaggagacc ataacccccca tccctaacc caaagctgcc attggccaaa cagagaggat gtcggacatc gatatacttc gagtcaataa gctttacaaa tgttga

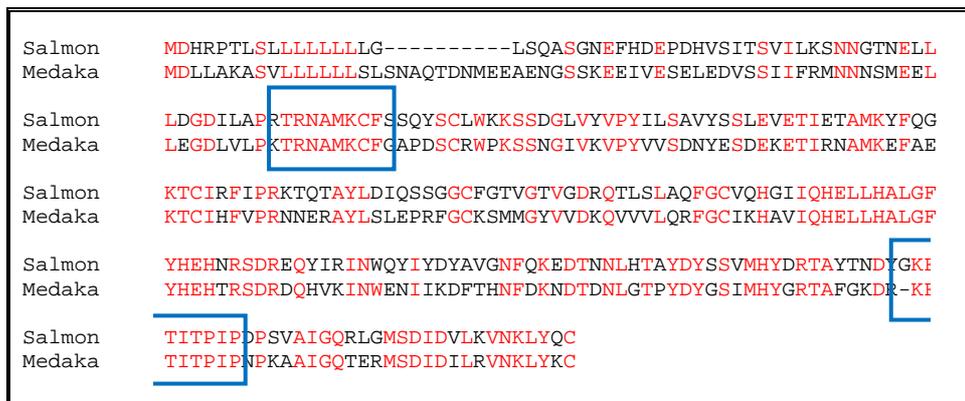
Tab. 3.1: a) amino acid sequence of medaka Choriolysin L; b) nucleotide sequence of medaka Choriolysin L.

Blast alignment between Salmon and Medaka Choriolysin showed a very significant similarity of 55% (Tab. 3.2).



Tab. 3.2 : Blast alignment outcome.

This medaka Choriolysin L sequence and the salmon protease were aligned by Blast2Sequences (Tab. 3.3) which showed the matches between the two protein sequences (medaka and salmon Choriolysin).



Tab. 3.3 : Blast2Sequence outcome by aligning medaka and salmon Choriolysins.

The alignment between medaka and salmon Choriolysin proteins showed two longer segments where the amino acid sequences matched perfectly (blue panel):

- sequence: TRNAMKCF (aa 70-77 in medaka and 60-67 in salmon).
- sequence: KETITPIP (aa 238-245 in medaka and 229-236 in salmon).

Medaka nucleotide sequence is known and corresponds to these two segments:

- sequence: accaggaatgccatgaagtgcttt (nt 208-231).
- sequence: aaggagaccataacccccatcct (nt 712-735).

The sequences were used as template to design primers for salmon Choriolysin. If both the species have the same amino acid sequence, it was possible that they have also the same nucleotide sequence.

After a careful selection, the resulted primers were:

ChF3: 5' – CCA GGA ATG CCA TGA AGT GCT TTG – 3'  
Length = 24nt,  $T_m = 52,3^\circ\text{C}$ , GC content = 50%

ChR3: 5' – AGG GAT GGG GGT TAT GGT CTC CT – 3'  
Length = 23nt,  $T_m = 53^\circ\text{C}$ , GC content = 57%

### 3.2.4 Amplification with Ch(F/R)3

The first PCR reaction with primers ChF3/ChR3, which should give an amplicon at about 531bp, took place with salmon stages 200dd (Fig. 3.2) and 370dd (Fig. 3.3) only, due to their major abundance and availability, in order to select the best PCR conditions to use also with the other stages.

The three products refer to different annealing temperatures that have been tried to find the best running conditions.

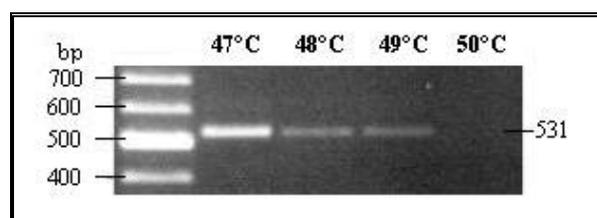


Fig. 3.2: 200dd stage PCR amplification.

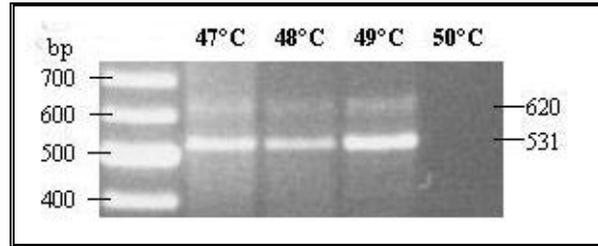


Fig. 3.3: 370dd stage PCR amplification.

PCR amplification yielded two products, the expected band at 531bp and another one around 620bp. Both were subject of sequence analysis in order to know the sequence of the first and to define the identity of the other one.

### 3.2.4.1 531bp sequence

More than one sample was sequenced and very few mismatches were found, probably due to mistakes of the sequencer. All the sequences were compared with CustalW to end up with a consensus sequence.

Choriolysin consensus sequence is reported in the table below (Tab. 3.4).

AAGTCATCTG	ACGGCTTGGT	GTACGTGCCT	TACATCCTCA	GCGCTGTATA	TTCCAGCTTG
GAGGTAGAGA	CTATTGAGAC	GGCCATGAAG	TACTTCCAAG	GGAAGACCTG	CATCCGCTTC
ATTCCACGTA	AGACACAGAC	TGCCTACCTG	GACATTCAGA	GCAGCGGCGG	GTGTTTTGGT
ACCGTGGGGA	CTGTTGGGGA	CAGGCAGACA	TTGTCTCTTG	CACAGTTTGG	CTGTGTTCAA
CATGGTATCA	TCCAGCATGA	GCTGCTTCAC	GCCCTGGGCT	TCTACCACGA	GCACAACAGG
AGTGACCGTG	AACAGTATAT	CAGGATCAAC	TGGCAATACA	TCTATGACTA	CGCCGTTGGG
AACTTCCAGA	AGGAGGACAC	CAACAACCTG	CACACTGCAT	ACGACTACTC	CTCTGTCATG
CACTATGATA	GAACAGCTTA	CACTAACGAC	TACGGAAAGG	AGACCATAAC	CCCCATCCCT

Tab. 3.4 : salmon Choriolysin consensus nucleotide sequence.

This sequence was translated into its amino acid sequence (Tab. 3.5) and compared by ClustalW with the Choriolysin protein sequence I had at the beginning of my study.

KSSDGLVYVP	YILSAVYSSL	EVETIETAMK	YFQKTCIRF	IPRKTQTAYL	DIQSSGGCFG
TVGTVGDRQT	LSLAQFGCVQ	HGIIQHELLH	ALGFYHEHNR	SDREQYIRIN	WQYIYDYAVG
NFQKEDTNLL	HTAYDYSSVM	HYDR TAYTND	YGKETITPIP		

Tab. 3.5 : salmon Choriolysin consensus amino acid sequence.

Compared with the previous one, it was found to be the same (Tab. 3.6).

Salmon	MDHRPTLSLLLLLLLLLGLSQASGNEFHDEPDHVSITSVILKSNNGTNELLLLDGDILAPRT
Consensus	-----
Salmon	RNAMKCFSSQYSCLWKKSSDGLVYVYPYILSAVYSSLEVETIETAMKYFQGKTCIRFIPRK
Consensus	-----KSSDGLVYVYPYILSAVYSSLEVETIETAMKYFQGKTCIRFIPRK
Salmon	TQTAYLDIQSSGGCFGTGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSDRE
Consensus	TQTAYLDIQSSGGCFGTGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSDRE
Salmon	QYIRINWQYIYDYAVGNFQKEDTNNLHTAYDYSSVMHYDR TAYTNDYGKETITP IPDPSV
Consensus	QYIRINWQYIYDYAVGNFQKEDTNNLHTAYDYSSVMHYDR TAYTNDYGKETITP IP----
Salmon	AIGQRLGMSDIDVLKVNKLYQC
Consensus	-----

Tab. 3.6 : ClustalW alignment between the beginner Choriolysin sequence [Salmon] and the amino acid consensus sequence [Consensus].

The protein sequence goes from the amino acid 76 to 196, which means that the nucleotide sequence goes to the nucleotide 226 to 646 for a total length of 480bp known of the whole 786nt sequence.

#### 3.2.4.2 620bp unknown band

This unspecific band was extracted from the agarose gel and sequenced (both nucleotide and amino acid sequences in Appendix). It was analyzed and compared by Blast and ClustalW but none of them gave back matches with any hatching enzyme. It was found to be an unknown sequence. Since it is not relevant for my study, I decide not to go forward.

#### 3.2.5 Primers Ch(F/R)4 and ChR5

Using the new consensus sequence, it was possible to draw new primers sequence specific. In order to perform RACE-PCR, two reverse primer were designed.

ChF4: 5' – AAG TCA TCT GAC GGC TTG GTG TAC G – 3'  
Length = 25nt,  $T_m$  = 60,7°C, GC content = 52%

ChR4: 5' – GAG GAG TAG TCG TAT GCA GTG TGC A – 3'  
Length = 25nt,  $T_m$  = 60,1°C, GC content = 52%

ChR5: 5' – ACT GTT CAC GGT CAC TCC TGT TGT G – 3'  
Length = 25nt,  $T_m$  = 60,9°C, GC content = 52%

### 3.2.6 RACE-PCR

With those new primers I tried to find the flanking sequences of the amplified region, or rather the 5'- and 3'- ends. Rapid amplification of cDNA ends (RACE) PCR showed, for the 3' end, amplicons of different sizes in different performance of this technique and one band for the 5' amplicon. According to the positions of the primer forward, only the bands higher than 600bp were extracted and sequenced: from ChF2 to the end of the sequence coding for the Choriolysin protein, the size is supposed to be 557bp and considering the poly(A) in the end, the sequence must have been longer.

In Fig. 3.4 is represented the only sequence between those selected (680bp for 5', 640bp, 690bp and 900bp for 3') which, once sequenced (sequence in Appendix), gave a match with Choriolysin sequence.

690bp band at 3' end was amplified using primers ChF4/oligo d(T).

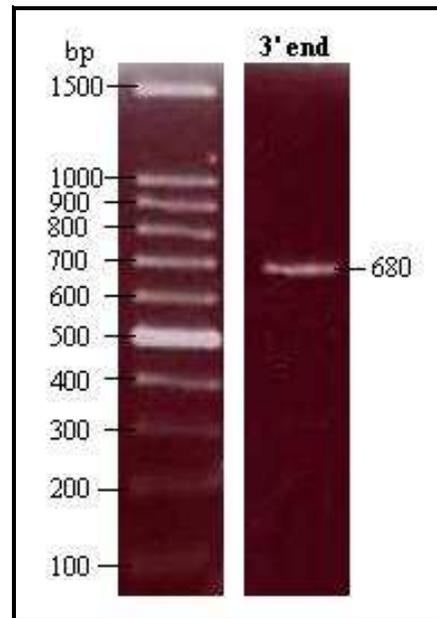


Fig. 3.4: RACE-PCR amplification

The new nucleotide sequence translated into its amino acid sequence is shown in Table 3.7.

```

QRCIFSLEVE TIETAMKYFQ GKTCIRFIPR KTQTAYLDIQ SSGGCFGTVG TVGDRQTLSTL
AQFGCVQHGI IQHELLHALG FYHEHNRSRDR EQYIRINWQY IYDYAVGNFQ KEDTNNLHTA
YDYSSVMHYD RTAYTNDYGK ETITPIPDPS VAIGQRLGMS DIDVLKVNKL YQC-EEERHC
-KCVMLDVLS CADVFYCWKF VCILLITLVI IKHGYGKKKR RVWVLPAL-P LHLPA-

```

Tab. 3.7: amino acid sequence of 3' end RACE-PCR product.

Protein sequences of Choriolysin sequence and this new one were compared by ClustalW (Tab. 3.8) and the two sequences had a perfect match till the end.

Choriol 3'end	MDHRPTLSLLLLLLLLLGLSQASGNEFHDEPDHVSITSVILKSNNGTNELLLLDGDILAPRT -----
Choriol 3'end	RNAMKCFSSQYSCLWKKSSDGLVYVPYIILSAVYSSLEVETIETAMKYFQGKTCIRFIPRK -----QRCIFSLLEVETIETAMKYFQGKTCIRFIPRK
Choriol 3'end	TQTAYLDIQSSGGCFGTGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSDR TQTAYLDIQSSGGCFGTGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSDR
Choriol 3'end	QYIRINWQYIYDYAVGNFQKEDTNNLHTAYDYSSVMHYDR TAYTNDYGKETITPDPDSV QYIRINWQYIYDYAVGNFQKEDTNNLHTAYDYSSVMHYDR TAYTNDYGKETITPDPDSV
Choriol 3'end	AIGQRLGMSDIDVLKVNKLYQC- AIGQRLGMSDIDVLKVNKLYQC-

Tab. 3.8 : Clustal W comparison of 3'end RACE.PCR product [3'end] and my protein sequence [Choriol].

The respective nucleotide sequences were compared too (sequence outcome of ClustalW in Appendix). Except for a few nucleotides, the sequences matched and the second one (after RACE) let us complete it till the final stop codon.

A final nucleotide Choriolysin sequence was made by adding the 3'end got after RACE-PCR (Tab. 3.9). This sequence is finally known from nucleotide 229 to 786 (+ 3nt stop codon).

AAGTCATCTG	ACGGCTTGGT	GTACGTGCCT	TACATCCTCA	GCGCTGTATA	TTCCAGCTTG
GAGGTAGAGA	CTATTGAGAC	GGCCATGAAG	TACTTCCAAG	GGAAGACCTG	CATCCGCTTC
ATTCCACGTA	AGACACAGAC	TGCCTACCTG	GACATTCAGA	GCAGCGGCGG	GTGTTTTGGT
ACCGTGGGGA	CTGTTGGGGA	CAGGCAGACA	TTGTCTCTTG	CACAGTFTGG	CTGTGTTCAA
CATGGTATCA	TCCAGCATGA	GCTGCTTCAC	GCCCTGGGCT	TCTACCACGA	GCACAACAGG
AGTGACCGTG	AACAGTATAT	CAGGATCAAC	TGGCAATACA	TCTATGACTA	CGCCGTTGGG
AACTTCCAGA	AGGAGGACAC	CAACAACCTG	CACACTGCAT	ACGACTACTC	CTCTGTCAATG
CACTATGATA	GAACAGCTTA	CACTAACGAC	TACGGAAAGG	AGACCATAAC	CCCCATCCCT
GACCCATCTG	TGGCCATTGG	ACAGAGACTG	GGCATGTCCG	ACATTGATGT	CCTGAAGGTC
AACAAGCTCT	ACCAATGCTA	A			

Tab. 3.9: Sequence known before the RACE-PCR (black); sequence got in addition after RACE which translation arrives till the end of Choriolysin sequence (green); stop codon (red).

The stretch between the second and the third stop codon is long enough to be some kind of product but comparing both its nucleotide and amino acid sequences in databases, they did not give out any match.

### 3.3 Choriolysin gene expression

#### 3.3.1 Amplification of Choriolysin by PCR

The Choriolysin gene amplification was performed then in all the stages I had at my disposal, to define in which stages it is present and study its expression:

Atlantic salmon <i>Salmo salar</i>	Zebrafish <i>Danio rerio</i>
Sperm	0 hours post fertilization
Non fertilized eggs	4 h
1dd embryos	6 h
200dd	9 h
370dd	12 h
550dd	18 h
Hatched embryos	24 h
	36 h
	48 h
	72 h
	120 h

The primer used to perform the PCR were Ch(F/R)4 in salmon (Fig. 3.5) and HE1F1/HE2R2 in Zebrafish (Fig. 3.6), which were supposed to amplify fragments at about 413bp and 498bp respectively.

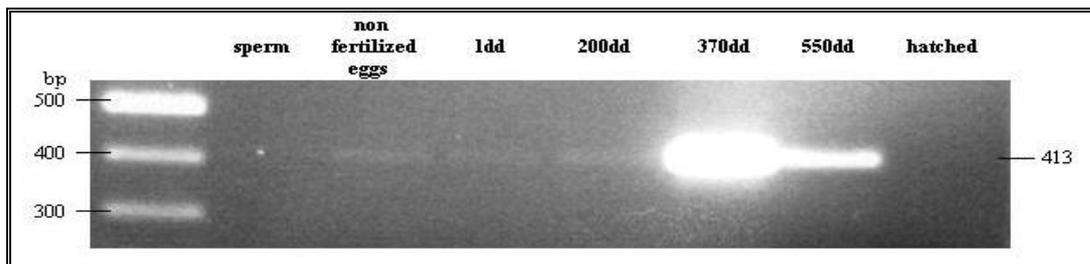


Fig. 3.5: Salmon PCR amplification.

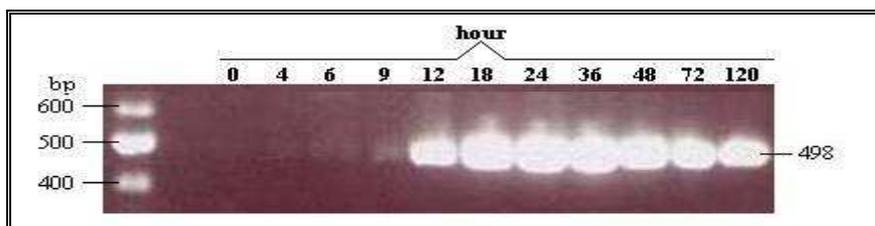


Fig. 3.6: Zebrafish PCR amplification.

### 3.3.2 Amplification of Choriolysin by Real time-PCR

Beside the usual PCR, I could perform the Real-time PCR in order to study the expression of Choriolysin in both the organisms Salmon and Zebrafish. This technique measures the concentration of the sample going up each cycle and allows to define the concentration of the starting cDNA sample (since it is related to the amount of RNA).

Once the standard curve was set up, the instrument extrapolated the concentration values of the samples (plot in Fig. 3.7) by detecting the absorbance of SYBR Green I bound to the DNA.

Salmon	
non fertilized eggs	$8,1 \cdot 10^{-8} \text{ ng}/\mu\text{l}$
1dd embryos	$1,19 \cdot 10^{-7} \text{ ng}/\mu\text{l}$
200dd	$7,8 \cdot 10^{-8} \text{ ng}/\mu\text{l}$
370dd	$2,48 \cdot 10^{-3} \text{ ng}/\mu\text{l}$
550dd	$5,75 \cdot 10^{-6} \text{ ng}/\mu\text{l}$
hatched	$3 \cdot 10^{-9} \text{ ng}/\mu\text{l}$

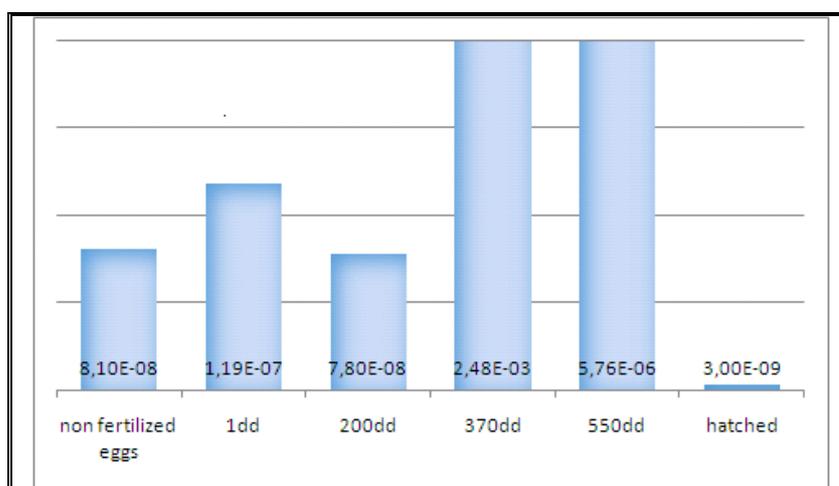


Fig. 3.7: concentration values of salmon stages

Another outcome beside these values is the melting pick(s) of each sample. More than one pick per sample means that there are many PCR products and this would be related to a low specificity of the reaction, due to the primers sequences as well as the PCR program.

Both values and plot (Fig. 3.8) support the proof that there is only one product, which melts around  $87^{\circ}\text{C}$ .

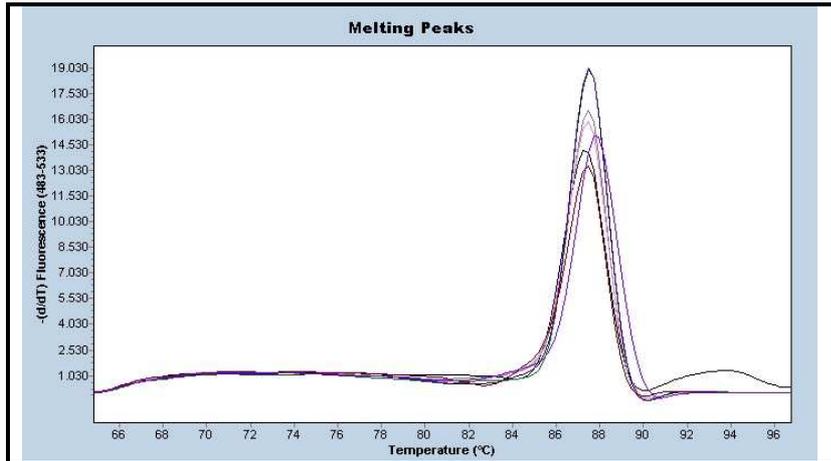


Fig. 3.8: melting picks of salmon stages

The same experiment was performed for zebrafish samples, so we got concentration values (plot in Fig. 3.9) and melting pick around 84°C (Fig. 3.10).

Zebrafish	
0 hours	$1 \cdot 10^{-14}$ ng/ $\mu$ l
4 hours	$1 \cdot 10^{-14}$ ng/ $\mu$ l
6 hours	$1 \cdot 10^{-14}$ ng/ $\mu$ l
9 hours	$1 \cdot 10^{-8}$ ng/ $\mu$ l
12 hours	$3,86 \cdot 10^{-10}$ ng/ $\mu$ l
18 hours	$3 \cdot 10^{-6}$ ng/ $\mu$ l
24 hours	$1,27 \cdot 10^{-5}$ ng/ $\mu$ l
36 hours	$4,59 \cdot 10^{-5}$ ng/ $\mu$ l
48 hours	$4,61 \cdot 10^{-10}$ ng/ $\mu$ l
72 hours	$1 \cdot 10^{-14}$ ng/ $\mu$ l
120 hours	$6,53 \cdot 10^{-10}$ ng/ $\mu$ l

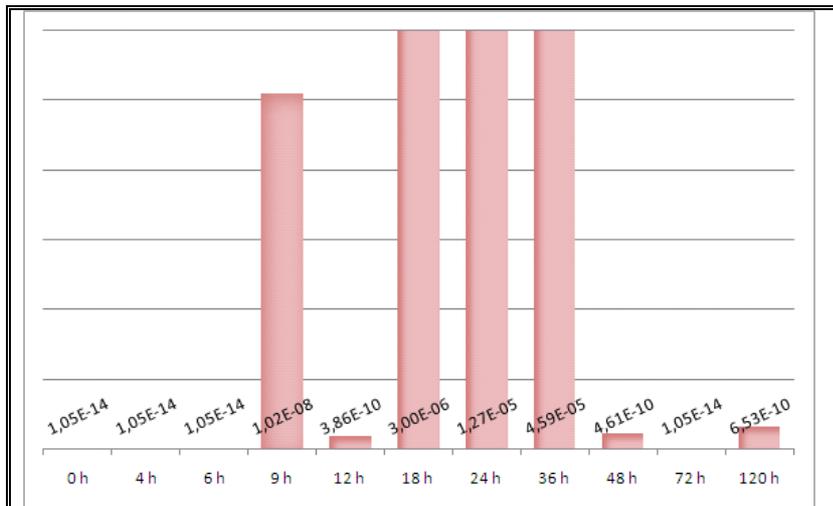


Fig. 3.9: concentration values of zebrafish stages

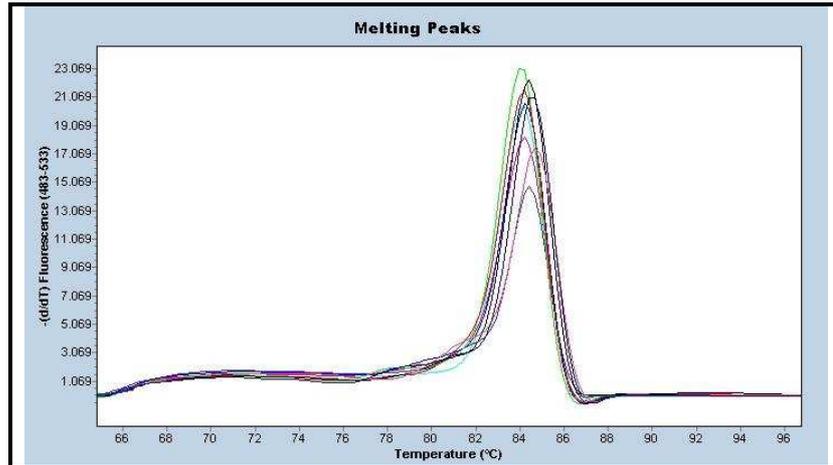


Fig. 3.10: melting picks of zebrafish stages

These results were checked by agarose gel (salmon Fig. 3.11 and zebrafish Fig. 3.12). It is possible to see one single band in every stage, an evidence that one melting peak refers to one product. Moreover, the intensity of each band shows the level of expression of the gene, supported by the real-time absorbance values.

Real time PCR was performed with primers ChF4/ChR5 in salmon and the expected amplified product size of 316bp was found. Also in zebrafish was found the right product size of 228bp, performed with primers HE1(F/R)1.

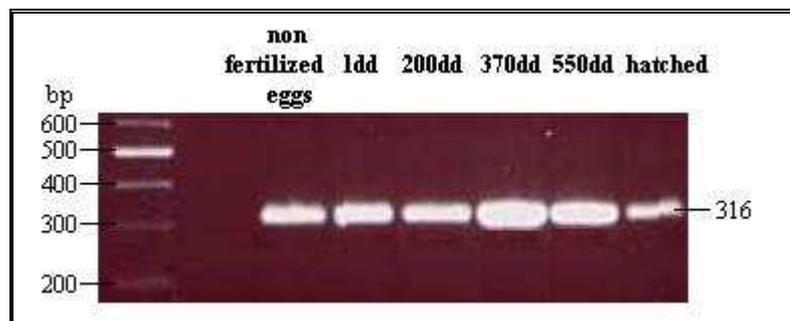


Fig. 3.11: Salmon Real-Time PCR amplification

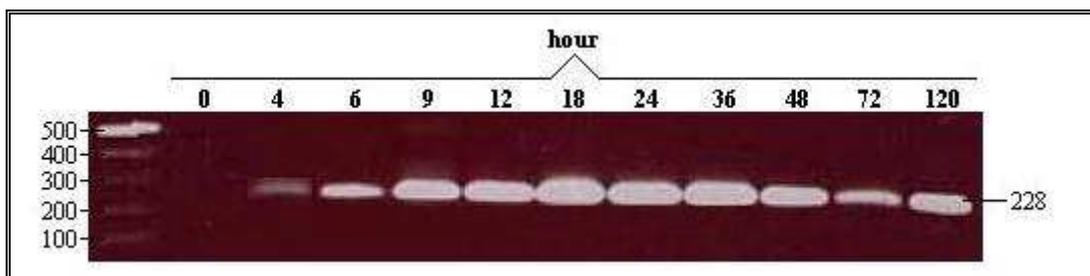


Fig. 3.12: Zebrafish Real-Time PCR amplification

Either the products of both salmon and zebrafish obtained by PCR and by Real-time PCR have been sequenced and they confirmed the contents of the samples as the known sequences of *Salmo Salar* Choriolysin and *Danio rerio* Hatching Enzyme 1.

## **4. DISCUSSION**

The purpose of this thesis was at first to determine the nucleotide sequence of *Salmo salar* Choriolysin, to be able in future to clone it and make recombinants to study its domain. During my study, the 3'-end and the core part of the sequence has been amplified and sequenced.

Choriolysin is an hatching enzyme, which is produced before hatching time for helping the breakdown of the eggshell and allowing the larva to get out. In order to understand when this protease starts to be transcribed and in which amount, its expression in embryo development has been checked, through its mRNA level.

### **4.1 Choriolysin nucleotide sequence**

#### **4.1.1 Amplification with Ch(F/R)1 and Ch(F/R)2**

Amplification of unknown sequence of a gene is considered a big challenge. The starting point of such a study is often the use of degenerate primers obtained from the known protein sequence [30]. Each amino acid is coded by 3 nucleotides, which means  $3^4$  possible combinations. Amino acids are only 20, so there can be several ways to combine nucleotides and end up with the same amino acid. A degenerated sequence can possibly translate into the right protein sequence but permits only guesses as to the right nucleotide sequence.

Based on the protein sequence of *Salmo salar* Choriolysin, I designed 2 couples of degenerated primers, Ch(F/R)1 and Ch(F/R)2.

The alignment of the whole Choriolysin nucleotide sequence obtained by retrotranscription of its protein matched with a microsatellite (sequence in Appendix). This kind of sequence is known to be a simple sequence repeats: it consists of short arrays of simple tandem repeats, dispersed throughout the

genome and often highly polymorphic. They're thought to have arisen mostly by replication slippage and they still have unknown significance.

This DNA has generally been identified in non-coding sequences, intergenic DNA or within the introns of genes. It is understandable that I should find this microsatellite instead of choriolysin sequence: the amplification is based on cDNA retrotranscribed from mRNA, but there is no chance to find a sequence which is placed in a non-coding region because there is no mRNA transcribed.

#### **4.1.2 Primers Ch(F/R)3**

Choriolysin enzyme is an hatching enzyme and its gene has been widely characterized in many species but not in Atlantic salmon *Salmo salar*. The only known sequence was the amino acidic one, which was purified in my laboratory by purification of Hatching Fluid. For this reason, prior to design working primers to use in PCR amplifications, I tried to get them from the sequence of another species which is evolutionarily closely highly related, medaka *Oryzias latipes* (Salmon – Medaka Choriolysin amino acid sequences = 55% similarity, Blast outcome).

Those two protein sequences were aligned by ClustalW and two particular fragments of 8 amino acids showed up without inner mismatches.

Since this family of proteases is highly conserved among species, it was possible that the same amino acid sequence might have also the same nucleotide sequence. With this strong opinion I designed primers Ch(F/R)3, based on medaka nucleotide sequence of the matching fragments.

There were other sequences I could use, for example the amino acids involved in the Zn-binding domain or in the Met-turn. These segments are highly conserved too in almost all the species, but their amino acid sequence resulted to have some more mismatches with my sequence, even if these changes don't compromise the 3D structure or the biochemical role.

### 4.1.3 Amplification with Ch(F/R)3

The PCR amplification with ChF3/ChR3 was performed only with 370dd stage, in order to check their function in a stage with larger amount of RNA and where this enzyme is supposed to be more abundant. The agarose gel performed with PCR products showed one strong and one weak band around 530 and 620 base pairs.

Primers Ch(F/R)3 were supposed to amplify a region of 531bp of my gene, Choriolysin and the sequencing of that sample obtained from agarose gel confirmed its identity.

Seven samples of the 531bp band have been sequenced (four were performed with forward primer ChF3 and three with reverse primer ChR3; comparison by ClustalW in Appendix) to be able to make a sort of consensus sequence, in order to avoid sequencer mistakes. At the beginning they had more “N” (undetected nucleotide) than what is reported in the results. Those sequences were reviewed because sometimes the given signal of a nucleotide might be unclear or covered by another one, and the sequencer has difficulties in reading it.

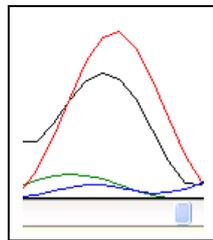


Fig. 4.1: double peak; the sequencer can not read properly but by eye one nucleotide is clearer than another nucleotide.

When this happens, the sequence comes out with an “N” in such positions instead of a specific nucleotide. Very often is possible to distinguish one nucleotide (one peak) from another one just reading by eyes the frame of peaks (Fig.4.1). Somehow the border sequences are always too difficult to read and I took them out, that is why I miss anyway the first part of the amplified sequence next to the forward primer ChF3.

The sequences, were compared with ClustalW. First the sequences obtained with reverse primer have been converted in their inverse complement. ClustalW analysis gave a great result because they differed for few nucleotides only, spread over the whole length. Have several sequences allows to analyze each position and make a consensus sequence.

The consensus nucleotide sequence resulted to be 480bp long, while the expected product size of the amplified fragment should be 531bp long. This is due to the many “N” (unknown nucleotides) present especially in one side of the

sequence, which make unreadable that first part. The last part indeed was specified till the last nucleotide used as primer (Tab. 4.1).

In the sequence length, few nucleotides showed up to be different in one of all the sequences I analyzed. Looking at the frame of peaks, those had the same intensity (high of them peaks) as the right ones should have, so it might easily be mistaken from the sequencer. Others instead were completely different, some determining amino acids changes and some not. Even this can be due to the sequencer, but could also be due to other copies of Choriolysin gene present in the genome, which can have little (and meaningless) differences in the nucleotide sequences. This should be object of further studies, to understand if the gene has one copy only (like medaka LCE has) or more than one copy (like zebrafish HE1).

The amino acid sequence obtained from the translation of this sequence, matched perfectly with the choriolysin protein sequence I had at the beginning of my study and this confirmed I amplified the right product.

The sequences above, Fig. 4.1, represent the choriolysin protein sequence I used to start this study (seq. 1) and the protein sequence obtained from the translation of the consensus sequence (seq. 2); it's quite clear that they are the same.

seq. 1	MDHRPTLSLLLLLLLLLGLSQASGNEFHDPEPDHVSITSVILKSNNGTNELLLLDGDILAPR <u>T</u>
seq. 2	-----
seq. 1	<u>R</u> NAMKCFSSQYSCLWKKSSDGLVYVPYILSAVYSSLEVETIETAMKYFQGKTCIRFIPRK
seq. 2	-----KSSDGLVYVPYILSAVYSSLEVETIETAMKYFQGKTCIRFIPRK
seq. 1	TQTAYLDIQSSGGCFGTVGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSRE
seq. 2	TQTAYLDIQSSGGCFGTVGTVGDRQTLSLAQFGCVQHGI IQHELLHALGFYHEHNRSRE
seq. 1	QYIRINWQYIYDYAVGNFQKEDTNLHTAYDYSSVMHYDR TAYTNDY <u>GKETITPIP</u> DPSV
seq. 2	QYIRINWQYIYDYAVGNFQKEDTNLHTAYDYSSVMHYDR TAYTNDY <u>GKETITPIP</u> ----
seq. 1	AIGQRLGMSDIDVLKVNKLYQC
seq. 2	-----

Tab. 4.1: starting choriolysin protein sequence [seq. 1]; consensus protein sequence [seq. 2]; in blue: sequences used as primers.

Concerning Blast analysis, the match in both medaka and zebrafish genomes confirmed that my amplified product belongs to the hatching enzymes family. Moreover, the first medaka sequence I found comparing the consensus sequence by BlastN, is Choriolysin L, the same I found comparing the starting salmon Choriolysin protein sequence, that I later used to design primers Ch(F/R)3. This

means that not only the protein sequence has high similarity but also the nucleotide sequence, which could lead to the high conservation between species of these genes.

The other product at 620bp was found unknown, after have been sequenced as well, but since it was not relevant for my study, I did not go forward to try to define the product.

There might be two reasons why some gene-specific primers can fail or give other products beside those expected. The main reason could be that primers Ch(F/R)3 were designed based on the sequence of another species, medaka, and even if the similarity of the fragments I used was 100% (no mismatches) and the whole protein sequence is shown to be highly conserved, I can not exclude that the nucleotide sequence differs a bit due to the degenerate code.

Consensus sequence I made after sequencing of 531bp band was aligned with medaka Choriolysin L nucleotide sequence and the right end, comprehensive of primer ChR3 sequence, matched 100%. Little unspecificity of ChF3 can be the reason why I always missed that first part of the sequence.

The second reason can be due to the PCR settings. Non-perfect annealing temperature or amount of cDNA can lead to lower specificity of this technique.

#### **4.1.4 RACE-PCR**

Rapid Amplification of cDNA Ends (RACE) -PCR takes advantage of the poly-A tail of a sequence to amplify unknown 3'- and 5'- ends. 3'-end is normally provided itself of a poly-A tail, since it is directly retrotranscribed from mRNA, while the 5'-end fragment has been added to a homopolymeric A-tail by a Terminal Transferase. Amplifications were performed with primers Ch(F/R)4 and ChR5, gene-specific primers based on the salmon Choriolysin sequence I got after the first amplification with primers Ch(F/R)3.

Bands intensity and size distribution of resulting products depend on the specificity of the gene-specific primers used (because all the cDNA sequences can anneal with the the oligo d(T) anchor primer) for the cDNA synthesis and the PCR, the complexity and relative abundance of target cDNA and the PCR conditions used. For these reason, products may vary from a single specific band to multiple discrete products or broad diffuse smear.

In view of the trouble that I might encounter, I decided to amplify only the 3'end.

In the end, adding the core part amplified with primers Ch(F/R)3, I obtained a sequence 558bp long instead of the total 786bp. This is enough to recognize the signature sequence and include it in the Astacin family, specifically among the Hatching Enzymes.

## 4.2 Choriolysin gene expression

The expression of a gene into development can be determined through its mRNA level, increasing or not, during a given period.

Choriolysin, as hatching enzyme, must be produced prior to hatching time, and the aim of this study was to understand exactly when and in which quantity it is expressed before this happening occurs. For this purpose I extracted the RNA from embryos at stages in early, middle and late development (sperm, non-fertilized eggs, 1day/degree, 200dd, 370dd, 550dd embryos and already hatched larvae).

Besides, I performed the same study on zebrafish embryos (stages 0 hours, 4h, 6h, 9h, 12h, 18h, 24h, 36h, 48h, 72h, 120h) to have a verification of my work. Choriolysin gene expression in this organism in fact has been larger characterized. It is known that zebrafish Hatching Enzyme 1 (the one of two more similar to salmon Choriolysin) starts to be expressed 45 minutes till 5 days after fertilization. It is reported that hatching process (breaking out of the eggshell) occurs after 52 hours post fertilization but the larvae maintain the hatching gland cells some days longer, in which time the hatching enzyme keeps being a little expressed.

Hatching enzymes are highly specific proteases and they might have another role in this organisms (as well as in the others), such as the degradation of the hatching gland cells. The death of those cells has been observed as apoptosis [24] and the hatching enzyme itself could be the protease that starts this process.

### 4.2.1 Amplification by PCR

The PCR amplification was performed with Ch(F/R)4 (expected product size at 413bp) and returned a specific product in all the stages I extracted the RNA from (non fertilized eggs, 1dd, 200dd, 370dd, 550dd embryos), except the hatched larvae (Fig. 3.5). Choriolysin is supposed to be used for hatching process, so it is clear why it is present in earlier stages but not in hatched larvae sample. Moreover, Fig. 3.5 gives an overview on how much this gene is expressed. The thickness of each band is related to the concentration of DNA inside the sample, which is related to the cDNA used for the PCR, which depends on the concentration of mRNA present into the extracted sample. Agarose gel shows that the amount of DNA is quite low in early stages, till 200dd, and then highly increases till 370dd, which means that the gene is at first only weakly transcribed and it increases getting closer to the hatching. Choriolysin needs as well time to be translated into its protein and this is reasonable why in the last stage, 550dd embryos, DNA (and so its RNA) looks less concentrated then 370dd (PCR amplification of 550dd stage, thinner band in Fig. 3.5).

I've called this 'overview' because the same experiment has been done in Real-Time PCR, which is much more specific. This tool comes up not only with a visible relation to the amount of expressed gene (if run sample in agarose gel), but also with concentration values of the starting cDNA amount of that gene. Normal PCR is instead reliable only to say if it is present, but not to define an expression assay.

A separated point concerns the sperm sample. A very weak band appeared in the agarose gel at the right size, 413bp, and I would be inclined to say that this is a false positive or non-specific signal. Unfortunately the concentration of DNA extracted from that band was not enough for sequencing and no more sample (cDNA, RNA or sperm itself) was available to complete this study. It would have been interesting to go through this because there would not be a reasonable motif for Choriolysin to be expressed in sperm.

Slightly different has been the amplification for salmon then for zebrafish. The same mechanism for starting apoptosis of hatching gland cells might be use but choriolysin seems to be expressed no longer then in zebrafish.

Zebrafish Hatching Enzyme 1 amplification was performed with primers HE1F1/HE1R2, which confirmed the knowledge of its presence from 45 minutes (present in 4 hours stage and not in 0 hours, Fig. 3.6) till 120 hours (Fig. 3.6). The amount looks very low until the 12 hours stage, where the amplicons seem to be more concentrated (stronger bands in the agarose gel). This is always related to its

role and the time it has to work. At the beginning of development (little post fertilization) the embryo does not need of any hatching enzyme. It might be low-level expressed in order to take advantage, but the most starts to be expressed about the middle of the embryo development stages, to have time to transcribe the gene and produce the protein, which is stored in granule until hatching time.

#### 4.2.2 Amplification by Real time PCR

Detect and quantify amounts of specific nucleic acids sequences was my aim to determine changes in gene expression as a result of development. This novel technique has been used to define specific amounts of Choriolysin in Atlantic salmon *Salmo salar* at several developmental stages, in order to understand its gene expression. The same study was performed for zebrafish *Danio rerio* Hatching Enzyme 1, a gene largely more characterized, which expression is known for every stage that it is supposed to be found (45 minutes to 5 days post fertilization).

If the reaction works properly, there will be twice as much specific dsDNA after each cycle of PCR. This means an exponential increase, which allows to calculate the initial concentration of an unknown sample. Actually, reactions do not maintain perfect efficiency because reactants within PCR are consumed after many cycles and the reaction will reach a plateau. Self-annealing of the accumulating product may also contribute to the plateau effect.

Using LightCycle 480 to perform a Real time PCR makes it possible to have two types of DNA quantitation:

- ~ Relative quantitation: levels of the gene of interest are related to an invariant control gene (housekeeping genes are not supposed to change during development). After PCR, the instrument returns a Cycle Time ( $C_T$ ) value, which occurs when fluorescence reaches a threshold level.  $C_T$  value is inversely proportional to the amount of a specific nucleotide sequence in the original sample. A difference of 1 between sample  $C_T$ s means that the sample with the lower  $C_T$  value had double the target sequence of the other sample: a change in  $C_T$  of 2 means a 4-fold of difference; changes of 3 means a 8-fold of difference. [ $\Delta C_T = 2^{-\Delta C_T}$  fold change].
- ~ Absolute quantitation: measures the actual nucleic acid copy number in a given sample. This requires a sample of known quantity of the gene of interest that can be diluted to generate a standard curve. This is an external “absolute”

standard. Unknown samples are compared with the standard curve for absolute quantitation. First challenge is to obtain an independent reliable standard.

Specificity of an amplified PCR product was assessed by performing a melting curve analysis on the instrument. The resulting melting curves allow discrimination between primer-dimers and specific product.

Dissociation of amplicons can be analyzed to determine the melting point. This is the temperature in which the dsDNA becomes ssDNA and it is unique of a specific length and composition in base pairs of a sequence. Smaller reaction volumes may result in melting temperature variations but usually one melting peak refers to one PCR product only, while more than one peak refers to more than one product. It suggests that the amplification was not specific for a single DNA target.

Real time PCR was performed with primers ChF4/ChR5 for salmon and HE1F1/HE1R1 for zebrafish. The standard curve was made by running a previous PCR with primers placed outside those I used this time, to make sure I have the sequence to amplify inside. The amplicons were diluted several times and with their known concentration I could build the standard curve.

The instrument measure the concentration of the sample through the florescence of SYBR Green I and entering those values in the standard curve it is possible to obtain the initial concentration of cDNA of Choriolysin.

Results quite reflect the ones obtained by usual PCR (Fig. 3.11). Salmon Choriolysin has low level in stages of non-fertilized eggs, 1dd and 200dd embryos, after which it increases a lot to end up at 370dd with the highest level of expression. Later in development, 550dd, this level lowers till the complete hatching of the larvae for a free life. Choriolysin is not supposed to be present in the larval stage but it showed up in Fig. 3.11. This might be due to a technical imprecision since I never had that result in other real-time or usual PCR I made before using other cDNA batches. Sample of sperm was not available, so the assay did not include it even though it would have given a negative control.

Zebrafish Hatching Enzyme 1 is instead a well known protease, expressed between 45 minutes and 5 days post fertilization in embryonic development. Absolute quantification through DNA levels and agarose gel after real-time PCR of all the stages confirm this point of view. In 0 hours in fact it is not present and from following stages it starts to increase until 36 hours. Hatching is supposed to occur at 48 hours and its level is expected reduced from that stage. Again the presence of it after hatching time (48 hours) is visible and probably related to another function of this protease. A bit unexpected is that in the last stage, 120

hours, the expression of Choriolysin increase. This may possibly be related to its potential role in hatching gland apoptosis.

The real-time PCR instrument also yielded the initial concentration of Choriolysin (values reported in 'Results' chapter) which reflects increases and decreases shown in agarose gels in Fig. 3.11 and Fig. 3.12.

Both salmon and zebrafish PCR products have been sequenced and they were found to match with the already known sequence.

### **4.3 Future works**

- a) Find the 5' end of *Salmo salar* Choriolysin gene sequence
- b) Make its recombinant and study different domain eventually present
- c) Analyze other sequences and amplify the gene from genomic DNA in order to find out if there are other copies and analyze the intron/exon structure
- d) Investigate whether choriolysin gene expression is present in sperm. This would provide a good negative control for the specificity of the probes used in this study
- e) Would be interesting also to see the variation of protein amount during embryo development, in comparison with its RNA.

## 5. Conclusions

Firstly, the purpose of this study was to define the Choriolysin sequence of *Salmo salar* Choriolysin gene and analyze its expression through embryo development.

This thesis presents the last 558 nucleotides, at the 3'end, which contains the signature sequence of the Astacin family HELLHALGFYHEHNRSDR, coding for the Zn-binding motif, and the SVMHY, coding for the Met-turn. Here it has been presented as a unique sequence (a sort of consensus sequence), which was not fully known in salmon.

Sequencing analysis anyway showed few indications that could suggest to the presence of more than one copy in the genome.

The sequence was amplified first from the nucleotide 229 to the nucleotide 709 with primers ChF3 (5' – CCA GGA ATG CCA TGA AGT GCT TTG – 3') and ChR3 (5' – AGG GAT GGG GGT TAT GGT CTC CT – 3') and then the 3'end was determined by RACE-PCR with primer ChF4 (5'– AAG TCA TCT GAC GGC TTG GTG TAC G – 3') and Oligo-dT.

The choriolysin gene was studied as to expression in embryo development (stages of non-fertilized eggs, 1dd embryo, 200dd, 370dd, 550dd and hatched larvae) was studied. Salmon Choriolysin was found to be expressed in a low level at the earliest stage, the non fertilized eggs, and it got higher between stages 200dd and 370dd. At 370dd it was in its highest expression and then it starts to lower until the time of hatch, where it is supposed to perform its work.

Expression of zebrafish Hatching Enzyme 1 was studied in stages of 0 hours post fertilization, 4 hours, 6 hours, 9 hours, 12 hours, 18 hours, 24 hours, 36 hours, 48 hours, 72 hours and 120 hours). Its RNA starts to be present in the second stage, 4 hours, but it seems to have a stepwise increase until hatching time. After that, it decreases. The reason for this finding has not been explained yet a further increase at 120 hours post fertilization, the last stage where previous works says that it should be expressed.



## References

- [1]. Bode W., Gomis-Ruth R.F., Huber R., Zwilling R., Stocker W. (1992). Structure of astacin and amplifications for activation of astacins and zinc-ligation of collagenases. *Nature*, 358:164-167.
- [2]. Bode W., Gomis-Ruth F.X., Stocker W. (1993). Astacin, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the “metzincins”. *FEBS Lett*, 331:134-140.
- [3]. Bond J.S., Beynon R.J. (1995). The astacin family of metalloendopeptidases. Protein science, Cambridge University Press, 4:1247-1261.
- [4]. Davis C. C. (1981). Mechanisms of hatching in aquatic invertebrate eggs. II. *Oceanogr. Mar. Biol. Ann. Rev.*, 19:95-123.
- [5]. Devenport J., Lonning S., Kjorsvik E. (1981). Osmotic and structural changes during early development of eggs and larvae of the cod, *Gadus morhua* L. *J. Fish Biol.*, 19:317-331.
- [6]. Dumermuth E., Sterchi E.E., Jiang W.P., Wolz R.L., Bond J.S., Flannery A.V., Beynon R.J. (1991). The astacin family of metalloendopeptidases. *J Biol Chem*, 266:21381-21385.
- [7]. Elaroussi M.A., De Luca H.F. (1994). A new member to the astacin family of metalloendopeptidases: a novel 1,25-dihydroxyvitamin D-3-stimulated mRNA from chorionallantoic membrane of quail. *Biochim Biophys Acta*, 1217:1-8.
- [8]. Groot E.P., Alderdice D.F. (1985). Fine structure of external egg membrane of five species of Pacific salmon and steelhead trout. *Can. J. Zool.*, 63:552-566.
- [9]. Hagenmaier H.E. (1973). The hatching process in fish embryos: the structure, polysaccharide and protein cytochemistry of the chorion of the trout egg, *Salmo gairdneri* (RICH). *Acta histochem.*, 47:61-69.
- [10]. Helvik J.V. (1991). Biology of hatching: mechanism and control of hatching in eggs of Halibut (*Hippoglossus hippoglossus*). Thesis submitted of the degree of Dr. Scient., University of Bergen.
- [11]. Hiroi J., Maruyama K., Kawazu K., Kaneko T., Ohtani-Kaneto R., Yasumasu S. (2004). Structure and development expression of hatching enzyme genes of the Japanese *Anguilla japonica*: an aspect of the evolution of fish hatching enzyme gene. *Dev. Genes Evol*, 214:176-184.
- [12]. Ishida J. (1944). Hatching enzymes in fresh-water siph, *Oryzias latipes*. *Annot. Zool. Jpn.* 22:137-154.

- [13]. Iuchi I., Yamamoto M., Yamagami K. (1982). Presence of Active Hatching Enzyme in the Secretory Granule of Prehatching Medaka Embryos. *Dev. Growth Differ.*, 24(2):135-143.
- [14]. Jiang W., Bond J.S. (1992). Families of metalloendopeptidases and their relationship. *FEBS Lett*, 312:110-114.
- [15]. Lee K.S., Yasumasu S., Nomura K., Iuchi I. (1994). HCE, a constituent of the hatching enzymes of *Oryzias latipes* embryos, releases unique proline-rich polypeptides from its natural substrate, the hardened chorion. *FEBS Lett*, 339:281-284.
- [16]. Lonning S., Kjorsvik E., Devenport J. (1984). The hardening process of the egg chorion of the cod, *Gadus morhua* L., and lumpsucker, *Cyclopterus lumpus* L.. *J. Fish Biol.*, 24:505-522.
- [17]. Needham J. (1931). "Chemical embryology". Cambridge Univ. Press, London and New York.
- [18]. Neurath H. (1989). The diversity of the proteolytic enzymes. In: Beynon RJ, Bond JS, eds. Proteolytic enzymes: A practical approach. Oxford, UK: IRL Press.
- [19]. Oppen-Berntsen D.O., Helvik J.V., Walther B.T. (1990). The major structural proteins of cod (*Gadus morhua*) eggshell and protein crosslinking during teleost egg hardening. *Dev. Biol.*, 137:258-265.
- [20]. Perona R.M., Wassarman P.M. (1986). Mouse blastocysts hatch in vitro by using a trypsin-like proteinase associated with cell of mural trophoctoderm. *Dev. Biol.*, 114:42-52.
- [21]. Rawlings N.D., Barrett A.J. (1993). Evolutionary families of peptidases. *Biochem J* 290:205-218.
- [22]. Schoots A.F.M., De Bont R.G., Van Eys G.J.J.M., Denucé J.M. (1982a). Evidence for a stimulating effect of prolactin on teleostean hatching enzyme secretion. *J. Exp. Zool.*, 219:129-132.
- [23]. Schoots A.F.M., Stikkelbroeck J.J.M., Bekhuis J.F., Denucé J.M. (1982b). Hatching in Teleostean Fishes: Fine structural changes in the egg envelope during enzymatic breakdown in vivo and in vitro. *J. Ultrastruct. Res.*, 80:185-196.
- [24]. Schoots A.F.M., Evertse P.A.C.M., Denucé J.M. (1983a). Ultrastructural changes in hatching-gland cells in pike-embryos (*Esox licius* L.) and evidence for their degeneration by apoptosis. *Cell Tissue Res.*, 229:573-589.
- [25]. Schoots A.F.M., Meijer R.C., Denucé J.M. (1983b). Dopaminergic regulation of hatching in fishes embryos. *Dev. Biol.*, 100, 59-63.
- [26]. Stocker W., Grams F., Baumann U., Reinemer P., Gomis-Ruth F.Z., McKey D.B., Bode W. (1995). The metzinins – Topological and sequential relation between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci*, 4, 823-840.

- [27]. Titani K., Torff H.J., Hormel S., Kumar S., Walsh K.A., Rodl J., Neurath H., Zwilling R. (1987). Amino acid sequence of a unique protease from crayfish *Astacus fluviatilis*. *Biochemistry*, 26, 222-226.
- [28]. Vogt G., Stocker W., Storch V., Zwilling R. (1989). Biosynthesis of *Astacus* protease, a digestive enzyme from crayfish. *Histochemistry*, 91, 373-381.
- [29]. Wallace R.A. (1985). Vitellogenesis and oocyte growth on nonmammalian vertebrates. I: Development biology, Vol. 1. L.W. Browder. Eds. Plenum Publishing, New York, pp. 127-177.
- [30]. Walther B.T., Andersen Ø., et al. (2001). Salmon Choriolysin Sequence. Internal document #11 in Aqua Bio Technology Ltd., N-5640 EIKELANDSOSEN, Norway.
- [31]. Willemse M.T.M., Denucé J.M. (1973). Hatching Glands in the Teleosts. *Brachydanio rerio*, *Danio malabaricus*, *Moenkhausia oligolepis* and *Barbus schuberti*. *Dev. Growth Differ.*, 15, 169-177.
- [32]. Wolpert L. (1998). Principles of development. Current Biology Ltd.
- [33]. Wourms J.P., Sheldon H. (1976). Annual fish oogenesis. II. Formation of the secondary egg envelope. *Dev. Biol.*, 50 355-366.
- [34]. Yamagami K. (1981). Mechanisms of hatching in fish: secretion of hatching enzyme and enzymatic choriolysis. *Amer. Zool.*, 21, 459-471.
- [35]. Yamagami K. (1988). "Mechanisms of hatching in fish". Academic Press.
- [36]. Yamamoto M. (1963). Electron microscopy of fish development. I. Fine structure of the hatching glands of embryos of the teleost, *Oryzias latipes*. *J. Fac. Sci., Univ. Tokyo, Sect. 4* 10, 115-121.
- [37]. Yamamoto M., Iuchi I., Yamagami K. (1979). Ultrastructural changes of the teleostean hatching gland cell during natural and induced precocious secretion. *Dev. Biol.* 68, 162-174.
- [38]. Yamamoto M., Yamagami K. (1975). Electron microscopic studies on choriolysis by the hatching enzyme of the teleost, *Oryzias latipes*. *J. Fac. Sci., Univ. Tokyo, Sect. 4* 10, 115-121.
- [39]. Yasumasu S., Katow S., Hamazaki T.S., Iuchi I., Yamagami K. (1992). Two constituent proteases of a teleostean hatching enzyme: Concurrent synthesis and packaging in the same secretory granules in discrete arrangement. *Dev. Biol.*, 149, 349-356.
- [40]. Yasumasu S., Iuchi I., Yamagami K. (1988). Isolation and some properties of Low Choriolytic Enzyme (LCE), a component of the Hatching Enzyme of the Teleost, *Oryzias latipes*. *J. Biochem.*, 105, 212-218.
- [41]. Yanay T. (1966). In: Hatching. – In: "Embryology of Vertebrates". Ed. M.Kume. eds. Baifukan, Tokyo., pp. 49-58.
- [42]. Yokoya S., Ebina Y. (1976). Hatching Glands in Salmon Fishes, *Salmo gairdneri*, *Salmo trutta*, *Salvelinus fontinalis* and *Salvelinus pluvius*. *Cell Tissue Res.*, 172, 529-540.



# Appendix

Reverse translation of *Salmo salar* Choriolysin protein  
(Cap. 3.2.2, pag. 41)

1	MetAspHisArgProThrLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeuGlyLeuSerGln ATGGATCATCGTCCTACTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTGGTCTTTCTCAA C CA C C CT CAGCT CT CT CT CT CT CT C CT CAGC G A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
21	AlaSerGlyAsnGluPheHisAspGluProAspHisValSerIleThrSerValIleLeu GCTTCTGGTAATGAATTTTCATGATGAACCTGATCATGTTTCTATTACTTCTGTTATCTT CAGC C C G C C C G C C C CAGC C CAGC C CT C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G
41	LysSerAsnAsnGlyThrAsnGluLeuLeuLeuAspGlyAspIleLeuAlaProArgThr AAATCTAATAATGGTACTAATGAACTTCTTCTTGATGGTGATATCTTGCTCCTCGTACT GAGC C C C C C GT CT CT C C C C CT C C CA C C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
61	ArgAsnAlaMetLysCysPheSerSerGlnTyrSerCysLeuTrpLysLysSerSerAsp CGTAATGCTATGAAATGTTTCTTCTCAATATTCTGTCTTTGGAAAAATCTTCTGAT A C C C G C CAGCAGC G CAGC CT C G GAGCAGC C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
81	GlyLeuValTyrValProTyrIleLeuSerAlaValTyrSerSerLeuGluValGluThr GGTCTTGTATGTTTCCTTATATTCTTTCTGCTGTTATCTTCTTCTTGAAGTTGAAACT CT C C C C C CT CAGC C C CAGCAGCT C G C G C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
101	IleGluThrAlaMetLysTyrPheGlnGlyLysThrCysIleArgPheIleProArgLys ATTGAAACTGCTATGAAATATTTTCAAGGTAAAACCTTGATTCGTTTATTCCTCGTAAA C G C C G C C G C G C C CA C C C CA C G A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
121	ThrGlnThrAlaTyrLeuAspIleGlnSerSerGlyGlyCysPheGlyThrValGlyThr ACTCAAACGCTTATCTTGATATTCAATCTTCTGGTGGTGGTTTTGGTACTGTTGGTACT C G C C CT C C C GAGCAGC C C C C C C C C C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
141	ValGlyAspArgGlnThrLeuSerLeuAlaGlnPheGlyCysValGlnHisGlyIleIle GTTGGTGATCGTCAAACCTTTCTTCTTGCTCAATTTGGTTGTGTTCAACATGGTATTATT C C CA C G CT CAGCT C C G C C C C G C C C C A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
161	GlnHisGluLeuLeuHisAlaLeuGlyPheTyrHisGluHisAsnArgSerAspArgGlu CAACATGAACTTCTTCATGCTCTTGGTTTTTATCATGAACATAATCGTTCTGATCGTGAA G C GT CT C C CT C C C C G C CA CAGC CA C G A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G
181	GlnTyrIleArgIleAsnTrpGlnTyrIleTyrAspTyrAlaValGlyAsnPheGlnLys CAATATATTCGTATTAATTTGGCAATATATTTATGATTATGCTGTTGGTAATTTTCAAAA G C CA C C C G C C C C C C C C C C G G A A A A A A A A A A A A A A A A G G G G G G G G G G G G G G G G

```

201  GluAspThrAsnAsnLeuHisThrAlaTyrAspTyrSerSerValMetHisTyrAspArg
      GAAGATACTAATAATCTTCATACTGCTTATGATTATCTTCTGTTATGCATTATGATCGT
      G C C C CT C C C C C C CAGCAGC C C C CA C
      A A A A A A A A A A
      G G G G G G G G G G

221  ThrAlaTyrThrAsnAspTyrGlyLysGluThrIleThrProIleProAspProSerVal
      ACTGCTTATACTAATGATTATGGTAAAGAACTATTACTCTATTCTGATCCTTCTGTT
      C C C C C C C C G G C C C C C C CAGC C
      A A A A A A A A A A A A A A A
      G G G G G G G G G G G G G G

241  AlaIleGlyGlnArgLeuGlyMetSerAspIleAspValLeuLysValAsnLysLeuTyr
      GCTATGGTCAACGCTTGGTATGCTGATATTGATGTTCTTAAAGTTAATAAACTTTAT
      C C C GA CT C C AGC C C C CT C G C C GT C C
      A A A A A A A A A A A A A A A
      G G G G G G G G G G G G G G

261  GlnCys
      CAATGT
      G C
  
```

\*\*\*\*\*

620bp band nucleotide consensus sequence.  
 (Cap. 3.2.4.2, pag. 47)

```

1  ATGTCTCATCAGCGCTCTTATCCCCGATAAACCTAGACCCACTCCAATTTGCATACCGCC
61  CCAACAGATCCACAGATGATGCAATCTCTATTGCACTCCACACTGCCCTTTCCACCTGG
121  ACAAAGGAACACTTATGTGAGAATGCTATTTCATTGACTACAGTTCAGCTTTCAACACCA
181  TAGTGCCCTCAAAGCTCATAACTAAAGTAAGGGGAGGGGACTAATGCTCCTTTTGCAA
241  CTGGATCCTGGACTCCTGGAAGGTCGCCAGGAGGTGAAAAAGAAGCTGGGATTCTACC
301  ACTAGCACAATCGTGAGACCGGGAGCCCTCTATGGGGTCTACTGGCACTCCTTCTGTG
361  ACTACGGCGATGGGAATTCTAGAAGGAGGACACGTTCTTCTGCACACTGAATATTAGT
421  ACTCCTCTGCAAAAAGAAAACGATAGAACCGCTTAGACTAACGACTACCGAAAGGAGACC
481  ATATCCCCATCCTTGACTGTGTTCCATAAAAAACAATATCCCTCAATGTAAAAAAA
541  AATAAGGAAGTTGATAGCACACTGCAGACAACCTAAGACCCTATCCTCTATATAAGACTTA
  
```

\*\*\*\*\*

620bp band amino acid consensus sequence.  
(Cap. 3.2.4.2, pag. 47)

1.	1	MSHQRSYPPIT-THSNLHTAPTDPQMMQSLHSTLFPPTWTKGTLM-ECYSLTTVQLSTP
	61	-CPQSS-LK-GGGD-LLLLQLDPGLLEGSPRR-KRSWDSTTSTNRETGSPLWGSTGTPSV
	121	TTAMGTSRRRTRSFCTLNISTPLQKKNDRTA-TNDYRKETISPILDCVP-KTTIPSM-KK
	181	NKEVDSTLQTTKTLSSI-DL
2.	1	CLISALIPR-PRPTPICIPPQQIHR-CNLYCTPHCPFPQKEHLCENAIH-LQFSFQHH
	61	SALKAHN-SKGEGTNCSFCNWILDSWKGRPGGEKEAGILPLAQIVRPGALYGLLALLL-
	121	LRRWELLEGGHVPSAH-ILVLLCKRKTIEPLRLTTTERRPYPPSLTVFHKKQLFPQCKKK
	181	IRKLIHCRQLRPYPLYKT
3.	1	VSSALLSPDNLDPLQFAYRPNRSTDDAISIALHTALSHLDRNTYVRMLFIDYSSAFNTI
	61	VPSKLITKVRGRGLIAPFATGSWTPGRVAQEVKKKLGFYH-HKS-DREPSMGVYWSFCDC
	121	YGDGNF-KEDTFLLHTEY-YSSAKEKR-NRLD-RLPKGDH I PHP-LCSIKNYSLNVKCK
	181	-GS--HTADN-DPILYIRL

Three amino acid sequences are reported due to the unknown starting point. Each of the first three nucleotides could be the first of the codon.

\*\*\*\*\*

3' RACE-PCR, consensus sequence of 680bp band. Nucleotide sequence.  
(Cap. 3.2.6, pag. 48)

1	CAGCGCTGTA	TATTCAGCTT	GGAGGTAGAG	ACTATTGAGA	CGGCCATGAA	GTA CTTCCAA
61	GGCAAGACCT	GCATCCGCTT	CATTCACAGT	AAGACACAGA	CTGCCTACCT	GGACATTCAG
121	AGCAGCGGCG	GGTGTTTGG	TACCGTGGGG	ACTGTTGGGG	ACAGGCAGAC	ATTGTCTCTT
181	GACAGTTTG	GCTGTGTTCA	ACATGGTATC	ATCCAGCATG	AGCTGCTTCA	CGCCCTGGGC
241	TTCTACCACG	AGCACAACAG	GAGTGACCGT	GAACAGTATA	TCAGGATCAA	CTGGCAATAC
301	ATCTATGACT	ACGCCGTTGG	GAACCTCCAG	AAGGAGGACA	CCAACAACCT	GCACACTGCA
361	TACGACTACT	CCTCTGTCAT	GCACTATGAT	AGAACCGCTT	ACACTAACGA	CTACGGAAAG
421	GAAACCATCA	CTCCCATCCC	AGACCCATCT	GTGGCCATTG	GACAGAGACT	GGGCATGTCC
481	GACATTGATG	TCCTGAAGGT	CAACAAGCTC	TACCAATGCT	AAGAGGAAGA	GCGCCATTGT
541	TGAAAATGTG	TGATGCTGGA	TGTGCTGTCA	TGTGCTGATG	TATTTTATTG	TTGGAAGTTT
601	GTATGTATCC	TTTTAATCAC	ATTGGTAATA	ATAAAGCATG	GTTATGGTAA	AAAAAACCG
661	CGAGTGTGGT	GGTACCCGC	GTTGTGACCG	CTACACTTGC	CAGCAGCTTA	G

\*\*\*\*\*

3' RACE-PCR, consensus sequence of 680bp band. Amino acid sequence.  
(Cap. 3.2.6, pag. 48)

1	QRCIFSLEVE	TIETAMKYFQ	GKTCIRFIPR	KTQTAYLDIQ	SSGGCFGTVG	TVGDRQTL	SL
61	AQFGCVQHGI	IHELLHALG	FYHEHNRSDR	EQYIRINWQY	IYDYAVGNFQ	KEDTNLHTA	
121	YDYSSVMHYD	RTAYTNDYGK	ETITPIPDPS	VAIGQRLGMS	DIDVLKVNKL	YQC-EEERHC	
181	-KCVMLDVLS	CADVFYCWKF	VCILLITLVI	IKHGYGKKKR	RVWWLPAL-P	LHLPAA-	

\*\*\*\*\*

ClustalW comparison of *Salmo salar* nucleotide sequence before and after RACE-PCR.

(Cap. 3.2.6, pag. 49)

before	AAGTCATCTGACGGCTTGGTGTACGTGCCTTACATCCTCAGCGCTGTATATTCCAGCTTG
after	-----CTCAGCGCTGTATATTC-AGCTTG *****
before	GAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGGAAGACCTGCATCCGCTTC
after	GAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGCAAGACCTGCATCCGCTTC *****
before	ATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTTTTGGT
after	ATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTTTTGGT *****
before	ACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGCACAGTTTGGCTGTGTTCAA
after	ACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGCACAGTTTGGCTGTGTTCAA *****
before	CATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAAACAGG
after	CATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAAACAGG *****
before	AGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGTTGGG
after	AGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGTTGGG *****
before	AACTTCCAGAAGGAGGACACCAACAACCTGCACACTGCATACGACTACTCCTCTGTCATG
after	AACTTCCAGAAGGAGGACACCAACAACCTGCACACTGCATACGACTACTCCTCTGTCATG *****
before	CACTATGATAGAACAGCTTACACTAACGACTACGAAAGGAGACCATAACCCCATCCCT
after	CACTATGATAGAACCGCTTACACTAACGACTACGAAAGGAAACCATCACTCCCATCCCA ***** ** *****
before	-----
after	GACCCATCTGTGGCCATTGGACAGAGACTGGGCATGTCCGACATTGATGTCTGAAAGTCT

```

before -----
after  AACAAAGCTCTACCAATGCTAAGAGGAAGAGCGCCATTGTTGAAAATGTGTGATGCTGGAT

before -----
after  GTGCTGTCATGTGCTGATGTATTTTATTGTTGGAAGTTTGTATGTATCCTTTTAATCACA

before -----
after  TTGGTAATAATAAAGCATGGTTATGGTAAAAAAAACGGCGAGTGTGGTGGTTACCCGCG

before -----
after  TTGTGACCGCTACACTTGCCAGCAGCTTAG

```

\*\*\*\*\*

Microsatellite sequence found after reverse translation of *Salmo salar* Choriolysin protein sequence.

(Cap. 4.4.1, pag. 55)

```

1  ctggaattgc cctcggaaact cagaaccaca tcttacatgc aatgtcaagt
61  ctgtcaggag agtctagagt ttaagtaaac aagtcattgg gtcacagagg
121 aaataattgt cttgagtcac ttctttgacg ttcaccaaca gtaatcagat
181 gctgccagag agttggtacg ggataaatat gaatattggg tgcgcaaaaag
241 agaaggagat gggagagatg agtgaaatag atgaataaga gaaagggaga
301 atatgaatga gtgagttgca ctgtgttcca gggtagagag agtgagtgag
361 tggggacgag agggaagtgc caacacatgt gctagtcatt taggaggcag
421 tttgctttac atctgttttc tgatctttca actaccccct ccttttctct
481 ctttttctct ctctatcttg ctctcgcttc ccctgaccct gctctttctc
541 ttgttcatca tgtttcccg c atgggacaca cacacacaca cacacacaca
601 cacacacaca cacacacaca cacacacaca cacacacaca cactggtagg
661 gttctctatg gaatgcagac atatcatccc cagtgatgta tcaaggatat
721 acattaccaa ctgcagcgtg cagcaatacg ccacaaatac atgtctgtct
781 caggtattct ggaatggaat gtgtcctgtg cgcaggtatt ctggaatgga
841 atgtgtccat gtctcaggta ttctggaatg aaatgtgtcc ctgtctcagg
901 tattctggaa tggaatgtgt ccctgtctca ggtattctgg aatggaatth
961 gtccatatct cagaatgcac atgtgtgttt agggccctct gggtatatac
1021 tctacattaa tcagtgtgcc atatttctct gtgtcgctgt gcggttggag
1081 taggacaacc atgtctatta gtcagcaaac tctcctgagc agacctatac
1141 tgccaatgct gcacatacct ccctcnag

```

\*\*\*\*\*

ClustalW comparison of *Salmo salar* Choriolysin nucleotide sequence, obtained after PCR amplification with primers Ch(F/R)3.  
(Cap. 4.1.3, pag. 57)

```

1F  --AACCNCNGNAANN---NGTGTACGTGCCTTACATCCTCA-G-CGCTGTATATTCCAG
2F  -TAACCANGNNANG---GNNGTANNTGCCTTACATCCTCA-G-CGCTGTATATTCCAG
3F  NNNNNCNACNNANGNA--GGTGTACGTGCCTTACATCCTCA-G-CGCTGTATATTCCAG
4F  NNNNCNCGNAANN---NGG-GTACGTGCCTTACATCCTCA-G-CGCTGTATATTCCAG
1R  -----CTGTATATTCCAG
2R  -NNAAGTATCTGACGGCTTGGTGTACGTGCCTTACATCACAAAGACGCTGTATATTCCAG
3R  -----GTGCCTTACATCAATCAGAACTGTATATTCCAG
                                     *****

1F  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGCAAGACCTGCATCCG
2F  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGGAAGACCTGCATCCG
3F  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGGAAGACCTGCATCCG
4F  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGGAAGACCTGCATCCG
1R  CTTGGAGGTAGAGGCTATTGAGACGGCCATGAAGTACTTCCAAGGCAAGACCTGCATCCG
2R  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGCAAGACCTGCATCCG
3R  CTTGGAGGTAGAGACTATTGAGACGGCCATGAAGTACTTCCAAGGCAAGACCTGCATCCG
*****

1F  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
2F  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
3F  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
4F  CTTCTCCACGTAAGACAAAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
1R  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
2R  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
3R  CTTCATTCCACGTAAGACACAGACTGCCTACCTGGACATTCAGAGCAGCGGCGGGTGTFTT
****

1F  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
2F  TGGTACCGTGGGGACTGTTGGGGACAGGGAGACATTGTCTCTTGACAGTTTGGCTGTGT
3F  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
4F  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
1R  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
2R  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
3R  TGGTACCGTGGGGACTGTTGGGGACAGGCAGACATTGTCTCTTGACAGTTTGGCTGTGT
*****

1F  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
2F  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
3F  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
4F  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
1R  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
2R  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
3R  TCAACATGGTATCATCCAGCATGAGCTGCTTCACGCCCTGGGCTTCTACCACGAGCACAA
*****

1F  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
2F  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
3F  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
4F  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
1R  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
2R  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
3R  CAGGAGTGACCGTGAACAGTATATCAGGATCAACTGGCAATACATCTATGACTACGCCGT
*****

```

