

CHARACTERIZATION AND QUANTIFICATION OF  
ATLANTIC SALMON, RAINBOW TROUT AND  
ATLANTIC COD VITELLOGENIN APPLYING  
COMBINED MASS SPECTROMETRIC TECHNIQUES

ALEJANDRO MARTIN COHEN







CHARACTERIZATION AND QUANTIFICATION OF ATLANTIC SALMON,  
RAINBOW TROUT AND ATLANTIC COD VITELLOGENIN APPLYING  
COMBINED MASS SPECTROMETRIC TECHNIQUES

PhD THESIS

by

Alejandro Martin Cohen

A thesis submitted to the  
School of Graduate Studies  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
Department of Biochemistry  
Memorial University of Newfoundland

May 2006

St. John's, Newfoundland. Canada



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-30425-9*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-30425-9*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## **Abstract**

Vitellogenin is a complex phosphoglycolipoprotein that is secreted into the bloodstream of sexually mature, female, oviparous animals in response to circulating estrogens. It is then incorporated into the ovaries by receptor mediated endocytosis, where it is further cleaved into smaller peptides which are constituents of the egg yolk proteins. It is generally accepted that these peptides serve as the main nutritional reserve for the developing embryo.

Quantification of vitellogenin in blood is useful for different purposes. The reproductive status and degree of sexual maturation of oviparous animals can be assessed according to the levels of vitellogenin in plasma. The expression of this protein can also be induced in males under the effect of estrogenic compounds. Relying on this observation, some studies have used vitellogenin as a biomarker of environmental endocrine disruption in many species.

The objective in this work was to develop a novel quantitative technique for measuring plasma levels of Atlantic salmon and rainbow trout vitellogenin using high performance liquid chromatography coupled to tandem mass spectrometry. These two species are of vital importance for the sustainable commercial fisheries and aquaculture activities in Atlantic Canada.

To achieve this goal, the first step of this project consisted of a comprehensive characterization and partial sequencing of Atlantic salmon and rainbow trout vitellogenin. The initial approach employed both chemical and enzymatic digestions of vitellogenin coupled with tandem mass spectrometry, one of the most powerful techniques available for sequencing, identification and characterization of proteins. The information from these studies was then used to select a 'signature peptide' characteristic for both species. This peptide was used as a surrogate of the precursor protein for quantification purposes. A detailed account is given on all the technical features of the quantitative technique. Special attention was placed on the tuning and testing of this system, focusing on the recovery, specificity and sensitivity of the method. Additional characterization and

sequencing was also performed for Atlantic cod vitellogenin. The results obtained in this work allow this method to be considered as an alternative to existing immunological assays.

## **Acknowledgements**

Traveling from Argentina with my wife and daughter, and leaving behind friends and family was made so easy and simple thanks to Dr. Joseph Banoub, my supervisor, and his family who made our arrival to Newfoundland so warm and friendly. Of course, his guidance and support was essential for me to complete my PhD program without major inconveniences.

Our stay in St. John's was possible thanks to the monetary support of Dr. Banoub's NSERC grant and Memorial University's School of Graduate Studies Fellowship awarded through the Department of Biochemistry. Research costs were also covered by Dr. Banoub's NSERC grant.

Additionally, I would like to mention the contribution of Dr. Atef Mansour, Research Scientist at Fisheries and Oceans' Northwest Atlantic Fisheries Center (NAFC) to my research, who provided the funds necessary for some laboratory equipment and took good care of the experimental fish in the tank room. His experience in veterinary was vital to keep the fish in good health during the long experimental protocols.

I would also like to thank my Supervisor Committee, Dr. David Heeley and Dr. Phil Davis, who have always been collaborative and willing to discuss some of the scientific issues related to my research.

Much of the initial work performed on rainbow trout vitellogenin was done with some semi purified samples obtained in Dr. Heeley's laboratory. Similarly, I appreciate the help of Donna Jackman, a technician in Dr. Heeley's lab, who was involved in the purification of those vitellogenin samples and helped me, especially at the beginning of my program, with some laboratory techniques like the gel electrophoresis experiments.

Some data analyzed by Qq-ToF MS presented in this thesis relies on the samples run in Dr. Pierre Thibault's laboratory. Thanks to his effort, I was able to start analyzing data that was obtained before Dr. Banoub acquired his own Qq-ToF mass spectrometer. The ELISA kit for rainbow trout vitellogenin was generously provided free of charge by Dr. Anders Goksoyr, from Biosense™ Laboratories.

During my PhD program, I attended two conferences organized by the American Society of Mass Spectrometry, held in Montreal and San Antonio, Texas, in 2003 and 2005, respectively. These trips were made possible thanks to the financial support given by my supervisor, the School of Graduate Studies, the Graduate Student Union and the Dean of Science.

A positive working environment plays a critical role on the student's life. For this I have to thank the staff and researchers at the NAFC, where I spent most of the time performing my experiments. Namely, George Sheppard for tutoring me with most technical issues regarding the mass spectrometers, Howard Hodder who always was around whenever I needed some help, especially giving me writing advice for my reports and papers, Anas El-Aneed also graduate student under the supervision of Dr. Banoub, Jean Hart for supplying me with much of the lab material required for my lab experiments, Janice Lannon and Marlene Carroll, two secretaries that made complicated paperwork look easy, Paula Hawkins and Nadine Templeman for the daily company during the long hours work (and the delicious ginger cookies). All these people made work both enjoyable and fun.

Although based at the NAFC, I frequently found myself visiting Memorial University's campus for both academic and non-academic purposes. In campus, I should be always appreciative to the help of Anne Sinott, Christine Squire and Betty Anne Lewis, all of them secretaries at the Biochemistry Department General Office, who have helped me in numerous occasions, always with a big smile on their faces.

Many other people made our stay in St. John's feel as though we were back in our homeland. The list of names is too long to enumerate, so I will group them: the St. John's International Football Team, the Biochemistry Department graduate students and the bunch of Latin St. John's residents, thank you for your daily company.

Finally, my admiration goes to my friend and wife, Pia Elustondo, and my daughter, Agustina Cohen, who have managed to cope with me and made my life meaningful during these marvelous years in St. John's, especially in the long, dark and cold months of winter.

## Table of Contents

Abstract	ii
Acknowledgements	iv
List of Tables	x
List of Figures	xii
List of Abbreviations and Symbols	xvi
List of Appendices	xvii
Chapter 1: Vitellogenin and the egg yolk proteins. Introduction and overview	1
1.1 The egg yolk proteins	2
1.2 Vitellogenesis	6
1.3 Vitellogenin size and structure	14
1.4 Vitellogenin as a sexual maturation indicator	24
1.5 Vitellogenin as a biomarker	26
1.6 Quantification of vitellogenin	35
1.7 Purpose and significance of study	42
Chapter 2: Application of mass spectrometry to protein chemistry	45
2.1 The mass spectrometer	45
2.2 Ionization sources for macromolecules	48
2.2.1 Electrospray ionization	48
2.2.2 Matrix-assisted laser desorption/ionisation	49

2.3	Tandem mass spectrometry	51
2.4	Mass spectrometry in proteomics	53
Chapter 3: Characterization of vitellogenin by mass spectrometry		57
3.1	Materials and methods	57
3.1.1	Vitellogenin induction protocol	57
3.1.2	Atlantic salmon and rainbow trout vitellogenin purification protocol	59
3.1.3	Solution digestion procedures	60
3.1.4	In-gel digestion procedures	61
3.1.5	HPLC-ESI-MS and MS/MS acquisition parameters	63
3.1.6	MALDI-MS and MS/MS acquisition parameters	64
3.1.7	Bioinformatics tools	66
3.1.7.1	Molecular weight calculation and isotopic cluster distribution	66
3.1.7.2	Protein digestion simulation	67
3.1.7.3	Peptide mass fingerprint database search	67
3.1.7.4	' <i>De novo</i> ' sequencing of tryptic peptides	68
3.1.7.5	Sequence alignment of 'de novo' sequenced peptides	69
3.2	ESI-MS and MALDI-MS of intact rainbow trout and Atlantic salmon vitellogenin	70
3.3	Digestion analyses of purified vitellogenin	75
3.3.1	MALDI-MS of the CNBr and trypsin derived peptides of rainbow trout vitellogenin	75
3.3.2	ESI-MS of the trypsin derived peptides of rainbow trout vitellogenin	83
3.3.3	ESI-MS/MS of the trypsin derived peptides of rainbow trout vitellogenin	86
3.3.4	MALDI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin	88

3.3.5 ESI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin	91
3.3.6 ESI-MS/MS of the trypsin derived peptides of Atlantic salmon vitellogenin	93
3.4 In-gel digestion analyses of vitellogenin	102
3.4.1 MALDI-MS of the trypsin derived peptides of rainbow trout vitellogenin	102
3.4.2 MALDI-MS/MS of the trypsin derived peptides of rainbow trout vitellogenin	105
3.4.3 MALDI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin	107
3.4.4 MALDI-MS/MS of the trypsin derived peptides of Atlantic salmon vitellogenin	111
3.4.5 MALDI-MS of the trypsin derived peptides of Atlantic cod vitellogenin	114
3.4.6 MALDI-MS/MS of the trypsin derived peptides of Atlantic cod vitellogenin	117
 Chapter 4: Development of a quantitative assay for vitellogenin using tandem mass spectrometry	 124
4.1 Quantitative methods in proteomics	124
4.2 The 'signature peptide' approach	127
4.3 Materials and methods	130
4.3.1 Pre-analytical sample preparation	130
4.3.2 Selection of the 'signature peptide'	136
4.3.3 Tuning and optimization of the HPLC-MS/MS acquisition parameters	138
4.4 Sample analysis	159
4.5 Recovery studies	166
4.6 Vitellogenin stability studies	172

Chapter 5: Discussion and conclusions	174
5.1 Similarities between Atlantic salmon and rainbow trout vitellogenin	174
5.2 The ‘signature peptide’ approach: The advantages and disadvantages	176
5.3 ELISA vs the ‘signature peptide’ approach: Which is better?	182
5.4 Atlantic cod vitellogenin sequence: Similarities to haddock and its implications	186
5.5 Application of the ‘signature peptide’ approach to future studies	188
5.6 Summary and recommendations	190
Reference List	191

## List of Tables

Table 3.1	Peptide mass fingerprint of the CNBr chemical degradation of RT Vtg	79
Table 3.2	Peptide mass fingerprint of the trypsin digestion of RT Vtg	80
Table 3.3	Optimization of peptide mass fingerprinting matching parameters	81
Table 3.4	Putative glycan structures found in RT Vtg	82
Table 3.5	Isobaric peptides found in the tryptic digestion of RT Vtg	84
Table 3.6	MALDI and HPLC-ESI-MS matching tryptic peptides of RT Vtg	85
Table 3.7	' <i>De novo</i> ' sequencing and sequence homology of RT Vtg tryptic peptides	87
Table 3.8	Peptide mass fingerprint of the trypsin digestion of AS Vtg	90
Table 3.9	Separation of AS Vtg tryptic peptides by HPLC-ESI-MS	92
Table 3.10	' <i>De novo</i> ' candidate sequences of the AS Vtg tryptic digestion	94
Table 3.11	' <i>De novo</i> ' sequencing and sequence homology of AS Vtg tryptic peptides	97
Table 3.12	Peptide mass fingerprint of the in-gel digestion of RT Vtg	104

Table 3.13	<i>'De novo'</i> sequencing and sequence homology of RT Vtg tryptic peptides obtained by in-gel digestion	106
Table 3.14	Peptide mass fingerprint of the in-gel digestion of AS Vtg	110
Table 3.15	<i>'De novo'</i> sequencing and sequence homology of AS Vtg tryptic peptides obtained by in-gel digestion	112
Table 3.16	Peptide mass fingerprint of the in-gel digestion of AC Vtg	118
Table 3.17	<i>'De novo'</i> sequencing and sequence homology of AC Vtg tryptic peptides obtained by in-gel digestion	120
Table 4.1	Solvent elution profile and flow rates for the chromatographic separation	132
Table 4.2	Results obtained for the plasma samples analyzed by the 'signature peptide' approach	165
Table 4.3	Statistical analysis of recovery studies	167
Table 4.4	Candidate peptides for the incomplete digestion of RT Vtg	169
Table 4.5	Effect of freeze-thaw cycles on final concentrations	173
Table 5.1	Modifications that could increase sensitivity and limit of detection of the 'signature peptide' approach	181
Table 5.2	Comparison of Vtg concentrations obtained from ELISA and HPLC-MS/MS using the 'signature peptide' approach	183

## List of Figures

Fig. 1.1	Scheme of the hypothalamic-pituitary-gonadal-liver axis	11
Fig. 1.2	Scheme of the global experimental design	44
Fig. 2.1	Basic principles of the electrospray ionization source	50
Fig. 2.2	Basic principles of the MALDI ionization source	50
Fig. 2.3	Basic components of the tandem mass spectrometer	52
Fig. 2.4	Collision induced dissociation of a peptide ion precursor	55
Fig. 2.5	Classical ‘bottom-up’ proteomic approach for identification of proteins	56
Fig. 3.1	Relative molecular mass and isotopic distribution predicted for RT Vtg	72
Fig. 3.2	MALDI-MS spectra of intact Vtg	73
Fig. 3.3	High mass range MALDI-MS acquisition of intact Vtg	74
Fig. 3.4	MALDI-MS spectra of the CNBr cleavage of purified RT Vtg	76
Fig. 3.5	MALDI-MS spectra of the trypsin digestion of purified RT Vtg	77
Fig. 3.6	Peptide map obtained for RT Vtg	82

Fig. 3.7	MALDI-MS spectra of the tryptic digestion of purified AS Vtg	89
Fig. 3.8	Graphical interpretation of CID-MS/MS spectra.	96
Fig. 3.9	Ambiguous graphical interpretation of CID-MS/MS spectra	101
Fig. 3.10	MALDI-MS spectra of the in-gel tryptic digestion of RT Vtg	103
Fig. 3.11	MALDI-MS spectra of the in-gel tryptic digestion of AS Vtg	108
Fig. 3.12	Induction of Vtg synthesis in AC fish	115
Fig. 3.13	MALDI-MS of the in-gel tryptic digestion of AC Vtg	116
Fig. 3.14	Product ion spectrum of the $[M+H]^+$ ion at $m/z$ 1089.54	121
Fig. 3.15	Product ion spectrum of the $[M+H]^+$ ion at $m/z$ 1106.58	121
Fig. 3.16	Product ion spectra of $[M+H]^+$ precursor ions at $m/z$ 942.49, 1009.53 and 1124.13	123
Fig. 4.1	Work scheme for the 'signature peptide' approach	129
Fig. 4.2	HPLC-MS/MS operated in the 'pseudo' SRM mode	135
Fig. 4.3	Full scan spectrum of the 'signature peptide' standard	140
Fig. 4.4	Product ion spectra of the 'signature peptide'	141
Fig. 4.5	Schematic diagram of an ESI-QqToF tandem mass spectrometer	143

Fig. 4.6	Optimization of the ionization energy, the nebulization gas and the auxiliary gas	145
Fig. 4.7	Optimization of the declustering potential (1 and 2) and the focusing potential	146
Fig. 4.8	Product ion spectra of the ‘signature peptide’ at varying collision energies	148
Fig. 4.9	Product ion spectra of the ‘signature peptide’ at varying collision energies	149
Fig. 4.10	Optimization of the collision energy for ‘pseudo’ SRM-MS/MS experiments	150
Fig. 4.11	Calibration of the ‘signature peptide’ by direct ‘loop’ infusion	151
Fig. 4.12	Full scan spectra of samples obtained by direct ‘loop’ infusion	152
Fig. 4.13	Analysis of a digested RT plasma by HPLC-MS and MS/MS	155
Fig. 4.14	Optimization of mass ranges for extracted ion chromatograms	157
Fig. 4.15	Full scan spectrum of an infused labeled ‘signature peptide’ standard	159
Fig. 4.16	Calibration of the ‘pseudo’ SRM-MS/MS experiments	161
Fig. 4.17	Calibration of the ‘pseudo’ SRM-MS/MS experiments	162

Fig. 4.18	Detection of the 'signature peptide' by HPLC-MS/MS	164
Fig. 4.19	Candidate peaks for incomplete digestion peptides of Vtg	170
Fig. 4.20	'Signature peptide' release during incubation with trypsin	171
Fig. 5.1	Sequence homology between RT and AS Vtg	175
Fig. 5.2	Comparison of the 'signature peptide' approach to classical immunoassay method development	178
Fig. 5.3	Comparative plots for samples analyzed by the 'signature peptide' approach and ELISA	184

## List of Abbreviations and Symbols

$[M+nH]^{+n}$	n protonated-n charged molecular species
AC	Atlantic cod ( <i>Gadus morhua</i> )
AS	Atlantic salmon ( <i>Salmo salar</i> )
BLAST	Basic local alignment search tool
CE	Collision energy
CID	Collision induced dissociation
Da	Dalton
DDT	Dichloro-diphenyl-trichloroethane
DNA	Deoxyribonucleic acid
DP	Declustering potential
DTT	Dithiothreitol
EIC	Extracted ion chromatogram
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
FP	Focusing potential
GtH	Pituitary gonadotropin
HPLC	High performance liquid chromatography
IDA	Information dependent acquisition
IE	Ionization energy
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LOD	Limit of detection
LOQ	Limit of quantification
<i>m/z</i>	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionisation
Mr	Relative molecular mass (molecular weight)
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
PCR	Polymerase chain reaction
PTM	Post-translational modification
QqQ	Triple Quadrupole analyzer
QqToF	Quadrupole/quadrupole-time of flight analyzer
RF	Radio frequency
RIA	Radioimmunoassay
RNA	Ribonucleic acid
RT	Rainbow trout ( <i>Oncorhynchus mykiss</i> )
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SRM	Selected reaction monitoring
TCDD	2,3,7,8-tetrachloro-dibenzo-p-dioxin
TIC	Total ion current
ToF	Time of flight analyzer
Vtg	Vitellogenin

## **List of Appendices**

Appendix 1. Single and three letter codes for the protein amino acids	216
Appendix 2. Amino acid sequences of Atlantic salmon and rainbow trout vitellogenin	217
Appendix 3. ELISA test for rainbow trout and Atlantic salmon vitellogenin	219

## **Chapter 1: Vitellogenin and the egg yolk proteins. Introduction and overview**

The work presented in this thesis will deal mainly with the nature of vitellogenin (Vtg), a complex phosphoglycolipoprotein, which is synthesized in the liver of all egg laying oviparous animals. As expected with any protein, structural and even functional differences are likely to be found among different species. In this particular work, a special focus was placed on rainbow trout (*Oncorhynchus mykiss*), Atlantic salmon (*Salmo salar*) and Atlantic cod (*Gadus morhua*) Vtg, although other species are occasionally mentioned whenever comparisons and analogies were considered relevant.

In this chapter, an introduction to the egg yolk proteins is given for the purpose of placing the reader into the historical context that led to the discovery of Vtg. The subsequent sections describe the biochemical aspects of Vtg synthesis, together with a detailed description of its structural characteristics.

Since its discovery, there has been a renewed interest in this protein, especially in the past decade when its use as biomarker was discovered. Moreover, in recent years, the use of this protein for assessing the sexual reproductive status of fish has been analyzed. Both of these issues are also discussed in this chapter.

A critical element in the nature of this work involves the quantification of Vtg. Throughout the years, different techniques have been developed in pursuit of obtaining reliable quantitative values of this protein in blood plasma. The advantages and drawbacks inherent to many of these techniques are described in section 1.6. Once a broad idea of the topic has been introduced, the purpose and significance of this work is presented in the last section of this chapter.

Finally, many of the Figures and Tables described in thesis deal with protein and peptide amino acid sequences. The single letter amino acid code was usually used to express these sequences. Appendix 1 contains the complete protein amino acid list, together with the single and three letter amino acid codes.

### **1.1 The egg yolk proteins**

The definition of the word ‘yolk’ according to the Third Edition of the Oxford Dictionary of Current English is:

Yolk: the yellow inner part of a bird’s egg, which is rich in protein and fat and nourishes the developing embryo.

This definition could easily be extended not only to birds, but in general to all oviparous animals. A microscopic look into the yolk unveils a granular constitution. These granules are large, membrane-bound, quasi-crystalline storage vesicles formed from the accretion of smaller pinocytic vesicles. As stated in the definition, the components of these granules are mainly proteins and lipids. The accumulation of this material and the mobilization of its constituents are vital requirements for the successful development of the embryo.

It is important to remember that the eggs are essentially closed systems. This means that during embryogenesis, the yolk granules will be the only source of nutrients

and provide the raw materials necessary for the developing embryo, until the hatchlings are free to seek their own food.

It should be noted that not all species produce 'yolky' eggs, as for example in the sea urchin. However, even in these species, small amounts of yolk-like proteins are present as reserve material. In these cases, eggs undergo quick development and hatch at an early stage. Much of the research on yolk proteins has been done on insects and a detailed description can be found in published reviews (Kunkel and Pan, 1976; Kunkel and Nordin, 1986).

Vitellin is the generic name given to the protein constituents of the yolk granules. The origin of these proteins was a puzzle that took a couple of decades to be solved and the development of chromatography, electrophoresis and immunological assays played a critical role in unveiling this mystery.

It is now widely accepted that the major precursors to the egg yolk proteins are the Vtg proteins synthesized in the liver of vertebrates and in the fat body cells of non-vertebrates such as insects. The accumulation of Vtg in the blood of ovariectomized insects was a strong evidence that supported this idea in early Vtg studies (Telfer and Telfer, 1954). However, numerous studies in insects have also demonstrated that proteins similar to the Vtgs are synthesized in a number of ovarian cell types (Srdic *et al.*, 1979; Jowett and Postlethwait, 1980; Brennan *et al.*, 1982).

The term vitellogenin was first coined in Pan *et al.* (1969), who refer to Vtg as the female specific proteins found in the blood of mature female insects. In this work, the female fat body of two insect species incorporated radio-labeled amino acids *in vitro* into substances which were precipitated by antibodies formed in response to the vitellogenic

blood proteins. This was not observed in the fat bodies of males or even females before the appearance of the Vtgs in blood (Pan *et al.*, 1969).

It is not surprising that much of the initial research on the composition and function of the yolk granules and its precursors was done in birds and amphibians, probably due to the large amounts and ease of accessibility to the studying material. This work was carried out at the beginning of the 20<sup>th</sup> century. In this regard, the first clues indicating a relation between the sexual maturation and the appearance of changes in the serum profiles were obtained by experiments which demonstrated that calcium and phosphorus levels were increased in female fish during the egg-laying season (Miescher, 1897; Hess *et al.*, 1928; Laskowski, 1935). Similar studies were demonstrated in frogs (Zwarenstein and Shapiro, 1933) and birds (Riddle and Reinhart, 1926; Hughes *et al.*, 1927; Buckner *et al.*, 1930; Charles and Hogben, 1933). These changes in the blood components would seem obvious nowadays considering the high concentration of both elements present in the egg shell.

Subsequent studies demonstrated that the same changes could be induced by the administration of estrogens (McDonald and Riddle, 1945). It was not until 1957, that changes in the serum protein concentrations were unexpectedly observed, while studying the effects of estradiol on calcium metabolism in goldfish (Bailey, 1957). This study showed that estradiol increased the concentration of calcium and phosphorus only in the non-ultrafilterable fraction, that is, the fractions bound to protein; changes which were the same as those found during the breeding season. However, the most remarkable observation was the concomitant increase in the total protein concentration in serum. At this time, it was hypothesized that most of the increase was due to an increase in

'vitellin', a protein believed to be synthesized in the liver and further deposited in the ovaries. This was the first evidence to support the idea that this particular protein was induced by the action of estradiol. Of course, at this time, the exact role and biochemistry of the sex hormones had not been completely elucidated. Interestingly, the results from this experiment were reproduced in both females and males; nevertheless this observation didn't catch the attention of the authors. As will be discussed later, this observation is one of the main pillars for the use of Vtg as an environmental biomarker for exposure to estrogenic compounds.

Stronger evidence was being built at that time through a study made on English sole and Pacific halibut (Utter and Ridgway, 1965). An antigenic serum component whose presence was related to age, length and maturity in maturing female pleuronectids was detected by immunodiffusion techniques. The factor which had a qualitative seasonal variation for English sole had its highest incidence during the spawning season and the lowest at midsummer and was detected in all the mature female halibut sampled. The factor was also detected in the serum of some immature females of both species during the spawning season. Evidence associating the synthesis of this factor with the production of estrogenic hormones was obtained when  $17\beta$ -estradiol was injected into male English sole and induced them to produce the factor. Interestingly, the word 'protein' was not used to describe this factor in the entire paper, since at that time, the nature of the described factor probably remained uncertain.

In spite of all the previous experiments, it was not until 1973 that the hypotheses involving the sources of Vtg and its fate in fish were demonstrated and widely accepted (Aida *et al.*, 1973). This study proposed strong evidence that in maturing female Ayu

*(Plecoglossus altivelis)*, female specific proteins formed outside the ovary, with the stimulation of 17 $\beta$ -estradiol, were incorporated into the oocyte with the commencement of vitellogenesis. Many of the early studies on the formation of yolk and differentiation in teleost fishes have been recompiled in a single book chapter (Ng and Idler, 1983).

## **1.2 Vitellogenesis**

According to the latest literature, vitellogenesis is defined as the estradiol-induced hepatic synthesis of Vtg, its secretion and transport in blood to the ovary, and its uptake into maturing oocytes (Mommsen and Walsh, 1988; Tyler and Sumpter, 1996). This definition implies a series of sequential complex biochemical processes which must be individually analyzed in order to fully understand their significance. It should be noted that literature published prior to approximately 1970 use this term referring to the stage during oogenesis where accumulation of yolk occurs. An extensive list of comprehensive reviews on Vtg and vitellogenesis can be found in the literature (Tata and Smith, 1979; Kunkel and Nordin, 1986; Arukwe and Goksoyr, 2003b).

Oocytes are the female ovarian cells that undergo meiosis to form eggs. They are derived from cells that develop from primordial germ cells that migrate into the ovaries early in embryogenesis. In female teleost fish, the oocytes in the ovary are arrested in the first meiotic prophase. Fertilization of these oocytes is only possible under hormonal stimulation, which results in the completion of the first meiotic division. At this stage

they are termed secondary oocytes; only after the second meiotic divisions they become mature eggs.

In fish, oocyte post-ovulatory ageing is associated with egg quality decrease. During this period, eggs are held in the body cavity bathed in a semi-viscous liquid known as coelomic fluid whose components are suspected of playing a role in maintaining oocyte fertility and developmental competence. The rainbow trout (RT) coelomic fluid proteome has been studied during the egg quality decrease associated with oocyte post-ovulatory ageing. The results from this study suggest that egg protein fragments accumulate in the coelomic fluid during the post-ovulatory period and could be used to detect egg quality defects associated with oocyte post-ovulatory ageing (Rime *et al.*, 2004).

The developing oocyte is situated in the centre of the ovarian follicle. The external part of this structure consists of an inner sublayer known as the granulose cell layer, and an outer sublayer, one or two cells wide, called the theca. These sublayers are separated by a basement membrane. The interface between the oocyte and the granulosa cell layer consists of an acellular eggshell or zona radiata. The zona radiata has also been referred to as pellucid, vitelline membrane, zona pellucida, chorion, eggshell and vitelline envelope. The proteins which constitute the egg shell are also incorporated from a precursor synthesized in the liver known as the zona radiata proteins (also referred to as choriogenin). The formation of the vitelline envelope precedes the active uptake of vitellogenin during oocyte development in the RT (Hyllner *et al.*, 1994). Similar results have been observed for Atlantic salmon (AS), where '*in vitro*' and '*in vivo*' experiments have demonstrated that zonagenesis precedes vitellogenesis (Celius and Walther, 1998).

The interior of the oocytes are filled with the yolk proteins during the oocyte development. These proteins are derived from the Vtgs, which are enzymatically cleaved into smaller components once incorporated inside the oocyte (Wallace, 1985). In RT the breakdown products of Vtg are known as lipovitelline proteins and phosvitin, a serine rich highly phosphorylated polypeptide. The different processes involved in the processing and mobilization of Vtg are not completely known, and according to literature, vary among species.

The identification and characterization of the proteases involved in specific proteolysis of vitellogenin and yolk proteins in salmonids has been studied (Hiramatsu *et al.*, 2002). A recent publication has demonstrated that in RT, certain enzymes like cathepsin B and lipoprotein lipase were over expressed exclusively during oogenesis while cathepsin D and cathepsin L were expressed constitutively throughout oogenesis and embryogenesis (Kwon *et al.*, 2001). The difference in the temporal expression pattern observed in these and probably other proteolytic enzymes indicate the carefully organized regulation during the processing of the yolk proteins. Degradation of Vtg in sea urchins seems to proceed by similar enzymatic routes (Yokota and Kato, 1988). With respect to yolk protein hydrolysis and oocyte hydration, cytoplasmic maturation appears to be controlled by a change in the pH (Matsubara *et al.*, 2003).

Both the zona radiata proteins and Vtg are synthesized in the liver of the fish under the hormonal control of the hypothalamic-pituitary-gonadal-liver axis. The principal external factors that affect the endocrine regulation of the oocyte maturation in fish are probably water temperature and the photoperiod cycles, which trigger in the hypothalamus the release of the gonadotropin-releasing hormone. In response to these the

pituitary gland secretes the gonadotropin hormones (GtH I and GtH II), which are structurally related to the follicle-stimulating hormone (FSH) and the luteinising hormone (LH) found in humans. GtH I (FSH) is involved in vitellogenesis and zonagenesis while GtH II (LH) is related to oocyte maturation and ovulation.

GtH I ultimately induces in the follicular cells the secretion of  $17\beta$ -estradiol into the bloodstream. Testosterone is initially synthesized by the theca cells and is further transformed to  $17\beta$ -estradiol in the granulosa cells by the action of cytochrome P450 aromatase (CYP19). Once exported to the blood, estradiol is transported to other organs by the sex-hormone binding protein, and is incorporated into the liver either by diffusion or receptor mediated uptake. According to a study,  $17\beta$ -estradiol appears also to be regulated by the concentration of Vtg in the oocyte (Reis-Henriques *et al.*, 1997). The dose-response relationships and pharmacokinetics of vitellogenin in RT after intravascular administration of  $17\alpha$ -ethynylestradiol have also been studied (Schultz *et al.*, 2001).

Once inside the hepatocyte, estradiol initiates a long chain of biochemical events which finally results in the activation or enhanced transcription of the Vtg and the zona radiata protein genes. Firstly,  $17\beta$ -estradiol is attached to a specific steroid-receptor, monomeric protein known simply as the estrogen-receptor. Vtg gene expression has been extensively studied in RT (Le Guellec *et al.*, 1988).

In AS, the structural modulation of the hepatic chromatin in response to  $17\beta$ -estradiol induced activation of the vitellogenin gene regions has been described (von der Decken and Waters, 1993). Furthermore, the estrogen-controlled gene expression induction of two estrogen-responsive genes in the liver of RT have been determined

(Chen *et al.*, 1989). Similar gene expression studies have demonstrated that temperature, calcium concentrations and calcium antagonists modulate the responsiveness of the liver to 17 $\beta$ -estradiol (Yeo and Mugiya, 1997). Temperature seems to affect only the stages which are independent of 17 $\beta$ -estradiol or estrogen receptor concentrations (Mackay and Lazier, 1993). Many other studies have demonstrated that the estrogen receptor plays an important role in the presence of different endocrine disruptors (Anderson *et al.*, 1996; Tollefsen *et al.*, 2002; Latonnelles *et al.*, 2002a; Bermanian *et al.*, 2004), and these will be discussed later in the section which deals with Vtg as a biomarker.

In the absence of ligand, the estrogen receptor is bound to the heat shock protein 90 (hsp90). When 17 $\beta$ -estradiol binds to the receptor, the hormone-receptor complex undergoes dimerization, enters the nucleus and binds to the DNA at the estrogen responsive elements situated upstream from the constitutive Vtg transcription genes. A summary of all the events from the release of the pituitary gonadotropins to the final synthesis of Vtg DNA can be seen in Figure 1.1. Interestingly, a study performed on goldfish also demonstrated the induction of Vtg synthesis by massive doses of androgens (Hori *et al.*, 1979).

Vitellogenin undergoes post-translational modification (PTM) by the addition of carbohydrate and phosphate functional groups to the amino acid backbone. Lipids are also incorporated into the protein, although evidence suggests that these are linked via non covalent bonding. Vtg also binds ions such as Ca<sup>2+</sup> and PO<sub>4</sub><sup>3-</sup> to its molecule, serving as a major source of minerals to the developing oocyte. Early studies suggested that the female-specific serum proteins also had iron binding properties (Kunkel and Pan, 1976).

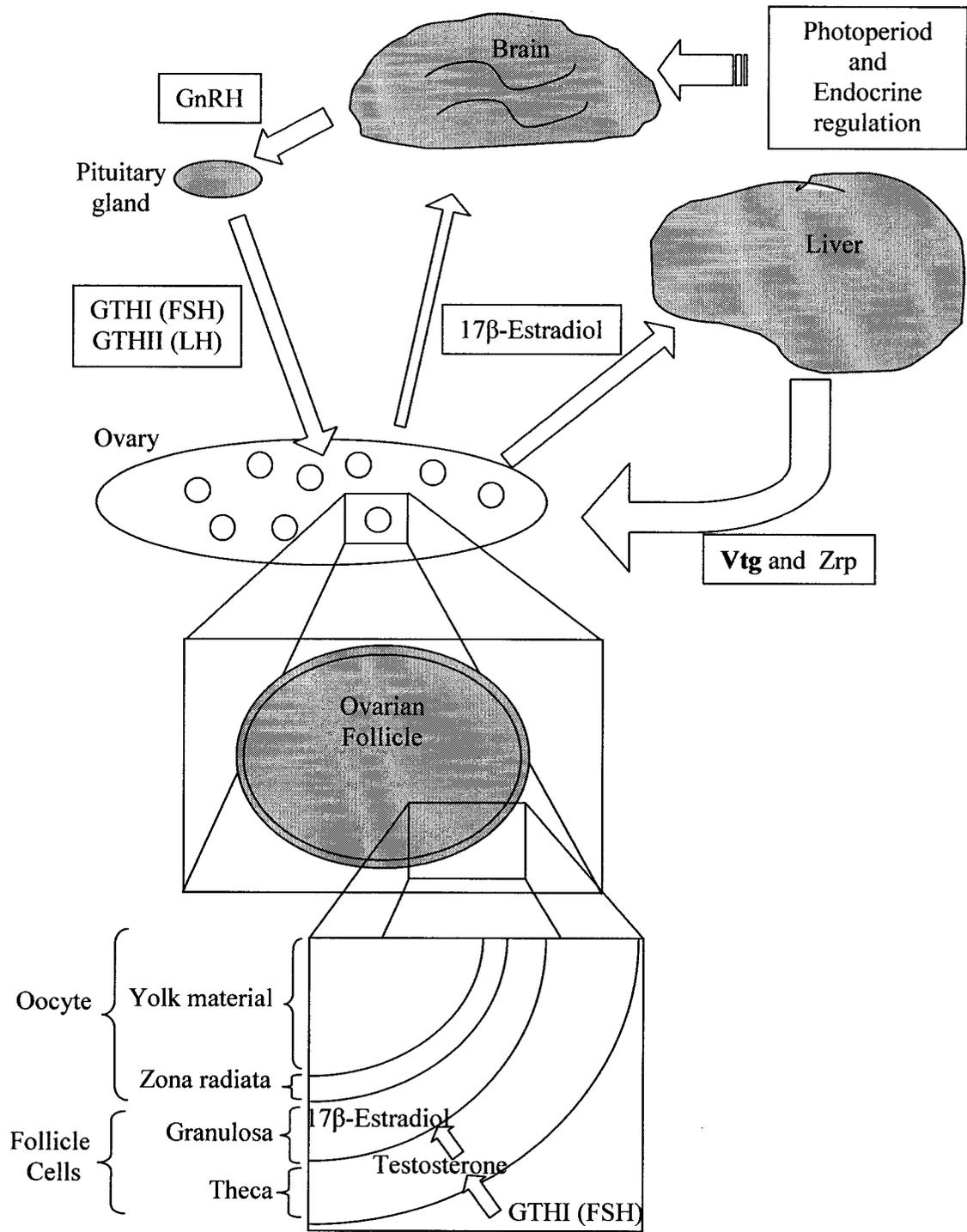


Figure 1.1 Scheme of the hypothalamic-pituitary-gonadal-liver axis

Furthermore, recent studies in different species suggest there might be a connection between retinol transport and Vtg (Lubzens *et al.*, 2003; Sammar *et al.*, 2005). Due to the complexity of Vtg, it would not be surprising to find in future studies, other lipid soluble vitamins and compounds present in this transport lipoprotein.

Once exported from the liver, Vtg circulates in the bloodstream and reaches the ovaries, where it is incorporated into the growing oocyte through receptor mediated endocytosis. The ovarian uptake and processing of Vtg in RT has been thoroughly studied (Tyler *et al.*, 1988; Tyler *et al.*, 1991). The Vtg receptors are distributed on the oocyte plasma membrane clustered in clathrin-coated pits (Le Menn *et al.*, 2000). Isolation and characterization studies of the RT Vtg receptors showed their molecular mass to be approximately 200kDa, probably comprising two 100kDa subunits (Tyler and Lancaster, 1993). These receptors also showed some affinity for Vtg of other species such as brown trout (*Salmo trutta*) and carp (*Cyprinus carpio*).

Cloning and sequencing experiments performed on the receptors revealed strong homology with the low density lipoprotein receptor (LDLR) (Schneider, 1996; Brown *et al.*, 1997b; Davail *et al.*, 1998). The Vtg receptor contained a cluster of 8 cysteine-rich repeating units close to its amino terminal ligand binding domain, also observed in the LDLR. Also, the short intracellular portion contained the internalization signal typical of the LDLR superfamily, Phe-Glu-Asn-Pro-Val-Try. However, ‘*in vitro*’ translation of the cloned cDNA produced a 97kDa protein, contrary to 200kDa observed by previous experiments (Tyler and Lancaster, 1993).

From an evolutionary point of view, the homology between the LDLR and the Vtg receptor strongly suggests that both have co-evolved from a common ancestor gene.

Interestingly LDLR is expressed not only in oviparous animals but in mammalian species also (Schneider, 1996). The Vtg receptor recognizes ligands that have similarities to structural elements found in the human apolipoproteins B-100 and E, as will be discussed below.

Modern cDNA cloning techniques have also allowed the determination of the amino acid sequence of Vtgs corresponding to different species. No sequence is available for AC Vtg, while AS Vtg has been partially sequenced (Yadatie *et al.*, 1999; Buisine *et al.*, 2002) and the complete amino acid sequence of RT Vtg has been elucidated (Mouchel *et al.*, 1996; Goulas *et al.*, 1996; Trichet *et al.*, 2000). These published sequences are shown in Appendix 2. As discussed later in Chapter 5.1, the RT and AS Vtg sequences bear high homology with each other. Genomic studies have demonstrated the complex evolution of the Vtg genes in salmonid fishes. Phylogenetic trees constructed from Vtg sequences revealed conflicting nodes with the consensus tree based on morphological and anatomical data (Buisine *et al.*, 2002). A comparative study with other marine teleost fishes has shown other common properties among these species (Covens *et al.*, 1987).

Atlantic salmon and RT Vtg also share regions of homology with other vertebrate Vtgs. However, their similarities with some mammalian proteins are even more striking. A sequence of approximately 670 amino acids located near the N-terminal region of Vtg is homologous to a similar sized region found in the same location of the microsomal triglycerol transfer protein (Shoulders *et al.*, 1993; Shoulders *et al.*, 1994) and the ApoB100 of LDL (Baker, 1988a; Baker, 1988b; Babin *et al.*, 1995; Li *et al.*, 2003). In LDL, this region is involved in lipoprotein assembly (Knott *et al.*, 1986). Surprisingly, it is not involved in receptor binding, in spite of the fact that both LDL and Vtg share the

same type of ligand receptors, as explained in the previous section. This task is performed by a region near the C-terminal part of ApoB100, curiously absent in Vtg (Milne *et al.*, 1989).

Additionally, a region of 150 amino acids near the C-terminal region of Vtg resembles a cysteine rich D domain found repeatedly in the pro-von Willebrand factor (Baker, 1988a) and the intestinal mucine 2 glycoprotein. This domain plays an essential role in the polymerization of these two proteins into large oligomers, via the formation of disulfide bonds. Furthermore, the CGLC motif found within the D domains is also conserved in most vertebrate Vtgs. In the pro-von Willebrand factor, this sequence is required for multimer assembly (Mayadas and Wagner, 1992). In the case of Vtg, this suggests that the same region could be involved with protein interaction functions such as dimer formation, reception binding to oocytes, sequestration in transitional yolk bodies or assembly and accretion of the yolk granules.

As observed from the homology studies, there seems to be much more to Vtg than just a nutritional reserve for the developing embryo. However, many of its functions are probably still to be discovered.

### **1.3 Vitellogenin size and structure**

According to the literature, it is generally accepted that Vtg is a phosphoglycolipoprotein. The structural differences found in this complex protein vary within different taxonomic classes. Vtg has been purified and characterized from a variety

of species ranging from sea urchins (*Strongylocentrotus purpuratus*) (Shyu *et al.*, 1987; Brooks and Wessel, 2004), nematodes (Spieth *et al.*, 1991), frogs (*xenopus laevis*) (Wiley and Wallace, 1981), goldfish (*Carassius auratus*) (De Vlaming *et al.*, 1980), coho salmon (*Oncorhynchus kisutch*) (Hara *et al.*, 1993), English sole (*Pleuronectes vetulus*) (Roubal *et al.*, 1997) to the exotic tuatara reptile (*Sphenodon punctatus*) (Brown *et al.*, 1997a). The main differences lie simply in the primary amino acid sequence, length and composition, although, as expected, evolutionary related species share many regions of conserved homology (Chen *et al.*, 1997). The post translational modifications observed in Vtgs of different species also vary considerably.

Because Vtg is generally considered a reserve of raw materials for the developing embryo, its amino acid composition is considered to be an issue of particular interest. Some researchers have considered this composition to be in strict relation to the essential biochemical requirements of the developing embryo for each species. For example, insects do not require large amounts of  $\text{Ca}^{2+}$  and  $\text{PO}_4^{3-}$  during embryogenesis, since bone synthesis is not necessary. Therefore, insect Vtg has not evolved as a major carrier of these ions. Conversely, vertebrate Vtgs contain a high number of serine residues which provide the sites necessary for phosphorylation and  $\text{Ca}^{2+}$  binding.

Some authors have tried to compare the amino acid composition of Vtg in terms of providing useful information on the nutritional requirements during embryogenesis in different species (Hagedorn and Kunkel, 1979). One work demonstrated that lysine is present in relatively high levels (8.44-8.74%) in ocean pout (*Macrozoarces americanus*), lumpfish (*Cyclopterus lumpus*) and AC Vtg. The authors suggest that this information could be considered when formulating a starter diet for young marine fish of aquaculture

importance (Yao and Crim, 1996). However, other studies that compared insect Vtgs to a large number of purified and sequenced proteins showed no significant difference in the observed average amino acid composition (Hagedorn and Kunkel, 1979). This work provides an interesting example in which composition and structure, but not sequence, are being conserved. The authors propose that if the precise amino acid sequence is unimportant and is being loosely conserved by evolutionary selection, random mutation could arrive at an average protein composition through the redundancy of the genetic code which biases all mutations towards a predictable average amino acid composition (King and Jukes, 1969). This view, however, seems to contrast with the selective force that different animal groups have, according to their nutritional needs. The idea of Vtg as a source of nutrients has even led some researchers to develop a novel strategy for genetically engineering a yeast strain (*Pichia pastoris*) to constitutively produce recombinant Vtg for application as an enriched feed (Ding *et al.*, 2004). The clones exhibited a significant increase in essential amino acids and long-chain polyunsaturated fatty acids, which are important for its application as a high-quality nutrient feed.

The calculated molecular mass for RT Vtg according to the 1659 amino acids (See Appendix 2) obtained from cloning and sequencing experiments is approximately 182 kDa. This value probably is an underestimate since it does not consider the additional mass corresponding to the post-translational modification.

On the other hand, most of the values obtained for the molecular mass of the native Vtg protein have traditionally relied on chromatographic and electrophoretic techniques which have been readily available in the past decades. The first electrophoretic

patterns of blood serum proteins from RT were done nearly 40 years ago (Thurston, 1967).

Some of these techniques have been used in pursuit of obtaining purified Vtg (Wiley *et al.*, 1979). Nevertheless it is surprising to find in the literature so many contradictory results with respect to the molecular mass of a protein, even for reports on the same species. For that reason it is very important to consider the experimental conditions under which these experiments have been conducted. For example, the molecular mass of the RT Vtg has been reported by different research groups. An early study showed that RT Vtg had a molecular mass of around 600 kDa by size exclusion chromatography on a Sepharose column (Hara and Hirai, 1978). The same study reported a molecular mass of 220 kDa when analyzed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Without much evidence, this work assumes Vtg to be a polymer of protein of molecular weight 220,000 Da bound with some lipids and carbohydrates, giving a molecule of molecular weight about 600,000 Da.

A study performed some years later, also on RT Vtg (Campbell and Idler, 1980) reported a molecular mass of 440 kDa by size exclusion chromatography on a Ultrogel AcA 22 column, 470 kDa by gradient gel electrophoresis and 342 kDa by ultracentrifugation. Another study also reported a molecular mass of 440 kDa, but performed on a Sepharose 6B column. However, these authors state that molecular masses obtained from studies done on lipoproteins by gel chromatography should be viewed with caution since these types of biomolecules exhibit 'aberrant behaviour' on this type of separation media due to chemical interactions between this protein and the stationary phase. This was the first paper to propose the dimeric nature of RT Vtg..

Further studies analyzed the molecular mass of Vtg by SDS-PAGE, obtaining values in the range of 170-180 kDa (Fremont and Riazi, 1988; Tyler, 1993; Silversand *et al.*, 1993; Kwon *et al.*, 1993; Folmar *et al.*, 1995). One of these studies still reported a molecular mass of 560 kDa when analyzed by different commercial size-exclusion chromatography media such as Sepharose CL6B, Biogel 1.5M, Sephacryl S300 and AcA 22 (Fremont and Riazi, 1988). In this contribution, the chromatographic profile was shown to be affected by the presence of  $\text{Ca}^{2+}$  in the mobile phase when AcA 22 was used as the stationary phase. One particular study was obtained from a culture of primary hepatocytes (Kwon *et al.*, 1993), in contrast with all the previous experiments in which Vtg had been obtained from serum or plasma. Similar experiments, carried out by another research group developing a two-step method for purification of RT Vtg, reported the molecular masses to be 442 kDa and 240 kDa by gel chromatography (Supedex 200HR) and SDS-PAGE, respectively (Brion *et al.*, 2000).

The dimeric nature of Vtg was also suggested in a study while validating a competitive enzyme-linked immunosorbent assay (ELISA). An interesting observation made by this study was the discovery of not one, but two bands present on SDS-PAGE, which were confirmed to be Vtg by western blotting. The molecular masses associated with these bands were 393 and 176 kDa, which the authors link to the dimeric and monomeric form of Vtg, respectively. It is surprising that none of the previous works had ever noted the band at higher molecular mass in the same type of experiments (Bon *et al.*, 1997). The band observed at 393 kDa implies that the dimer was able to withstand the non-denaturing conditions of SDS performed in this experiment. These conditions might have not been met in the previous referenced studies.

The lowest molecular mass obtained by size exclusion chromatography was 383 kDa. Interestingly, this group obtained 154 kDa by SDS-PAGE, also the lowest figure by this technique (Watts *et al.*, 2003).

This is just a fraction of the published literature which covers the characterization of intact RT Vtg. It is obvious, as stated previously, that the experimental conditions are critical for determination of the molecular mass of intact Vtg. However, even under the non-denaturing conditions employed for the different types of size-exclusion chromatography, the results obtained varied from 383 to 600 kDa. On the other hand, the results obtained from SDS-PAGE varied from 154 to 393 kDa. The denaturing conditions conferred by the SDS disrupt the non-covalent interaction, thus, the protein loses its native secondary and tertiary structure, and any lipids non-covalently attached to it. There are some inherent causes that might explain the discrepancies within the same type of techniques. The limitation for accurately determining the band positions and the poor band resolution can probably account for some variability observed in the SDS-PAGE experiments, even in the order of 10kDa.

Even greater differences were encountered in the size-exclusion chromatography experiments. The mass resolution for this technique is even lower than the SDS-PAGE. Furthermore, the differences reported indicate that certain interactions might have existed between the Vtg molecule and the stationary phases employed.

What most research groups observed, independent of the techniques employed, was a disparity of results when comparing the molecular masses obtained in denaturing and non-denaturing conditions. Some authors suggest that this difference is only due to the dimeric nature of Vtg, which unfolds into subunits in denaturing conditions (Norberg

and Haux, 1985; Bon *et al.*, 1997; Brion *et al.*, 2000). In most of these experiments, mass calibration of size-exclusion chromatography was performed by relating the molecular masses of the standards to their elution volumes or times. Alternatively, a group working with winter flounder (*Pleuronectes americanus*) Vtg, used the same laboratory procedure, but made their calibrations by relating the Stokes radius ( $r = f / 6\pi\eta$ , where  $f$  stands for frictional coefficient and  $\eta$  for the viscosity) of their standards to the corresponding column partition coefficients. Based on their results, they also concluded that Vtg in serum exists as a dimer (Hartling *et al.*, 1997).

Other groups simply state that Vtg circulates as a polymeric protein, without specifying the number of subunits present (Hara and Hirai, 1978; Fremont and Riazi, 1988). Nevertheless, a close inspection of their results reveals some intriguing facts. For example, the study that determined a native molecular mass of the RT Vtg to be 560 kDa by gel filtration, also determined that the monomer was 170 kDa by SDS-PAGE (Fremont and Riazi, 1988). However, the chemical analyses of Vtg showed that 79% of the native molecule corresponded to protein, 19% to lipids, 0.3% to carbohydrates and 0.7% to phosphate and calcium. The molecular mass does not agree with the proposal of a dimeric structure for Vtg, since the sum of both monomers is far from the mass obtained for the native Vtg, even with the addition of the non-protein components of the molecule. However, a trimeric situation is not convincing.

Fewer studies have been carried out on AS Vtg. A publication referenced earlier (Watts *et al.*, 2003) reported a molecular mass of 557 kDa for the native protein by size-exclusion chromatography and a major band of 159 kDa by SDS-PAGE. Another study

reported a molecular mass of 495 and 520 kDa using gel chromatography and gradient gel electrophoresis, respectively.

Atlantic cod Vtg is the least studied protein of the three species covered in this thesis. Its molecular mass was reported to be 485 and 167 kDa according to studies carried out by gel filtration and SDS-PAGE, respectively (Silversand *et al.*, 1993; Yao and Crim, 1996).

From the disparity of results obtained from various experiments carried out to determine the size and quaternary structure of Vtg, one can draw two conclusions: the techniques employed in different experimental conditions are difficult to reproduce and/or the heterogeneity exhibited by this particular protein is much higher than expected for the expression of a gene from a single species. Regardless, this shows that much light has yet to be shed on this matter. Similar situations can be recompiled from literature for other species.

The biochemical and structural characteristics of Vtg not only depend on its primary structure, but also on its post translational modifications. Again, literature in this area is vast and overwhelming. In this case, the discrepancy might not only be due to analytical limitations, but to actual differences in the materials studied since some of the post translational modifications are known to be temporary signaling groups in the lifespan of a protein.

The first studies focused on unraveling the biochemical complexity of RT Vtg were performed in the late 70's (Hara, 1976; Hara and Hirai, 1978) based on earlier studies performed on other species. One of these studies, using disc electrophoresis, confirmed the presence of lipids and carbohydrates by staining techniques with Sudan

Black B and Schiff's reagent, respectively. A later study performed quantitative analyses on these components and discovered that the total lipid and phosphorus contents were 21.5 and 0.6 % by weight, respectively. These values have been reproduced by other authors in subsequent publications (Norberg and Haux, 1985). The lipids in Vtg are non-covalently attached to the protein component of the molecule and its content is likely to vary slightly, as with other lipoproteins. Even so, the lipid values are considerably lower than those found in other lipoproteins. For example, the normal lipid content values for RT VLDL, IDL, LDL and HDL are 95, 87, 70, and 55%, respectively (Babin and Vernier, 1989). This makes Vtg the molecule with the highest density among the lipoproteins ( $d > 1.210$ ).

A study that assayed the compositional analyses of the lipid fraction of RT Vtg reported the total lipid content to be 18%, phospholipids 11%, triglycerides 4% and cholesterol 2% of the total weight of the purified Vtg (Norberg and Haux, 1985). These values are very similar to those obtained by other researchers who also analyzed their distribution in the phospholipid classes, their fatty acid composition and the carbohydrate composition (Fremont and Riazi, 1988). A comparison of the lipid content in ocean pout, lumpfish and Atlantic cod seems to agree with those found in RT Vtg although the phosphorus content appears to be slightly lower in the latter (Yao and Crim, 1996).

Not much is known about RT Vtg regarding the carbohydrates moieties. Two consensus motifs for potential asparagine linked glycosylations at positions 1089 and 1627 were found by sequence homology studies. More information on the Vtg oligosaccharides is available in insects (Kunkel and Nordin, 1986) and crustaceans (Khalaila *et al.*, 2004). These studies indicate that most, if not all of their carbohydrate

content is present as high mannose type oligosaccharides. Since insect glycans do not contain sialic acid (Ng and Dain, 1976) it is improbable that complex types of oligosaccharides will be found. Also, experiments have shown a high micro-heterogeneity in the oligosaccharide fraction of newly secreted Vtg molecules, even within a single specimen.

In contrast, the oligosaccharide moiety of the African clawed frog (*Xenopus laevis*) Vtg is a complex type glycan, linked to the protein via an N-glycosydic bond between an asparagine residue and N-acetylglucosamine (Gottlieb and Wallace, 1982). The same authors have determined the mechanisms of Vtg phosphorylation in *Xenopus laevis* (Gottlieb and Wallace, 1981).

X-ray crystallography and nuclear magnetic resonance (NMR) are two techniques that allow determination of the secondary and tertiary structures of proteins. Proteins must first be crystallized before they can be subjected to x-ray crystallography. Vtgs have not been amenable to this technique because researchers have not been able to obtain crystallized samples. The main reason for this is their varying amount of heterogeneous lipid. However, lipovitelline from lamprey was crystallized with a substantial portion of associated heterogeneous lipid (Meininger *et al.*, 1984) and subjected to X-ray analysis. The visualization of the three dimensional image indicates that the protein is organized into a single domain with a large cavity enclosed by 3 anti-parallel  $\beta$ -sheets and one helical domain. This cavity was proposed to be the lipid binding site of lipovitelline, which is probably conserved in Vtg. The preponderance of hydrophobic residues along the inner surface of this cavity supports its lipid binding nature. However, the total absence of electron-density for any bound lipid in this cavity leaves unanswered any

questions regarding their orientations and protein interactions. The structure also revealed two additional sites of ordered lipid molecules and an additional site for  $Zn^{2+}$  binding (Anderson *et al.*, 1998). Some of the observations obtained from the X-ray crystallographic investigations have been confirmed by NMR studies (Banaszak and Seelig, 1982). The results of the  $^{31}P$  and  $^2H$  NMR experiments correlated with the low-resolution model of the crystalline lipovitellin complex obtained by diffraction studies. Such studies proposed that each subunit of lipovitellin contains a microdomain of phospholipid in a bilayer-like arrangement.

Interestingly, the studies on the structure of Vtg have gone beyond any interest ever imagined by the initial researchers of this protein. The structural information derived from the x-ray crystallographic studies of lipovitellin have provided a model for the assembly and secretion of atherogenic lipoproteins (Mann *et al.*, 1999). This work demonstrates that the structural and functional evolution of Vtg provides a unifying scheme for the invertebrate origins of the major vertebrate lipid transport system.

#### **1.4 Vitellogenin as a sexual maturation indicator**

Determining the reproductive status of caged and wild fish stocks is extremely important from various points of view. For example, some species which are endangered due to natural predation or excessive fishing pressure exhibit an early onset of sexual maturation and increased frequencies of the reproductive cycle. This reaction can be interpreted as a response of the species to adverse survival conditions which would lead

to earlier spawning and higher chances to restore the fish population. Therefore, relating fish age to the onset of reproductive maturation is a good indicator of the selective pressure on that species, considering the reference values are known for each particular ecosystem. For example, the plasma Vtg levels during the annual reproductive cycle of the female RT has been established (Bon *et al.*, 1997). These values oscillate between 0.5 and 60 mg/mL during the reproductive cycle. These levels can be considered as reference values for mature females and used for comparison purposes.

In aquaculture, a different situation is encountered. The objective of this commercial activity is to obtain large fish with a maximum ratio of flesh to body weight. Therefore, it is expected for feed to be transformed to muscle synthesis with a high efficiency. High muscle content translates into increased commercial value given primarily by the fillet market. However, sexual maturation is an expensive metabolic process, where much of the food energy is transformed into the development of the gonads, especially in females where the fully grown gonads account for up to about 25% of the total weight of a mature fish. For this reason, it is very important to detect the onset of sexual maturation to determine the optimal slaughter time in terms of maximizing the profits.

Thus, estimation of the sexual maturation is an important objective. Observation of increased gonad size is an obvious way to determine sexual status, although macroscopic changes are usually the last indicators. Microscopic histopathology is an earlier indicator, although both imply destructive techniques. More recently, invasive but non-destructive techniques have been proposed. Serum calcium, total phosphoprotein phosphorus and alkali-labile phosphoprotein phosphorus have been shown to

increase during vitellogenesis (Nagler *et al.*, 1987). Other studies additionally propose serum Mg levels to increase during this stage (Wiegand, 1982; Bjornsson *et al.*, 1986). All these parameters can be used as indicators of plasma Vtg levels (Arukwe and Goksoyr, 2003a). Alternatively, the hormones of the hypothalamic-pituitary-gonadal-liver axis can be measured. However, direct measurement of plasma Vtg levels seems to be the most appropriate indicator since it is the last in a series of biochemical pathways leading to sexual maturation.

### **1.5 Vitellogenin as a biomarker**

Vitellogenin synthesis is induced by  $17\beta$ -estradiol, the physiological ligand of the estradiol receptor in the liver of vertebrate oviparous animals. In invertebrates, this synthesis can occur in analogue tissues or organs such as the fat body of insects or the hepatopancreas of crustaceans. There are, however, other molecules that can interact with the estradiol receptor too, producing an agonistic effect on Vtg synthesis. These compounds are given the generic name of “xenoestrogens”. The basic meaning of this word refers to any chemical foreign to that particular organism, which will have an estrogenic effect. Xenoestrogens are one of the many types of endocrine disruptors or modulators. These compounds affect not only wildlife, but also humans, as discussed in some publications (Toppari *et al.*, 1996; Damstra *et al.*, 2002a).

The xenoestrogens originate from diverse sources, and can be classified on different bases. Some occur naturally in the environment, such as the phytoestrogens.

Others are of anthropogenic nature, chemically synthesized by humans and released into the environment by agricultural and industrial activities. It is common to find in literature other names that refer to this type of compounds, namely, environmental estrogens, eco-estrogens, xenobiotic estrogens, environmental hormones, or hormone related toxicants, among others.

The exact mechanism of action of the xenoestrogens is not completely understood although it is generally accepted that most of them act as estradiol mimics and trigger the cascade of biochemical events, such as Vtg synthesis. The zona radiata proteins are also altered by the presence of these compounds and its increased expression is also considered an important biomarker for environmental estrogens (Arukwe *et al.*, 1997). However, other mechanisms of Vtg induction have also been proposed. In this respect, a study with nonylphenol suggested that this molecule is involved in a post-transcriptional regulation process (Ren *et al.*, 1996).

Other types of molecules also bind to the estrogen receptor producing an inhibition or antagonistic effect. These types of endocrine disruptors are named antiestrogens (Safe, 1995; Safe and Krishnan, 1995; Ahlborg *et al.*, 1995). An example of this type of inhibitors is 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). Presence of TCDD inhibits Vtg induction probably through the action of the aryl hydrocarbon receptor in complex with its nuclear translocator, and possibly by direct interference with the autoregulatory transcriptional loop of the estrogen receptor. In addition, 17 $\beta$ -estradiol does not interfere with TCDD induced CYP1A gene expression, suggesting that cross-talk between the estrogen receptor and the aryl hydrocarbon receptor signaling pathway is unidirectional (Bermanian *et al.*, 2004). Cadmium has also been discovered to inhibit Vtg

expression and will be discussed later. Also, studies comparing the ligand-binding properties of estrogen receptors to endogenous steroids and environmental estrogens have demonstrated similar affinities for both RT and AS (Tollefsen *et al.*, 2002).

As a result of exposure to xenoestrogens, Vtg is over-expressed in sexually mature females and in young immature females. Nevertheless, the most striking finding is that Vtg is also found circulating in blood of males exposed to either estrogenic hormones or xenoestrogens. Endogenous  $17\beta$ -estradiol levels in males are usually undetectable, however, the biochemical machinery required to synthesize Vtg is readily available given the proper stimulation.

For all these reasons, Vtg has been used as a biomarker for environmental xenoestrogens and measurement of Vtg plasma levels has consequently formed a central component of a wide spectrum of investigations. The values of plasma Vtg obtained when fish are exposed to xenobiotic estrogens depend on the nature of chemical, its concentration and the route of exposure. For example, a study on the estrogenicity of butylparaben in RT exposed via food and water showed Vtg values ranging from 1 to 100 $\mu$ g/mL, according to the doses of fed butylparaben (Alslev *et al.*, 2005). Another study investigating the estrogenic activity of a wide variety of chemicals such as nonylphenol, bisphenol A and phthalates, showed Vtg levels in immature RT vary from 1  $\mu$ g/mL to 1 mg/mL when injected intraperitoneally with single doses of the test compounds (Christiansen *et al.*, 1998).

According to the nature of the source to which the species are exposed, these investigations can be classified into two main groups. One type of experiment is that performed mainly by environmental scientists concerned with contamination or pollution of the environment. In these studies, Vtg is measured '*in vivo*' in indigenous species and compared to control levels to assess the degree of xenobiotic estrogens present in the environment. Alternatively, the estrogenic activity of the environmental samples can be analyzed in sentinel species. In this case, the test animals in the laboratory are exposed to the environmental samples for a period of time, after which Vtg levels are measured. An '*in vitro*' variation of this approach has also been done by replacing the sentinel species with hepatocyte cultures. This type of investigation provides very important information. Although the presence of some xenoestrogens can be individually tested, the composition of the environmental sample is never completely known. Thus, Vtg plasma levels represent a valuable indicator for the combined effect of the environmental estrogens present.

Examples of this approach can be seen in studies performed on different sources suspected with estrogenic activity such as municipal sewage, wastewater treatment works and effluents derived from diverse industries such as paper mills and oil refinery treatment works (Knudsen *et al.*, 1997; van den Heuvel and Ellis, 2002; van den Heuvel *et al.*, 2002; McClain *et al.*, 2003; Nakari, 2004; Jobling *et al.*, 2004; Tyler *et al.*, 2005; Johnson *et al.*, 2005). These studies have been done both '*in vitro*' and '*in vivo*' to either cell cultures or animals of different species.

The other type of experiment is done usually by eco-toxicologists and biologists interested in the estrogenic potency of known chemicals. In these studies, a special focus

is placed on assessing the developmental, reproductive and health effects produced by the xenoestrogens in wildlife and laboratory animals (Lam, 1983; Colborn and Clement, 1992; Kime, 1995; Arukwe and Goksøyr, 1998; Pedersen *et al.*, 1999; Pedersen *et al.*, 2000; Damstra *et al.*, 2002a; Damstra *et al.*, 2002b; Goksøyr *et al.*, 2003; Madsen *et al.*, 2004). The Vtg assay is one of the many biochemical and biological parameters measured to assess the health of the organisms studied. The experimental procedure is analogous to the one described above, however, in this case, the composition of the chemicals in the administration source is known beforehand. Thus, Vtg levels after exposure are proportional to the estrogenic potency of the studied chemicals. This approach can also be done both '*in vivo*' using a sentinel species, or '*in vitro*' via hepatocyte cultures (Radice *et al.*, 2002).

Extensive studies have been performed on a variety of organic compounds with estrogenic activity, including polychlorinated biphenyls, polycyclic aromatic hydrocarbons, polychlorinated dibenzodioxins, alkylphenols, alkylhydroxy benzoates and synthetic estrogens (Sumpter and Jobling, 1995; Palmer *et al.*, 1998; Bjerregaard *et al.*, 1998; Christiansen *et al.*, 1998; Yadetie *et al.*, 1999; Lindholst *et al.*, 2000; Lindholst *et al.*, 2001; Boon *et al.*, 2002; Verslycke *et al.*, 2002; Yadetie and Male, 2002; Okoumassoun *et al.*, 2002; Thomas-Jones *et al.*, 2003; Tollefsen *et al.*, 2003; Pedersen *et al.*, 2003; Bjerregaard *et al.*, 2003; Lindholst *et al.*, 2003; Jobling *et al.*, 2004; Meucci and Arukwe, 2005; Alslev *et al.*, 2005). These compounds are usual components of surfactants, plasticizers, flame retardants, pesticides, herbicides, fungicides, paints, cosmetics, pharmaceuticals and food preservatives used in agriculture and industries.

Most of these chemically synthesized compounds will eventually end up in the environment.

As an example, a study performed in an AS hepatocyte culture showed the estrogenic effect of both endogenous steroids ( $17\beta$ -estradiol, estriol, estrone and  $17\alpha$ -estradiol) and estradiol mimics (ethynylestradiol, diethylstilbestrol, zearlalenone, bisphenol A, 4-t-octylphenol, 4-n-nonylphenol and 2-chloro,4-chloro-diphenyltrichloroethane (o,p'-DDT)). However, this study did not show response to dieldrin and toxaphene, two putative estrogenic compounds. Other investigations have also demonstrated differences between '*in vitro*' and '*in vivo*' studies (Folmar *et al.*, 2002). These studies have demonstrated that '*in vitro*' bioassays underestimate the results obtained in '*in vivo*' experiments. This raises the issue of the limitations of the '*in vitro*' bioassays, especially for proestrogens, substances that require metabolic activation. Furthermore, '*in vitro*' experiments are not capable of providing other physiological alterations. Consequently, the same authors recommend that these assays supplement '*in vivo*' studies, much to the dislike of animal rights supporters. Other scientific papers also observed differences in both types of assays and even marked notorious interspecies dissimilarities (Latonnelle *et al.*, 2002b).

Another work studied the effect on the reproductive impairment of medaka (*Oryzias latipes*) based on the fecundity, fertility and hatching success of eggs from fish that had previously been exposed to different concentrations of diphenyltrichloroethane (DDT). This study concluded that Vtg expression may be interpreted as a warning of reproductive consequences, but its absence does not imply a deficiency of consequences (Cheek *et al.*, 2001).  $\beta$ -sitosterol and other wood derived phytoestrogens present in pulp

and paper mill effluents have also shown some estrogenicity in RT exposure experiments (Mellanen *et al.*, 1996). Binding affinities of these phytoestrogens for the hepatic nuclear estrogen receptors has also been calculated (Latonnelle *et al.*, 2002a).

Other compounds like 4-tert-octylphenol, 4-nonylphenol, bisphenol and dibutyl phthalate have also been tested in RT and zebrafish (*Danio rerio*) (Van den *et al.*, 2003). All compounds exhibited estrogenic effects except for dibutyl phthalate. No interspecies differences were observed except for nonylphenol, for which zebrafish was five times less sensitive.

Long term effects of nonylphenol on gonadal development in juvenile RT showed no effect on sexual differentiation in the concentrations normally found in sewage treatment effluents, although Vtg was over expressed (Ackermann *et al.*, 2002). Nevertheless, a trans-generational study done with the same compound raised a serious issue. The published results showed not only a reduction in reproduction success indicated by decreased hatching rates but also hormonal imbalances in the offspring of exposed fish (Schwaiger *et al.*, 2002).

As mentioned, some studies have tried to determine biomarker lowest-observed-effect concentrations (Thomas-Jones *et al.*, 2003). Conclusions drawn by these experiments are very clear under controlled laboratory conditions, however, there is a clear uncertainty if these values are of any use for assessing the estrogenic potency of environmental samples. The relative potencies and combination effects of a mixture of steroidal estrogens have been measured using Vtg induction in a 14-day '*in vivo*' juvenile RT screening assay (Thorpe *et al.*, 2003). This study confirmed the ability of each individual estrogen to contribute to the overall effect of a mixture in a way predicted by a

concentration addition model. This is particularly interesting since most estrogens in domestic effluents are usually present as mixtures. This means that the overall effect of a mixture can be harmful, even if each individual estrogen is present in no effect concentrations. It should be noted that this study was carried out using natural steroids like 17 $\beta$ -estradiol and estrone. Although these compounds would not fit into the definition of xenoestrogens, their presence in domestic effluents converts them into potential natural endocrine disruptors (Routledge *et al.*, 1998).

Binary mixtures using other compounds such as nonylphenol and methoxychlor were also assayed yielding similar results (Thorpe *et al.*, 2001). Again, the model of concentration addition accurately predicted the effects on Vtg induction for compounds sharing the same mechanism of action. On the contrary, cadmium was shown to inhibit the transcription of Vtg when co-injected in RT with 17 $\beta$ -estradiol (Olsson *et al.*, 1995; Vetillard and Bailhache, 2005), although opposite results have also been obtained in similar experiments (Ait-Aissa *et al.*, 2003).

Vtg in body surface mucus has also been measured as an alternative sample to serum or plasma. The detectable values of Vtg and zona radiata proteins found after exposure to waterborne nonylphenol suggest that this noninvasive and nondestructive example could be used for the detection of endocrine disrupting chemicals and related pollutants in the environment (Meucci and Arukwe, 2005). Also, estrogenicity in bile of juvenile RT has been studied as a measure of exposure and potential effects of endocrine disruptors (Allard *et al.*, 2004).

In a different approach, other studies have addressed the problem of endocrine disruptors by analyzing differential gene expression in fish by Northern blotting,

quantitative reverse transcriptase polymerase chain reaction (PCR), subtractive hybridizations, and gene arrays. Each of these systems showed advantages and disadvantages. However, the experiments with each of these methods provided important information about the molecular mechanisms that resulted from exposure to endocrine disruptors and this information can be used in risk assessment of polluted environmental sites (Larkin *et al.*, 2003).

There have been innumerable investigations to determine the biological effects of the environmental estrogens. In spite of all these efforts, a comprehensive review published recently suggests that the evidence supporting the impact of these compounds on the reproductive health and sustainability of indigenous fish populations is not very convincing (Mills and Chichester, 2005). According to this review, there appears to be sufficient evidence from laboratory experiments to link the exposure to endocrine disrupting chemicals with reproductive impairment in many fish species. However, studies of the impact of these chemicals on real fish populations are lacking. Furthermore, this review suggests that a reliable method or indicator to determine the reproductive success of fish has yet to be developed.

## 1.6 Quantification of vitellogenin

As described in the previous sections, different experimental objectives have lead scientists to be involved in measuring Vtg plasma levels. Over the last two decades, a wide spectrum of analytical methods has been developed to meet this purpose. The first indirect estimates of Vtg were performed colorimetrically by determining the alkaline-labile phosphorous content of fish plasma (Wallace R.A. and Jared D.W., 1968; Verslycke *et al.*, 2002). Other techniques have also been used, such as immunoagglutination (Le Bail and Breton, 1981), densitometry following electrophoresis (van Bohemen *et al.*, 1981; van Bohemen and Lambert, 1981), immunoblotting (eg western blot) (Kwon *et al.*, 1993), radial immunodiffusion (Hara and Hirai, 1978) and by liver tissue slice immunoassays (Schmieder *et al.*, 2000).

Vitellogenin synthesis can also be estimated by quantification of Vtg mRNA levels in liver tissues and different approaches are possible for this task. mRNA can be qualitatively assayed by direct northern blotting while more sophisticated methods such as reverse transcriptase PCR and quantitative PCR provide more accurate results. Quantitative PCR is also known by the name of real time PCR, although this term usually is mistaken for reverse transcriptase-PCR.

A detailed protocol to develop a reverse transcriptase-PCR method against Vtg mRNA can be found in the literature (Bowman and Denslow, 1999). Similar reverse transcriptase-PCR techniques have been used in experiments that measured the effects of nonylphenol, (Ackermann *et al.*, 2002), benzo[*a*]pyrene, 17 $\beta$ -estradiol and cadmium (McClain *et al.*, 2003) on RT Vtg induction. Both reverse transcriptase and quantitative

PCR have been used in studies to determine Vtg mRNA of male mummichog (*Fundulus heteroclitus*) exposed to 17 $\beta$ -estradiol and common carp (*Cyprinus carpio*) in a river known for its high levels of estrogenic alkylphenols.

Sample preparation for PCR techniques is very critical. One of the major obstacles in RNA preparation is the RNAases since these ubiquitous enzymes are very difficult to inactivate. Also, extreme laboratory precautions must be taken to prevent external contaminations. However, the main drawback to this approach is that sample handling and the procedure to obtain the liver tissue homogenates is more cumbersome, costly and time consuming than the techniques for measuring Vtg in serum. Consequently, this approach is prohibitive for large scale sampling and high throughput analyses.

The advantages of the techniques that measure mRNA are their increased sensitivity and their potential to detect earlier the estrogenic response, since mRNA synthesis always precedes protein synthesis. Measurements done on the mRNA indicate recent exposure, since mRNA is quickly degraded, with a half life of approximately 2 days (Denslow *et al.*, 1999). This characteristic is particularly valuable for biomarker ‘*in vitro*’ studies performed on hepatocyte cultures to assess the estrogenic potency of endocrine disruptors.

An alternative to PCR for the detection and quantification of mRNA is the chemiluminescent hybridization assay. This technique relies on the hybridization of acridinium ester-labeled oligonucleotide probes to target RNA. In this duplex, the labeled oligonucleotides are resistant to the hydrolytic step that follows. Detection is done by chemiluminescence which is measured in a luminometer and the signal is proportional to the original amounts of RNA. This extremely sensitive and specific technique has been

used to study the dynamics of estrogen biomarker responses in RT exposed to 17 $\beta$ -estradiol and 17 $\alpha$ -ethinylestradiol (Thomas-Jones *et al.*, 2003).

It was not until the development of radioimmunoassays (RIA) and ELISA that plasma or serum Vtg levels were first accurately and rapidly measured. Both techniques have had a wide acceptance in the last decades, mainly due to the very low limits of detection (LOD) they have attained, in the order of ng/mL, and their high reproducibility.

The latest publications describing quantitative analyses of Vtg indicate that there has been a declining trend in the use of RIA, probably due to the radioactivity hazards, their higher regulatory costs and the significant disposal problems caused by radiochemicals. Because of these concerns, RIAs have continuously been replaced by ELISAs, which have reached similar levels of sensitivity.

The first RIA for RT Vtg was developed in 1980 by raising antibodies against Vtg purified from the plasma of estrogen induced fish. Little cross reactivity with AS Vtg was found (Campbell and Idler, 1980). RIA techniques are so sensitive that Vtg has even been detected in the plasma of male and immature females, although several methodological issues have been raised regarding the validity of these results (Copeland *et al.*, 1986). RIAs have also been developed with antibodies raised against AS Vtg purified from plasma (So *et al.*, 1985; Sumpter and Jobling, 1995). Again, this method demonstrated poor cross-reactivity with purified Vtg of RT and other teleosts. It is interesting to observe this high immunospecificity between such high-molecular mass proteins in such closely related species as RT and AS. RIAs for other species such as brown trout (*Salmo trutta*) have also been described.

The first ELISAs for RT Vtg were developed approximately one decade after the appearance of the RIAs. Before the commercial ELISA kits were introduced, researchers prepared their own procedures using polyclonal antibodies raised against purified samples of Vtg. The first ELISAs described were ‘sandwich’ type methods, where the Vtg molecules were ‘sandwiched’ between the raised antibodies coated on the microtitre plates and the labeled antibodies used for color development (Crowther, 2000). Validation of the techniques was usually done with well understood systems such as the induction of Vtg by 17 $\beta$ -estradiol in primary hepatocyte cultures (Kwon *et al.*, 1993). Other sandwich type ELISAs were tested for amphibian models (*Xenopus laevis*) (Palmer *et al.*, 1998).

A simplified version of the ‘sandwich’ type approach has been reported for winter flounder Vtg (Hartling *et al.*, 1997). In this ‘direct’ assay, protein solutions from standards and unknowns were dot blotted onto a nitrocellulose membrane. Blots were first incubated with the rabbit antiserum followed by the secondary conjugated goat anti rabbit IgG. The color reaction was quantified by densitometry of a video camera acquired image. An alternative approach to this direct assay has been described by coating the samples and standards on microtitre plate wells instead of blotting on a membrane (Denslow *et al.*, 1999). In this case, the results can be measured colorimetrically by a microplate reader. Other ‘sandwich’ type ELISAs have been developed for a variety of species such as the common carp (*Cyprinus carpio*), fathead minnow (*Pimephelas promelas*), zebrafish (*Danio rerio*) and Japanese medaka (*Oryzias latipes*) (Nilsen *et al.*, 2004). An immunochemical assay of medaka Vtg has also been documented (Hara *et al.*, 1983).

In addition to the 'sandwich' type, competitive ELISAs have also been developed. In this type of ELISA, purified Vtg coated on the microtitre plates is incubated with the specific antiserum and serial dilutions of samples or standards. After incubation, color is revealed by means of a secondary labeled antiserum. Vtg concentration in samples or standards is inversely proportional to color intensity. Again, calibration by this technique showed very low LOQ, with the most reliable results ranging between 20 and 320 ng/mL (Bon *et al.*, 1997). In contrast with some RIA assays, male serum was shown not to contain Vtg.

Both polyclonal and monoclonal antibodies have been used for these techniques. Polyclonal antibodies are easier to prepare and offer higher signal to noise ratios. Nonetheless, ELISAs using these antibodies are usually less specific and harder to validate. Monoclonal antibodies directed to a single epitope present higher specificity but are more expensive to produce and may be less sensitive (Denslow *et al.*, 1999). However, the benefit of the increased specificity of monoclonal antibodies is questionable. Some researchers have taken advantage of the lower specificity of polyclonal antibodies by developing ELISAs which could be used for Vtgs of related species.

The question of whether a universal monoclonal assay could be developed was also addressed (Heppell *et al.*, 1995). Promising results were obtained from both monoclonal and polyclonal antibodies. Monoclonal antibodies generated against purified RT Vtg recognized other estrogen-inducible proteins in serum of species representing four of the vertebrate classes (fish, amphibians, reptiles and birds). Similar results were obtained from polyclonal antibodies raised against a synthetic consensus peptide

representing the conserved N-terminal amino acid sequence of Vtg from phylogenetically diverse teleost fish.

Other examples of cross reactivity are extensively described in the literature. RT Vtg has been measured using an ELISA with monoclonal antibodies raised against AS Vtg (Tyler *et al.*, 2002). This work is particularly interesting because it shows no significant difference when the same samples are analyzed with a homologous polyclonal RT ELISA. Another study showed variable results in cross reactivity when comparing ELISAs from greenback flounder (*Rhombosolea tapirina*), RT and AS. Also, fathead minnow (*Pimephales promelas*) Vtg has been quantified using an ELISA based on carp (*Cyprinus carpio*) (Tyler *et al.*, 1999).

The quest for measuring Vtg has even gone beyond traditional chemical and physical analytical techniques. A specific immunosensor for carp (*Carassius auratus*) has been developed by coating a conducting polymer electrode with anti Vtg monoclonal antibodies. The antigen-antibody interactions were measured by means of a quartz crystal microbalance and impedance spectroscopic techniques. The sensor showed a dynamic range from 1 to 8 µg/L with a 0.41 µg/L LOD and no cross reaction with non-target proteins such as IgG, horseradish peroxidase and bovine serum albumin (Darain *et al.*, 2004).

In spite of the many advantages offered by immunoassays, certain issues need to be considered before starting to develop a new technique. The first step in the process is obtaining a highly purified homogenous standard. In the case of Vtg, extra precautions should be taken since this is a thermo-labile and susceptible to the action of proteases. Purification of this protein usually yields degraded or aggregated species (Kanungo *et al.*,

1990; Tao *et al.*, 1993; Silversand *et al.*, 1993; Specker and Sullivan, 1994). Some researchers have avoided these problems by raising the antibodies against purified yolk proteins (Idler *et al.*, 1979; Campbell and Idler, 1980; Kunkel and Nordin, 1986; Brown *et al.*, 1997b). The purification of this high quality standard is critical for the following step, the production of polyclonal or monoclonal antibodies. This is probably the most labor intensive and important step. Increased non-specific binding and background noise are the consequence of poor protein purification.

Once issues regarding cross reactivity and false positive results are resolved, the technique must be validated using a standard. Here again, the issue regarding a reliable standard plays a key role. Usually, Vtg is purified after an extensive laboratory procedure. The concept of 'pure' is subject to many interpretations according to the analytical procedure performed. The amount of protein obtained can be calculated by simple weight or by assessing concentration via colorimetric protein assays or quantitative amino acid composition analysis. Colorimetric protein assays also raise the issue as to what standard should be used for calibration. It is usually assumed that the protein to be analyzed will behave the same way as the standard, which usually is bovine serum albumin, ovoalbumin or immunoglobulin G, however, this is usually not the case for Vtg. Moreover, the non-proteinaceous parts of Vtg should also be taken into account for the final calculations. In spite of the variety of PTM described for this protein, it is usually accepted that the protein part of the Vtg molecule accounts for 65-75% of the total molecular weight.

After validation, the technique must be tried on serum or plasma from real samples, during which the dynamic range of Vtg concentrations found in fish creates

another problem. These values can vary from the ng/mL of male or immature females to the mg/mL found in fully mature female RT (Wallaert and Babin, 1994). For this reason serial dilutions for each unknown sample must be assayed in order to obtain signal values that fall within the optimum working range of the technique.

### **1.7 Purpose and significance of study**

The objective of this study was to develop a novel ‘signature peptide’ approach for quantification of Vtg using combined techniques of mass spectrometry. For this purpose, a diagnostic proteolytic peptide was chosen as a surrogate of the precursor protein (Vtg) and monitored by HPLC-ESI-MS/MS for quantification. To attain this objective, this study was organized in three main parts. A summary of all the work involved in this thesis is portrayed in the scheme shown in Figure 1.2.

The first part consists of the characterization of Vtg from the three studied species by different mass spectrometric techniques and involves the highly careful and descriptive experiments aimed at exploring the primary structure of Vtg. The information obtained from these experiments was essential to elaborate a list of ‘signature peptides’ for each species, necessary to conduct the subsequent experiments.

The second part of this study was mainly methodological and consisted of optimizing and tuning the analytical instrumentation used for the quantitative analysis. The experimental conditions of high performance liquid chromatography coupled to the

tandem mass spectrometer were optimized to enhance the level of detection and the reproducibility capabilities of the system.

Finally, the analytical system was used to assay real biological samples. Vtg levels were compared in fish prior to and following estrogenic induction with 17 $\beta$ -estradiol. RT, AS and AC were the species chosen for this study. The quantification technique was set and completed for RT and AS Vtg. AC Vtg was characterized, and a 'signature peptide' was also generated for this species, although no quantitative studies were performed.

These three species are considered priority species at the NAFC, due to their important role in Atlantic Canada. RT and AS are used in aquaculture, while AC is a key species in the fisheries industries. Whether caged or in wild, these species are essential to maintain the fragile ecosystems found in this region. The development of this quantitative technique provides researchers with an alternative for measuring Vtg in plasma samples.

The approach explored in this work offers an alternative to developing immunoassays, especially when no commercial kits are available for a particular species. The 'signature peptide' approach circumvents the need of protein purification and antibody production, two time consuming and costly steps required for immunoassay development.

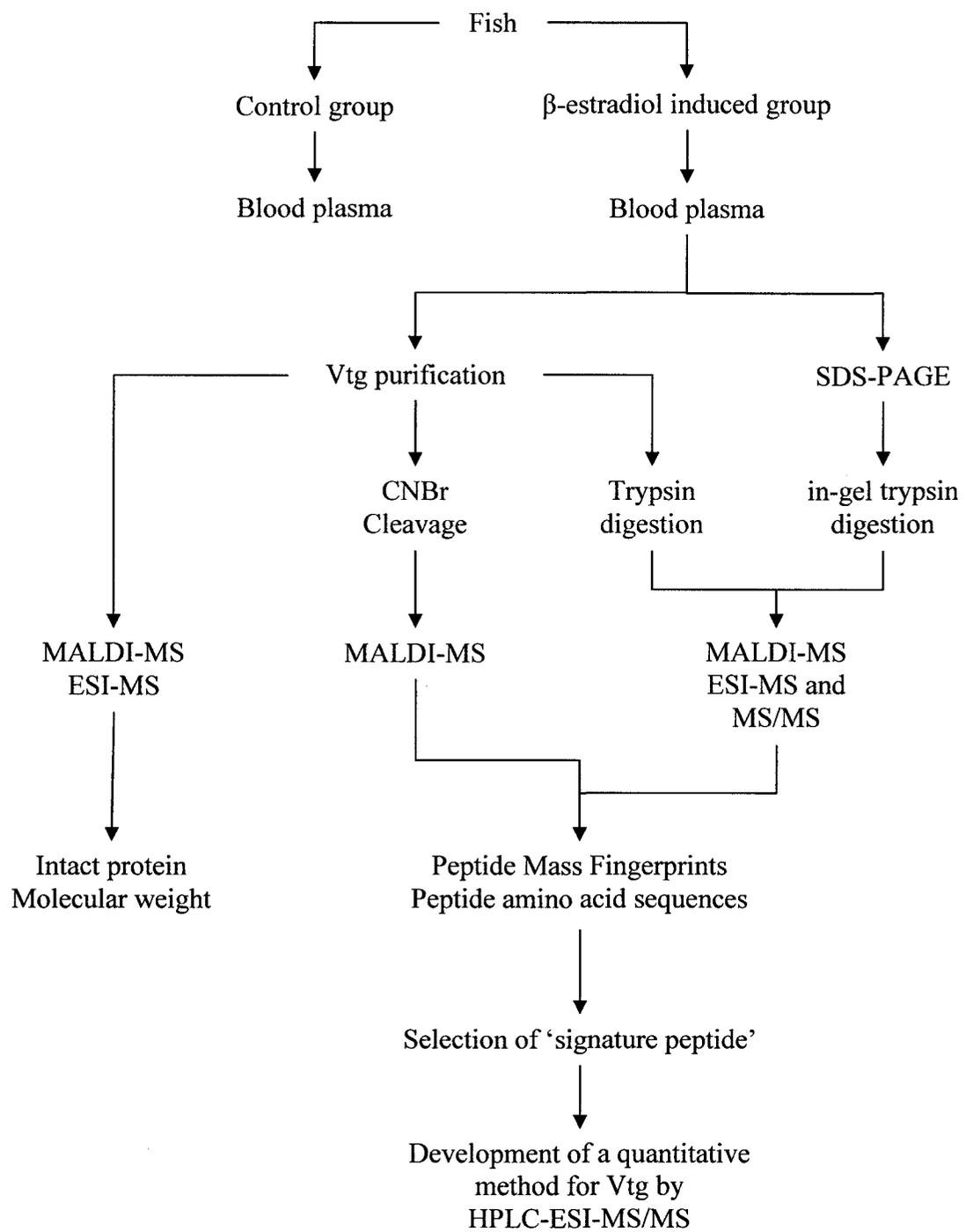


Figure 1.2 Scheme of the global experimental design

## **Chapter 2: Application of mass spectrometry to protein chemistry**

The experimental work done in this thesis relies mainly on mass spectrometric analyses. This powerful tool provides researchers with qualitative information regarding the nature of the samples studied. Coupled to chromatographic techniques, mass spectrometry also provides valuable quantitative information.

Until recently, mass spectrometry was exclusively employed for the analysis of inorganic ions or small volatile organic compounds. However, the development of electrospray ionization (ESI) and matrix-assisted laser desorption/ionisation (MALDI), two gentle non-destructive ionization sources, has permitted this technique to be employed for large biomolecules, including proteins and DNA.

This chapter is not intended to be a detailed tutorial on mass spectrometry. A brief description of the outstanding features of this technique is given to provide the concepts necessary to understanding and interpreting the results obtained from the diverse mass spectrometric methods used in this work.

### **2.1 The mass spectrometer**

Over the past two decades, mass spectrometry has evolved at a fascinating rate. Analysis speed, mass resolution and sensitivity levels have increased, while costs have continuously decreased. Regardless, the basic configuration of the mass spectrometer remains unchanged. No matter the degree of sophistication, all mass spectrometers

consist of four basic modules: an ionization source, the mass analyzer, the detector and the data integrator.

The ionization source, as its name indicates, produces ions from the analytes of interest. Ionization sources were traditionally under vacuum, however new sources that operate at atmospheric pressures are now available. Examples of low pressure or vacuum sources are electron ionization (EI), chemical ionization (CI), fast atom bombardment (FAB) and MALDI (Muddiman *et al.*, 1995). On the other hand, ESI, atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI) are some examples of ionization sources that work under atmospheric pressure (Bruins, 1991). There is also a modification of the MALDI source that can be operated at atmospheric pressure.

The mass analyzer is the core component of the mass spectrometer. It is responsible for separating or sorting out the ions based on their mass to charge ( $m/z$ ) ratio. Strictly speaking, mass spectrometers should really be called mass to charge spectrometers. The theory behind them is quite complex and goes beyond the scope of this introduction, however, detailed descriptions of each one can be found in the literature. Mass analyzers are classified according to the physical principles employed for ion separation. Magnetic sectors, electric sectors, quadrupoles (Dawson, 1986), ion-traps (Todd, 1991; March R.E., 2000; Douglas *et al.*, 2005), time-of-flight (Wollnik, 1993; Guilhaus, 1995) and ion cyclotron resonance analyzers (Köster *et al.*, 1992; Marshall A.G. *et al.*, 1998) are examples of these.

One of the parameters measured for each analyzer is the resolution, or the resolving power. Literature varies on the definition of this parameter, sometimes referring to it as equivalents. Resolution (R) is generally defined as:

$$R=m/\Delta m$$

Where  $m$  is the mass at the center of the peak and  $\Delta m$  is the width of the peak at half height. According to their R, mass analyzers are usually classified as low resolution or high resolution mass spectrometers, although this definition is quite subjective and has changed for the same analyzers with the development of more powerful instruments. For example, ToF analyzers are currently working at  $R\sim 10000-15000$ , while the latest ICR-FT-MS are capable of  $R\sim 300000$  and higher. However, caution should be taken when comparing R because this parameter might vary, even for the same instrument, along the  $m/z$  range.

The detector simply transforms the arrival of the ions into a measurable signal which is finally acquired and processed by the data integrator. This last component is nowadays a computer loaded with analytical software used to assist the operator in the interpretation of the data.

As mentioned earlier, until the last decade, mass spectrometry was exclusively used for the analysis of small volatile organic molecules and inorganic ions. The major breakthrough in this area arrived with the invention of ESI and MALDI. Since then, large biomolecules have been amenable to mass spectrometric analysis and this has been the starting point for a new era in protein biochemistry and proteomics. These ionization

sources in conjunction with mass analyzers such as the time-of-flight, permit the detection of macromolecules that go well beyond the mega Dalton mass range.

## **2.2 Ionization sources for macromolecules**

### **2.2.1 Electrospray ionization**

This ionization source was introduced by John Fenn, who received the 2002 Nobel Prize in chemistry for his invention (Fenn *et al.*, 1990). Surprisingly, this ionization technique is by far one of the simplest to understand. Samples are usually dissolved in a buffer or solvent which is introduced into the mass spectrometer in the form of a spray (Smith *et al.*, 1991). A high voltage electric field is applied between the tip of the nebulizing needle and the inlet (skimmer) of the analyzer. A carrier gas, usually air or nitrogen at elevated temperatures, is used to assist in the formation of the spray. This environment aids in the evaporation of the solvent, reducing the size of the micro droplets. The charged particles within the droplets are brought closer together until the repulsive electrostatic force expels them into the gas phase (see Figure 2.1)

A characteristic of ESI for proteins is that it usually produces multi-charged species, according to the nature of the analyzed sample. These species are represented as:  $[M+nH]^{n+}$  or  $[M-nH]^{n-}$ , where M represents the molecular species and n the number of protons (added or lost) and charges. The distribution of these species, i.e. the amount and charge, and the presence of adducts (eg.  $[M+Na]^+$ ) can be experimentally modified

by changing the composition of the solvent or the pH. Some structural protein characterization can also be deduced from the charge distribution of these multiple charged species (Grandori, 2003).

The production of multiple charged species also has an additional advantage. Their presence permits the analysis of high molecular mass biomolecules by lowering the  $m/z$  to a value compatible with the transmission range of the quadrupole mass analyzer.

### **2.2.2 Matrix-assisted laser desorption ionization**

Koichi Tanaka also shared the 2002 Nobel Prize in chemistry for his contribution in developing the MALDI source (Tanaka *et al.*, 1988). Nonetheless, much of the credit given to the development of this technique should also go to Michael Karas and Franz Hillenkamp, who developed the ideas of laser desorption techniques (Karas and Hillenkamp, 1988; Hillenkamp and Karas, 1990; Karas *et al.*, 1991; Hillenkamp *et al.*, 1991; Glückmann and Karas, 1999). This technique has permitted accurate mass measurements of peptides and proteins to be achieved with unprecedented success (Russell, 1997).

The principle behind this source is quite unique. The samples of interest are co-crystallized with a matrix compound onto a plate. This plate is placed in the low pressure compartment of the mass spectrometer. Ionization of the analyte is achieved by a laser beam that is aimed towards the plate (See Figure 2.2).

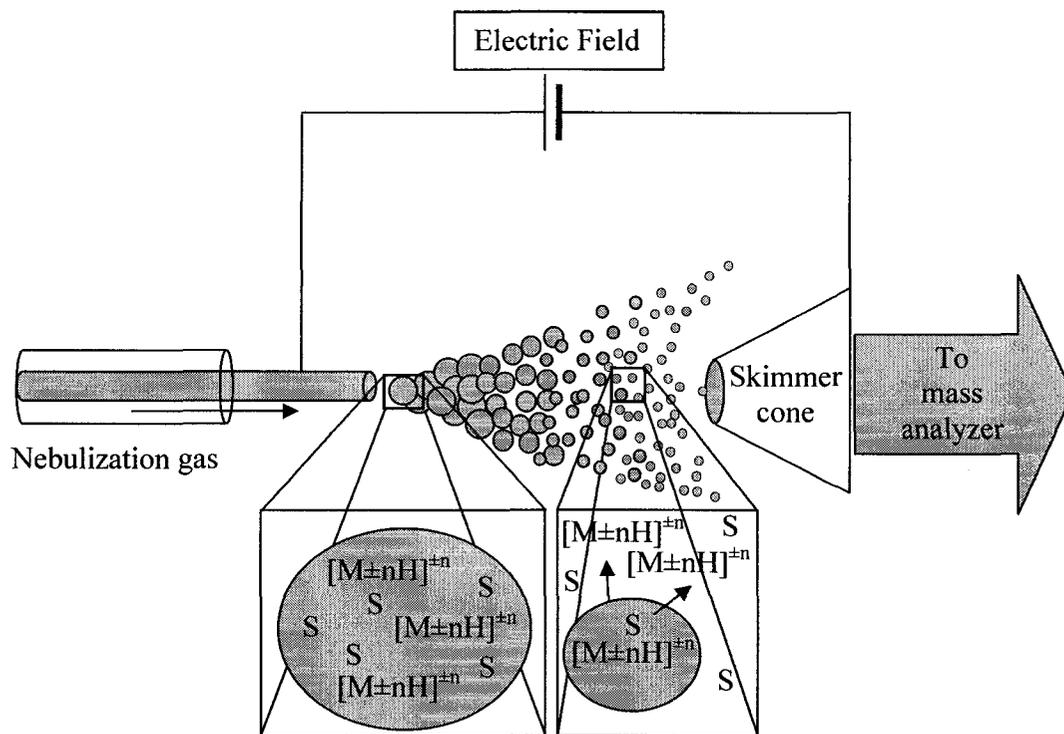


Figure 2.1 Basic principles of the electrospray ionization source  
Solvent and analyte molecules are indicated with the letters 'S' and 'M', respectively.

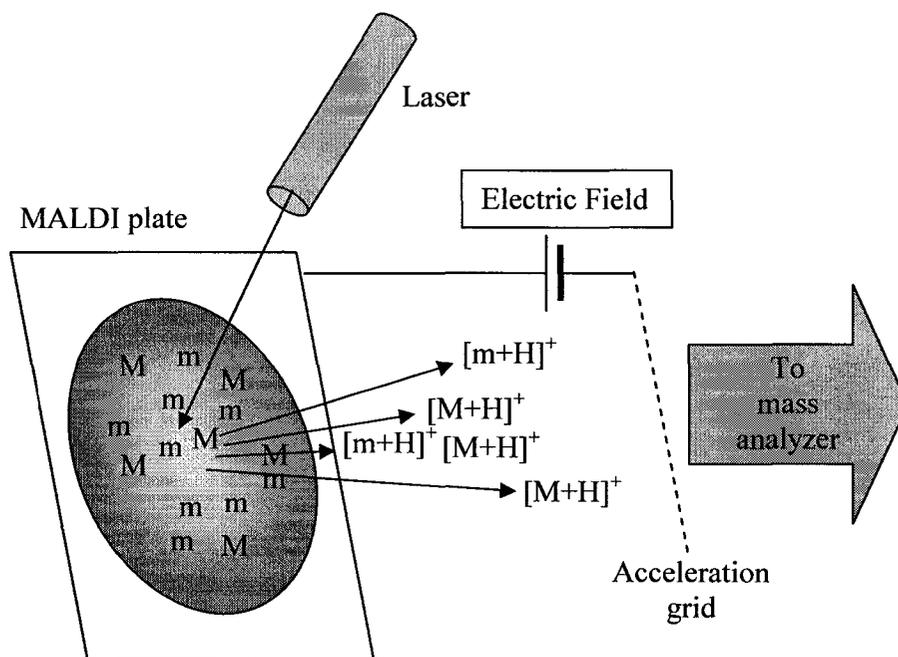


Figure 2.2 Basic principles of the MALDI ionization source  
Matrix and analyte molecules are indicated with the letters 'm' and 'M', respectively.

The exact principle of ionization is not clear yet, although it is generally accepted that the energy from the laser is transmitted to the matrix ions, which vaporize together with the analyte molecules. A transfer of protons towards the analyte molecules also occurs in this process. These are finally accelerated towards the analyzer.

An outstanding feature of the spectra produced by MALDI analysis is that the singly charged species  $[M+H]^+$  are the most prevalent ions present, even for large biomolecules such as proteins and DNA. Many hypotheses have been proposed to explain this remarkable observation, yet none seem to be conclusive (Karas *et al.*, 2000).

### **2.3 Tandem mass spectrometry**

As its name suggests, tandem mass spectrometry (MS/MS) is the result of performing two or more sequential separations of ions, usually coupling two or more mass analyzers. Some mass analyzers, such as the quadrupole ion traps can perform this task within the same mass analyzer. When coupled to ESI, tandem mass spectrometry results in a very powerful tool for a wide range of biochemistry applications (Griffiths *et al.*, 2001).

The power of MS/MS is due to the capability of producing a fragmentation step between the mass analyzers. This is usually done in a collision cell filled with an inert gas situated between the analyzers (See Figure 2.3), in a process called collision induced dissociation (CID) (Shukla and Futrell, 2000).

The configuration of the tandem mass spectrometer allows different types of experiments to be performed. The one most commonly used in this work is the product ion scan. In this mode of operation, the first mass analyzer is set to a mass range, usually fixed to a window of 1 or 2 Da, thus acting as a filter. In this way, a precursor ion is selected for CID. The products obtained from the collision are then monitored in the second mass analyzer. The product ion spectrum is characteristic of the precursor selected and from this spectrum important structural features can be inferred. Most of the experiments performed in this thesis using tandem mass spectrometry were carried out on a QqToF-MS/MS configuration (Chernushevich *et al.*, 2001).

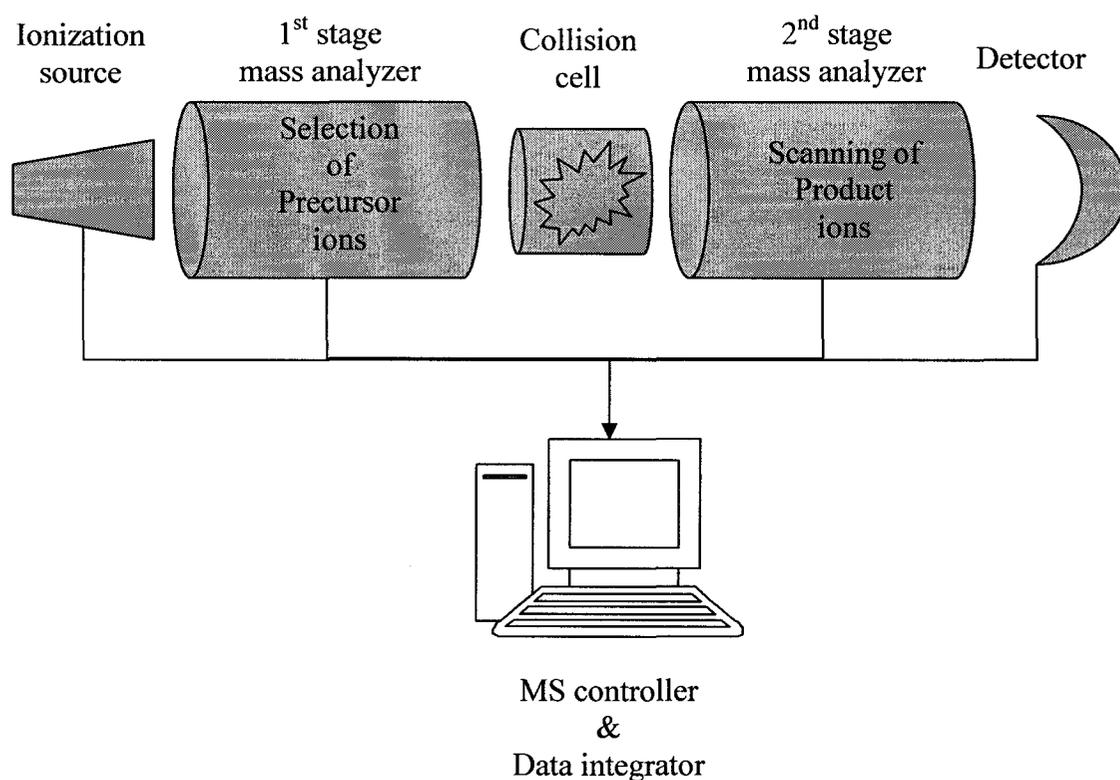


Figure 2.3 Basic components of the tandem mass spectrometer

## 2.4 Mass spectrometry in proteomics

The proteome is defined as the set of all expressed proteins in a cell, tissue or organism. In contrast to the genome, the proteome is dynamic, even for the same cell type. Recent advances in MS and bioinformatics have permitted the systematic analysis of all expressed cellular proteins. In this way, proteomics has been revolutionized by MS and MS/MS, which have enabled the identification of proteins in complex mixtures at a high output rate (Yates, III, 1998; Gygi *et al.*, 2000; Gygi and Aebersold, 2000; Godovac-Zimmermann and Brown, 2001; Aebersold and Goodlett, 2001; Griffin and Aebersold, 2001; Peng and Gygi, 2001; Gygi *et al.*, 2002; Jennie, 2003; Zimmer *et al.*, 2006). An integral and comprehensive description of mass spectrometric techniques applied to proteomic studies is available (Kinter and Sherman, 2000), and this book was used as the source of information for most of the the methods and techniques used in this project.

Different strategies have been developed for protein identification (Hernandez *et al.*, 2006). The ‘top/down’ and the ‘bottom/up’ are the two general approaches used in proteomics for identification and characterization purposes (Bogdanov and Smith, 2005). The ‘top/down’ approach is very straightforward. The intact protein molecules are introduced into the mass spectrometer, in which a series of fragmentation experiments are performed. The analyses of these spectra permit complete or partial sequencing to the primary structure of the protein. This information may be used for protein identification if the sequence is available in the protein databases, or simply for characterization purposes. Unfortunately, the ‘top/down’ approach is generally limited to the very costly ion-cyclotron resonance MS/MS.

The 'bottom/up' approach is used more extensively in laboratories involved with proteomics projects. This approach uses a chemical or enzymatic digestion step prior to MS analyses (Jonsson *et al.*, 2001; van Montfort *et al.*, 2002). Within the 'bottom/up' approach there are basically two strategies to follow. Protein semi purification can be done by one or two dimension gel electrophoresis, followed by an in-gel digestion step. The digest is analyzed by MALDI-MS to produce a characteristic mass spectrum called the peptide mass fingerprint. A list of peaks is obtained from this spectrum to perform a database search, from which a list of candidate proteins is retrieved.

Alternatively, the digestion can be done directly on the protein mixture. Each of the resulting peptides is selected for CID-MS/MS. A product ion spectrum is acquired for each precursor peptide and analyzed. Product ion spectra of peptides show characteristic peaks corresponding to peptide ions which are cleaved at the amide bond during CID, which are named b and y ions.

The analysis of these ions permits the total or partial amino acid sequence to be deduced and this information is then used for protein identification via database searches. A schematic diagram of this process is shown in Figure 2.4. It is also possible to correlate the uninterpreted tandem mass spectra of peptides produced under low energy collision conditions with amino acid sequences in the protein databases (Eng *et al.*, 1994). A summary of the most common methods employed in the 'bottom-up' approach for protein identification in proteomics is depicted in Figure 2.5.

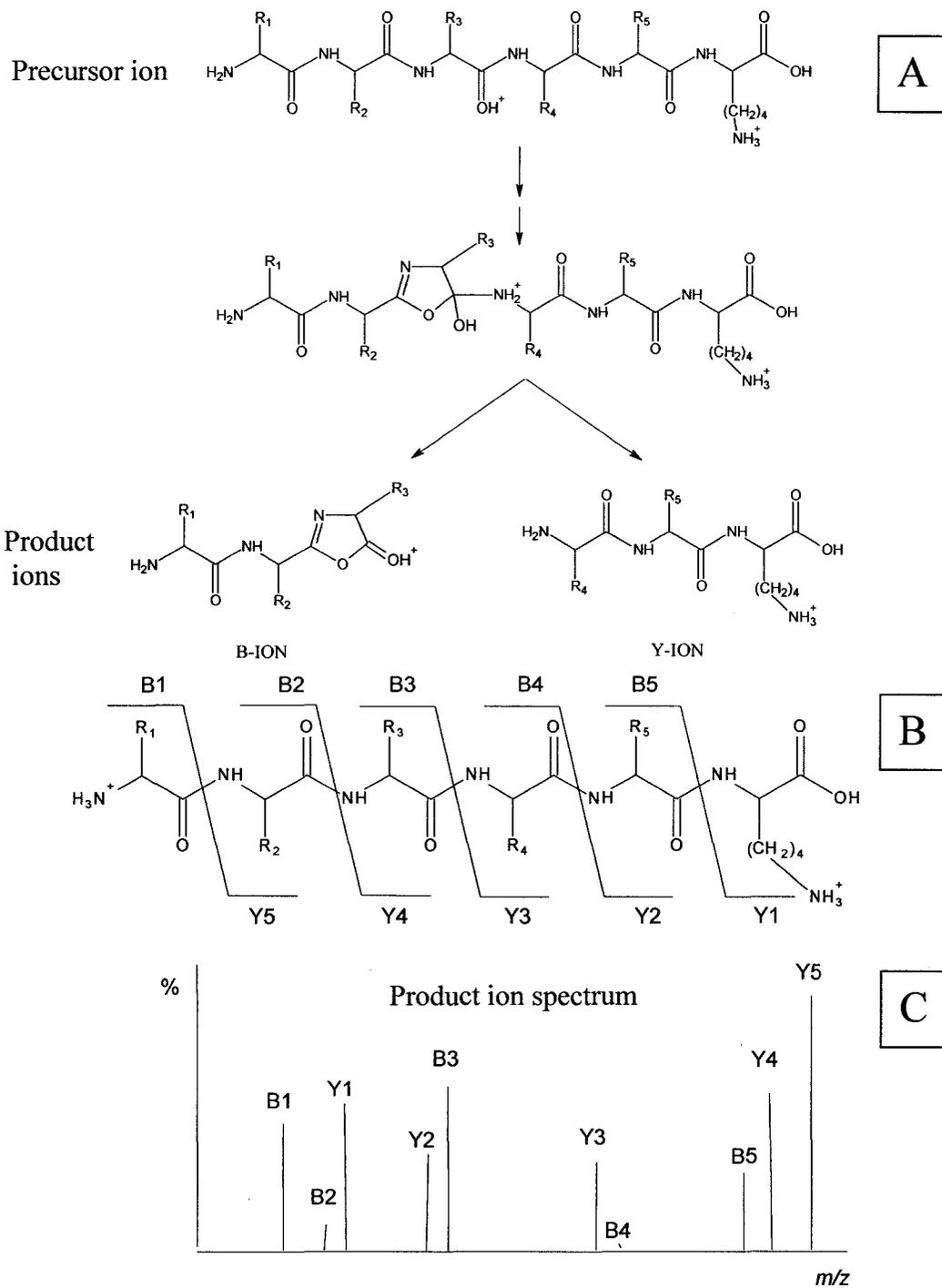


Figure 2.4 Collision induced dissociation of a peptide ion precursor

The top scheme (A) shows the chemical fragmentation route of a peptide ion precursor leading to the formation of the corresponding B and Y ions. The center scheme (B) shows all the possible B and Y ions generated from a generic peptide chain. The bottom scheme (C) shows a schematic representation of a product ion spectrum obtained from the peptide described above.

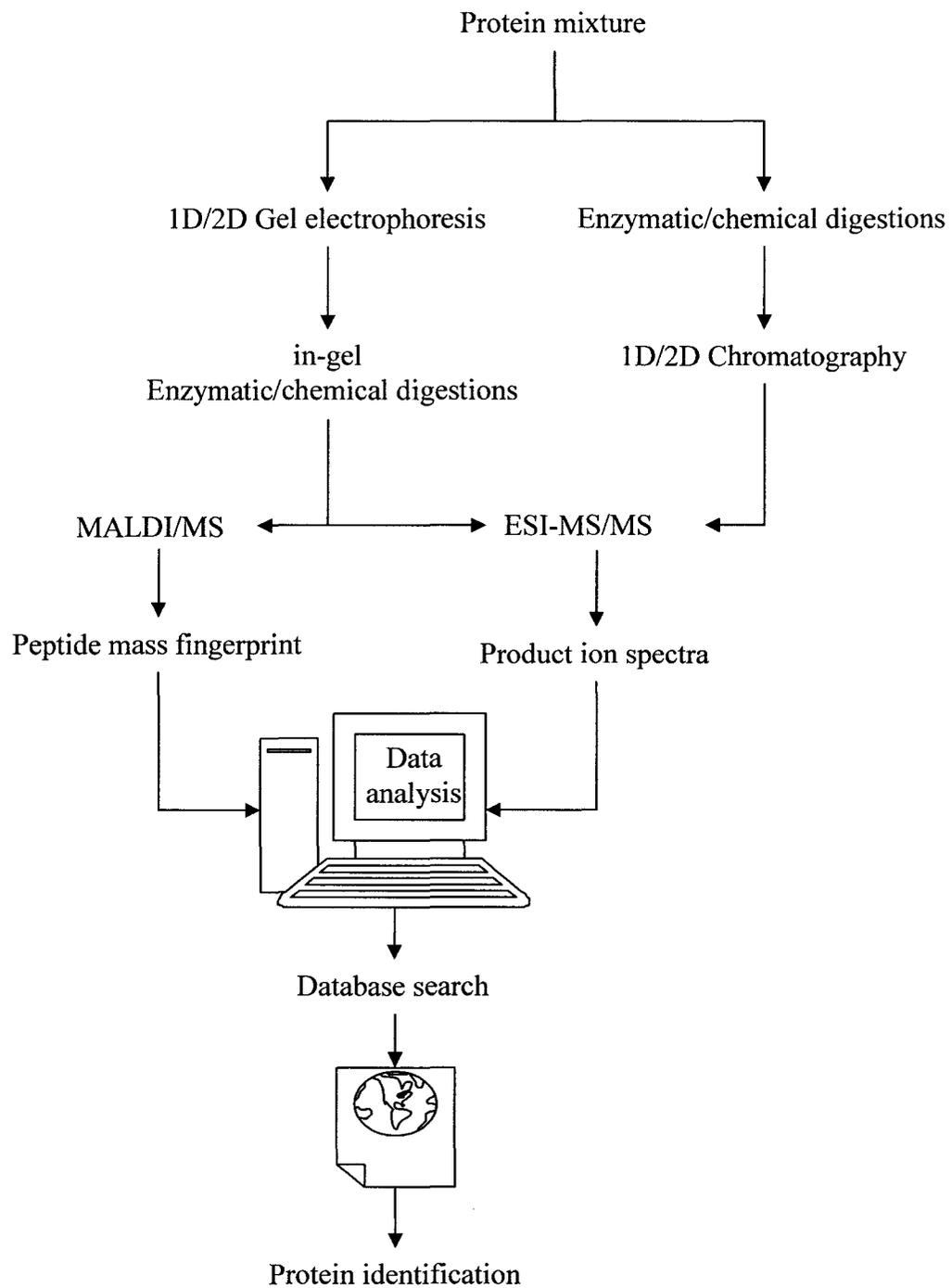


Figure 2.5 Classical 'bottom-up' proteomic approach for identification of proteins

## **Chapter 3: Characterization of vitellogenin by mass spectrometry**

The starting point of this work was based on the structural analyses of Vtg by various mass spectrometric techniques. To provide enough starting material, Vtg synthesis was induced *'in vivo'* to obtain high plasma concentrations of this protein. Details of this procedure can be read in Section 3.1.1. Purification of RT Vtg was carried out by a series of filtration steps, size exclusion and ion exchange chromatography. Atlantic salmon Vtg was purified by size exclusion chromatography. These methods are described in detail in Section 3.1.2. Determination of intact protein molecular masses was performed by MALDI-ToF-MS. The partial amino acid sequence of Vtg was obtained by MS/MS analysis of the digested protein. The purified Vtg was subjected to solution digestions, while in-gel digestions were performed on the plasma samples. The results of these experiments are given in the following sections.

### **3.1 Materials and methods**

#### **3.1.1 Vitellogenin induction protocol**

Thirteen juvenile RT (purchased from Rainbow Springs Hatchery, Thamesford, On, Canada), ten adult AS and twelve adult AC were kept in separate 60 gallon tanks. The RT were maintained with running fresh dechlorinated city water, while the AS and AC were kept with running sea water. All fish were fed *'ad libitum'* with commercial

pellets (Vigor Trout/Char Grower 4 Corey Aquaculture fish feed, Corey Feed Mills Ltd., Fredericton, NB, Canada).

A twelve hour light-darkness cycle was programmed for each of the fish tanks. All fish were identified by intra-muscular injection of a radio tag under the dorsal fin. After one month of acclimatization, experimental fish were injected intraperitoneally four times, at one week intervals, with a dose of 10mg of  $17\beta$ -estradiol (Sigma, Prod. No. E8875. Sigma, St. Louis, MO, USA) per kilogram of fish body weight using a 10 mg/ml solution in oil/acetone (9:1, v/v). Control fish were injected only with the oil/acetone solution.

One week after the last  $17\beta$ -estradiol injection, fish were anesthetized using tricaine methanesulfonate (TMS) and bled from the caudal vein using heparinized syringes. Blood samples from each fish were divided in two 1mL aliquots. Twenty microliters of a 4 TIU/mL aprotinin (Sigma, Prod. No. A1153) solution were added to one of the aliquots. All samples were collected in 1mL plastic tubes on ice and centrifuged at 800 g (3000rpm on a Beckman GS-15R Centrifuge. Rotor F2404 Head. Beckman-Coulter, Fullerton, CA, USA) for 5 minutes at 0°C. Plasma was separated and kept frozen at -80°C until further analysis. All live fish experimentation was approved by the Animal Care Committee of the Northwest Atlantic Fisheries Center and the procedures were conducted according to its guidelines.

### **3.1.2 Atlantic salmon and rainbow trout vitellogenin purification protocol**

Extraction and purification of RT Vtg was done in Dr. Heeley's laboratory at the Department of Biochemistry, Memorial University of Newfoundland. One milliliter of plasma was added to an equal volume of a 50 mM TRIS, 0.5 M NaCl, 10 mM EDTA, 100 TIU/L aprotinin, pH 8.0 at room temperature, then mixed and centrifuged at 16,000g (14,000rpm on an Eppendorf 5415D Microcentrifuge. Hamburg, Germany) for 3 minutes. The supernatant was loaded onto a Sephacryl S300 column (120 cm x 1.5 cm) at 4°C and 2.5 mL fractions were collected at a flow rate of 30mL/hr, using the same buffer as described above. The protein content of each fraction was assayed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Electrophoresis was conducted using 8% (m/v) polyacrylamide of at a constant voltage of 80V for the first 10 minutes, and 180V for the following 50 minutes. The fractions 74 to 132 contained a band compatible with that of Vtg and were pooled together. This band corresponded to the over-expressed band in 17 $\beta$ -estradiol treated fish.

Clean-up of the pooled fractions was done by ultrafiltration using an Amicon Y100 membrane to remove proteins of less than 100 kDa. The sample was then equilibrated against 50 mM TRIS, 6 M Urea, 1 mM EDTA, 0.5 mM dithiothreitol (DTT), pH 8,0 at room temperature, and subjected to Ion Exchange Chromatography using an 80mL bed volume Fast Q Sepharose column. Fast-flow proteins were collected using a salt gradient of 0.0 to 0.4 M NaCl at a flow rate of 40mL/hr in 3 ml fractions. The collected fractions were analyzed by SDS-PAGE in the conditions described above (Laemmli, 1970).

Fractions 88 to 120, which contained the purified Vtg, were pooled and dialyzed extensively against distilled water and finally freeze-dried.

Atlantic salmon Vtg had been originally purified by Dr. Idler<sup>†</sup> from Memorial University of Newfoundland for the the NAFC, according to So *et al.*, 1985. Briefly, 5 mL of plasma was thawed and diluted with an equal volume of 50 mM TRIS, 0.5 M NaCl, 10 mM EDTA, 3% aprotinin, pH 8.0 at room temperature and centrifuged at 0°C. The supernatant was then chromatographed at 4°C on a 2.6 x 86 cm column of Sephacryl S-300 which had been previously equilibrated with the same buffer used to dilute the plasma. Fractions were collected at 1.5 mL per tube at a flow rate of 15 mL/h. Tubes with absorbance at 280 nm were assayed by SDS-PAGE and Coomassie blue staining to locate Vtg elution. Those containing Vtg, eluting between fractions 140 and 160 were pooled and dialyzed extensively against distilled water and finally freeze-dried and stored at the NAFC for further analyses.

### **3.1.3 Solution digestion procedures**

The trypsin solution digestion of Vtg was carried out in the laboratory of Dr. Pierre Thibault, University of Montreal, QC, Canada. Five hundred µg of purified protein was reduced and alkylated. The samples were previously denatured in 500 µL of 6.0 M guanidine/HCl, 0.1 M Tris-HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.5 at room temperature. Argon gas was gently blown over the solution surface for 10 min. Reduction was carried out by the addition of 50 µL of 4 mM DL-DTT and incubation for

2 hours at 37°C. Alkylation was performed by adding dropwise 40 µL of 50 mM iodoacetamide to the mixture, followed by incubation in darkness for 1 hour at 37°C. The excess iodoacetamide was quenched by adding 1 µL of beta-mercaptoethanol and subsequently dialyzed against 500 mL of 50 mM ammonium bicarbonate for a minimum of 24 hours with periodic change of dialysis buffer. The solution was then lyophilized and redissolved in 0.5 mL of a 0.1 M ammonium bicarbonate and 0.1 mM calcium chloride (at pH 8.0) solution. The tryptic digestion was carried out at 37°C for 12 hours using a substrate-to-enzyme ratio of 25:1 (w/w). When not analyzed immediately, samples were stored at -20°C for subsequent analysis.

The cyanogen bromide degradation of purified rainbow trout vitellogenin was performed according to the following procedure: 10 mg of purified protein were dissolved in 1 mL of 70% (v/v) formic acid. 1 crystal (approximately 5 mg) of CNBr was added, swirled until dissolution and left 12 hours in the dark under a fume hood. 15 mL of deionized water were added. The solution was frozen in liquid nitrogen and finally lyophilized until further analysis.

#### **3.1.4 In-gel digestion procedures**

Separation of Vtg by SDS-PAGE (Laemmli, 1970) and subsequent in-gel digestion was performed following the procedure described in detail by Kinter *et al.* with minor modifications (Kinter and Sherman, 2000). Five microliters of thawed plasma from both experimental and control fish were mixed and loaded onto a 6% (w/v)

polyacrylamide SDS-PAGE and analyzed in the same conditions as those described in Section 3.1.2. The gel was stained using the colloidal Coomassie blue protocol. The band corresponding to the induced Vtg was easily recognized by comparison of the protein banding from both control and experimental fish. This band was cut into 1mm<sup>3</sup> pieces using surgical blades and placed into 2 mL plastic vials for enzymatic digestion. An unstained piece of gel from one corner of the gel slab was simultaneously processed for use as a blank. Samples were destained with serial washes of 50% (v/v) methanol and 5% (v/v) acetic acid over a period of 24 hours. These steps are critical for removing the excess of SDS and salts in the gel pieces which interfere in most MALDI and ESI analysis. Subsequently, the gel pieces were reduced with 30µL of a 10 mM solution of dithiothreitol (DTT) (Sigma, Prod. No. D9163) and alkylated with 30µL of a 100mM iodoacetamide (Sigma, Prod. No. I1149) solution. Finally, the gel pieces were dehydrated by the addition of 200µL of acetonitrile. The gel pieces were further dried in a vacuum centrifuge at room temperature.

Enzymatic digestion was performed by adding 30µL of a 20ng/mL solution of trypsin (Sigma, Prod. No. T6567) to the dried gel pieces and letting the tubes stand in a water bath at 37°C for 24 hours. The tryptic peptides were extracted three times using 30 µL of a 50% v/v solution of acetonitrile containing 5% v/v formic acid. The final volume was reduced to <20 µL by evaporation in a vacuum centrifuge at ambient temperature. Sample volume was finally adjusted to ~20 µL with 1% acetic acid. When not analyzed immediately, samples were stored at -20°C for subsequent analysis.

### 3.1.5 HPLC-ESI-MS and MS/MS acquisition parameters

Chromatographic separation of tryptic peptides for sequencing analysis was achieved using a HP1100 (Hewlett Packard, Palo Alto, CA) high performance liquid chromatograph equipped with a 15 cm x 0.32 mm PepMap capillary column (LC Packings, San Francisco, CA, USA) using a binary linear gradient elution of 5 to 95 % (v/v) acetonitrile in 35 min (Complementary solvent: 0.2 % formic acid). A 100:1 flow splitter was mounted before the injector such that only a 3-4  $\mu\text{L}/\text{min}$  flow rate was introduced in the capillary column. The sample was manually injected via a Rheodyne six-port valve (Rohnert Park, CA, USA) fitted with a 5  $\mu\text{L}$  sample loop

A Micromass ESI-qTOF instrument coupled in-line to the HPLC was used to analyze the tryptic digest. The mass spectrometer was operated in electrospray positive ion mode with a source temperature of 80°C, a desolvation temperature of 80°C, desolvation gas flow rate of 300 L/hr and a potential of 3.5 kV applied to the capillary. MS data was acquired and integrated every 2 seconds. Conventional mass spectra were obtained by operating the quadrupole in a RadioFrequency (RF)-only mode while a pusher electrode was pulsed to transfer all ions to the time-of-flight analyzer. Data was acquired and processed by the Mass Lynx Window NT based data system.

Tandem mass spectrometric experiments were performed with on a Micromass ESI-qTOF instrument coupled to the HPLC system described above. The mass spectrometer was operated in electrospray positive ion mode with a source temperature of 80°C, a desolvation temperature of 80°C, desolvation gas flow rate of 300 L/hr and a potential of 3.5 kV applied to the capillary. Product ion spectra of mass selected ions

were induced by CID using argon as the collision gas with a 25 V offset between the DC voltage of the entrance quadrupole and the RF-only hexapole cell. MS spectra were obtained over a mass range varying from  $m/z$  500 to 1700. Mass spectral resolution was typically 4,000-5,000. For MS/MS experiments, precursor ions were automatically selected from the MS survey by running an Information Dependent Acquisition program (IDA). Selection of precursor ions was performed by the first quadrupole while a pusher electrode was pulsed to transfer fragment ions formed in the RF only hexapole cell to the TOF analyzer, which scanned the fragment ions in a range of  $m/z$  50 to 2000. A scan duration of one and two seconds was set for conventional and MS/MS mass spectral acquisition, respectively. Data was acquired and processed by the Mass Lynx, Windows NT based data system.

### **3.1.6 MALDI-MS and MS/MS acquisition parameters**

The purified intact rainbow trout and Atlantic salmon vitellogenin protein was analyzed by MALDI-MS. Ten milligrams of purified sample of intact protein was dissolved in 1mL water, desalted using Millipore VS membrane filters and mixed with equal volumes (1:1, v/v) of a 10 mg/mL solution of sinapinic acid matrix . The sample was then analyzed by a MALDI-TOF System (PE Voyager Biospectrometry Workstation), operating in the linear mode using the Post Acceleration Detector (PAD) (also known as Delayed Extraction Technology). External calibration was done using BSA (Bovine Serum Albumin).

The CNBr and tryptic solution digestions of RT and AS Vtg were analyzed by MALDI-MS. Mass spectra of digests were acquired on a PerSeptive Biosystems Elite-STR (Framingham, MA, USA) equipped with delayed extraction technology. One  $\mu\text{L}$  aliquots of both tryptic and CNBr digests were mixed with 2  $\mu\text{L}$  of a saturated solution of 2,5-dihydroxybenzoic acid (2,5-DHB) MALDI matrix, originally prepared as 0.2 M in methanol:deionized water (50:50, v/v). 0.5  $\mu\text{L}$  aliquots of this final solution were applied to the MALDI target and allowed to evaporate at room temperature prior to mass spectral analysis. Reflectron mode was utilized to enhance mass resolution. For the analysis of large peptide fragments, 1  $\mu\text{L}$  of sample was mixed with 10  $\mu\text{L}$  of sinapinic acid (SA) matrix solution, prepared as 10 mg/mL in a mixture of methanol:acetonitrile:water (1:1:1, v/v/v). An aliquot of 0.5  $\mu\text{L}$  was deposited on the MALDI plate. External mass calibration was carried out using the singly-protonated ions of insulin and apo-myoglobin and the mass spectra were acquired using linear mode.

The in-gel digestions of RT and AS VTG were analyze by MALDI-MS. Mass spectra of digests were acquired on a QSTAR-XL (Applied Biosystems, Framingham, MA, USA) quadrupole-time-of-flight (QqToF) hybrid tandem mass spectrometer equipped with a UV nitrogen laser MALDI source (oMALDI 1, from Applied Biosystems, Framingham, MA, USA). Ten  $\mu\text{L}$  of the peptide extract belonging to one of the experimental fish and 10  $\mu\text{L}$  of 2,5-dihydroxybenzoic acid (2,5-DHB) MALDI matrix were mixed together, spotted on the MALDI plate in 1  $\mu\text{L}$  aliquots and allowed to evaporate at room temperature prior to mass spectral analysis. Upon optimization, typical acquisition parameters for the full scan positive ion mode were: Laser frequency 25Hz, Laser power 30%, declustering potential (DP) -5V, focusing potential 255V and DP2

10V. Spectra were acquired over a one minute period by manually centering the laser beam on the 'sweet spots' of the crystallized samples on the MALDI plate. All data acquired was processed on the AnalystQS software.

The MALDI-MS/MS of the same in-gel digestions were acquired on the same system as described above for the MALDI-MS for PMF. The acquisition parameters for the full scan positive ion mode were: Laser frequency 25Hz, Laser power 30%, declustering potential (DP) -5V, focusing potential 255V and DP2 10V The product ion spectra were obtained by manually 'ramping' the collision energy from 30V to 120V during the acquisition until the spectra showed an adequate pattern of complete fragmentation, shown by evenly distributed peaks across the complete *m/z* range being analyzed. Each acquisition lasted typically between 3 and 5 minutes. Argon gas an arbitrary value of 6 was used for the low energy collision induced dissociation (CID) MS/MS experiments. All data acquired was processed using the AnalystQS software.

### **3.1.7 Bioinformatics tools**

#### **3.1.7.1 Molecular weight calculation and isotopic cluster distribution**

The relative molecular mass and the isotopic distribution pattern of Vtg were calculated using an online internet calculator application available at the following address: <http://www.ch.cam.ac.uk/magnus/MolWeight.html>. The molecular formula used in the calculations was: C<sub>8138</sub>H<sub>13040</sub>N<sub>2196</sub>O<sub>2360</sub>S<sub>69</sub>. This formula was deduced from the RT

Vtg amino acid sequence found as entry Q92093 in the Swiss-Prot Database (Appendix 2).

### **3.1.7.2 Protein digestion simulation**

A computer based tryptic simulated digestion of rainbow trout and Atlantic salmon Vtg was performed with Micromass MassLynx v3.3 Protein Digestion Simulation Software. One missed cleavage was allowed and all the Cys residues were replaced by carboxiamidomethyl cysteine to take into account the alkylation process with iodoacetamide.

### **3.1.7.3 Peptide mass fingerprint database search**

Identification of the Vtg and its derived peptides was achieved using MASCOT (<http://www.matrixscience.com>), a free access internet search engine tool. For each search, a list of peptides was built from all the monoisotopic peaks which were clearly distinguished from the baseline noise (signal to noise >10) in each of the MALDI-ToF PMF spectra. The Swiss-Prot database was chosen for protein identification purposes. All search queries were completed considering a  $\pm 2.0$   $m/z$  of peptide mass tolerance and one missed cleavage for the matching parameters. No taxonomic constraints were selected. For the CNBr derived PMFs, homoserine lactone was introduced as a fixed modification.

For the tryptic digest, carboxyamidomethyl cysteine (Cys CAM) was set as the only fixed modification.

#### **3.1.7.4 'De novo' sequencing of tryptic peptides**

Interpretation of the ESI low energy CID-MS/MS obtained from product ion spectra of the tryptic peptides was performed using a computer based sequencing algorithm, called Lutefisk. In contrast to other software programs, this program can be obtained from its author at no cost. The results from this program were comparable to those obtained by other software (eg, Masslynx Protein Analysis Software, Micromass). However, processing time was considerably shorter using Lutefisk. Furthermore, the output file created by Lutefisk can be easily merged into CIDentify, a sequence homology program, described below. The sequencing parameters on this program were strictly optimized for the ESI-QqToF configuration, which was used for the experiments. The peptide mass tolerance was set to 0.50, the fragment ion tolerance to 0.25 and the final fragment error to 0.05. All cysteine residues were considered as carboxyamidomethyl cysteine for sequencing purposes.

Alternatively, the interpretation of the MS/MS spectra was performed with BioAnalyst, the sequencing algorithm provided with the Bioanalyst software (AppliedBiosystems, Foster City, CA, USA). The product ion spectra obtained from the MALDI-MS/MS of the tryptic in-gel digestions of AS, RT and AC Vtg were analyzed using particular program since Lutefisk is specially designed towards the interpretation of

ESI-MS/MS data. The precursor ion was automatically selected by the software. The peptide and fragment mass tolerance was set to 0.1 Da. All cysteine residues were considered as carboxyamidomethyl cysteines. In the candidate sequence list, leucine (L) and isoleucine (I) are both labeled with the letter L.

### **3.1.7.5 Sequence alignment of 'de novo' sequenced peptides**

CIDentify, a homology-based search algorithm program, was used to compare the sequenced peptides obtained from Lutefisk with the following sequences: the complete rainbow trout Vtg sequence on Swiss-Prot (See Appendix 2), together with the variant conflicts found in the four references reported for this protein (Le Guellec *et al.*, 1988; Mouchel *et al.*, 1996; Goulas *et al.*, 1996; Ren *et al.*, 1996), the two partial amino acid sequences of AS Vtg (See Appendix 2), the sequence of trypsin (Hermodson *et al.*, 1973) (entry P00761 on Swiss-Prot) and the aprotinin sequence (Creighton and Charles, 1987). The trypsin sequence was included to account for possible autolysis fragments. Aprotinin was used as a constituent of the buffers used in the Vtg purification process.

Alternatively, the alignment of the sequenced peptides obtained from the MALDI-MS/MS of the tryptic in-gel digestions of AS, RT and AC was performed using an internet application of FASTA (Compare Two Sequences with FASTA):

[http://fasta.bioch.virginia.edu/fasta\\_www/cgi/search\\_frm2.cgi](http://fasta.bioch.virginia.edu/fasta_www/cgi/search_frm2.cgi)

The sequenced peptides obtained from the in-gel digestions of RT and AS Vtg were compared to the RT and AS sequences. The peptides obtained from AC Vtg were

compared to haddock Vtg A and Vtg B (entries Q98T86 and Q98T87 on the TrEMBL database). For scoring purposes, the parameters chosen were Ktup=2, Scoring Matrix: Blosum 50, gap:-10 and ext: -2.

### **3.2 ESI-MS and MALDI-MS of intact rainbow trout and Atlantic salmon vitellogenin**

Attempts to measure the molecular mass of RT Vtg by ESI-MS did not show the expected molecular fragment around 180-200 kDa. However, a series of multicharged ions,  $[M+5H]^{5+}$  at  $m/z$  1303.01,  $[M+6H]^{6+}$  at  $m/z$  1086.22 and  $[M+7H]^{7+}$  at  $m/z$  931.26, were observed. The molecular mass of this fragment was transformed and deconvoluted into a single mass peak of 6510.8 Da by the Maximum Entropy algorithm available on the “Biolynx Program”. This peak was compatible with the phosvitin segment of RT Vtg, whose sequence is NGTRASSSSS SSSSSSRSS SSRSRKSR SSSSSSSSSS RISKRDGPFY YNPNDRKFKKN (Calculated Mr: 6511.76 Da), although its identity should be confirmed by MS/MS.

The results obtained from the MALDI-ToF-MS experiments provide information on both the molecular mass and structural characteristics of the intact Vtg. For high molecular mass measurements, where the isotopic cluster of the analyzed ions can not be resolved by the mass analyzer, the apex of the peak represents the relative molecular mass. The broadness of the peaks obtained also has a structural implication. The peak width is given by the isotopic contribution of all the atomic constituents of the molecular

ions analyzed. When the empirical formula of the analyzed sample is known, the peak width can be predicted. MALDI-ToF-MS spectra also show an increase in the peak width due to the inherent energetic population spread of the ions produced in the laser plume of the ionization source. This spread can be partly corrected by the delayed extraction technology offered by some MALDI-ToF-MS, used in the analysis performed in this work.

The theoretical isotopic envelope and the calculated relative molecular mass for RT Vtg, are shown in Figure 3.1. These figures were obtained using a free access molecular weight calculator (see Section 3.1.7.1) available on the internet. However, this value (Mr 181616) does not include any possible mass contribution due to the post translational modifications. No molecular mass for AS Vtg could be calculated since only partial amino acid sequences are known for this species.

The spectra obtained for both AS and RT Vtg samples subjected to MALDI-ToF-MS can be seen in Figure 3.2. The methodological details and the acquisition parameters for the MALDI-ToF-MS of the intact Vtg analyses were described in Section 3.1.6. The intact purified RT Vtg showed two peaks around  $m/z$  181938 and 94279 corresponding to the single charged  $[M+H]^+$  and double charged  $[M+2H]^{2+}$  molecular species. Similarly, the AS Vtg showed a spectrum containing two peaks identified as the  $[M+H]^+$  at  $m/z$  177153.1 and the  $[M+2H]^{2+}$  at  $m/z$  91250.4. The molecular mass values corresponding to the singly charged species are similar to those found in the SDS-PAGE experiments in this work and by other researchers.

The MALDI-ToF-MS spectra corresponding to RT and AS intact Vtg showed broad peaks for both species. The widths observed exceeded by 2 orders of magnitude the

theoretical calculated values observed in Figure 3.1 according to the empirical formula of RT Vtg. For both RT and AS Vtg, the  $[M+H]^+$  ranged from  $m/z$  170 to 210 kDa. This suggests the presence of sample microheterogeneity within the purified Vtg samples, which can be attributed to the presence of post translational modifications, or protein sequence variants, none of which can be resolved in this MS or typical SDS-PAGE experiments. This means that the peak observed in the MALDI-MS represents a family of Vtg molecular species, each contributing to the overall peak shape. Therefore, in the presence of such broad peaks, the apex of the peaks should be considered as an estimate of the relative molecular mass of all molecular species present in the samples.

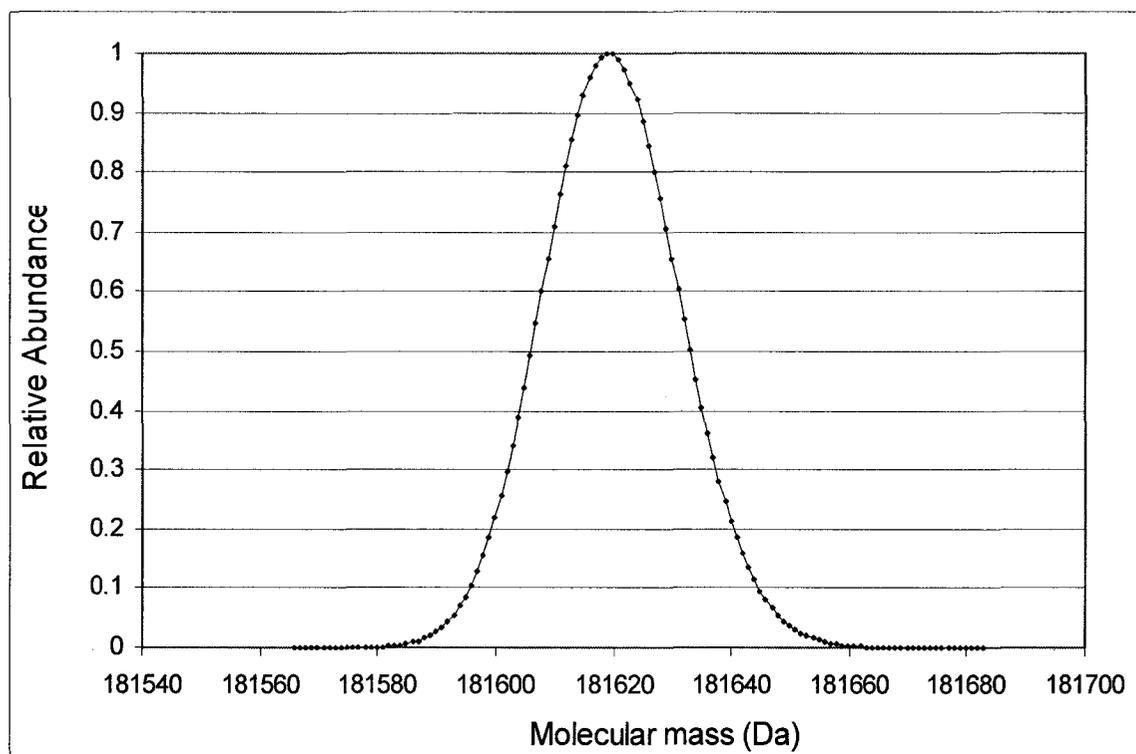


Figure 3.1 Relative molecular mass and isotopic distribution predicted for RT Vtg. The molecular weight of RT Vtg was calculated based on a molecular formula of  $C_{8138}H_{13040}N_{2196}O_{2360}S_{69}$ , obtained from its corresponding sequence (Appendix 2: Entry Q92093).

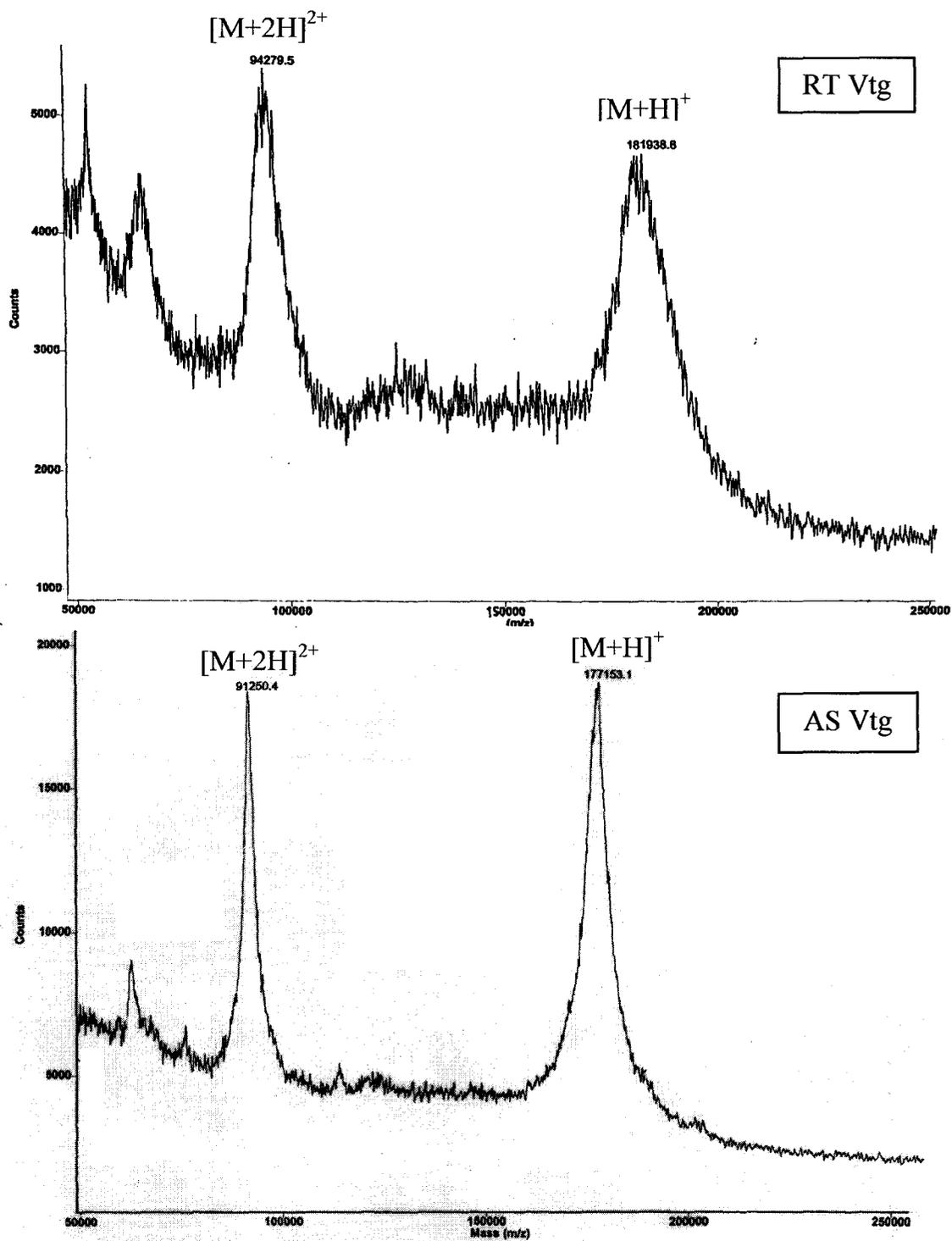


Figure 3.2 MALDI-MS spectra of intact Vtg  
The top and bottom spectra show the peaks obtained for intact RT and AS Vtg, respectively.

The MALDI-ToF-MS spectra of AS Vtg showed an additional interesting feature. A small peak in the high mass range was observed at  $m/z$  346713 (Figure 3.3). This peak, whose mass is nearly double the mass of the singly charged species, was attributed to the presence of the  $[2M+H]^+$  species. This observation further supports the hypothesis that Vtg circulates in the bloodstream as a dimer. This fascinating observation is only possible due to the extremely soft ionization conditions offered by the MALDI source. Although the molecular mass of this dimer is still below that obtained from other studies using size exclusion chromatography under non denaturing conditions, it is much closer to the 393 kDa obtained by previous workers using SDS-PAGE. Nonetheless, it is not clear if such differences in molecular masses can be attributed exclusively to variations in the PTMs or the lipid content in this complex biomolecule.

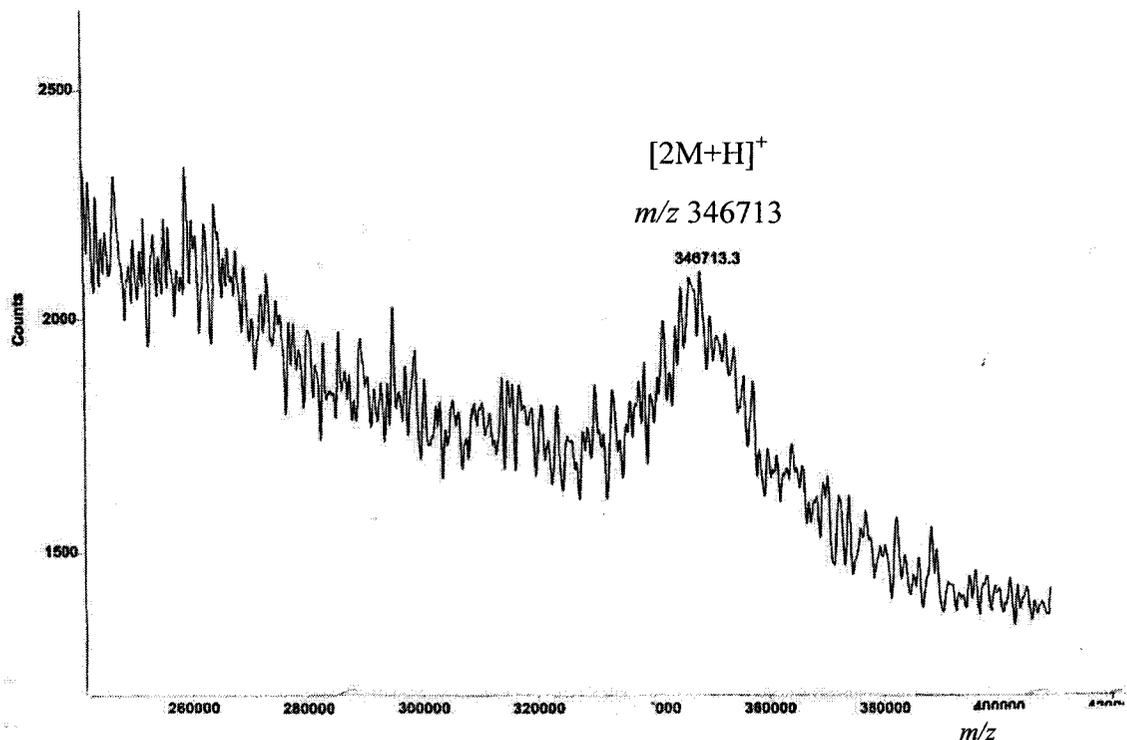


Figure 3.3 High mass range MALDI-MS acquisition of intact Vtg  
The peak observed at  $m/z$  346713 suggests the presence of a dimeric form of AS Vtg.

### **3.3 Digestion analyses of purified vitellogenin**

#### **3.3.1 MALDI-MS of the CNBr and trypsin derived peptides of rainbow trout vitellogenin**

The purified RT Vtg was subjected to CNBr and trypsin digestions and the digested samples were subjected to MALDI-ToF-MS. CNBr cleaves proteins at the C-terminal side of all methionine residues, affording homoserinelactone as a modified amino acid residue. Trypsin cleaves at the C-terminal side of both arginine and lysine residues, except when followed by proline. The methodology used for these reactions was described in Section 3.1.3, while instrumental acquisition parameters were described in Section 3.1.6. The spectra obtained for the CNBr cleavage (Figure 3.4) and the tryptic digestion (Figure 3.5) confirmed the complexity of the samples analyzed. In spite of this, these spectra yielded a suitable set of peaks to perform the peptide mass fingerprint database search, as described in Section 3.1.7.3. The MASCOT (Perkins *et al.*, 1999) search results of the CNBr cleavage identified this protein as RT Vtg, entry Q92093 on the Swiss-Prot Database. Twenty out of the 78  $m/z$  values searched matched the predicted digestion, which allowed 36% sequence coverage. A list of these matching peptides and the complete set of searched fragments can be seen in Table 3.1. Identical results were obtained from the tryptic digest search. In this case 55 out of the 72  $m/z$  values searched matched the predicted digestion, which allowed 37% sequence coverage. The results for this search are shown in Table 3.2.

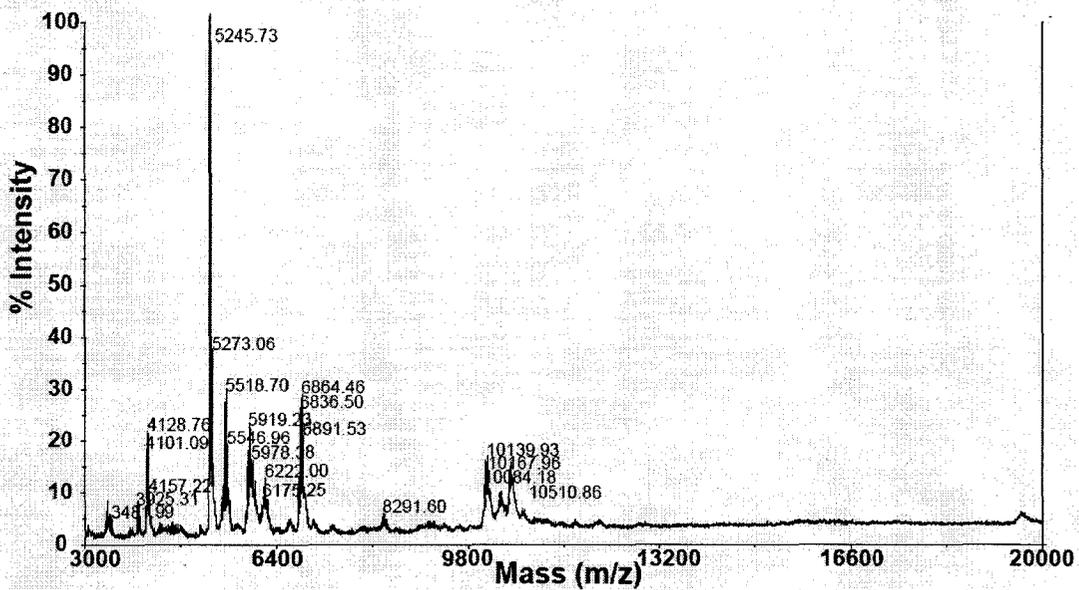
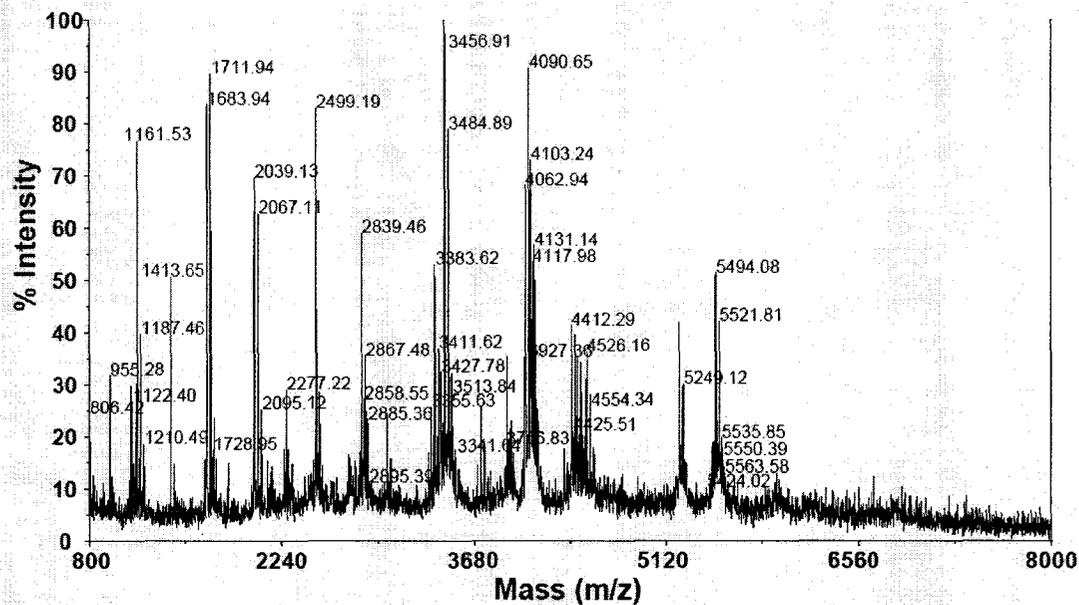


Figure 3.4 MALDI-MS spectra of the CNBr cleavage of purified RT Vtg. The top spectrum was acquired using dihydroxybenzoic acid as matrix, scanning from  $m/z$  800 to 8000. The bottom spectrum was acquired using sinapinic acid as matrix, scanning from  $m/z$  3000 to 20000.

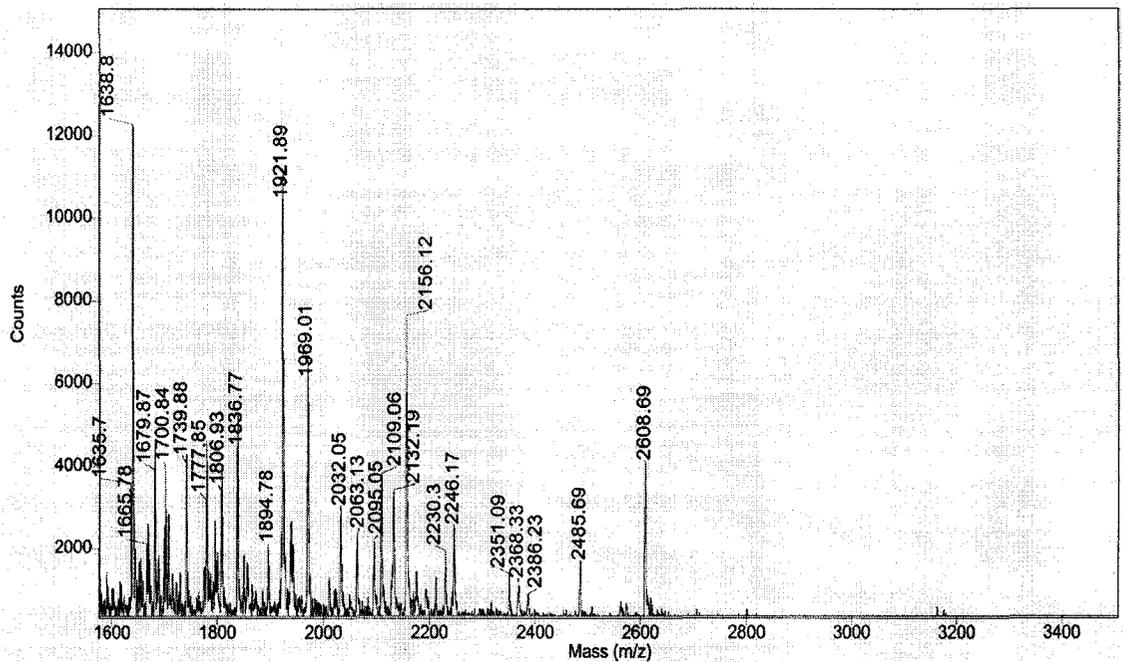
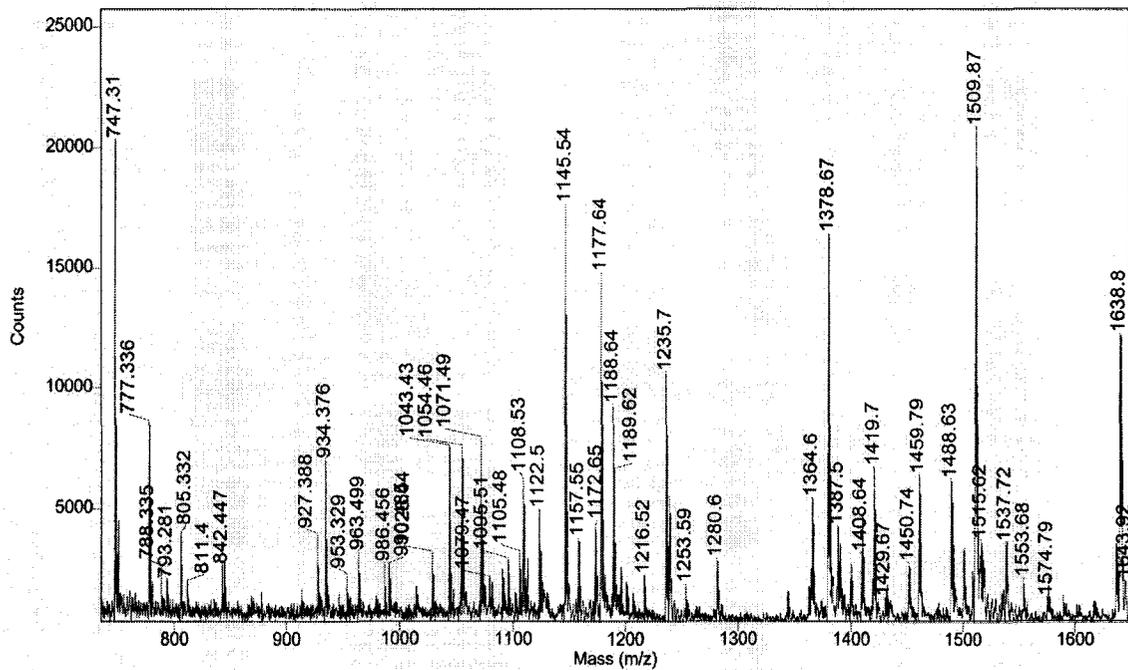


Figure 3.5 MALDI-MS spectra of the trypsin digestion of purified RT Vtg. The abscissa of the spectrum was split into two  $m/z$  ranges to facilitate the visualization of the peaks. The top and bottom spectra show the  $m/z$  ranges from 700 to 1650 and 1600 to 3500, respectively. The MALDI-MS was operated in reflectron mode using DHB as matrix.

Special precautions should be exercised when completing the PMF search queries, since unexpected results can be obtained if the search parameters are not correctly evaluated. For example, when  $\pm 3$  peptide mass tolerance and 2 missed cleavages were allowed, Q92093 no longer stood at the top score of the probable protein hits. As seen in Table 3.3, increasing the number of allowed missed cleavages did not increase significantly the number of matching sequences in the both types of digestions. This suggests that practically complete cleavage of the protein had taken place. Increased peptide mass tolerance resulted in a higher number of matching peptides, although the identity of these peptides should be confirmed in the case of increasing differences between the theoretical and the observed  $m/z$  values.

A peptide map was built to show the distribution of the matching peptides of both digestions with respect to their positions in the intact Vtg sequence (Figure 3.6). The tryptic digest showed a more regular distribution of matching peptides along the main protein chain. Interestingly, the gaps between the matching peptides, indicated a large number of motifs, that were compatible with likely phosphorylation and glycosylation sites as predicted by the SCAN PROSITE (Gattiker *et al.*, 2002; Sigrist *et al.*, 2002; Hulo *et al.*, 2004) and GlycoMod Tool programs (Cooper *et al.*, 2001). The SCAN PROSITE program allows the user to scan a protein sequence for the occurrence of patterns and profiles stored in the database. Two possible N-glycosylation sites were identified in positions 1089-1092 (NGTR) and 1627-1630 (NRSE). A series of interlaced peaks found in the CNBr digestion of RT Vtg were compatible with a candidate peptide (positions 1573 to 1643 of RT sequence) with different hybrid/complex glycan structures. These glycosylated peptides were tentatively assigned in Table 3.4.

Table 3.1 Peptide mass fingerprint of the CNBr chemical degradation of RT Vtg

The observed  $m/z$  corresponds to the monoisotopic peak observed in the MALDI-MS spectrum for this protein. The position indicates the location of the matching peptide in the amino acid sequence of RT Vtg, entry Q92093 in the Swiss-Prot Database (See Appendix 2). The "\*" shows the presence of an internal methionine (M) residue, indicating a missed cleavage site.

Observed $m/z$	Calculated $m/z$	$\Delta m/z$	Sequence	Position
955.28	955.44	-0.16	GAATYNYLM	216-224
1122.4	1122.51	-0.11	QVASFTYSHM	582-591
1161.53	1161.63	-0.1	DKALHPELRM	543-552
1187.46	1187.63	-0.17	GVNTAFIQAAIM	857-868
1413.65	1413.77	-0.12	LACIVLFETKPPM	553-565
1683.94	1683	0.94	GLVITLASILKTEKNM	566-581
2039.13	2038.1	1.03	VRDTLTFNNKKYKINM	1394-1410
2499.19	2499.2	-0.01	KDFGLAYTEKCVCECRQRGEALM	194-215
2839.46	2838.42	1.04	KPADNGALILEATVTELHQFTPFNEM	225-250
3383.62	3383.65	-0.03	KPADNGALILEATVTELHQFTPFNEMSGAAQM*	225-256
3484.89	3483.64	1.25	KDFGLAYTEKCVCECRQRGEALMGAATYNYLM*	194-224
3925.31	3925.25	0.06	VIIIFRLVRADHKIEGYQVTAYLNKATSRLQIIM	1189-1222
4062.94	4062.04	0.9	VSKYCVHEPNCPAELVKPIHELAVQAVANSKFEELSM	443-479
4101.09	4101.17	-0.08	LTFVEIKKDPIIVPDNNYVHRGSIRYEFATEILQM	262-296
4412.29	4410.29	2	VLLKKDHASEQNHINVKISDIDVDLYTEDHGVIVKVNEM	1430-1468
4495.6	4494.56	1.04	IGKHSVLISVKPSASEPAIERLEFEVQVGPKAAEKIIVITM	1022-1063
5217.87	5217.97	-0.1	PIQLLKISNARAQAVKILNHLVYNTAPVHEDAPLKFLQFIQL LRM	297-342
5546.95	5547.95	-1	QKDKNNEKQIQFTVVATSERITLDVILKTPKMTLYKLGVNLP CSLPFESM*	1317-1365
10056.36	10055.61	0.75	RRVIKALSDWRSLATSKPLASIYVKFFGQEIAFANIDKSIIDQA LQLANSPSAHALGRNALKALLAGATFQYVKPLLAEEVRRIF PTAVGLPM	708-800
10510.86	10511.56	-0.7	AALDENDNWKLCADGVLLSKHKVTAKIAWGAECKDYNTFI TAETGLVGPSPAVRLRLSWDKLPKVPKAVWRYVRIVSEFIPG YIPYYLADLVPM	1223-1316

Table 3.2 Peptide mass fingerprint of the trypsin digestion of RT Vtg

Refer to legend in Table 3.1 for details. The '\*' sign shows the presence of an internal lysine (K) or arginine (R) residue, indicating a missed cleavage site.

Observed <i>m/z</i>	Calculated <i>m/z</i>	$\Delta$ <i>m/z</i>	Sequence	Position
747.31	747.37	-0.06	DKNNEK*	1319-1324
777.34	777.40	-0.06	VVDWMK	1516-1521
788.34	788.44	-0.10	THYVIR	164-169
811.4	811.53	-0.13	ILLPDLK	1082-1088
934.38	934.45	-0.07	IAWGAECK	1249-1256
953.33	953.41	-0.08	DASECLMK	1566-1573
963.5	963.58	-0.08	IQTHPVLR	425-432
986.46	986.57	-0.11	LSWDKLPK*	1279-1286
1028.44	1028.50	-0.06	NVEDVPAER	918-926
1043.43	1043.51	-0.08	DFGLAYTEK	195-203
1054.35	1053.51	0.84	DTLTTFNNK	1396-1404
1071.49	1071.55	-0.06	CLPGCLPVR	1601-1609
1079.47	1079.55	-0.08	LPPAPENADR	694-703
1095.51	1095.58	-0.07	FEELSMVLK	474-482
1108.53	1108.61	-0.08	QMLTFVEIK	260-268
1122.5	1122.56	-0.06	CYSVEPVLR	1592-1600
1145.54	1145.62	-0.08	LEFEVQVGPK	1043-1052
1157.55	1157.63	-0.08	LVESLAFNHK	415-424
1172.65	1172.73	-0.08	LAPMVILFR	1185-1194
1177.64	1177.71	-0.07	FLQFIQLLR	333-341
1188.64	1188.62	0.02	QVIVDDRESK*	1582-1591
1216.52	1214.67	1.85	LEKQVIVDDR*	1579-1588
1235.7	1235.75	-0.05	ITPLIPAQGVAR	927-938
1378.67	1378.74	-0.07	QIQFTVVATSER	1325-1336
1387.5	1387.59	-0.09	DGPDQPYNPNDR	1134-1145
1408.64	1408.71	-0.07	DSQSTSNVISRSK*	1153-1165
1419.7	1419.69	0.01	DLNCCQQRIMK*	184-194
1429.67	1429.71	-0.04	ELTMLGYGTMVSK	433-445
1459.79	1459.85	-0.06	LTSALAAQFSIPIK	95-108
1509.87	1509.92	-0.05	LLPVFGTAAAALPLR	499-513
1515.62	1515.68	-0.06	DGPDQPYNPNDRK*	1134-1146
1537.72	1537.78	-0.06	MVQEVAVQLFMDK	532-544
1553.68	1551.78	1.90	DPIIHPDNNYVHR	270-282
1574.79	1574.90	-0.11	ALGNAGHPASIKPITK	483-498
1635.7	1635.75	-0.05	ACFEFTSESAAFIR	1001-1014
1638.8	1638.85	-0.05	TYFAGAAADVLEVGVR	668-683
1665.78	1665.78	0.00	VNEMEISNDNLPLYK	1465-1478
1679.87	1679.89	-0.02	KDPIIHPDNNYVHR*	269-282
1700.84	1700.92	-0.08	YEALLGGLPEEGLAR	34-49
1739.88	1738.80	1.08	VVDWMKQTCGLCGK*	1516-1530
1777.85	1777.92	-0.07	LVNPEIFEYSGVWPK	72-86
1806.93	1806.99	-0.06	HWILDAVPSIGSSVAVR	363-379
1894.78	1894.90	-0.12	TQNVYELQEAGAQQVCK	147-163
1921.89	1921.89	0.00	SSVSFAHSWVLPDSCR	1549-1565
1969.01	1969.02	-0.01	IGAAASAFYINDAATLFPR	642-660
2032.05	2033.01	-0.96	LGVKACFEFTSESAAFIR*	997-1014
2063.13	2062.10	1.03	SIIDQALQLANSPSAHALGR	746-765
2109.06	2108.06	1.00	DYNTFITAETGLVGPSPAVR	1257-1276
2132.19	2131.22	0.97	ALLAGATFYVVKPLLAEEVR	770-789
2156.12	2155.09	1.03	TPITIGFHCLPVDSNLNR	1610-1628
2230.3	2229.08	1.22	LSCHFSAIHLDAYSNPLR	623-641
2246.17	2245.22	0.95	NVEDVPAERITPLIPAQGVAR*	918-938
2351.09	2350.28	0.81	GSIRYEFATEILQMPIQLLK*	283-302
2386.23	2385.18	1.05	RLSCHFSQAIHLDAYSNPLR*	622-641
2608.69	2607.33	1.36	FEAFPVSPPEHIAAAHIETFAVAR	894-917

Table 3.3 Optimization of peptide mass fingerprint matching parameters

The top and bottom tables correspond to the CNBr and trypsin digestions, respectively. The number of missed cleavages and the peptide mass tolerance were the two parameters which were adjusted for the database search. The number of matching peptides and the sequence coverage of RT Vtg are indicated by the letters 'm' and 'sc', respectively.

CNBr		Number of missed cleavages					
		0		1		2	
		m	sc	m	sc	m	sc
Peptide mass tolerance (m/z)	0.5	9	11%	10	13%	10	13%
	1	15	27%	16	27%	16	27%
	2	21	33%	25	36%	25	36%
	3	26	38%	31	41%	31	41%

Trypsin		Number of missed cleavages					
		0		1		2	
		m	sc	m	sc	m	sc
Peptide mass tolerance (m/z)	0.5	37	25%	41	27%	43	28%
	1	41	30%	48	32%	51	33%
	2	47	36%	56	37%	59	39%
	3	49	38%	59	38%	63	40%

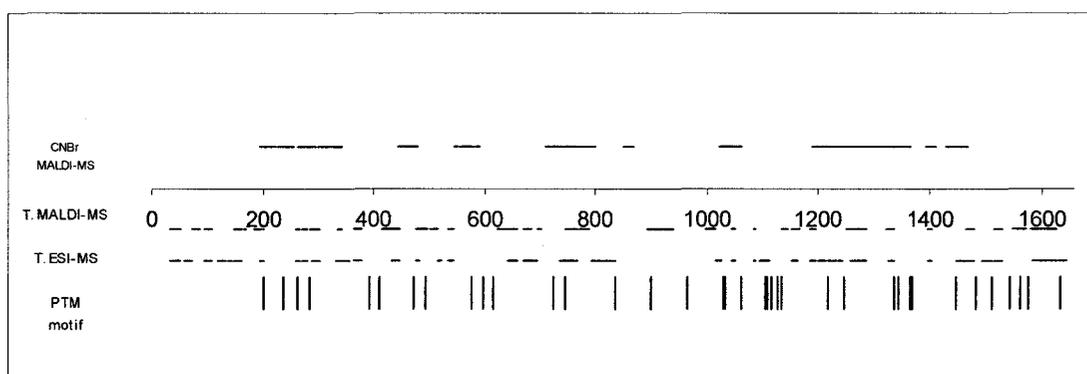


Figure 3.6 Peptide map obtained for RT Vtg

The horizontal central line represents the sequence of RT Vtg. The numbers indicate the amino acid positions along this protein. The horizontal segments represent the peptides found in the peptide mass fingerprints of the CNBr and trypsin (T) digestions obtained from the MALDI-MS and ESI-MS experiments. The positions of post-translational modification (PTM) consensus motifs are shown with vertical lines.

Table 3.4 Putative glycan structures found in RT Vtg

Proposed structures of different glycans compatible with the peaks found in the MALDI-MS of the CNBr cleavage of RT Vtg. Hex stands for hexose, HexNAc for N-acetylhexose amine, NeuGc for N-glycolyl neuraminic acid, Phos for phosphate, Man for mannose and GlcNAc for N-acetylglucosamine.

Glycoform mass	Glycan structures	Calculated $[M+H]^+$	Observed $[M+H]^+$	$\Delta m/z$
1927.552	(Hex) <sub>1</sub> (HexNAc) <sub>2</sub> (NeuGc) <sub>1</sub> (Phos) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	9825.65	9825.33	-0.32
2157.783	(Hex) <sub>1</sub> (HexNAc) <sub>4</sub> (NeuAc) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	10055.88	10056.36	0.48
2268.673	(Hex) <sub>5</sub> (HexNAc) <sub>2</sub> (Phos) <sub>2</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	10166.77	10167.17	0.40
2294.771	(HexNAc) <sub>5</sub> (NeuGc) <sub>1</sub> (Phos) <sub>1</sub> + (Man) <sub>3</sub> (GlcNAc) <sub>2</sub>	10192.87	10193.23	0.36

### **3.3.2 ESI-MS of the trypsin derived peptides of rainbow trout vitellogenin**

The separation of the RT Vtg tryptic peptides followed by MS analysis using a HPLC-ESI-QqToF MS showed a total ion chromatogram (TIC) with numerous overlapping peaks. The analytical parameters selected for the HPLC-ESI-QqToF MS acquisition were described in Section 3.1.5. A perusal of the different full scan mass spectra revealed a total of eighty-one peptides matching the predicted digestion of RT Vtg (Refer to Section 3.1.7.2 for details on the Vtg digestion simulation). Matching was done by considering both the mass and charge of each of the peaks obtained in the spectra. The total number of peptides generated by the simulated digestion of RT Vtg was 377. In this digestion simulation, pairs of peptides were noticed, which, in spite of having different empirical formulae, had similar molecular masses. For this reason, it was not surprising to find a couple of spectra which contained peaks that could be assigned to more than one candidate tryptic peptide, as exemplified in Table 3.5. Eight pairs of isobaric peptides that matched peaks obtained in the spectra from the TIC chromatograms were found. For these particular isobaric peptides, sequence confirmation can only be achieved through MS/MS experiments. When comparing the MALDI-MS and ESI-MS results of the tryptic digestion, 29 matching peaks were found. For this, the  $[M+H]^+$  of the ESI-MS peaks were calculated for comparison because most of them were found as multiple charged species (Table 3.6).

Table 3.5 Isobaric peptides found in the tryptic digestion of RT Vtg

The following peaks identified by HPLC-ESI-MS matched isobaric tryptic peptides predicted by the simulated digestion of RT Vtg.

Sequence	Formula	Theoretical [M+H] <sup>+</sup> m/z	Calculated [M+H] <sup>+</sup> m/z	Δm/z
TLDVILK	C <sub>37</sub> H <sub>68</sub> N <sub>8</sub> O <sub>11</sub>	801.51	801.36	-0.15
AGVKVISK	C <sub>36</sub> H <sub>68</sub> N <sub>10</sub> O <sub>10</sub>	801.52		-0.16
QVIVDDR	C <sub>35</sub> H <sub>61</sub> N <sub>11</sub> O <sub>13</sub>	844.45	844.29	-0.16
DPFVPAAK	C <sub>40</sub> H <sub>61</sub> N <sub>9</sub> O <sub>11</sub>	844.46		-0.17
DSQSTSNVISRSK	C <sub>55</sub> H <sub>97</sub> N <sub>19</sub> O <sub>24</sub>	1408.70	1408.43	-0.27
FAAQLDIANGNFK	C <sub>64</sub> H <sub>97</sub> N <sub>17</sub> O <sub>19</sub>	1408.72		0.01
IHYLFSEVNAVK	C <sub>67</sub> H <sub>102</sub> N <sub>16</sub> O <sub>18</sub>	1419.76	1419.49	-0.27
DLNNCQQRIMK	C <sub>56</sub> H <sub>98</sub> N <sub>20</sub> O <sub>19</sub> S <sub>2</sub>	1419.68		-0.19
DASECLMKLESVK	C <sub>62</sub> H <sub>108</sub> N <sub>16</sub> O <sub>23</sub> S <sub>2</sub>	1509.73	1509.61	-0.12
LLPVFGTAAAALPLR	C <sub>72</sub> H <sub>120</sub> N <sub>18</sub> O <sub>17</sub>	1509.92		-0.31
HSVLSVKPSASEPAIER	C <sub>84</sub> H <sub>142</sub> N <sub>24</sub> O <sub>27</sub>	1920.06	1919.69	-0.37
EPRMVQEVAVQLFMDK	C <sub>84</sub> H <sub>138</sub> N <sub>22</sub> O <sub>25</sub> S <sub>2</sub>	1919.97		-0.28
ILNHLVTYNTAPVHEDAPLK	C <sub>102</sub> H <sub>161</sub> N <sub>27</sub> O <sub>30</sub>	2245.20	2244.77	-0.43
NVEDVPAERITPLIPAQGVAR	C <sub>98</sub> H <sub>165</sub> N <sub>29</sub> O <sub>31</sub>	2245.23		-0.46

Table 3.6 MALDI and HPLC-ESI-MS matching tryptic peptides of RT Vtg  
Refer to legend in Table 3.1 for details. The  $m/z$  values for the peaks obtained from the ESI experiments correspond to calculated  $[M+H]^+$  species.

ESI $m/z$	MALDI $m/z$	$\Delta m/z$	Sequence	Position
811.33	811.40	-0.07	ILLPDLK	1082-1088
986.35	986.46	-0.10	LSWDKLPK	1279-1286
1043.20	1043.43	-0.23	DFGLAYTEK	195-203
1071.33	1071.49	-0.16	CLPGCLPVR	1601-1609
1095.35	1095.51	-0.16	FEELSMVLK	474-482
1108.43	1108.53	-0.10	QMLTFVEIK	260-268
1122.35	1122.50	-0.15	CYSVEPVLR	1592-1600
1145.41	1145.54	-0.13	LEFEVQVGPK	1043-1052
1172.47	1172.65	-0.18	LAPMVILFR	1185-1194
1177.49	1177.64	-0.15	FLQFIQLLR	333-341
1188.45	1188.64	-0.19	QVIVDDRESK	1582-1591
1235.51	1235.70	-0.19	ITPLIPAQGVAR	927-938
1378.73	1378.67	0.06	QIQFTVVATSER	1325-1336
1408.43	1408.64	-0.21	DSQSTSNVISRSK FAAQLDIANGNFK	1153-1165 881-893
1419.49	1419.70	-0.21	IHYLFSEVNAVK DLNNCQQRIMK	1379-1390 184-194
1429.47	1429.67	-0.20	ELTMLGYGTMVSK	433-445
1459.59	1459.79	-0.20	LTSALAAQFSIPIK	95-108
1509.61	1509.87	-0.26	LLPVFGTAAAALPLR DASECLMKLESVK	499-513 1566-1578
1537.45	1537.72	-0.27	MVQEVAVQLFMDK	532-544
1638.49	1638.80	-0.31	TYFAGAAADVLEVGVR	668-683
1665.45	1665.78	-0.33	VNEMEISNDNLPYK	1465-1478
1679.60	1679.87	-0.27	KDPIIVPDNNYVHR	269-282
1700.57	1700.84	-0.27	YEALLGGLPEEGLAR	34-49
1739.54	1739.88	-0.34	VVDWMKGQTJGLJGK	1516-1530
1777.63	1777.85	-0.22	LVNPEIFEYSGVWPK	72-86
1806.61	1806.93	-0.32	HWILDAVPSIGSSVAVR	363-379
1894.61	1894.78	-0.17	TQNVYELQEAGAQGVCK	147-163
1921.46	1921.89	-0.43	SSVSFAHSWVLPDSCR	1549-1565
1968.69	1969.01	-0.32	IGAAASAFYINDAATLFPR	642-660

### 3.3.3 ESI-MS/MS of the trypsin derived peptides of rainbow trout vitellogenin

The same instrumentation described in the previous section was used to perform the MS/MS experiments. The acquisition and instrument parameters were set as described in Section 3.1.5. An Information Dependent Acquisition (IDA) was performed to automatically select the most abundant peaks found in the full scan MS. These precursors were then automatically selected for collision induced dissociation (CID) MS/MS. Unfortunately, only ten product ion spectra contained information of diagnostic value. Interpretation of these spectra was done by a computer based sequencing algorithm, called Lutefisk (Taylor and Johnson, 1997; Taylor and Johnson, 2001), described in Section 3.1.7.4. Sequence alignment between the '*de novo*' sequenced tryptic peptides and the published sequence of RT Vtg was performed using Cidentify (Section 3.1.7.5). The top scored sequenced peptides and their alignment with RT Vtg are shown in Table 3.7. The selected precursor peptide ions at  $m/z$  816.69, 819.75, 849.24, 850.79, 912.91 and 984.85 showed partial sequence homology with RT Vtg, shown in boldface font in the last column of Table 3.7.

Some sequence variants were observed when comparing the sequences obtained by Lutefisk to the published RT Vtg sequence. Sometimes, the information provided by the product ion spectrum is insufficient to provide complete '*de novo*' sequencing and in these cases the software showed the masses of these missing gaps in brackets. Other explanations of sequence variants can be attributed to isobaric sequences of amino acids, which sometimes render very similar mass patterns. An example of both of these situations can be observed for the product ion spectrum of the precursor ion at  $m/z$  819.27

where the [264.1] gap corresponds to the same mass as the [TY] fragment, which is present in the database sequence. In this same sequence, the internal K residue assigned by Lutefisk is isobaric with the [GA] dipeptide fragment observed in the database sequence while the remainder of the sequence bears perfect homology with RT Vtg. Similar analogies can be made for the rest of the precursor ion spectra.

Table 3.7 'De novo' sequencing and sequence homology of RT Vtg tryptic peptides  
Mr indicates the relative molecular mass of the precursor ions. The numbers between brackets under the Sequence column indicate mass gaps that could not be assigned to any particular amino acid sequence. The matching amino acid residues indicating homology to the RT Vtg sequence are shown in bold font.

Precursor ion (m/z)	Charge (+)	Mr	Sequence	Sequence homology
816.69	2	1631.37	WALMGAATYNYLMK	<b>GEALMGAATYNYLMK</b>
819.72	2	1637.43	[264.1]FAKAADVLEVGVR	<b>TYFAGAAADVLEVGVR</b>
828.83	2	1655.65	W[287.1]HYFY[216.0]LPK	No match
849.14	2	1696.27	MA[291.0]LLNALWAEFK	<b>MASSETINAIWAEFK</b>
850.75	2	1699.49	[292.1]ALLLGGL[231.0]MYAR	<b>YEALLLGGLPEEGLAR</b>
858.55	3	2572.64	[274.0][244.0]GD[212.1]KYNFLA VFAYHGK	No match
912.91	2	1823.81	[243.1]DEPGMCSGAAKMEAK	<b>FTPFNEMSGAAQMEAK</b>
984.77	2	1967.53	[270.1]LSAFYLNDAATLFPR	<b>IGAAASAFYINDAATLFPR</b>
1044.26	2	2086.51	[232.0]NEL[346.1]D[320.0]TLDG ANK	No match
1064.06	2	2126.11	[269.2]S[158.0]S[283.2]LN[213.1] LSS[168.1]TAR	No match

An interesting case is illustrated by the sequencing of the precursor ion at  $m/z$  912.91. There was no peptide that matched the mass of this precursor in the digestion simulation of RT Vtg, however, a significant sequence homology was found to a portion of this protein. This suggests that a sequence variant that produced a different tryptic peptide may be present.

### **3.3.4 MALDI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin**

The spectrum obtained by MALDI-ToF-MS for the tryptic digestion of the purified AS Vtg is shown in Figure 3.7. The major peaks from this spectrum were used for protein identification using the PMF search in MASCOT (Section 3.1.7.3), however none of the proteins in the database showed a statistically significant hit from this list of peaks. This was expected, since the complete sequence of AS has not yet been reported.

However, RT Vtg (See Appendix 2: entry Q92093 on the Swiss-Prot Database) was the protein with the most matching peptides for this peak list. In fact, thirty-one searched peaks matched known peptides resulting from the tryptic digestion of the RT Vtg and a list of these can be seen in Table 3.8. Note that the last column shown in this table corresponds to the position along the RT Vtg sequence.

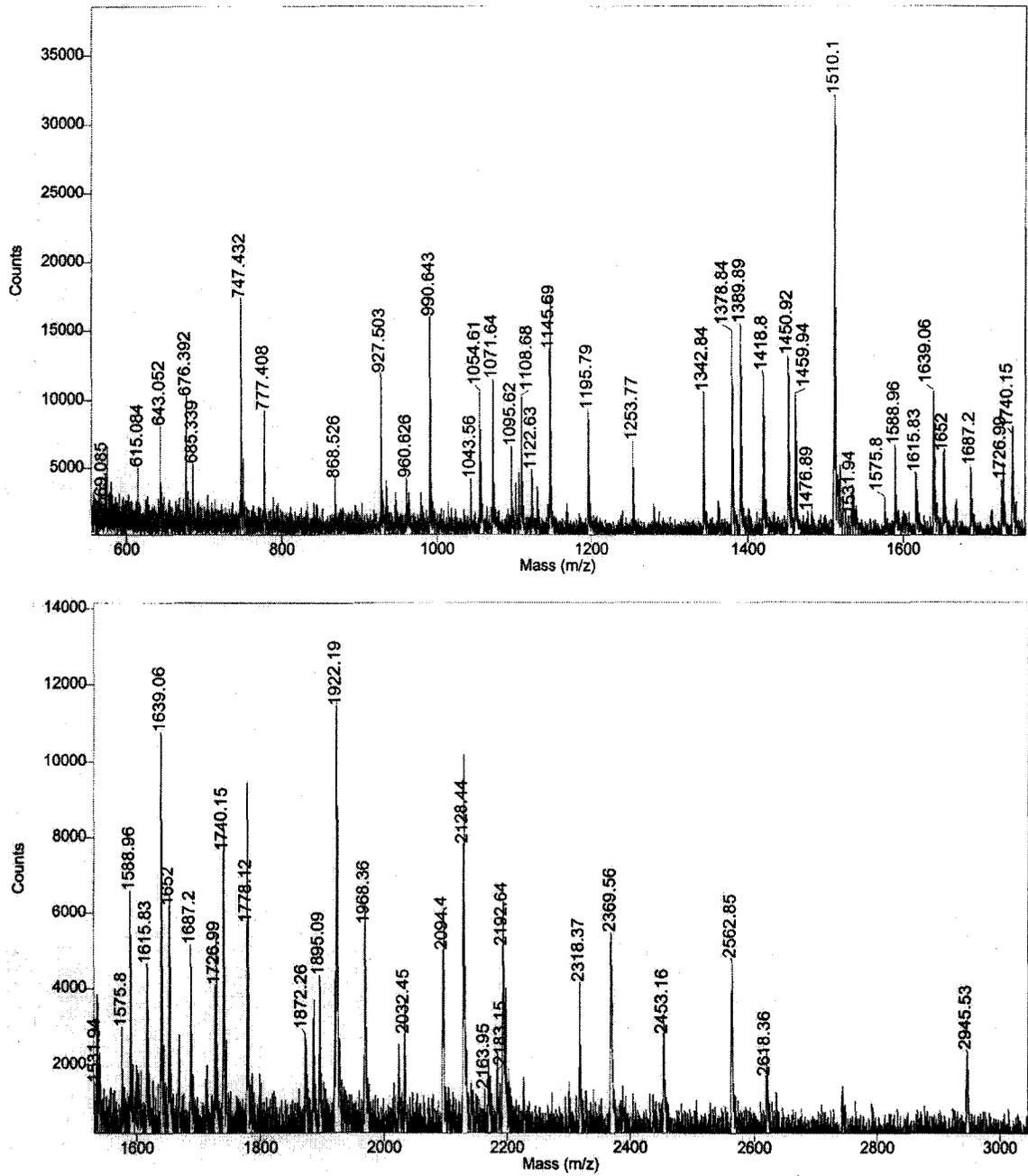


Figure 3.7 MALDI-MS spectra of the tryptic digestion of purified AS Vtg  
 The abscissa of the spectrum was split into two  $m/z$  ranges to facilitate the visualization of the peaks. The top and bottom spectra show the  $m/z$  ranges from 550 to 1750 and 1550 to 3050, respectively.

Table 3.8 Peptide mass fingerprint of the trypsin digestion of AS Vtg

Refer to legend in Table 3.2 for details.

Observed m/z	Calculated m/z	$\Delta m/z$	Sequence	Position
685.34	683.42	1.92	HKVTAK*	1243-1248
747.43	747.38	0.05	ALSDWR	713-718
777.41	777.40	0.01	VVDWMK	1516-1521
1043.56	1043.51	0.05	DFGLAYTEK	195-203
1054.61	1055.62	-1.01	VQADAVLALR	514-523
1071.64	1071.55	0.09	CLPGCLPVR	1601-1609
1095.62	1095.58	0.04	FEELSMVLK	474-482
1108.68	1108.61	0.07	QMLTFVEIK	260-268
1122.63	1122.56	0.07	CYSVEPVLR	1592-1600
1145.69	1145.62	0.07	LEFEVQVGPK	1043-1052
1195.79	1197.59	-1.80	SSASSFHAIYK	1166-1176
1378.84	1378.74	0.10	QIQFTVVATSER	1325-1336
1389.89	1389.61	0.28	KSESSSSSSSSSR*	1116-1129
1418.80	1419.69	-0.89	DLNNCQQRIMK*	184-194
1459.94	1459.85	0.09	LTSALAAQFSIPIK	95-108
1476.89	1477.87	-0.98	SLATSKPLASIYVK	719-732
1510.10	1509.92	0.18	LLPVFGTAAAALPLR	499-513
1531.94	1533.89	-1.95	VLISAVAENTYLLK	58-71
1575.80	1575.72	0.08	VITMNEEEEAPEGK	1060-1073
1588.96	1589.73	-0.77	SESSSSSSSSSRISK*	1117-1132
1639.06	1638.85	0.21	TYFAGAAADVLEVGVR	668-683
1687.20	1686.80	0.40	CSMVRDTLTTFNK*	1391-1404
1740.15	1738.81	1.34	VVDWMKGQTCGLCGK*	1516-1530
1778.12	1777.92	0.20	LVNPEIFEYSGVWPK	72-86
1895.09	1894.90	0.19	TQNVYELQEAGAQGVCK	147-163
1922.19	1922.08	0.11	ATLTPALPETFHAAQLLK	818-835
1968.36	1969.02	-0.66	IGAAASAFYINDAATLFPR	642-660
2032.45	2033.02	-0.57	LGVKACFEFTSESAAFIR*	997-1014
2128.44	2130.10	-1.66	VITMNEEEEAPEGKTVLLK*	1060-1078
2163.95	2162.15	1.80	TEGIQEALLKLPPAPENADR*	684-703
2945.53	2944.57	0.96	IFPTAVGLPMELSSYYTAAVAKAYVNVR*	791-817

No match was found for the tryptic peptide ions at  $m/z$  569.09, 615.08, 643.05, 676.39, 868.53, 927.50, 960.63, 990.64, 1253.77, 1342.84, 1450.92, 1615.83, 1652.00, 1726.99, 1872.26, 2094.40, 2183.15, 2192.64, 2318.37, 2369.56, 2453.16, 2562.85 and 2618.36. These remaining non-identified peaks may tentatively be assigned to either exclusive AS Vtg tryptic peptides or post-translational modified peptides, which are not taken into account for the peptide mass fingerprint.

### **3.3.5 ESI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin**

The analysis of the tryptic digestion of AS Vtg by HPLC-ESI-QqToF MS (Section 3.1.5) showed numerous full scan spectra containing peaks which matched the masses of those predicted by the tryptic simulation (see Section 3.1.7.2) of both the RT and AS Vtg. Matching was done by comparing the monoisotopic  $m/z$  values seen in the spectra with the predicted multiply charged peptide ions considering a  $m/z$  peak tolerance of  $\pm 0.5$ . The charge of the peptide ions was confirmed by resolving the peak isotopic cluster. These matching peptides are presented in Table 3.9.

Table 3.9 Separation of AS Vtg tryptic peptides by HPLC-ESI-MS

The retention time (Rt) of the peptides is given minutes (min). The observed (Obs) and calculated (Cal) *m/z* values are shown for each particular charged molecular species.

Peptide	Species	Rt min	[M+H] <sup>+</sup>		[M+2H] <sup>2+</sup>		[M+3H] <sup>3+</sup>	
			Obs m/z	Cal m/z	Obs m/z	Cal m/z	Obs m/z	Cal m/z
TLDVILK	RT			801.5		401.3		
AGVKVISK	RT	18.8	801.4	801.5	401.2	401.3		
DTLTTFNK	RT	22.5	1054.3	1053.5	527.6	527.3		
VQADAVLALR	RT	19.2	1055.4	1055.6	528.2	528.3		
FEELSMVLK	RT	21.0	1095.3	1095.6	548.2	548.3		
TEGIQEALK	RT	18.9	1101.4	1101.6	551.2	551.3		
QMLTFVEIK	RT	21.1	1108.4	1108.6	554.7	554.8		
LEFEVQVGPK	RT	19.6	1145.4	1145.6	573.2	573.3		
SEGLSSFYEK	AS	18.7	1146.3	1146.5	573.6	573.8		
GILNILQLNIK	RT	23.9	1238.5	1238.8	619.8	619.9		
QIQFTVVATSER	RT	21.9	1378.4	1378.7	689.7	689.9		
LTSALAAQFSIPIK	RT	21.8	1459.5	1459.9	730.3	730.4	487.2	487.3
DASECLMKLESVK	RT			1509.7		755.4		503.9
LLPVFGTAAALPLR	RT	23.5	1509.6	1509.9	755.3	755.5	503.9	504.0
DGPDQYPNPDRK	RT			1515.7		758.4		505.9
DASECLMTFESVK	AS	21.0	1516.4	1516.7	758.7	758.8	506.1	506.2
VLISAVAENTYLLK	RT	21.7	1533.5	1533.9	767.3	767.5	511.9	512.0
VITMNEEEEAPEGK	RT	16.6			788.2	788.4		
SESSSSSSSSSRISK	RT	20.3			794.8	795.4	530.2	530.6
TYFAGAAADVLEVGVR	RT	22.2	1638.5	1638.9	819.8	819.9	546.8	547.0
ADGEVRQEYSTPSGR	AS	23.2	1651.5	1651.8	826.3	826.4	551.2	551.3
AIVVHLEETICVER	RT	22.6			834.3	833.4	556.5	555.9
CSMVRDTLTTFNK	RT	22.0	1686.4	1686.8	843.8	843.9	562.9	562.9
VVDWMKGQTCGLCGK	RT + AS	19.9			870.3	869.9	580.5	580.3
LVNPEIFEYSGVWPK	RT					889.5		593.3
STAPDFASVAAACNVAVK	RT	23.1			889.3	889.9	593.2	593.6
HSVLISVKPSASEPAIER	RT					960.5		640.7
EPRMVQEVAVQLFMDK	RT	26.5			959.8	960.5	640.2	640.7
SSVSFAHSWVLPDSCR	RT + AS					961.5		641.3
ATLTPALPETFHAAQLLK	RT	19.8			961.2	961.5	641.2	641.4
IGAAASAFYINDAATLFPR	RT	22.7			984.3	985.0	656.5	657.0
ISDIDVDLYTEDHGVIVK	RT	22.1			1016.3	1016.0	677.9	677.7
TPITIGFHCLPVDSNLSR	AS	20.6			1064.3	1064.6	709.9	710.0
MASSETINAIWAEFKAKPAY R	RT					1192.6		795.4
INMPLSCYQVLAQDCTTELK	RT	22.4			1193.4	1192.6	795.9	795.4
RLSCHFSQAIHLDAYSNPLR	RT					1193.1		795.7
QKGEGVSLYAPSHGLQEVYF DK	AS	19.2			1226.3	1226.6	817.9	818.1
AQAVKILNHLVTYNTAPVHE DAPLK	RT	25.2			1370.8	1371.8	914.2	914.8
IFPTAVGLPMELSSYYTAAVA KAYVNVR	RT	24.9			1471.9	1472.8	981.6	982.2

In this digestion simulation, pairs of peptides were noticed which, in spite of having different amino acid composition, had similar molecular masses. For this reason, six doubly charged peptide ions observed at  $m/z$  401.16, 755.28, 758.67, 889.25, 959.79 and 961.24 were assigned to two candidate isobaric tryptic peptides predicted by the digestion of AS and RT Vtg, as shown in Table 3.9. In fact, one ion at  $m/z$  1193.35 could be assigned three isobaric tryptic peptides. It should also be noted that two  $[M+2H]^{2+}$  peaks at  $m/z$  870.29 and 961.24 were assigned to peptides that were present in both AS and RT Vtg.

### **3.3.6 ESI-MS/MS of the trypsin derived peptides of Atlantic salmon vitellogenin**

The instrumental parameters selected for this particular analysis were identical to those used for the ESI-MS/MS experiments performed on RT Vtg, described in section 3.3.3. The results from these experiments revealed and confirmed the identity of some of the tryptic peptides found in the previous section. Unfortunately, a chromatographic tailing effect was observed for some peaks. This caused some precursor peptide ions to be repeatedly selected throughout the chromatogram during the IDA experiments which led to a reduced list of precursors selected for CID-MS/MS.

Furthermore, because of the strict parameters chosen for Lutefisk analyses (See Section 3.1.7.4 for the details on 'de novo' sequencing of peptides obtained from MS/MS spectra) many of the product ion spectra had to be discarded for sequencing purposes. The results of the interpretation of all the MS/MS spectra are shown in Table 3.10.

Table 3.10 'De novo' candidate sequences of the AS Vtg tryptic digestion

The following candidate sequences were obtained from the Lutefisk analyses of the product ion spectra generated by the HPLC-ESI-MS/MS experiments on the AS Vtg tryptic digestion. The letters between brackets correspond to pairs of amino acids that match a mass gap, but whose order could not be assigned. The numbers between brackets under the Candidate sequences column indicate mass gaps that could not be assigned to any particular amino acid sequence.

precursor ion (m/z)	Charge (+)	Calculated Mr	Candidate sequences
788.28	2	1574.47	[212.15]TMNEEEEAPEGK [212.15]TMNEEEAAPWK [212.15]TMNEEEAEPGK
573.66	2	1145.31	[216.08]GLSSFYEK
700.2	2	1398.39	[242.10]AMLYG[PV][194.08]R [242.10]AMLYGTL[204.09]K
794.75	2	1587.55	[184.08]AVQLEDTLCAER [184.12]AVKLEDTLCAER [184.12]AVGALEDTLCAER
834.31	2	1666.61	[276.14][242.14]VLEALYWKK [276.14]ALYLAELV[244.12]QK [276.14][242.12]VLEALYWQK
889.28	2	1776.55	[212.15]NEPLFEYSRWPK [212.15]NEPLFEYS[GV]WPK [212.15]NEPLFEYSR[186.07]PK [212.15]N[202.10]HFEYSRWPK [212.15]NEPLFEYS[GV][186.10]PK
819.79	2	1637.57	[TY]FAGAAADVLEVGVR [TY]FAGAAADVLEVRR [TY]FAGAAADVLEV[184.12]K
561.68	2	1121.35	CYSV[DH]WR
865.27	2	1728.53	[241.14]NLLAALDESDNWK [241.14]NLLAALDESDN[186.07]K
843.81	2	1685.61	[EY]ALVLG[DH]LL[RC]K [EY]AVLLG[DH]VLLA[274.07]
572.2	2	1142.39	LPTVTE[243.08]LK
795.18	1	794.17	M[244.12]QWS
893.28	2	1784.55	[212.15]TNDDH[204.09]EDW[LP]K
791.05	4	3160.19	[MT][225.15]LMEFE[WW][212.15]AAVHFNLEPCANK [MT]NELLEFET[RD][212.15]AAVHFNLEPCANK [199.14]EELMEFE[WW][212.15]AAVHDMAGTLNENK

Less stringent sequencing parameters would have yielded a higher number of matching peaks, however, this would have compromised the sequencing quality, increasing the number of candidate peptides for each product ion spectrum. Thus, only fourteen product ion spectra were completely sequenced by Lutefisk. An example of two product ion scans obtained from doubly charged precursors at  $m/z$  819.79 and  $m/z$  788.24 are shown in Figure 3.8. B- and Y- type fragment ions are indicated above the peaks with their corresponding single letter amino acid code sequence.

As with other sequencing programs, Lutefisk generates a list of sequence candidates. For that reason, some spectra were assigned more than one sequence, as seen in Table 3.10. However, it should be noted that many of them are usually simply inversion of dipeptides or isobaric amino acid replacements. In addition, incomplete sequence data provided the location and mass of a dipeptide, but not the exact sequence. In these cases, the sequencing program showed the nominal mass of the corresponding dipeptide between brackets, in the position along the sequence. When the mass of the dipeptide is accurate enough, its composition is shown between brackets, although the order is unknown. The isomeric pair, leucine and isoleucine are not distinguishable under the experimental conditions selected for this experiment. For this reason they are both designated with the letter L by the sequencing algorithm.

The results of the alignment of the sequenced peptides obtained by tandem mass spectrometry, with the six proteins chosen for comparison using CIDentify (Taylor and Johnson, 2001) (Section 3.1.7.5), are shown in Table 3.11.



Eleven of them showed at least some degree of homology with the published RT Vtg sequence. Two of the sequenced peptides obtained from the precursor ions at  $m/z$  573.66 and  $m/z$  561.68 also showed homology with the published AS sequence. In fact, the sequence from the precursor ion at  $m/z$  573.66 proved to be a better match with the AS Vtg than the RT Vtg. Three sequences from the precursor peptide ions at  $m/z$  795.18, 893.28 and  $m/z$  791.05 did not show any homology with the selected proteins. Finally, none of the sequences generated corresponded to trypsin autolysis peptides or aprotinin tryptic peptides.

Table 3.11 '*De novo*' sequencing and sequence homology of AS Vtg tryptic peptides

Mr indicates the relative molecular mass of the precursor (Prec) ions. The numbers between brackets under the Lutefisk output column indicate mass gaps that could not be assigned to any particular amino acid sequence. The matching amino acid residues indicating homology to the RT Vtg sequence are shown underlined. In bold font appear the conserved substitutions found in the analyzed peptides. RT 1,2,3 and 4 correspond the four variant sequences of RT Vtg found in the Swiss-Prot database.

Prec. (m/z)	Calc. Mr	Lutefisk output	Match	Database sequence	Mr
788.3	1574.5	[212.15]TMNEEEEAPEGK	RT	EKIIK <u>VITMNEEEEAPEGK</u> TVLLK	1574.7
573.7	1145.3	[216.08]GLSSFYEK	AS	SNLSR <u>EGLSSFYEK</u> SVDLR	1145.5
			RT	SNLNR <u>SEGLSSIYEK</u> SVDLM	1111.5
700.2	1398.4	[242.10]AMLGYGTL[204.09]K	RT	HPVLR <u>ELTMLGYGTMVSK</u> YCVEH	1428.7
794.8	1587.6	[184.12]AVELEDTLCAER	RT	KPIIK <u>AIVVHLEETICVER</u> LGVKA	1666.9
834.3	1666.6	[276.14]ALYLAELV[244.12]QK	RT1,3,4	EFIPGYIPYY <u>LADLVPMQK</u> DKNNE	2555.3
			RT 2	EFIPGHIPYY <u>LADLVPMQK</u> DKNSE	2529.3
889.3	1776.6	[212.15]NEPLFEYS[GV]WPK	RT	TYLLK <u>LVNPEIFEYS</u> GVWPK DPFVP	1776.9
819.8	1637.6	[TY]FAGAAADVLEVGVR	RT	VAKAR <u>TYFAGAAADVLEVGVR</u> TEGIQ	1637.8
561.7	1121.4	CYSV[DH]WR	RT and AS	DRESK <u>CYSVEPVL</u> R CLPGC	1121.6
865.3	1728.5	[241.14]NLLAALDESDNWK	RT1,3,4	KATSR <u>LQHMAALDEN</u> DNWK LCADG	1772.9
			RT2	TTSDEPQI <u>TAALDEN</u> DNWK LCADG	2260.1
843.8	1685.6	[EY]AVLLG[DH]VLLA[274.07]	RT	TYVYK <u>YEALLL</u> GGLPEEGLAR AGVK	1699.9
572.2	1142.4	LPTVTE[243.08]LK	RT	GALILEAT <u>VTEL</u> HQFTPFN	3630.7
795.2	794.2	M[244.12]QWS	none		
893.3	1784.6	[212.15]TNDDH[204.09]EDW[LP]K	none		
791.1	3160.2	[MT]NELLEFET[RD][212.15]AAVHFNLEPCANK	none		

A thorough inspection of the alignments (Table 3.11) permitted the elucidation of some of the differences found between the sequenced peptide ions and the database sequences. The CID-MS/MS of the selected precursor peptide ion at  $m/z$  788.28 showed a gap at the beginning of the sequence of 212.15. This mass is compatible with the dipeptide VI, which agrees with the sequence found for the RT Vtg. A similar situation has been observed in the CID-MS/MS of selected precursor peptide ions at  $m/z$  573.66, 700.20, 794.75, 834.31, 889.28 and  $m/z$  865.27, in which the 216.08, 242.10, 184.08, 276.14, 212.15 and 241.14 gaps could be replaced by the dipeptides SE, EL, AI, YI, LV and LQ, respectively, confirming protein homology. Although not completely sequenced, dipeptides [GV], [TY] and [EY] were correctly assigned in the precursor peptide ions at  $m/z$  889.28, 819.79 and  $m/z$  843.81. Additionally, the precursor peptide ion at 834.31 possessed a gap of 244.12 near the carboxyl terminal which, interestingly, agrees with the PM dipeptide, in which the methionine has been oxidized.

It is noteworthy that some non-matching amino acids found between the sequenced precursor peptide ions and the database sequences are replacements with others of similar physicochemical properties. Some of these sequence variants are supported by the difference between the calculated Mr of the precursor peptide ions and the theoretical Mr of the tryptic peptides found in the database, as seen in Table 3.11. These conservative substitutions can be seen in the precursor peptide ions at  $m/z$  834.31, where an aspartic acid (D) residue has been found in place of a glutamic acid (E). Another D-E replacement and an alanine (A) in place of a valine (V) in two other sites can be seen in the peptide ion at  $m/z$  794.75 (see bold font on sequences in Table 3.11). A peak found in the MALDI-ToF-MS at  $m/z$  1588.96 supports this last hypothetical

sequence. Surprisingly, this same peak was assigned to another sequence by the peptide mass fingerprint as seen in Table 3.8. This example raises the issue of peptide identification when performing peptide mass fingerprinting, especially in the case of very large proteins, where the large amount of peaks obtained in a spectrum increases the probability of erroneous peptide assignments.

The MS/MS of the precursor peptide ion at  $m/z$  700.20 is an interesting case where the non-matching sequence output L[204.09] has a mass difference of only 0.03 with respect to the sequence MVS found in the database. This is an example where this sequencing algorithm may encounter problems in the interpretation of the spectra. Nevertheless, the alanine found in the third position of this fragment differed from the threonine found in the database, suggesting the possibility of a sequence variant found in the AS Vtg. Note that the expected Mr of this RT Vtg tryptic peptide, differs from the one calculated from its precursor peptide ion. A similar case has been observed for the MS/MS of the precursor peptide ion at  $m/z$  561.68. In this case, the sequenced [DH]W tag found near the carboxyl terminal of the peptide, has a mass of 438.16, while the EPVL sequence found in the database has a mass of 438.25. Again, this CID-MS/MS has an ambiguous interpretation, as seen in Figure 3.9. B- and Y- type fragment ions are indicated above the peaks with their corresponding single letter amino acid code sequence. Spectral peaks are compatible with both the sequences of EPVLR (top spectrum in Figure 3.9) and DHWR (bottom spectrum in Figure 3.9) obtained from the database and peptide sequencing, respectively.

Coincidentally, the sequenced fragment and the peptide sequence found in the database have a Mr difference of only 0.08. Furthermore, the peaks observed in this CID-

MS/MS can be tentatively assigned to both sequences. In this last case, there is not enough evidence to support one structure over the other.

An interesting observation for the CID-MS/MS of the selected precursor peptide ion at  $m/z$  834.31 is that although a match can be observed with the RT Vtg, the database sequence does not contain either a residue of lysine (K) or arginine (R) preceding the residues matching the beginning of the sequenced peptide. This may indicate that a new cleavage site has been introduced next to the N terminal of this peptide on the AS sequence, although a non-specific cleavage during sample preparation can not be ruled out.

The selected precursor peptide ions at  $m/z$  788.28, 889.28, 819.70 and 561.68 used in the ESI-MS/MS analyses were also found in MALDI-ToF-MS analyses as  $[M+H]^+$  species. These CID-MS/MS sequenced peptides had a perfect match with the RT Vtg sequence, with the exception of the precursor peptide ion at  $m/z$  561.68, as described above, and a dipeptide inversion in positions 3 and 4 of the sequenced precursor peptide ion at  $m/z$  889.28. The CID-MS/MS of the selected precursor peptide ions at  $m/z$  843.81 and  $m/z$  572.20 showed only to share a small tag with respect to the database search.

Finally, a remark should be made about the sequenced peptides obtained from the CID-MS/MS of the selected precursor peptide ions at  $m/z$  795.18, 893.28 and  $m/z$  791.05. These are apparently exclusive of the AS Vtg, possibly indicating regions of sequence variability. Further characterization of these and other non-identified fragments should allow post translational modifications and genetic variants to be confirmed.

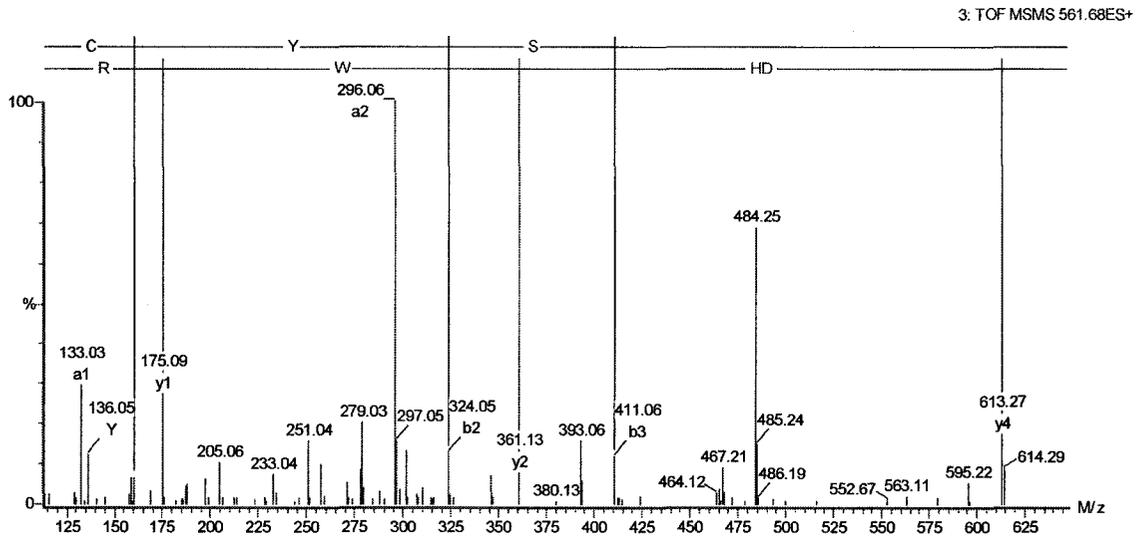
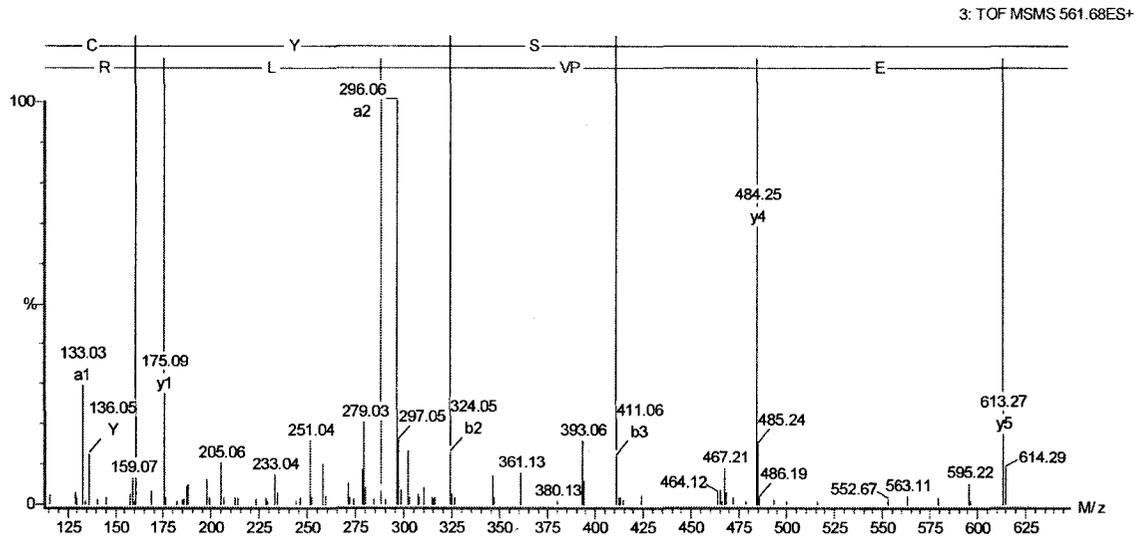


Figure 3.9 Ambiguous graphical interpretation of CID-MS/MS spectra

Graphical interpretation of the CID-MS/MS spectra obtained from the precursor ion at  $m/z$  561.68. The amino acid sequence of the peptides is given in the single letter code above the corresponding Y and B product ions. As shown in this example, the same peaks can be tentatively assigned to different amino acid sequences.

### **3.4 In-gel digestion analyses of vitellogenin**

#### **3.4.1 MALDI-MS of the trypsin derived peptides of rainbow trout vitellogenin**

In-gel trypsin digestions of Vtg were assayed directly from plasma as an alternative to digestions of the purified Vtg in solution. The procedure for in-gel digestions was described in Section 3.1.4. The rationale behind doing this type of digestion was that it avoided the time consuming process of protein purification, thus reducing the possibilities of protein degradation or modification. Control and 17 $\beta$ -estradiol induced plasma samples were analyzed in parallel by SDS-PAGE. After staining with Coomassie Blue, the Vtg band was easily distinguished by comparing the intensity of the bands between both samples. This is a routine analytical technique which permits the perusal of qualitative and semi-quantitative results obtained in protein biochemistry experiments. In a classical biochemistry approach to protein characterization and identification, these bands could be further blotted onto membranes for amino acid sequencing by the Edman degradation reaction. Alternatively, thanks to the sensitivity offered by mass spectrometry, they can be enzymatically or chemically digested for additional identification and characterization experiments.

The bands corresponding to Vtg were excised and subjected to in-gel digestion with trypsin and the resulting peptides were analyzed by MALDI-MS (see Section 3.1.6). The MALDI-MS spectrum obtained for RT Vtg is shown in Figure 3.10. The details of the PMF are shown in Table 3.12.

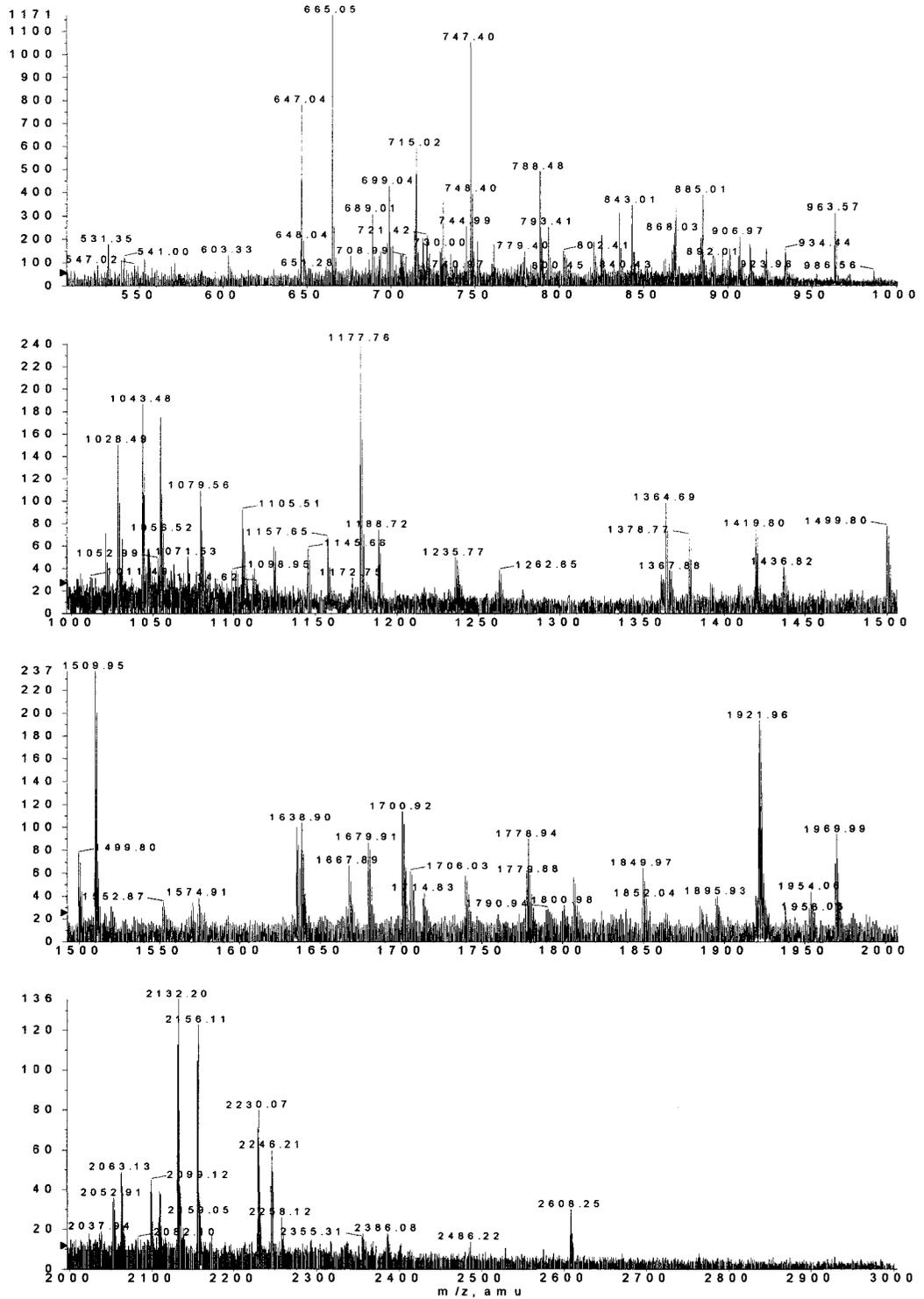


Figure 3.10 MALDI-MS spectra of the in-gel tryptic digestion of RT Vtg  
 The abscissa of the spectrum was split into four  $m/z$  ranges to facilitate the visualization of the peaks.

Table 3.12 Peptide mass fingerprint of the in-gel digestion of RT Vtg

The Start and End column indicate the position along the RT Vtg sequence (See Appendix 2). The experimental (expt) and calculated (calc) relative molecular masses (Mr) are shown. Miss indicates the presence of missed cleavages. The ‘.’ in the Sequence column indicate the cleavage sites.

Start	End	Observed	Mr (expt)	Mr (calc)	$\Delta$ Mass	Miss	Sequence	
380	-	384	665.05	664.04	663.40	0.65	1	R.FIKEK.F
713	-	718	747.40	746.39	746.37	0.02	0	K.ALSDWR.S
164	-	169	788.46	787.45	787.43	0.02	0	K.THYVIR.E
1639	-	1645	820.45	819.44	820.40	-0.96	0	K.SVDLMEK.A
1510	-	1515	825.03	824.02	823.46	0.56	1	K.YSWKIK.V
545	-	551	835.49	834.48	834.47	0.01	0	K.ALHPELR.M
1646	-	1653	913.44	912.43	912.42	0.01	0	K.AEAHVACR.C
1522	-	1530	922.96	921.95	922.40	-0.45	0	K.GQTCGLCGK.A
1249	-	1256	934.45	933.44	933.44	0.00	0	K.IAWGAECK.D
425	-	432	963.57	962.56	962.57	0.00	0	K.IQTHPVL.R
1279	-	1286	986.55	985.54	985.56	-0.02	1	R.LSWDKLPK.V
918	-	926	1028.56	1027.55	1027.49	0.06	0	R.NVEDVPAER.I
195	-	203	1043.51	1042.50	1042.50	0.01	0	K.DFLAYTEK.C
1601	-	1609	1071.56	1070.55	1070.54	0.02	0	R.CLPGCLPVR.T
694	-	703	1079.56	1078.55	1078.54	0.01	0	K.LPPAPENADR.I
1629	-	1638	1111.61	1110.60	1111.54	-0.94	0	R.SEGLSSIYEK.S
1043	-	1052	1145.67	1144.66	1144.61	0.05	0	R.LEFEVQVGP.K
415	-	424	1157.67	1156.66	1156.62	0.04	0	K.LVESLAFNHK.I
333	-	341	1177.77	1176.76	1176.70	0.06	0	K.FLQFQLLR.M
1582	-	1591	1188.78	1187.77	1187.61	0.16	1	K.QVIVDDRESK.C
927	-	938	1235.80	1234.79	1234.74	0.05	0	R.ITPLIPAQGVAR.S
182	-	191	1262.65	1261.64	1261.58	0.06	1	K.SKDLNQCQR.I
184	-	194	1361.73	1360.72	1361.65	-0.93	1	K.DLNQCQRIMK.D
1134	-	1145	1387.78	1386.77	1386.58	0.19	0	R.DGPDQYPNPDR.K
1435	-	1446	1391.72	1390.71	1390.66	0.05	0	K.DHASEQNHNK.I
1379	-	1390	1419.79	1418.78	1418.76	0.03	0	K.IHYLFSEVNAV.K
733	-	745	1499.79	1498.78	1498.75	0.04	0	K.FFGQEIAPANIDK.S
499	-	513	1509.96	1508.95	1508.91	0.04	0	K.LLPVFGTAAALPLR.V
270	-	282	1551.84	1550.83	1550.78	0.05	0	K.DPIHVPDNNYVHR.G
483	-	498	1574.92	1573.91	1573.89	0.02	0	K.ALGNAGHPASIKPITK.L
1001	-	1014	1635.79	1634.78	1634.74	0.04	0	K.ACFFTSESAIFIR.N
668	-	683	1638.87	1637.86	1637.84	0.02	0	R.TYFAGAAADVLEVGVR.T
983	-	996	1667.90	1666.89	1666.87	0.02	0	K.AIVVHLEETICVER.L
269	-	282	1679.93	1678.92	1678.88	0.04	1	K.KDPIHVPDNNYVHR.G
34	-	49	1700.95	1699.94	1699.91	0.03	0	K.YEALLGGLPEGLAR.A
119	-	134	1706.00	1704.99	1704.95	0.04	0	K.VLAPTAVSETVLNVHR.G
72	-	86	1777.94	1776.93	1776.91	0.02	0	K.LVNPEIFEYSVWP.K
363	-	379	1807.01	1806.00	1805.98	0.02	0	R.HWILDAVPSIGSSVAVR.F
1198	-	1213	1849.99	1848.98	1848.94	0.05	1	R.ADHKIEGYQVTAAYLNK.A
147	-	163	1894.92	1893.91	1893.89	0.02	0	K.TQNVYELQEAGAQGVCK.T
1549	-	1565	1921.97	1920.96	1920.88	0.08	0	K.SSVSFAHSWVLPDSCR.D
287	-	302	1937.94	1936.93	1936.04	0.89	0	R.YEFATEILQMPIQLL.I
642	-	660	1969.04	1968.03	1968.01	0.02	0	R.IGAAASAFYINDAATLFP.R
965	-	982	2051.92	2050.91	2050.10	0.81	0	R.SSEILYSDLPSNFKPIK.A
746	-	765	2062.10	2061.09	2061.10	0.00	0	K.SIIDQALQLANSPSAHALGR.N
95	-	113	2098.13	2097.12	2097.15	-0.03	1	K.LTSALAAQFSIPIKFEYAK.G
1257	-	1276	2108.06	2107.05	2107.06	-0.01	0	K.DYNTFITAETGLVGPSPAVR.L
770	-	789	2131.22	2130.21	2130.22	-0.01	0	K.ALLAGATFQYVKPLLAEEVR.R
1610	-	1628	2155.10	2154.09	2154.09	0.00	0	R.TTPTITIGFHLCPVDSNLNR.S
623	-	641	2229.10	2228.09	2228.08	0.01	0	R.LSCHFSQAIHLDAYSNPLR.I
918	-	938	2245.48	2244.47	2244.22	0.25	1	R.NVEDVPAERITPLIPAQGVAR.S
29	-	49	2355.28	2354.27	2354.25	0.02	1	K.TYVYKYEALLGGLPEGLAR.A
622	-	641	2385.14	2384.13	2384.18	-0.05	1	R.RLSCHFSQAIHLDAYSNPLR.I
147	-	169	2607.30	2606.29	2606.29	0.00	1	K.TQNVYELQEAGAQGVCKTHYVIR.E

The increased background and low signal-to-noise observed in this spectrum is probably due to the low concentration of peptides extracted from the in-gel digestion step. The peaks obtained from this spectrum were selected to perform a PMF search using MASCOT (Section 3.1.7.3). RT Vtg was the only significant hit reported in the results of this search. No matches were found for the other peaks observed in figure 3.10.

#### **3.4.2 MALDI-MS/MS of the trypsin derived peptides of rainbow trout vitellogenin**

The product ion spectra of the peptides extracted from the in-gel digestions were acquired in low energy CID-MS/MS mode with the same instrumentation as described for the previous experiments. The parameters chosen for these studies are described in Section 3.1.6. The product ion spectra were analyzed using the '*de novo*' sequencing algorithm provided by the Bioanalyst software (see Section 3.1.7.4). Additionally, the spectra were visually inspected to assess the suitability of the sequences suggested by the sequencing algorithm. Table 3.13 shows the sequences obtained from the analyses of these spectra, together with the matching sequences to RT Vtg (Appendix 2). The matching sequences to RT Vtg are shown underlined next to the candidate sequences, while discrepancies in amino acid sequences are emphasized in boldface font. For these particular peptides, sequence alignment was carried out using the FASTA alignment program available on the internet (see Section 3.1.7.5). The inspection of Table 3.13 shows some interesting features. The peptide ions at  $m/z$  963.57, 1043.51, 1079.56 and

1177.77 were unequivocally sequenced and had perfect alignment with the published RT Vtg sequence.

Some of the candidate sequences had slight differences with the sequence of RT Vtg, exemplified by the peptide ions at  $m/z$  1028.56, 1419.79 and 1509.96. However, in all these examples, the sequenced peptides had the same molecular mass as those predicted from the RT Vtg sequence. This suggests that ambiguous interpretation of the product ion spectra by the sequencing software occurred (shown in boldface font on Table 3.13).

Table 3.13 '*De novo*' sequencing and sequence homology of RT Vtg tryptic peptides obtained by in-gel digestion

Homology between the Candidate sequence and the Matching sequence is underlined. The non-matching sequences are showed in bold font. The Position column indicates the location of the matching peptide in the amino acid sequence of RT Vtg, entry Q92093 in the Swiss-Prot Database (See Appendix 2).

$[M+H]^+$ $m/z$	Candidate sequence	Matching sequence	Position
963.57	<u>LQTHPVLR</u>	K <u>IQTHPVLR</u> E	425-432
1028.56	<u>NVEDVEPAR</u>	R <u>NVEDVPAER</u> I	918-926
1043.51	<u>DFGLAYTEK</u>	K <u>DFGLAYTEK</u> C	195-203
1054.53	HGGGCAGSGPVR	No match	
1079.56	<u>LPPAPENADR</u>	K <u>LPPAPENADR</u> I	694-703
1105.56	<b>QQHVEPVLR</b>	K <u>CYSVEPVLR</u> C	1595-1600
1177.77	<u>FLQFLQLLR</u>	K <u>FLQFIQLLR</u> M	333-341
1235.8	QAPGALCGLLHR	No match	
1361.73	<b>DSTVPTMAGAAANR</b>	<b>T RSTAPDFASVAAACNVAVK</b> M	597-607
1365.57	HGGSYGAFGGVAASK	No match	
1419.79	<u>HLYLFSEVNAAR</u>	K <u>IHYLFSEVNAVK</u> C	1381-1390
1509.96	<u>LLASLATEAAALPLR</u>	K <u>LLPVFGTAAAALPLR</u> V	499-513
2607.3	(1891.99)APSLNSK	No match	

For example, for the peptide ion at 1028.56, the sequence EPA is a result of the amino acid inversion PAE found in the RT Vtg sequence. This type of amino acid inversion is not uncommon when performing ‘*de novo*’ sequencing from product ion spectra. For the ions at  $m/z$  1419.79 and 1509.96, small tags of isobaric amino acid sequences are replaced: AR for VK and ASLA-E for PVFG-A, respectively. This is also a common occurrence when the information provided by the product ion spectrum is insufficient to provide a detailed fragmentation pattern for the B and Y ions present in that particular region of the peptide. The peptide ions at  $m/z$  1105.56 and 1361.73 had only partial sequence homology with the RT Vtg sequence. Furthermore, these peptides do not agree with the theoretical molecular mass of the corresponding tryptic peptide, suggesting these might possibly be sequence variants found for Vtg in this sequence. Similarly, no matching sequences were found for the peptide ions at  $m/z$  1054.53, 1235.80, 1365.57 and 2607.30.

### **3.4.3 MALDI-MS of the trypsin derived peptides of Atlantic salmon vitellogenin**

The AS Vtg was analyzed in the same way as described for RT Vtg in section 3.4.1. The MALDI-ToF-MS spectrum obtained for AS Vtg is shown in Figure 3.11. The top, center and bottom spectra show the mass scale from  $m/z$  800-1500, 1500-2000 and 2000 to 3000, respectively. The peaks obtained from this spectrum were selected to perform a PMF search using MASCOT (Section 3.1.7.3).

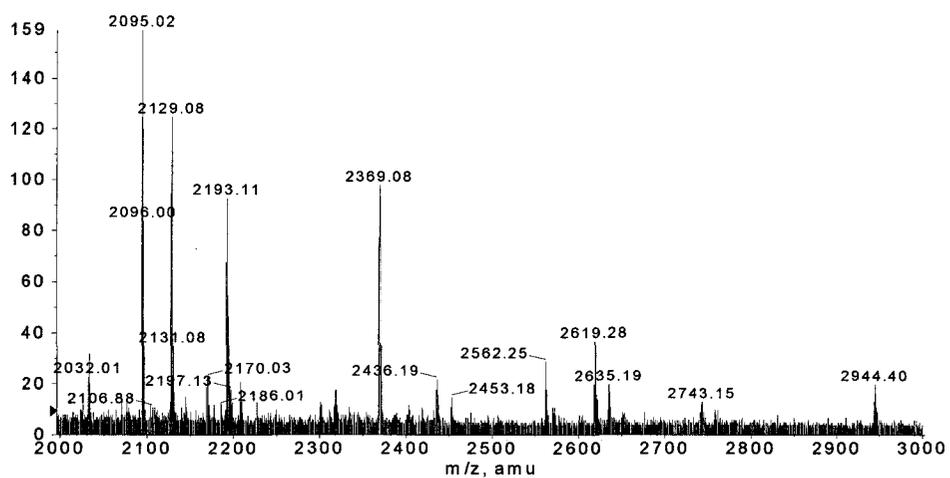
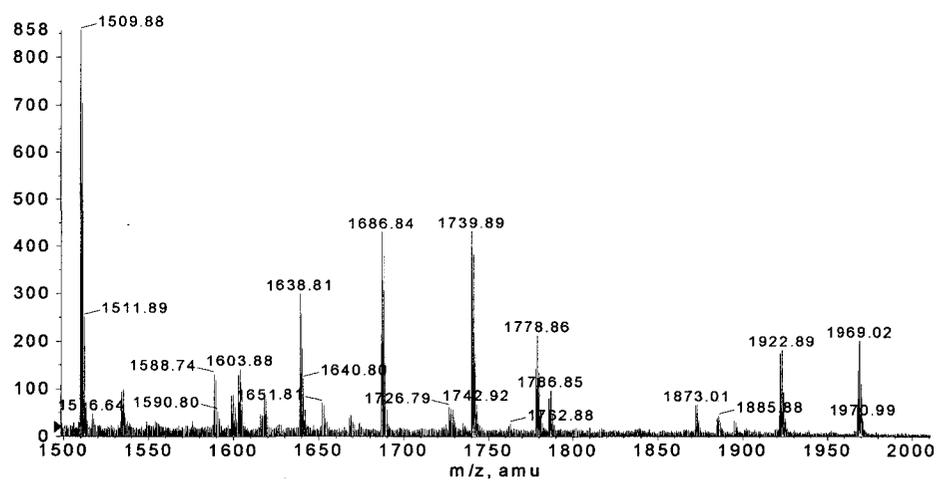
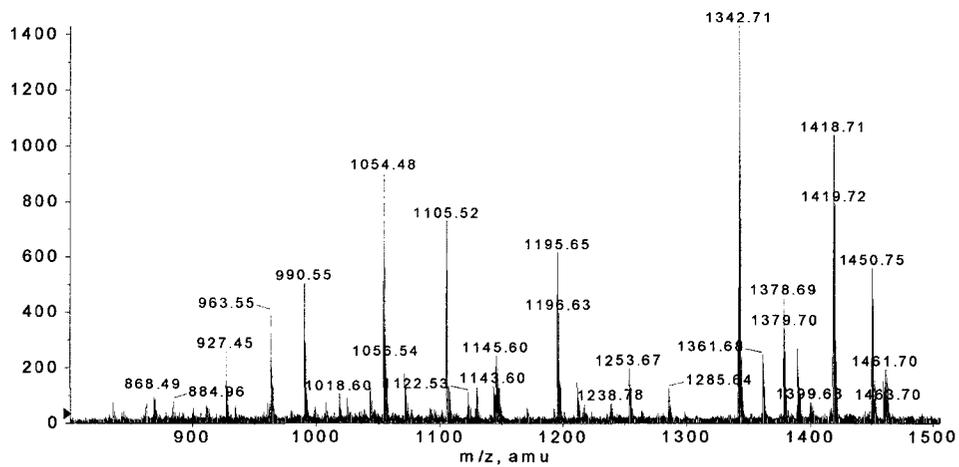


Figure 3.11 MALDI-MS of the in-gel tryptic digestion of AS Vtg

The abscissa of the spectrum was split into four  $m/z$  ranges to facilitate the visualization of the peaks.

As with the solution digestion of AS Vtg, none of the proteins in the database showed a significant hit from this list of MALDI-MS peaks. This was expected, since the complete sequence of AS has not yet been reported. However, RT Vtg (Appendix 2: entry Q92093 on the Swiss-Prot Database) was the protein with most matching peptides for this peak list. The details of the PMF are shown in Table 3.14. It should be noted that in this case the matching peptides and their relative position correspond to those found in RT Vtg.

These results again show evidence of the striking similarity for this protein between these two species. No match was found for the peaks at  $m/z$  868.56, 911.24, 927.44, 1105.52, 1129.58, 1143.58, 1195.65, 1211.65, 1253.67, 1285.63, 1342.71, 1450.75, 1461.73, 1588.78, 1598.63, 1726.79, 1739.91, 1872.01, 1884.88, 2094.01, 2128.05, 2169.00, 2192.11, 2208.11, 2300.09, 2318.12, 2368.11, 2435.17, 2634.21, 2636.20 and 2943.44. The peaks at  $m/z$  963.55, 1043.49, 1071.53, 1145.60, 1361.68, 1509.88, 1638.81, 1777.88, 1894.89 and 1921.89, which were assigned to tryptic peptides in the RT sequence, were also found in the in-gel digestion of RT Vtg.

Table 3.14 Peptide mass fingerprint of the in-gel digestion of AS Vtg

Refer to legend in Table 3.12 for details.

Start	End	Observed Mr (expt)	Mr (calc)	$\Delta$ Mass	Miss	Sequence	
545	- 551	835.46	834.45	834.47	-0.02	0	K.ALHPELR.M
425	- 432	963.55	962.54	962.57	-0.02	0	K.IQTHPVL.R
184	- 191	990.54	989.53	989.43	0.10	0	K.DLNNCQQR.I
195	- 203	1043.49	1042.48	1042.50	-0.01	0	K.DFGLAYTEK.C
1396	- 1404	1054.49	1053.48	1052.51	0.97	0	R.DTLTTFNNK.K
1601	- 1609	1071.53	1070.52	1070.54	-0.01	0	R.CLPGCLPVR.T
1592	- 1600	1122.55	1121.54	1121.55	-0.01	0	K.CYSVEPVL.R
1043	- 1052	1145.60	1144.59	1144.61	-0.02	0	R.LEFEVQVGPK.A
184	- 194	1361.68	1360.67	1361.65	-0.98	1	K.DLNNCQQRIMK.D
1325	- 1336	1378.68	1377.67	1377.73	-0.05	0	K.QIQFTVVATSER.T
1116	- 1129	1389.73	1388.72	1388.60	0.12	1	R.KSESSSSSSSSSR.T
184	- 194	1418.71	1417.70	1418.68	-0.97	1	K.DLNNCQQRIMK.D
95	- 108	1459.88	1458.87	1458.84	0.03	0	K.LTSALAAQFSIPIK.F
499	- 513	1509.88	1508.87	1508.91	-0.04	0	K.LLPVFGTAAALPLR.V
58	- 71	1533.86	1532.85	1532.88	-0.03	0	K.VLISAVAENTYLLK.L
668	- 683	1638.81	1637.80	1637.84	-0.04	0	R.TYFAGAAADVLEVGVR.T
1646	- 1659	1651.83	1650.82	1650.64	0.19	1	K.AEAHVACRCSEQCM.-
1391	- 1404	1686.87	1685.86	1685.79	0.08	1	K.CSMVRDTLTTFNNK.K
72	- 86	1777.88	1776.87	1776.91	-0.04	0	K.LVNPEIFEYSGVWPK.D
947	- 964	1785.87	1784.86	1783.88	0.99	0	K.LTSMIADSAASFAGLSR.S
147	- 163	1894.89	1893.88	1893.89	-0.01	0	K.TQNVYELQEAGAQGVCK.T
1549	- 1565	1921.89	1920.88	1920.88	0.00	0	K.SSVSFAHSWVLPDSCR.D
642	- 660	1968.02	1967.01	1968.01	-1.00	0	R.IGAAASAFYINDAATLFPR.T
1447	- 1464	2032.01	2031.00	2030.02	0.98	0	K.ISDIDVDLYTEDHGVIVK.V
1406	- 1427	2561.35	2560.34	2560.24	0.11	1	K.YKINMPLSCYQVLAQDCTTELK.F
1406	- 1427	2618.20	2617.19	2617.26	-0.07	1	K.YKINMPLSCYQVLAQDCTTELK.F
1549	- 1573	2742.21	2741.20	2741.22	-0.02	1	K.SSVSFAHSWVLPDSCRDAECLMK.L

#### **3.4.4 MALDI-MS/MS of the trypsin derived peptides of Atlantic salmon vitellogenin**

The peptides obtained from the in-gel trypsin digestion of AS Vtg were also subjected to MALDI-MS/MS, as described for the RT Vtg peptides in section 3.4.2. The 'de novo' sequencing and the alignment results are portrayed in Table 3.15. The matching sequences to RT or AS Vtg are shown underlined next to each candidate sequence. AS1 and AS2 stand for Vtg entries Q800N6 and Q8UWG2, respectively (See Appendix 2). Discrepancies in amino acid sequences are emphasized in boldface font.

As expected, the alignment studies shows fewer candidate sequences because of the sequence variants found in AS Vtg with respect to RT Vtg and because only two partial sequences of AS Vtg are known at the present time. Therefore, most non-matching sequenced peptides are believed to be sequence variants found in the non-sequenced regions of AS Vtg. However, as shown in this table, some peptides exhibited a marked homology with the RT Vtg sequence.

Some interesting cases found in this table are worth mentioning for discussion. The ions at  $m/z$  927.44 and 1739.91 were found to partially align with sequences of both RT and AS Vtg. Interestingly the sequence homology with RT was better than with AS, for both cases. Furthermore, the molecular mass of the peptides did not agree with that predicted from the sequence of RT or AS Vtg.

Table 3.15 'De novo' sequencing and sequence homology of AS Vtg tryptic peptides obtained by in-gel digestion

Homology between the Candidate sequence and the Matching sequence is underlined. The non-matching sequences are showed in bold font. The Species column indicates whether the matching sequence corresponds to rainbow trout (RT, entry Q92093) or one of the two published sequences of Atlantic salmon Vtg (AS1, entry Q800N9 and AS2, entry Q8UWG2 in the Swiss-Prot Database). The position indicates the location of the matching peptide in the amino acid sequence of the corresponding proteins.

$[M+H]^+$ <i>m/z</i>	Candidate Sequence	Matching sequence	Species	Position
911.24	<b>CHN</b> <u>VEPGR</u>	K <u>CYSVEPVLR</u> C	RT	1592-1597
			AS2	108-113
927.44	<b>EGVA</b> <u>HCGVR</u>	K <b>YEALLGGLPEEGLARAGVK</b> V	RT	45-52
		K <b>YEALLGGLPEEGLSRAGIK</b> V	AS1	45-52
990.53	<b>GGLGGAGHLPR</b>	K <u>ALGNAGHPASIK</u> P	RT	484-489
1043.48	<b>DFGLAYTEK</b>	K <u>DFGLAYTEK</u> C	RT	195-203
		K <u>EFGLAYTENCVECCQK</u> G	AS1	195-202
1054.50	<b>AAGGNP</b> <u>CLPVR</u>	R <u>CLPGCLPVR</u> T	AS2	121-125
			RT	1605-1609
1105.52	<b>QQH</b> <u>VEPVLR</u>	K <u>CYSVEPVLR</u> C	RT	1595-1600
			AS2	111-116
1195.65	<b>FQMFLQLLR</b>	K <u>FLQFIQLLR</u> M	RT	333-341
1342.71	<b>FAANPLPANQGSR</b>	No match		
1361.68	<b>FGADTAVDGHSEK</b>	No match		
1378.68	<b>QSVPEASCNAAQR</b>	R <b>AVVLALTLALVASQSVNFAPDFAASK</b> T	RT	16-28
1418.71	<b>CALPSGAAQADLFR</b>	No match		
1450.75	<b>HLTSYLADLFDR</b>	No match		
1509.88	<b>LLASLL</b> <u>TGAAALPLR</u>	K <u>LLPVFGTAAAALPLR</u> V	RT	499-513
1638.81	<b>FYTAGAAWVLEVGVR</b>	R <b>TYFAGAAADVLEVGVR</b> T	RT	669-683
1686.86	<b>YEALVVQMNLQSHR</b>	No match		
1739.91	<b>AMD</b> <u>AVVSETVLGGVHR</u>	K <u>VLAPTAVSETVLN-VHR</u> G	RT	125-134
		K <u>VFAPTGVSETVLN-IHR</u> G	AS1	125-134
1777.89	<b>WGWVGS</b> <u>LVGGLHVWPK</u>	K <u>LVNPEIFEYSGVWPK</u> D	RT	83-86
		K <u>LVDPEIFEYSGVWPK</u> D	AS1	83-86
1872.00	<b>RPYEDVKPQVSHMMR</b>	K <b>YCV EHPNCPAELVKPIHELAVQAVANSK</b> F	RT	454-460
1921.88	(951.3158) <b>WSHVVGMR</b>	No match		
1968.00	<b>NAGPGV</b> <u>AHMFDPAAATLFPR</u>	R <b>IGAAASAFYINDAATLFPR</b> T	RT	648-660
2094.01	(1227.4888) <b>LVGSPAAAR</b>	K <b>DYNTFITAETGLVGPSPAVR</b> L	RT	1268-1276
2128.05	(1437.6974) <b>DSNPCR</b>	R <b>LSCHFSAIHLDAYSNPLR</b> I	RT	637-641
2368.11	(1376.5099) <b>LVQPLHER</b>	K <b>YCV EHPNCPAELVKPIHELAVQAVANSK</b> F	RT	457-463

Similar examples where the molecular mass of the sequenced peptides did not agree with the molecular masses of the predicted RT and AS Vtg tryptic peptides were found for the peptide ions at  $m/z$  911.24, 990.53, 1054.50, 1105.52, 1195.65, 1378.68, 1739.91, 1872.00, 2094.01, 2128.05 and 2368.11. These discrepancies are likely to occur due to sequence variants within the analyzed peptide and/or to the introduction of new enzymatic cleavage sites which produces peptides which differ in length, but conserve sequence homology.

The peptide ion at  $m/z$  1043.48 matched the sequence of RT Vtg perfectly. However, the published sequence of AS Vtg contains a similar sequence, but lacks the terminal lysine residue found in the analyzed peptide. The presence of this peptide suggests a sequence variant to the entry Q800N6 found in the protein database. The peptide ions at  $m/z$  1509.88, 1638.81 and 1968.00 also matched the molecular mass of the tryptic peptides predicted from this digestion, however, only partial amino acid sequences were found to match the sequenced peptides.

Another interesting example is that of the peptide ion at  $m/z$  1777.89. The four C-terminal amino acids found in this peptide matched sequences found for RT and AS Vtg. However, the database sequence for this tryptic peptide in both species differs by one amino acid in position 3, where Asn is found in RT, while Asp is found in AS. In this case, the molecular mass of the analyzed peptide coincides with the RT Vtg tryptic peptide. This finding suggests that this particular site might be susceptible to sequence variants or posttranslational amidation of the Asp residue. Also, inconsistencies in the protein databases are likely to exist due to errors in cloning and sequencing experiments.

### **3.4.5 MALDI-MS of the trypsin derived peptides of Atlantic cod vitellogenin**

The efficiency of Vtg induction with 17 $\beta$ -estradiol was assessed by SDS-PAGE as seen in Figure 3.12. The strong induced band which appeared in the experimental fish was excised and subjected to in-gel trypsin digestion as described in Section 3.1.4.

The peptides extracted from the digestion were subjected to MALDI-ToF-MS (Figure 3.13) as described in Section 3.1.6. Both spectra correspond to the same acquisition and the axes have been cut to improve peak visualization. The top and bottom spectra correspond to a  $m/z$  range of 800-1800 and 1700-3000, respectively.

Although no sequence for Atlantic cod Vtg is currently available, the peptide ions obtained in the enzymatic digestion were compared with those generated ‘in-silico’ in the protein databases, since Vtg sequences from other species have already been reported. The  $[M+H]^+$  ions present in the spectra were used for protein identification using the Mascot Peptide Mass Fingerprint database search engine (Section 3.1.7.3).

Interestingly, the two known different forms of haddock Vtg (VtgA and Vtg B, entries Q98T86 and Q98T87 on the TrEMBL database, respectively) gave statistically significant hits with the search query. Atlantic cod and haddock are two ground fish species which are evolutionary related to one another. Both species belong to the family Gadidae, subfamily Gadinae (Nelson, 1984). As described in literature, haddock, in common with other species, co-expresses more than one form of Vtg (Reith *et al.*, 2001) which have been shown to have different biological roles after being incorporated into the oocyte.

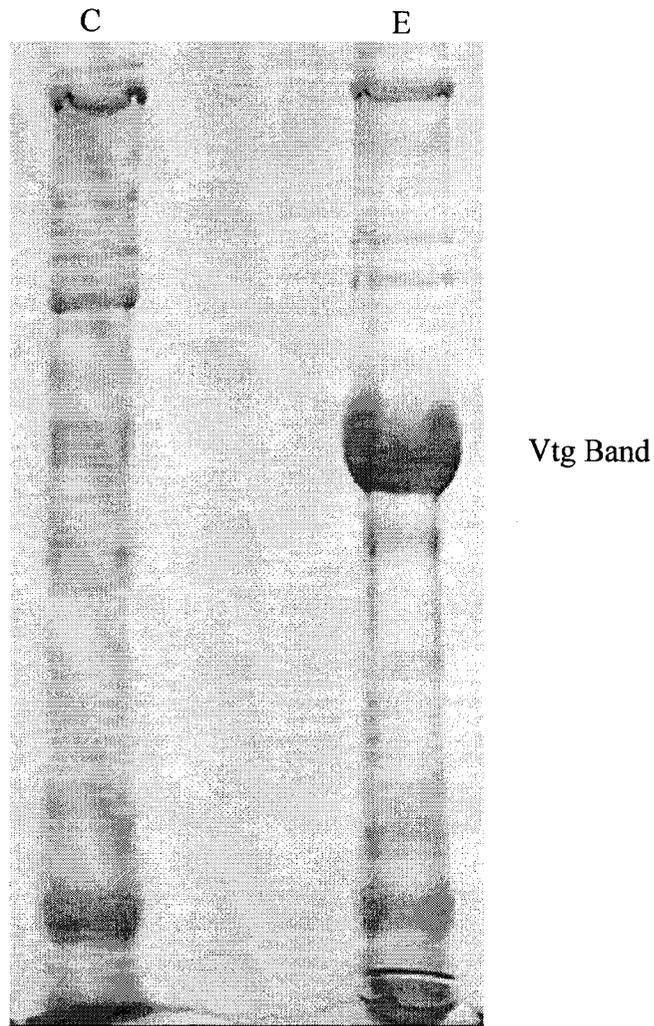


Figure 3.12 Induction of Vtg synthesis in AC fish

Plasma from control (C) and experimental (E) fish was subjected to SDS-PAGE. Five  $\mu\text{L}$  of plasma were loaded into each lane and electrophoresed in the conditions described in Section 3.1.2. The banding profile indicates the increased synthesis of Vtg in the fish treated with  $\beta$ -estradiol.

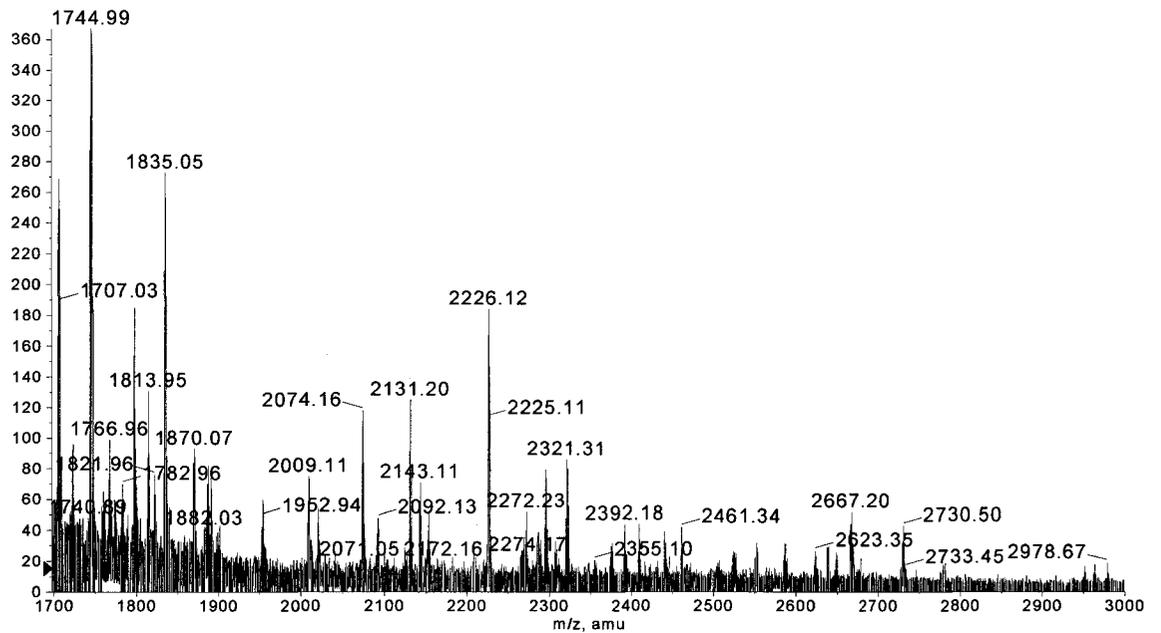
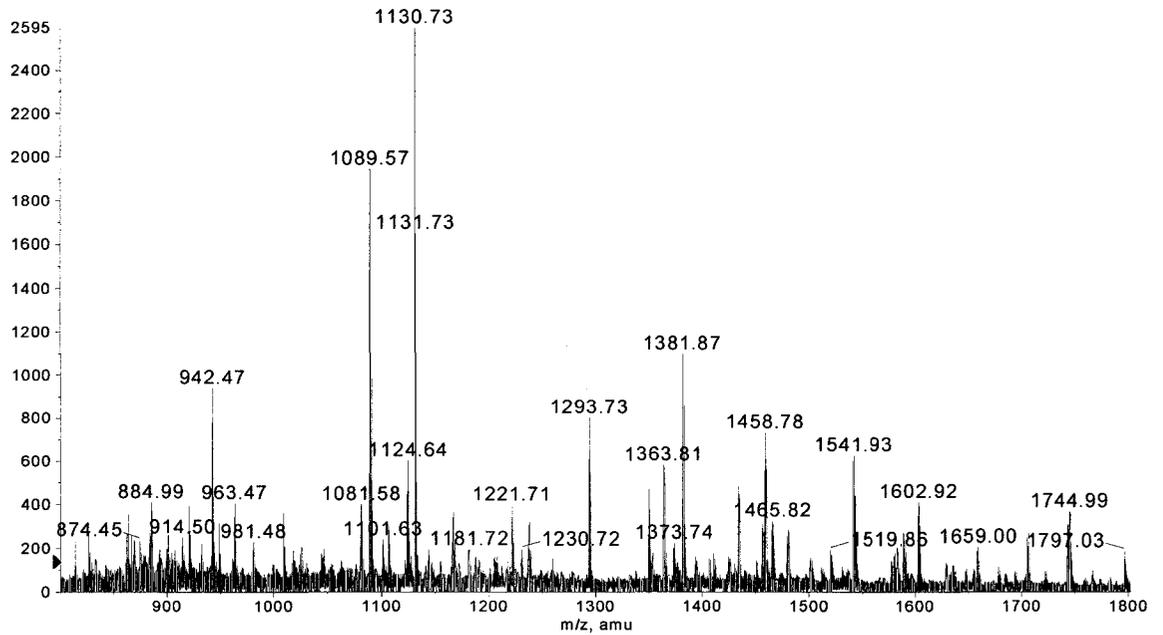


Figure 3.13 MALDI-MS of the in-gel tryptic digestion of AC Vtg  
 The abscissa of the spectrum was split into two  $m/z$  ranges to facilitate the visualization of the peaks.

A careful analysis of the MALDI-ToF-MS matching peptides between both species showed that some of the Atlantic cod Vtg peptides matched those of Vtg A, and others of Vtg B, while some were common to both Vtg A and B, as shown in Table 3.16. The sequences that correspond to haddock Vtg A and B are marked in superscript 1 and 2 respectively.

A close look at Table 3.16 shows some interesting information. The peptide ions at  $m/z$  1009.53 and 1101.63 corresponded to pairs of peptides which were not identical, but contained leucine-isoleucine substitutions. The ion at  $m/z$  1130.72 matched a pair of peptides in which two amino acids have switched positions. Alternatively, the ions at  $m/z$  1089.56 and 1743.13 were assigned to pairs of isobaric peptides with absolutely no sequence homology between them.

#### **3.4.6 MALDI-MS/MS of the trypsin derived peptides of Atlantic cod vitellogenin**

Product ion spectra were acquired by MALDI-QqToF MS/MS, as described in Section 3.1.6, to confirm the identity of both matching and non-matching peptide ions. The results of these analyses are shown in Table 3.17. As mentioned earlier, ‘*de novo*’ sequencing of peptides from product ion spectra is an intensive, time-consuming procedure. Computer software has made this task faster, yet not necessarily more accurate. When such programs are used, a list of candidate peptides is shown in score order. Such a score depends, among other parameters, on the number of matching fragment ions and their respective mass error.

Table 3.16 Peptide mass fingerprint of the in-gel digestion of AC Vtg

The headings Vtg A and Vtg B correspond to the entries Q98T86 and Q98T87 of haddock Vtg found in the Swiss-Prot database. In the Sequence column, superscript 1 and 2 refer to the Q98T86 and Q98T87 entries, when a match was found for both sequences.

[M+H] <sup>+</sup> m/z	Vtg A	Vtg B	Vtg A+B	Sequence
813.50		x		IKHFIR
942.47	x			YESFAVAR
948.46		x		VAWGIDCK
963.46	x			DLNNCQEK
1009.53	x	x	x	DLGLAYTEK <sup>1</sup> and DIGLAYTEK <sup>2</sup>
1081.57		x		LPQAPVDADR
1089.56	x	x	x	LCADGILLSK <sup>1</sup> and INLNAFAFKK <sup>2</sup>
1101.63	x	x	x	TEGLQEALLK <sup>1</sup> and TEGIQEALLK <sup>2</sup>
1106.59		x		CFSVEPVLR
1124.62	x			VHVDAILAMR
1130.72	x	x	x	FLELIQVLR <sup>1</sup> and FVELIQLLR <sup>2</sup>
1166.61	x			LESEDASFIR
1221.71		x		MAAALVLFETR
1293.73		x		FTWDKLPTSAK
1349.79	x			IVKDLGLAYTEK
1363.81		x		ITAATVETFAIAR
1372.70	x			DNIEQNWIVK
1458.78		x		FFGQEIGFASIDK
1465.82	x			FLGQEIAFANLDK
1479.84	x			TPIAPVNAQYLHR
1743.13	x	x	x	IVPIPAEILEPLIGR <sup>1</sup> and LTLALTSKTLNIVLK <sup>2</sup>
1796.01	x			VYSPEGISTTVLNIFR
1834.05		x		VQLILANLVEENHWR
2008.11	x			ASATPALPQNFLWTHLLK
2020.06		x		VHEDAPLKFVELIQLLR
2073.16		x		AHDQLIQVATGPSVATYGR
2225.11		x		LASNSVSYAQSWVIPAESCR
2294.29	x			LESEDASFIRNTPLYQLIGK
2320.29		x		GSLQYEFATELLQTPIQLLR
2665.17	x			DAVSYAHSWVIPAENCQDASECR

Nevertheless, each candidate sequence was manually checked since it is very common to find errors in peak assignment, especially with regard to the correct identification of the monoisotopic peak. Therefore, the candidate sequences shown in Table 3.17, correspond to those which best matched the MS/MS spectra, but did not necessarily match with the highest score.

Unfortunately, not all the ions found in the peptide mass fingerprint were suitable for MS/MS analysis since some spectra did not yield sufficient diagnostic peaks to permit full sequencing of the peptides. Nevertheless, some partial sequences could be obtained. In these cases a mass value between brackets is shown to account for the amino acid sequence gaps.

Examination of Table 3.17 reveals some very interesting information. The product ion spectrum of the precursor ion at  $m/z$  1089.54 (see Figure 3.14) was shown to have a sequence similar to that found in Vtg B of haddock, yet the peptide mass fingerprint had erroneously assigned this ion to other haddock peptides. The expected fragmentation pattern is shown in the figure inset while the observed diagnostic B and Y peptide product ions are outlined in boxes. There is evidence to support this new sequence (FDDVEPVLR), since the mass of this peptide differs from the theoretical mass of the corresponding haddock peptide (CFSVEPVLR, Mr 1105.56). Interestingly, the product ion spectrum of the ion at  $m/z$  1106.58 (see Figure 3.15) showed no apparent similarity to the peptide assigned to it on the peptide mass fingerprint. These two proposed peptides should be considered tentative, since not all the expected B and Y ions were found.

Table 3.17 '*De novo*' sequencing and sequence homology of AC Vtg tryptic peptides obtained by in-gel digestion

The number between bracket under the Proposed sequence column corresponds to mass gaps that could not be assigned to any particular amino acid sequence. Comments A and B mean a perfect match was found with respect to haddock Vtg sequence A and B, respectively (See legend in Table 3.16 for details). Comments C and D mean that partial sequence homology was found with respect to haddock Vtg sequence A and B, respectively. Comment E was used when no homology was found to haddock Vtg.

Precursor <i>m/z</i>	Proposed sequence	Comment	Haddock Vtg sequence
942.49	YESFAVAR	A	
1009.53	DLGLAYTEK	A + B	
1081.56	QPLMNDAGR	D	LPQAPVDADR
1089.54	FDDVEPVLR	D	CFSVEPVLR
1106.58	HLGLTTPPDR	E	
1124.62	VHVDALLAMR	A	
1130.71	FLEVLQLLR	C + D	FLELIQVLR and
1166.60	LESEPLPLR	C	LESEDASFIR
1221.67	ACVALVLFETR	D	MAAALVLFETR
1237.70	RPGVVWDPVGR	E	
1293.64	WDTFQMQLPK	D	FTWDKLPTSAK
1363.75	SSPPGMETFALAR	D	ITAATVETFAIAR
1381.81	VPLPQFLPLAMR	E	
1433.75	ALAAGVDLTEGNFR	E	
1458.74	FFGAWVAFAEADK	D	FFGQEIGFASIDK
1465.82	FLGGPCLAFANLDK	C	FLGQEIAFANLDK
1479.81	EPQAAPVNTFLHR	C	TPIAPVNAQYLHR
1519.76	(805.41)EWSHR	E	
1540.90	GVAAGFLPLAASLTPR	E	
1588.90	(789.3369)VLELGLR	E	
1602.90	TTLGSFNVDALLPVR	E	
1704.95	HSAATVLESWTFTVR	E	
1743.13	(833.51)LLEPLLGR	C	IVPIPAEILEPLIGR
1796.01	(1133.59)LNLFR	C	VYSPEGISTTVLNIFR
1834.97	HPHQRPDLLENHWR	D	VQLILANLVEENHWR
2073.99	(975.5507)NPYQVAGGHR	E	
2131.02	NAVSGSYNHTFEHATVATVR	E	
2226.00	(1410.56)LPASSGGAR	E	

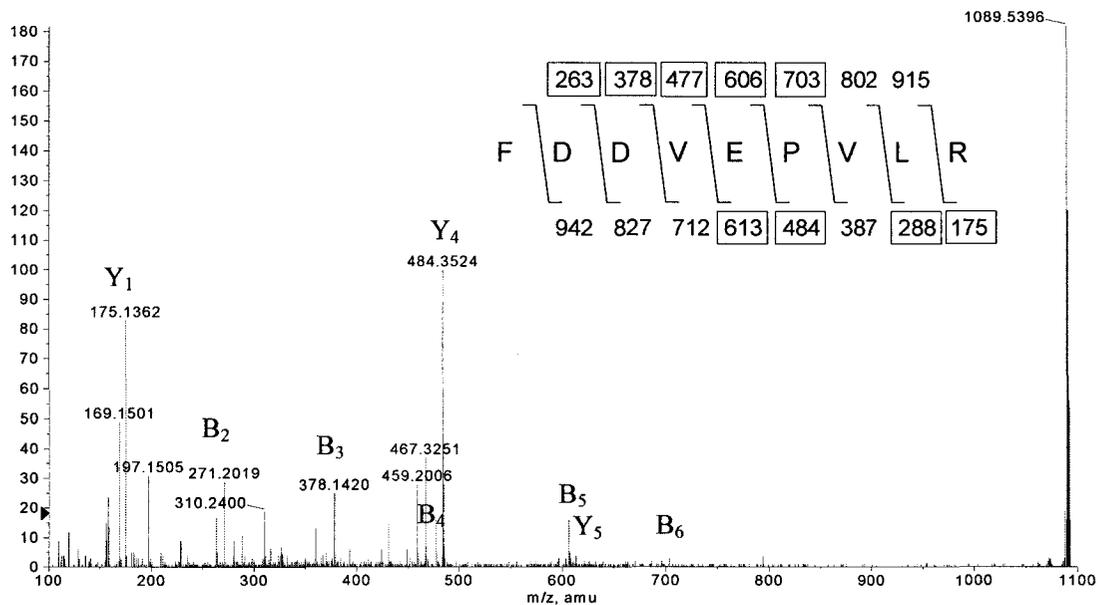


Figure 3.14 Product ion spectrum of the  $[M+H]^+$  ion at  $m/z$  1089.54

The inset shows the fragmentation pattern indicating the formation of the expected B and Y ions. The product ions which were found are labeled on the peaks and with boxes in the inset.

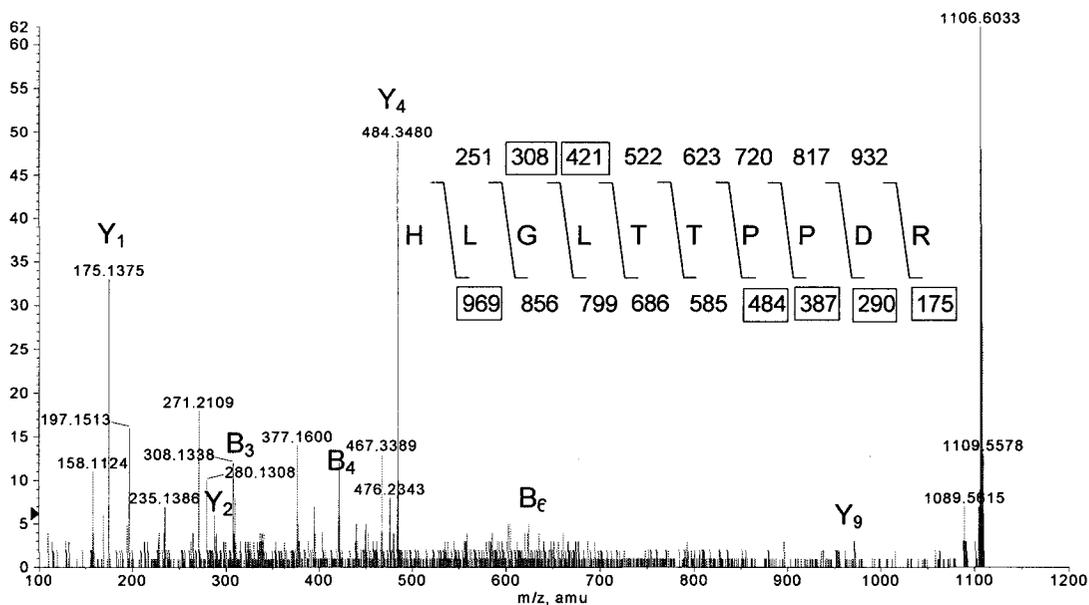


Figure 3.15 Product ion spectrum of the  $[M+H]^+$  ion at  $m/z$  1106.58

See legend in Figure 3.14 for details.

The product ions spectra of ions at  $m/z$  942.49, 1009.53 and 1124.62, shown in Figures 3.16A, B and C were correctly assigned by the Bioanalyst sequencing algorithm to the theoretical amino acid sequence corresponding to the haddock tryptic peptides. This is shown by the matching B and Y ions found in the spectra. Note that leucine and isoleucine cannot be distinguished from each other for sequencing purposes and are marked with the letter L by the sequencing software to indicate both amino acids.

The product ion spectrum of the selected precursor ion at  $m/z$  1743.13, which had been assigned to two different peptides in Table 3.16, was partially sequenced to match the Vtg A of haddock Vtg. The remaining ions showed either partial or no sequence homology with the haddock Vtg.

It is important to mention that there was a series of ions obtained in the conventional MALDI-ToF-MS at  $m/z$  826.47, 884.98, 920.43, 1123.61, 1952.94, 2142.12, 2391.19, 2408.18, 2440.36, 2460.33, 2585.56, 2637.35, 2729.49, 3387.66, 3435.72 and 3450.78, which did not match any of the predicted haddock Vtg tryptic peptides with the chosen database search parameters. These ions were tentatively designated as diagnostic for AC Vtg.

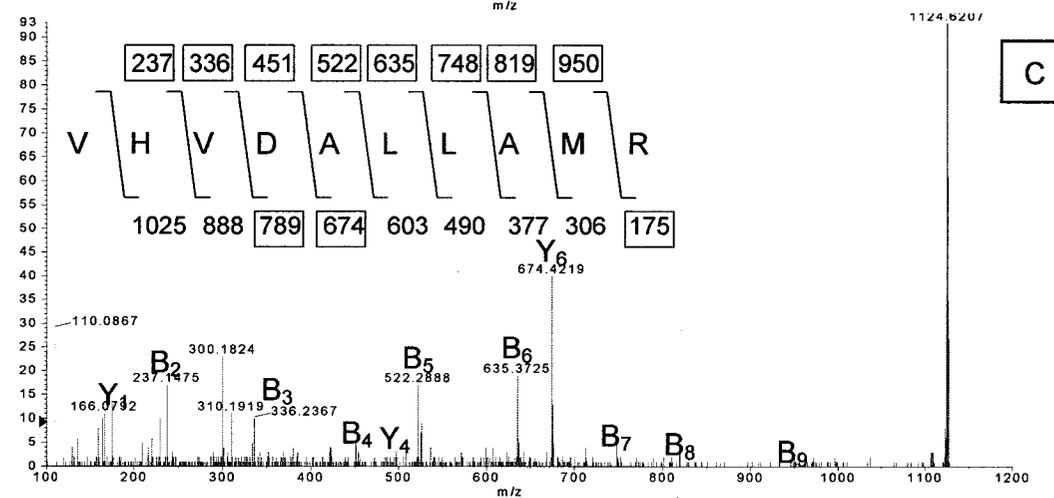
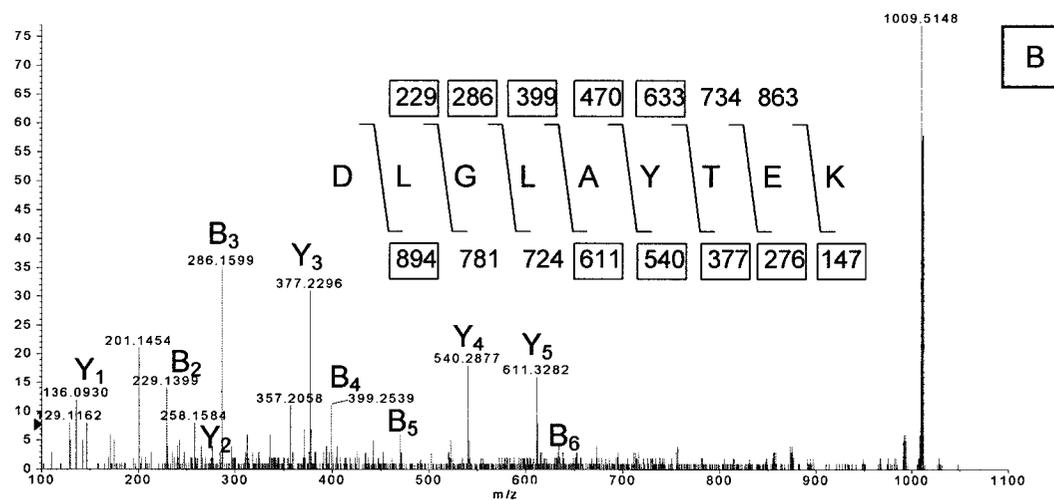
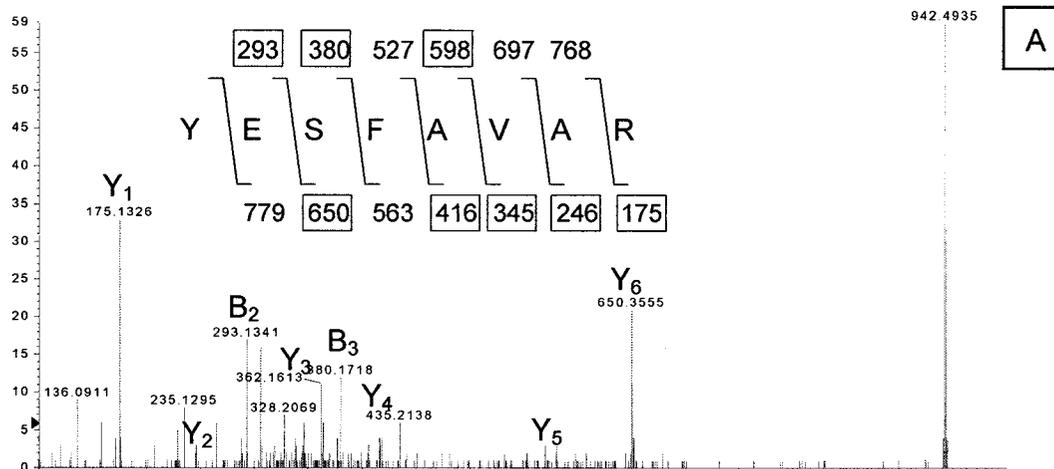


Figure 3.16 Product ion spectra of  $[M+H]^+$  precursor ions

Figures 3.16 A, B and C correspond to the product ion spectra obtained from precursors at  $m/z$  942.49, 1009.53 and 1124.13, respectively. See legend in Figure 3.14 for remaining details.

## **Chapter 4: Development of a quantitative assay for vitellogenin using tandem mass spectrometry**

The previous chapters have laid the basis for the final objective of this work, that is, developing a simple mass spectrometric quantitative technique for measuring the plasma levels of AS and RT Vtg. The technique chosen to accomplish this task is known as the ‘signature peptide’ approach. The following sections describe all the steps that were necessary for the selection of the ‘signature peptide’. Also, a detailed account of the tuning and optimization of the analytical methods involving the HPLC-ESI-QqToF-MS/MS is described.

### **4.1 Quantitative methods in proteomics**

Recent advances in mass spectrometry have provided researchers with powerful quantitative techniques in the area of proteomics (Zhang *et al.*, 2004b). Quantification of proteins can be achieved by using either a relative or an absolute technique. Relative quantification is usually sufficient in those studies which focus on differential protein expression between control and experimental groups. The objective of this kind of studies is to determine whether protein expression in an experimental group varies with respect to a control group, under the effect of an external stimulus, which can be either physical (eg. temperature change) or chemical (eg. administration of a therapeutic drug). The final result is either increased, decreased or no change in protein expression. Although

semiquantitative results may be obtained, usually, no units are expressed in these type of studies. Different strategies have been adopted for this type of quantification. For example, some authors have used isotopically labeled functional groups in proteolytic peptides by means of derivatization agents (Ji *et al.*, 2000; Geng *et al.*, 2000; Riggs *et al.*, 2001). Similar techniques use more sophisticated isotope-coded affinity tags (Gygi *et al.*, 1999; Gygi *et al.*, 2002; Ong *et al.*, 2002; Yu *et al.*, 2004; DeSouza *et al.*, 2005), many of which are commercially available. Relative quantification can be done by  $^{18}\text{O}$  labeling of enzymatically digested proteins with  $\text{H}_2^{18}\text{O}$  (Yao *et al.*, 2001; Yao *et al.*, 2003). All these techniques rely on selectively tagging the analyzed protein in the experimental group with a specific mass label which can be selected and monitored in a mass spectrometer, and compared relatively to the non-labeled protein expressed in the control group. Alternatively, relative quantification can be done by measuring accurate mass and time tags (Masselon *et al.*, 2000; Conrads *et al.*, 2000; Smith *et al.*, 2002; Wang *et al.*, 2003; Strittmatter *et al.*, 2003; Pasa-Tolic *et al.*, 2004; Elias *et al.*, 2005; Norbeck *et al.*, 2005; Qian *et al.*, 2005; Silva *et al.*, 2005). In this approach, the proteins expressed in the control and experimental groups are proteolitically cleaved and analyzed by HPLC coupled to mass spectrometry. Each biomarker peptide is assigned a characteristic mass and time pair, which serves as a coordinate for the study of the complete proteome of a specific organism. This technique does not require tagging of the studied proteins and can be scalable to less sophisticated MS-only spectrometers. However, reproducible chromatographic conditions are critical, and a high resolution mass spectrometer is required for the assignment of the accurate mass values.

Absolute protein quantification requires a different approach. In these types of studies, the absolute concentration of proteins in a specific sample (eg. blood plasma) is expressed in some kind of units (eg. mg/mL, mM, etc). Usually, these studies are done to compare the protein expression with a reference value. Biomarkers are frequently assayed by this approach to determine if their concentrations are above or below a threshold value, usually indicating the limits of an altered or pathological condition.

In this study, an absolute quantification method was selected for Vtg using the signature peptide approach. The rationale for this approach is that peptides generated by the digestion of complex mixtures may be used as analytical surrogates for the protein from which they are derived. This approach is based on the fact that single peptides are in many cases easier to separate and identify than the intact proteins. Furthermore, the synthesis of custom made peptide standards and internal standards is occasionally easier than the isolation and purification of the intact proteins for calibration purposes. The internal standards usually consist of homologues of the analyzed peptides in which one or more atoms have been replaced by stable isotopes (e.g.  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$  etc). Variations of this approach have been applied to different proteins: fathead minnow (*Pimephales promelas*) Vtg, rhodopsin, apolipoprotein A-1 and a hemoglobin  $\beta$ -chain derived glycosylated hexapeptide have been quantified using triple quadrupole mass spectrometers (Barr *et al.*, 1996; Jeppsson *et al.*, 2002; Barnidge *et al.*, 2003; Zhang *et al.*, 2004a). This strategy for quantification has also been described by other authors, and is known as the AQUA (Absolute QUAntification) strategy (Gygi and Aebersold, 1999; Gerber *et al.*, 2003; Kirkpatrick *et al.*, 2005). A variation of this method, the SISCAPA (Stable Isotope

Standards and Capture by Anti-Peptide Antibodies) approach, uses an additional step, where the peptide of interest is isolated by antibody-affinity chromatography prior to the HPLC separation, conferring additional sensitivity to the technique by enriching the concentration of the peptide and reducing the background noise level (Anderson *et al.*, 2004).

A different approach for absolute quantification of fathead minnow Vtg using HPLC-MALDI-MS has recently been published (Wunschel *et al.*, 2005). However, in this technique, the intact protein was analyzed and quantified by HPLC using a UV detector, and MALDI-MS was used to confirm the identity of the eluting proteins by peptide mass fingerprinting. In this study, the minimum detectable level of Vtg injected that produced a visible peak was 1 µg and the samples analyzed were all in the order of mg/mL.

#### **4.2 The 'signature peptide' approach**

Fried tongues of cod fish are a culinary delicacy in St. John's, Newfoundland. According to historians, many decades or even centuries ago, fishermen used to cut the tongues of the cod as soon as the nets were brought on deck. In doing so, the catch could be easily estimated for commercial purposes by counting the 'small' tongues collected instead of having to deal with the huge amount of fish stored in the ship's hold (Chrestien, 2005). Quantification of proteins by the 'signature peptide' approach is essentially the same, instead of 'counting' big protein molecules in a complex protein mixture, the focus is placed on one of their small pieces.

As described in previous chapters, a wide spectrum of analytical methods has been developed for quantification of Vtg levels. Indirect measurement of Vtg has been achieved by colorimetrically determining the alkaline-labile phosphorous content of fish plasma (Wallace R.A. and Jared D.W., 1968). More recently, a number of immunoassays, such as the competitive enzyme-linked immunosorbent assays (ELISA) (Palmer *et al.*, 1998; Denslow *et al.*, 1999; Tyler *et al.*, 2002) and radioimmunoassays (RIA) (Campbell and Idler, 1980; So *et al.*, 1985; Norberg and Haux, 1988) have been developed. Similarly, Vtg mRNA levels have been assayed (Bowman and Denslow, 1999; Garcia-Reyero *et al.*, 2004) although the use of liver homogenate samples in this technique is prohibitive for large scale animal monitoring.

In this study a mass spectrometric technique is presented, by which both plasma Vtg levels of RT and AS can simultaneously be quantified using an adaptation of the 'signature peptide' approach. The simplicity of this approach relies on a single step 'one pot' trypsin digestion of the serum, without the need for protein isolation and purification. A characteristic 'signature' proteolytic peptide, common to both RT and AS Vtg was monitored by high performance liquid chromatography (HPLC) coupled to an electrospray ionization quadrupole-time-of-flight (ESI-QqToF-MS/MS) tandem mass spectrometer. The concentration of peptide present in the digestion will be proportional to the concentration of Vtg in plasma. Absolute concentration values obtained from the 'signature peptide' approach were calculated considering a 1:1 mole ratio of the 'signature peptide' to the Vtg molecule, their individual molecular masses and the dilution factors involved in sample processing.

The rationale behind choosing a common peptide for both RT and AS Vtg as standard was to use the same standard calibration curve to analyze samples of both species in one batch. This feature resembles the ‘cross reactivity’ of immunological assays which allows homologous proteins sharing common epitopes to be detected using the same polyclonal and/or monoclonal antibodies. A diagram of the work scheme is summarized in Figure 4.1.

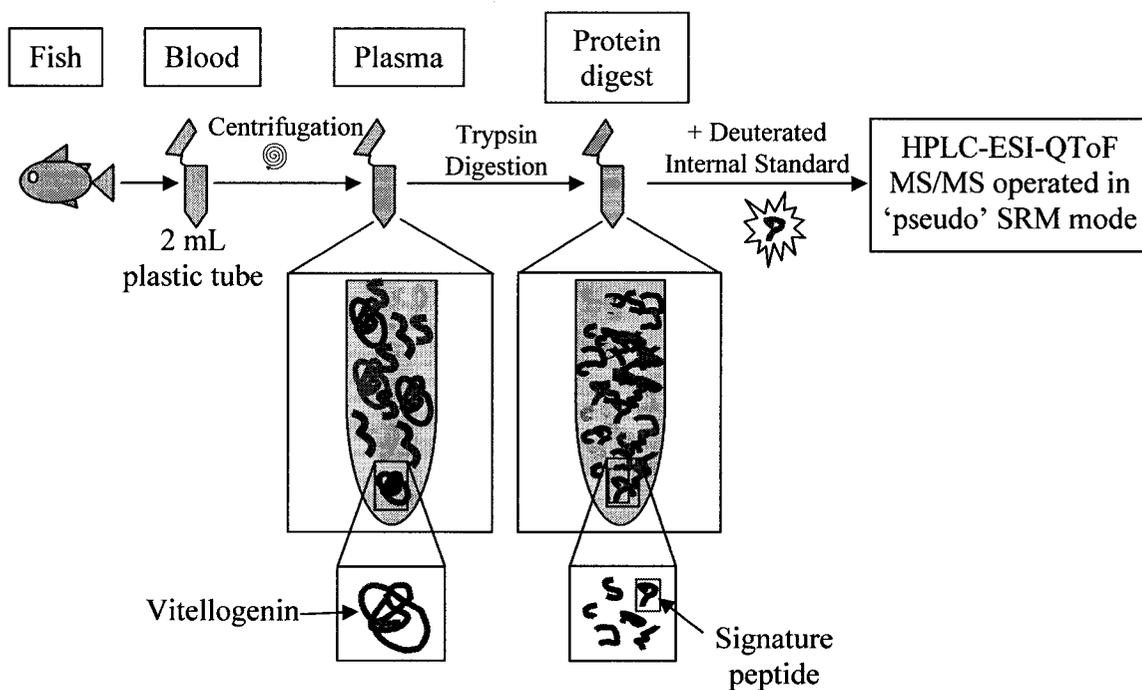


Figure 4.1 Work scheme for the ‘signature peptide’ approach.

## **4.3 Materials and methods**

### **4.3.1 Pre-analytical sample preparation**

Blood was extracted from fish and centrifuged to obtain the plasma. Five microliters of plasma were mixed with 100  $\mu\text{L}$  of 8M urea, 4mM dithiothreitol denaturing solution and heated at 65°C. Then, 300  $\mu\text{L}$  of a 50mM solution of  $\text{NH}_4\text{HCO}_3$  were added, followed by 95  $\mu\text{L}$  of a 20 $\mu\text{g}/\text{mL}$  trypsin (Sigma, Prod. No. T6567) solution which was freshly prepared for each batch analysis. The unused trypsin solution was separated in aliquots and frozen at -20°C for upto one month for subsequent digestions. The amount of enzyme was calculated based on the recommendation of the trypsin supplier which recommends a 1:100 w/w ratio of enzyme to total protein content. Considering that the average total protein concentration in RT plasma is approximately 43 g/L (Rehulka *et al.*, 2005), 5  $\mu\text{L}$  of plasma would contain approximately 215  $\mu\text{g}$  of total protein. Thus, 95  $\mu\text{L}$  of a 20 $\mu\text{g}/\text{mL}$  trypsin solution would contain 1.9  $\mu\text{g}$  of trypsin. Calibration curves were constructed by preparing and injecting 0.05, 0.10, 0.25 and 0.50  $\mu\text{g}/\text{mL}$  solutions of signature peptide (Genemed Synthesis Inc, San Fransico, CA, USA) in  $\text{H}_2\text{O}:\text{acetonitrile}:\text{acetic acid}$ , 70:30:0.1 (v/v/v) into the HPLC-MS/MS. All standards were spiked with internal deuterated signature peptide standard (Genemed Synthesis Inc, San Fransico, CA, USA) to produce a final concentration of 0.10 $\mu\text{g}/\text{mL}$  of the deuterated signature peptide.

Digestion was carried out at 38°C for 24 hours. Aliquots of 100  $\mu\text{L}$  of the digested samples were mixed with 380  $\mu\text{L}$  of  $\text{H}_2\text{O}$ :acetonitrile:acetic acid, 70:30:0.1 (v/v/v) and spiked with 20  $\mu\text{L}$  of 5  $\mu\text{g}/\text{mL}$  deuterium-labeled standard solution. Finally, samples were passed through 0.45  $\mu\text{m}$  filters (Millipore, Cat. No.SJHV004NS. Millipore, Billerica MA, USA,) prior to injection into the HPLC. After the labeled peptide was added to the trypsin digestion, the samples were analyzed using an HPLC coupled to a QqToF-MS/MS. Samples and standards were analyzed with a HPLC system (HP 1050, Hewlett-Packard, Palo Alto, CA, USA) coupled in-line to a QqToF (QStar XL, Applied Biosystems, Foster City, CA, USA) tandem mass spectrometer equipped with an ESI ion source (TurboIonSpray). Reverse phase chromatographic separation was achieved on a 30mm length, 2mm internal diameter, Gemini 5 $\mu$  C18 110Å column (Phenomenex, Torrance CA, USA). All samples and standards were loaded into a 10  $\mu\text{L}$  loop connected to a 10 port Valco (Houston, TX,USA) switching valve. The mobile phases labeled A, B and C contained distilled water, acetonitrile and acetic acid in 95:5:0.1, 70:30:0.1 and 5:95:0.1 (v/v/v) proportions, respectively. Table 4.1 gives the step-wise solvent elution profile used for all chromatographic separations. The parameters shown in this table were chosen to optimize the signals which were monitored during the MS/MS acquisition. Note that the flow rate was reduced from the 4<sup>th</sup> minute to the 10<sup>th</sup> minute, at which time the ‘signature peptide’ eluted from the chromatography column.

The product ion spectra used in quantification were acquired in a ‘pseudo’ SRM mode. The word ‘pseudo’ is used in terms of comparing this mode of analysis to that performed traditionally in triple quadrupole mass spectrometers. The mass spectrometer was programmed to alternatively monitor the selected product ions from the precursor

ions at  $m/z$  819.9 and 824.4, corresponding to the doubly charged species  $[M+2H]^{2+}$  of the ‘signature peptide’ and its deuterated isotopic homologue, respectively. This cycle was automatically repeated 750 times throughout the 25 minute chromatographic run.

Table 4.1 Solvent elution profile and flow rates for the chromatographic separation.

Time (min)	Flow rate (mL/min)	Mobile phase (%)		
		A	B	C
0.0-0.5	0.2	100	0	0
0.5-4.0	0.2	0	100	0
4.0-10.0	0.05	0	100	0
10.0-15.0	0.4	0	0	100
15.0-25.0	0.2	100	0	0

The ToF analyzer was set to scan the product ions produced from  $m/z$  800 to 1300. The tuning parameters of the mass spectrometer were optimized in terms of maximizing the signals produced by the  $m/z$  819.9→(957.5 - 1028.57 - 1156.6) and the  $m/z$  824.4→(960.5 - 1034.61 - 1165.7) transitions. The diagnostic product ions monitored correspond to the  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  fragments of both the ‘signature peptide’ and its labeled homologue (Biemann, 1990) and were selected as product ions for quantification, as explained in the following sections. These ions were additionally enhanced by setting the ion release delay (IRD) and the ion release width (IRW) parameters to 102.0 and 44.2 ms, respectively, to increase the duty cycle of the ToF detector for these particular ions. The

remaining tuning parameters used were: 4700 V for ionization energy, 50 psi for nebulizer gas, 80 psi for auxiliary gas, 300°C for probe temperature, 40 V for declustering potential (DP), 190 V for focusing potential, 15 V for DP2 and 36 V for the collision energy. The collision gas used was nitrogen at an arbitrary value of 5 set on the AnalystQS software. The first selecting quadrupole was operated at low resolution to further increase the product ion signal.

For quantification purposes, the extracted ion chromatograms of the diagnostic product ions listed above were generated using a 2 Da mass range window around the selected product ions and manually integrated. The ratio between these two peak areas was taken as the dependant variable for the calibration curve and the same ratio was used for the digested plasma samples. The values of signature peptide in plasma were interpolated from the calibration curve, corrected using the corresponding dilution factors and finally converted to mg/mL of Vtg assuming a 1:1 mole ratio of 'signature peptide' to Vtg molecule. The relative molecular masses considered for Vtg were 181938 and 177153 Da for rainbow trout and Atlantic salmon, respectively. These values were obtained from MALDI-MS experiments on the native intact Vtg proteins, reported in Section 3.2 (Figure 3.2) and recently published (Banoub *et al.*, 2003; Banoub *et al.*, 2004). Results were also expressed in  $\mu\text{M}$  units, considering a relative molecular mass for the 'signature peptide' of 1638.819 Da.

Quantification by mass spectrometry has commonly been performed on triple quadrupole (QqQ) tandem mass spectrometers. These instruments have shown exceptional results when operated in Selected Reaction Monitoring (SRM) mode. Alternatively, in this work, the use of the hybrid QqToF-MS/MS instrument has been

explored for quantification purposes (Chernushevich, 2000; Chernushevich *et al.*, 2001). Although the QqToF-MS/MS configuration has already been used for quantitative analysis of a variety of compounds ranging from cocaine, tyrosine to goserelin, an analogue of luteinizing hormone-releasing hormone (Jeanville *et al.*, 2000; Marvin *et al.*, 2003; Michalet *et al.*, 2003; Luo *et al.*, 2004), no research has yet been published using this system for the quantification of such a complex protein.

This method, developed by adjusting and optimizing the tuning parameters of the QqToF analyzer was termed as ‘pseudo’ SRM. This CID-MS/MS method was programmed to operate in a mode that monitors three diagnostic product ions of both the ‘signature peptide’ and its deuterated isotopic homologue, used as an internal standard. The ratio of each of the diagnostic product ions to its isotopic homologue was used as the analytical quantitative response signal for identification purposes.

Figure 4.2 shows a schematic diagram of the MS/MS operating in the ‘pseudo’ SRM mode. Tryptic peptides separated by the HPLC system are analyzed ‘in line’ by MS/MS. The first mass analyzer alternatively selects the ‘signature peptide’ and its isotopic homologue while the product ions of these precursors are monitored by the second stage mass analyzer. The ratio of the signals produced by these two product ions is considered as the analytical response signal for calibration and quantification purposes. A signal is obtained for each pair of diagnostic product ions (product ions at  $m/z$  957.5 - 1028.57 - 1156.6) and their corresponding isotopic homologue (960.5 - 1034.61 - 1165.7). These three pairs are individually monitored and used to obtain three separate calibrations curves obtained from the standards, from which the unknown samples are interpolated. Finally, the three concentrations are averaged to obtain the final result.

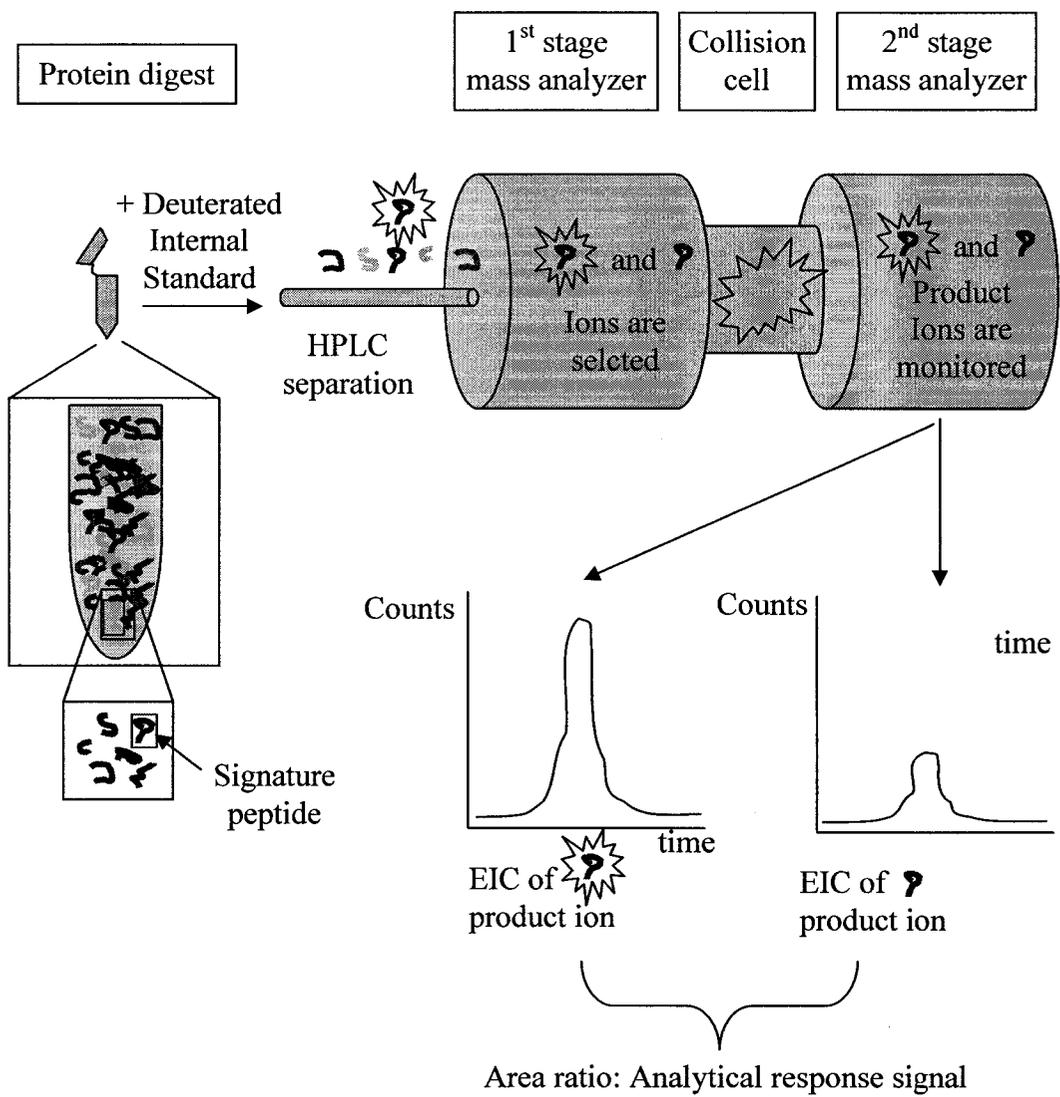


Figure 4.2 HPLC-MS/MS operated in the 'pseudo' SRM mode. EIC stands for Extracted Ion Chromatogram. During analysis, individual pairs of EICs are obtained for each one of the three diagnostic product ions monitored.

### 4.3.2 Selection of the ‘signature peptide’

The sequence of the selected ‘signature peptide’ was the following:

Thr-Tyr-Phe-Ala-Gly-Ala-Ala-Ala-Asp-Val-Leu-Glu-Val-Gly-Val-Arg.

The selection of this ‘signature peptide’ was based on previous ‘*de novo*’ sequencing studies performed on Vtg tryptic peptides described in Chapter 3 and recently published (Banoub *et al.*, 2003; Banoub *et al.*, 2004). Basically, a long list of candidate peptides from both species was elaborated. Many of the peptides belonging to RT were corroborated from the sequences found in the protein databases (See Appendix 2, entry Q92093), since Vtg from this species had been completely sequenced. On the other hand, only a few peptides of the AS Vtg matched the database search since only the 284 N terminal (Appendix 2: entry Q800N6) and the 175 C terminal (Appendix 2: entry Q8UWG2) amino acids of this species have been sequenced. Nevertheless, only a few peptides were found common to both species, reducing significantly the list of candidate peptides.

A critical question to answer when selecting this ‘signature peptide’ was whether an identical peptide sequence could be found in nature in any protein other than Vtg. Although the genome of both species has not yet been totally elucidated, the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990; McGinnis and Madden, 2004) was used against all known protein and DNA databases with no taxonomic restriction and provided no significant matches.

The amino acid composition of the candidate sequences was also considered influential in the final selection. Peptides containing methionine and cysteine were discarded because of their propensity to form oxidized residues under different experimental conditions. Therefore, the presence of less-reactive residues such as Ala and Gly was preferred. Furthermore, tryptic peptide sequences starting with or having proline following the terminal lysine or arginine, i.e. X-X-Arg/Lys-Pro-X, X being any other amino acid, were ignored because, when found at these positions, these amino acids are known to block the proteolytic action of trypsin.

The presence of post translational modification in the 'signature peptide' location of the protein would also weaken the technique in two ways: it would increase the probability of peptide heterogeneity and increase enzymatic blockage during sample digestion. The selected 'signature peptide' lacked both glycosylation and phosphorylation consensus motifs, reducing the chances of these post translational modifications to occur.

A prerequisite for the synthesis of the peptide standard was the incorporation of stable isotopes for the internal standard. Ideally, this internal standard should have a molecular mass high enough to distinguish itself clearly from the isotopic cluster corresponding to the non-labeled standard. This was achieved by incorporating three consecutive L-alanine-3,3,3-d<sub>3</sub> into the 'signature peptide' sequence. The fact that the same deuterated amino acid was used and that they were situated in adjacent positions significantly reduced its synthesis cost.

Finally, the selection of the 'signature peptide' was based on empirical circumstances. There are many experimental parameters which can affect the final results of this method. Slight modifications in the sampling, digestion or MS conditions can

result in changes that can affect the outcome of the experiments. Sometimes, these changes are known or can be predicted and in these cases, the results can be accordingly corrected to reflect the changes in the final results. However, very often, the experimental variations in these conditions are unseen or even unknown. The fact that the selected ‘signature peptide’ consistently appeared in all our previous characterization experiments proves it to be very robust under a variety of experimental conditions. This peptide appeared in the PMF of both solution and in-gel digestions of RT (Tables 3.2, 3.6 and 3.12) and AS (Tables 3.8, 3.9 and 3.14) Vtg. It was also successfully sequenced in the MS/MS experiments from both species using MALDI and ESI as ionization sources (Tables 3.7, 3.10, 3.11 and 3.15). Thus, if this peptide had ‘survived’ such variety of experimental conditions, it was clearly the winner of the ‘signature peptide’ contest. It was of course present in the RT Vtg sequence found in the Swiss-Prot Database (See position 668 in entry Q92093 of Appendix 2). However, while this peptide is not found in any of the published sequences of AS Vtg (Appendix 2: entries Q800N9 and Q8UWG2), strong evidence from all the previous experiments was sufficient to claim its presence.

#### **4.3.3 Tuning and optimization of the HPLC-MS/MS acquisition parameters**

Initially, the dilemma of which precursor ion,  $[M+H]^+$  or  $[M+2H]^{2+}$ , was going to be selected for the CID-MS/MS experiments had to be resolved. A standard solution of 5  $\mu\text{g/mL}$  of the ‘signature peptide’ dissolved in  $\text{H}_2\text{O}:\text{acetonitrile}:\text{acetic acid}$ , 70:30:0.1

(v:v:v) was directly infused in the MS and analyzed in full scan mode, and this is shown in the top spectrum of Figure 4.3.

The bottom left and right spectra show the enlarged peaks corresponding to the  $[M+2H]^{2+}$  and  $[M+H]^+$ , respectively. The doubly charged species  $[M+2H]^{2+}$  was found at  $m/z$  819.94. This monoisotopic peak was only 0.01 Da apart from the predicted  $m/z$  at 819.93, corresponding to a mass accuracy of only 15ppm. This peak exhibited a much stronger ion count than that of the singly charged species  $[M+H]^+$  at  $m/z$  1638.9. From this observation, the  $[M+2H]^{2+}$  ion was selected as the precursor for the CID-MS/MS experiments. The distribution of peptide ions observed in this spectrum was expected, since this peptide contains two functional groups likely to be protonated; one at the amino group at the N-terminal of the signature peptide and the other at the guanidinium group of the C-terminal arginine residue. The intensity of each precursor is proportional to the abundance of each protonated species in the nozzle tip of the ESI capillary.

Figure 4.4 shows the CID-MS/MS spectra obtained from the  $[M+2H]^{2+}$  species of the 'signature peptide' standard (top spectrum) and its deuterated homologue (bottom spectrum) when infused directly into the mass spectrometer. The insets indicate the b- and y-type product ions expected from the classical peptide fragmentation routes (Biemann, 1990). Also, highlighted are the diagnostic  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  product ions at  $m/z$  957.5, 1028.57, 1156.6 and at  $m/z$  960.5, 1034.61, 1165.7 which were monitored during the 'pseudo' SRM-MS/MS experiments. Also, all the product ions selected contained the deuterated moieties.

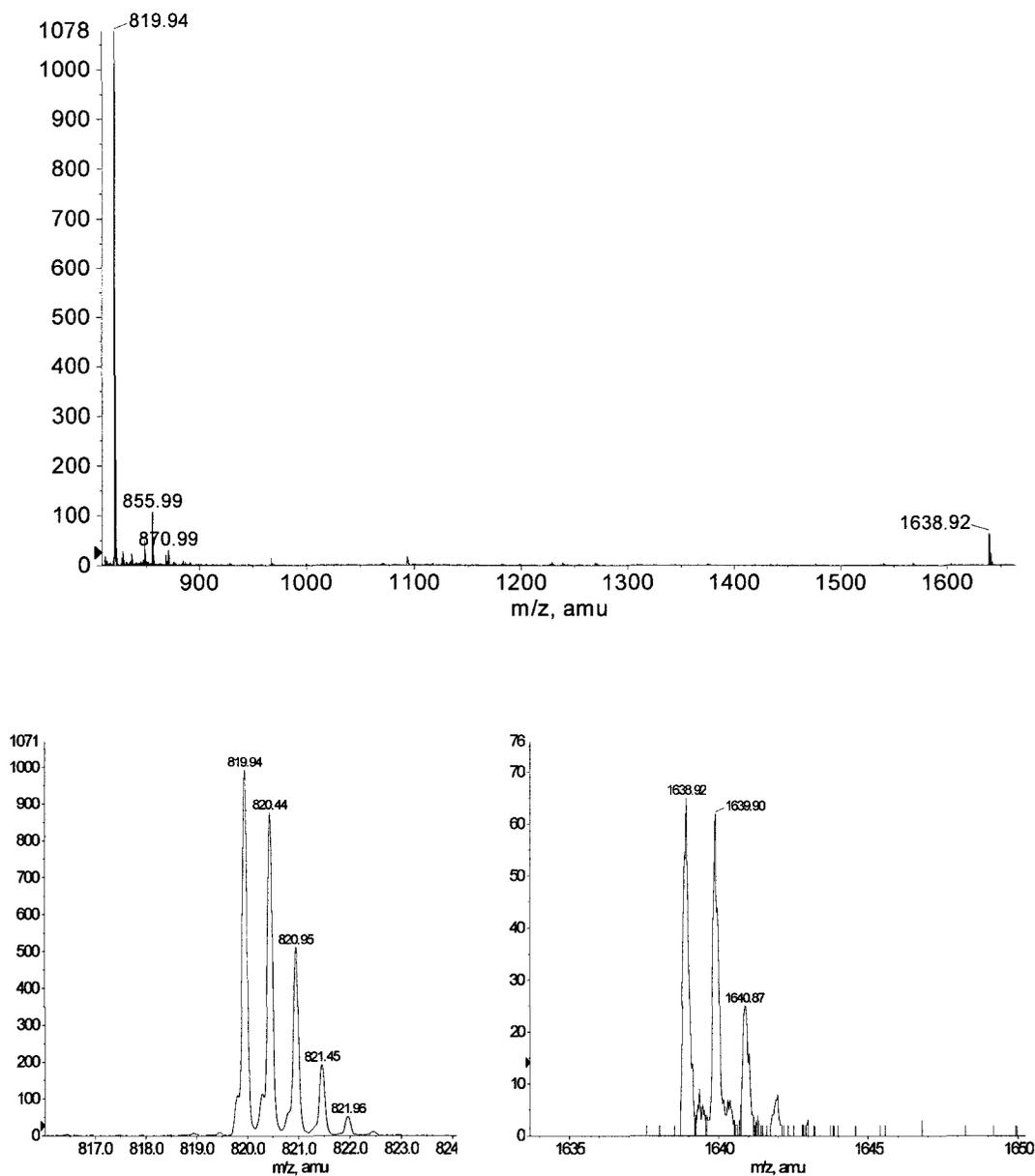


Figure 4.3 Full scan spectrum of the 'signature peptide' standard  
 The top spectrum shows the  $[M+H]^+$  and the  $[M+2H]^{2+}$  at  $m/z$  1638.9 and 819.9, respectively. The bottom left and right spectra show the same peaks in an enlarged scale.

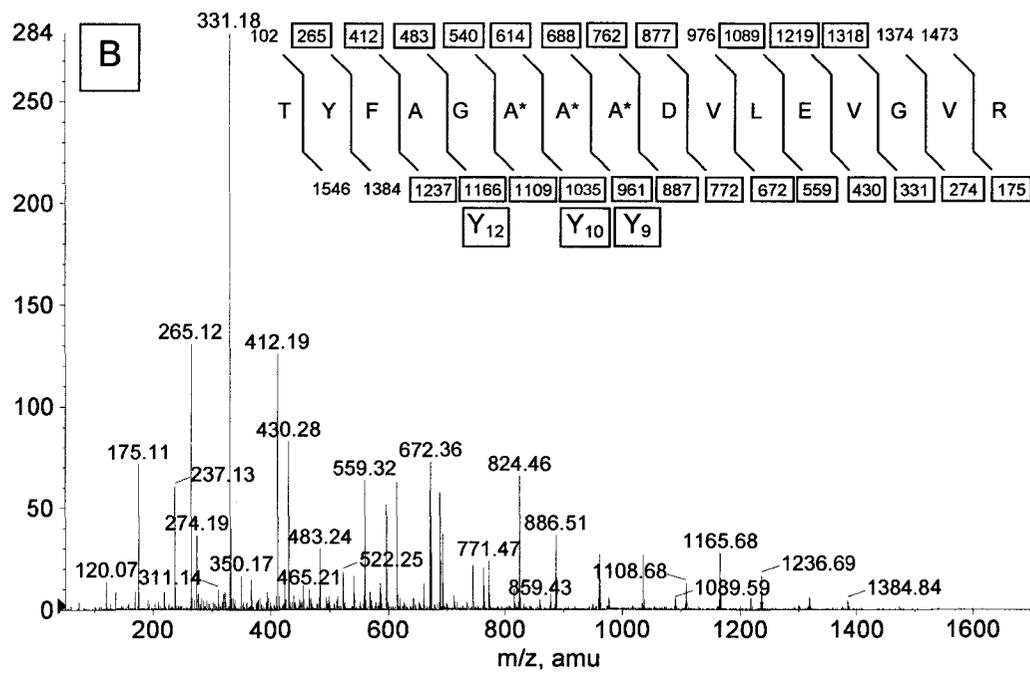
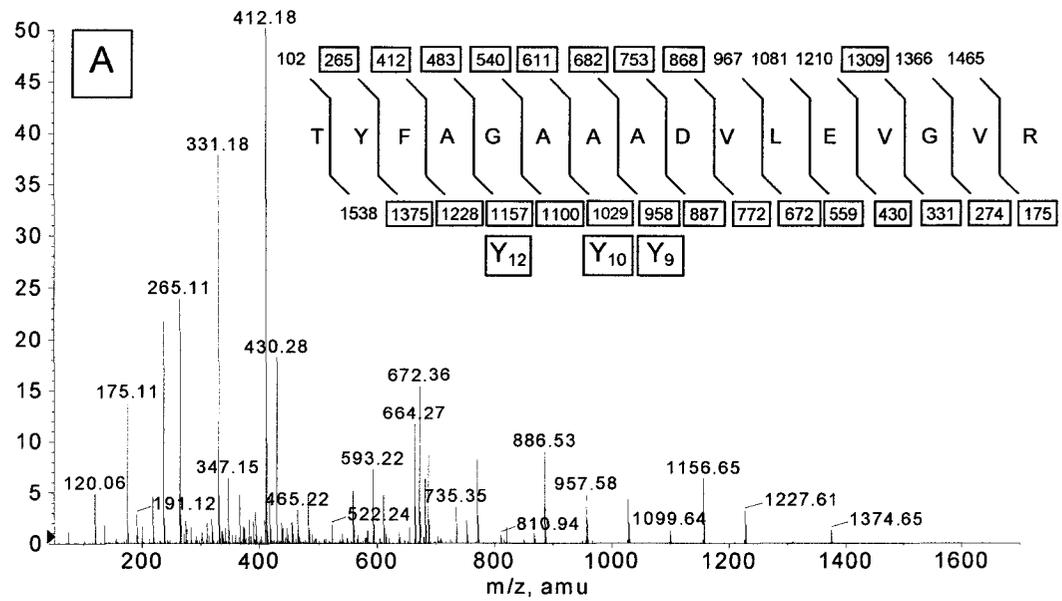


Figure 4.4 Product ion spectra of the ‘signature peptide’

The top (A) spectrum corresponds to the signature peptide and the bottom (B) to its deuterated isotopic homologue. The amino acid A\* stands for the deuterated amino acid L-Alanine-3,3,3-d<sub>3</sub>. See legend in Figure 3.14 for remaining details.

The decision to monitor these particular ions was based on a trade-off between sensitivity and specificity. The fact that the product ions monitored had higher  $m/z$  values than those of the selected doubly charged precursor ions allowed for an increase in the specificity of the method. Under this criterion, any singly charged ion precursors with the same  $m/z$  as the 'signature peptide's were neglected. However, this decision compromised the sensitivity of the method, given that some less specific product ions below  $m/z$  400 would have rendered a much higher intensity signal as shown in Figure 4.4.

Once the monitored product ions were selected, the acquisition parameters of the tandem mass spectrometer had to be tuned and optimized. By default, these instrumental parameters are typically set to encompass a wide range of biomolecules with different physicochemical properties over a wide  $m/z$  range. However, to obtain optimum results, each one of these parameters had to be adjusted for every particular compound, in this case, a tryptic peptide.

Although there are at least twenty parameters that can be modified in the mass spectrometer, only seven of them were shown to significantly affect the signal produced by the product ions. A diagram of the general components of the ESI-QqToF-MS/MS is shown in Figure 4.5. The parameters optimized and their location is also illustrated in this figure. The open brackets indicate the positions where the electric fields are applied. The ionization energy (IE), the nebulization gas (GAS 1) and the auxiliary gas (GAS 2) are critical parameters associated with the ESI ionization source and optimization of these translated into high ionization efficiency of the precursor ions and proper focusing into the mass spectrometer.

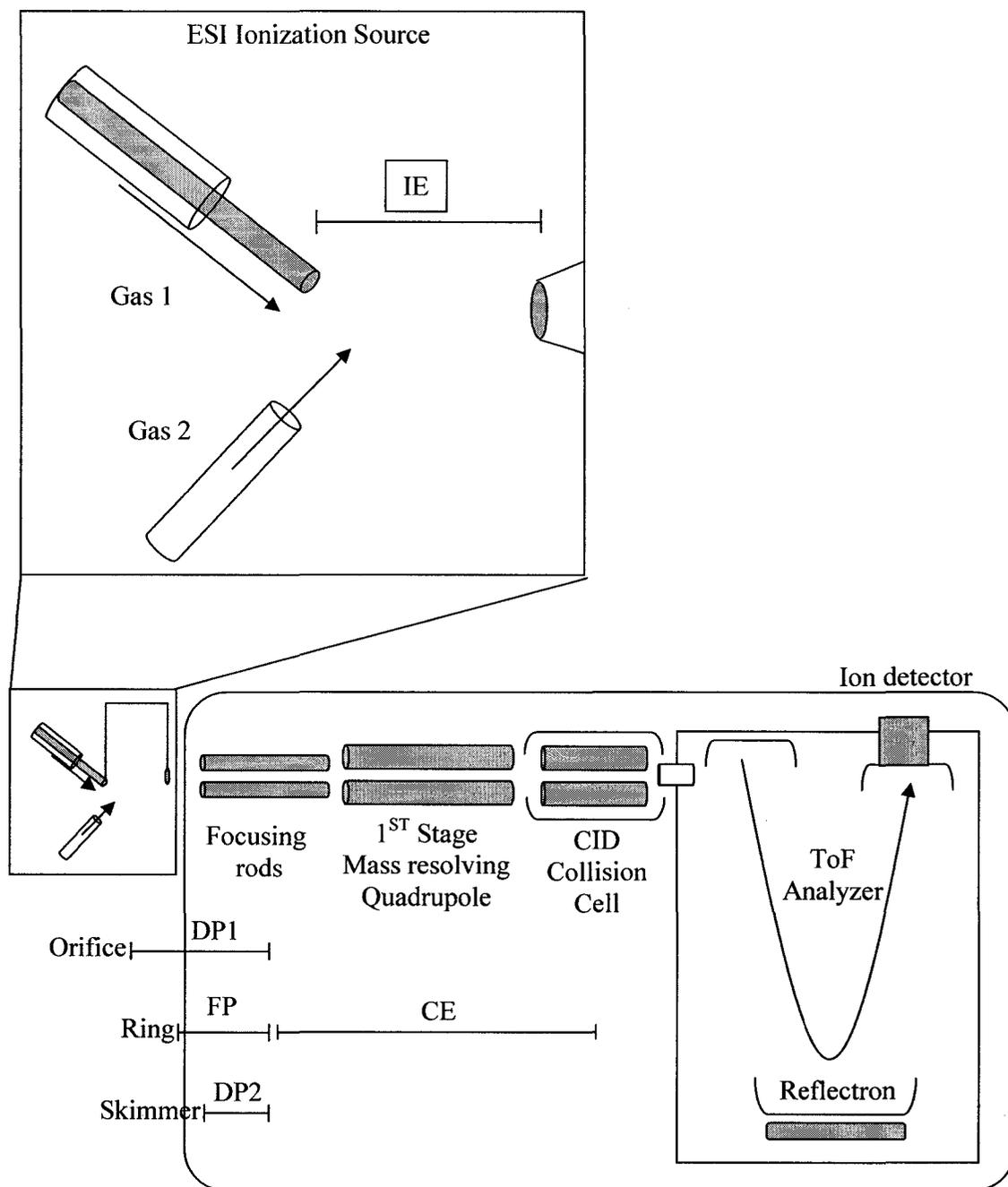


Figure 4.5 Schematic diagram of an ESI-QqToF tandem mass spectrometer

The diagram includes the instrumental parameters which were tuned for the 'pseudo' SRM-MS/MS acquisitions: Gas 1, Gas 2, the ionization energy (IE), the declustering potential 1 and 2 (DP1 and DP2), the focusing potential (FP) and the collision energy (CE). The horizontal brackets indicate the location where the electric potentials are applied.

The declustering potentials one and two (DP1 and DP2, also known as the Cone Voltage) and the focusing potential (FP) play a critical role in the steering of the ions within the mass spectrometer and should be carefully tuned to avoid ‘in-source’ fragmentation of the precursor ions. All these parameters were optimized and adjusted by direct infusion of a 5 µg/mL solution of the standard ‘signature peptide’ while monitoring the  $m/z$  1156 signal corresponding to the Y<sub>12</sub> product ion. The results of this tuning procedure are graphically illustrated in Figures 4.6 and 4.7. Each parameter was manually ramped, while the remaining parameters were fixed at the following values: Gas 1: 40, Gas 2: 50, IE: 4700V, DP1: 40V, DP2: 10V, FP: 200V and CE: 34V. The Gas values are given in arbitrary units displayed in the AnalystQS software program.

The collision energy imparts the energy to the precursor ion for CID. In the case of quadrupole mass analyzers, this process is called ‘low-energy’ CID because of the relatively small amount of energy which the precursor ions possess before collision with the inert collision gas. Modification of the CE produces distinctive mass spectral profiles. With low CE, the precursor ion can survive the CID and appear in the product ion spectrum. Also, the low CE spectra have a predominance of larger fragments. As the CE is increased, the intensity of the precursor ion in the product ion spectrum starts to decrease with a concomitant increase in the amount of smaller product ions (Data not shown).

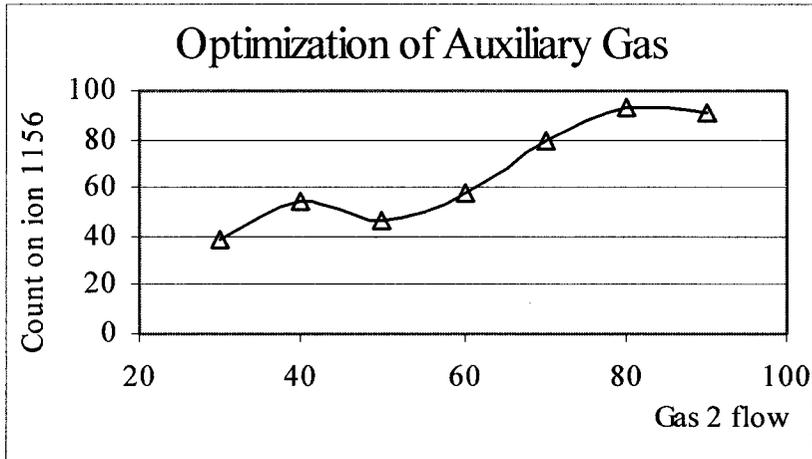
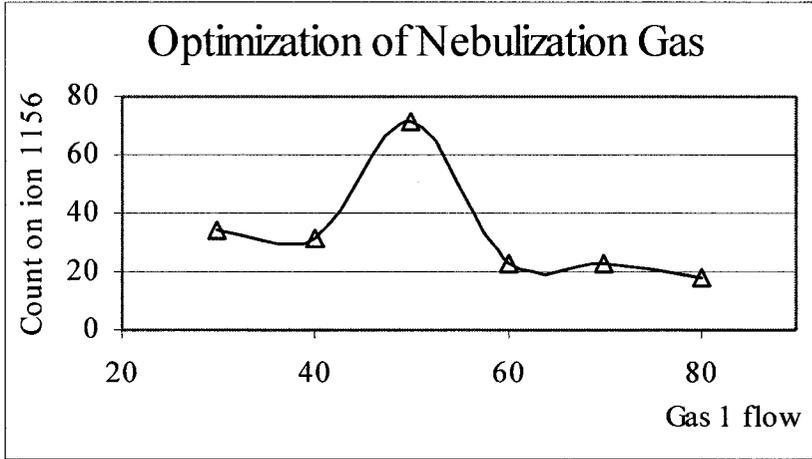
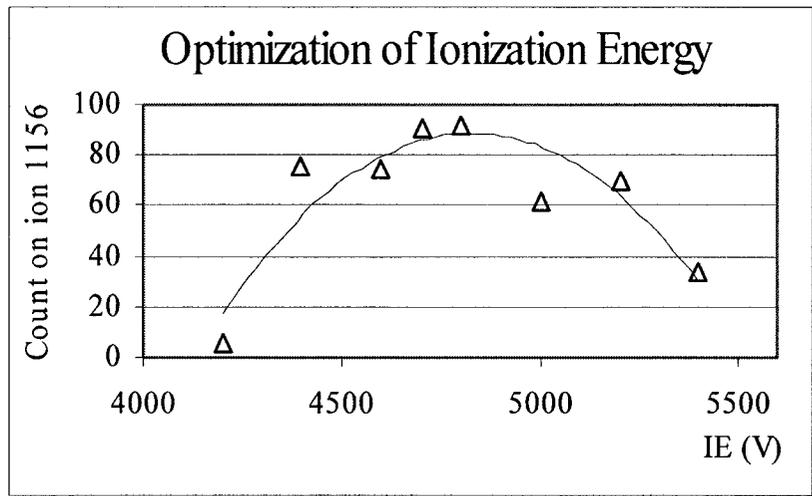


Figure 4.6 Optimization of the ionization energy, the nebulization gas and the auxiliary gas.

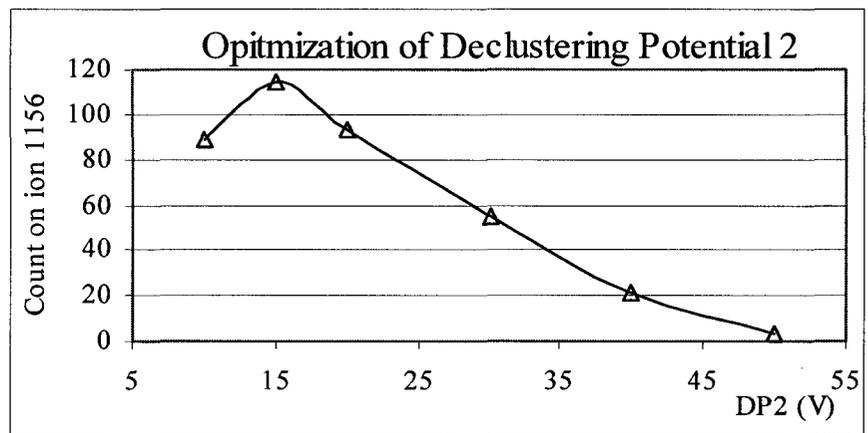
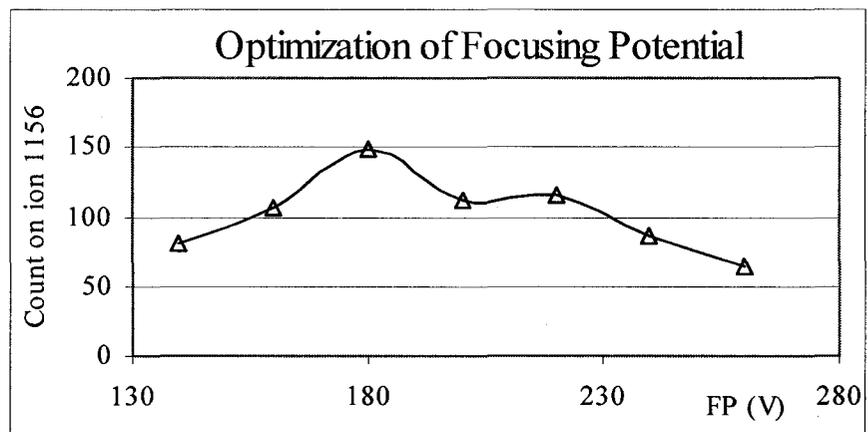
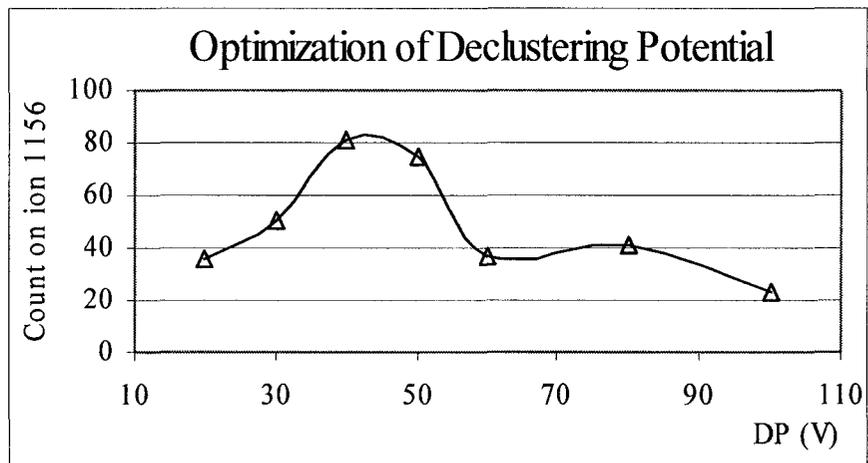


Figure 4.7 Optimization of the declustering potential (1 and 2) and the focusing potential.

The spectra acquired at different collision energies are shown in Figures 4.8 and 4.9. For particular experimental purposes in this work, a compromise was selected where the CE was increased to maximize the intensity of the signal at  $m/z$  1156 and reduce the signal of the doubly charged precursor ion  $[M+2H]^{2+}$  at  $m/z$  819.9. This situation is portrayed graphically in Figure 4.10. The graph shows a steady decline in the intensity of the precursor ion with increasing CE. At a CE of 36 V, a maximum of product ion is produced while the signal produced by the precursor ion is close to baseline noise.

Once the parameters for the MS/MS acquisitions were optimized, a calibration curve was prepared and a digested plasma sample was analyzed by direct injection into the mass spectrometer to test if it could be used for quantification purposes, without the use of the HPLC. The rationale behind this idea lies in the fact that during MS/MS acquisition, the first mass resolving analyzer (a quadrupole mass analyzer in the instrument used) is fixed at a specified mass window. This could have been sufficient to isolate the 'signature peptide'.

Ten microliters each of standards and plasma samples were loaded into the injection loop and sequentially injected. The product ions were also monitored in 'pseudo' SRM mode. Certainly, this approach worked adequately for the calibration standards and it even made sample analysis much faster, reducing the acquisition times from 25 to 3 minutes (See Figure 4.11). The top figure shows the TIC 'chromatogram' obtained from a 0.25 ug/mL standard. The two center figures show the product ion spectra obtained for the 'signature peptide' (center top) and the internal standard (center bottom). The calibration curve is shown in the bottom of Figure 4.11.

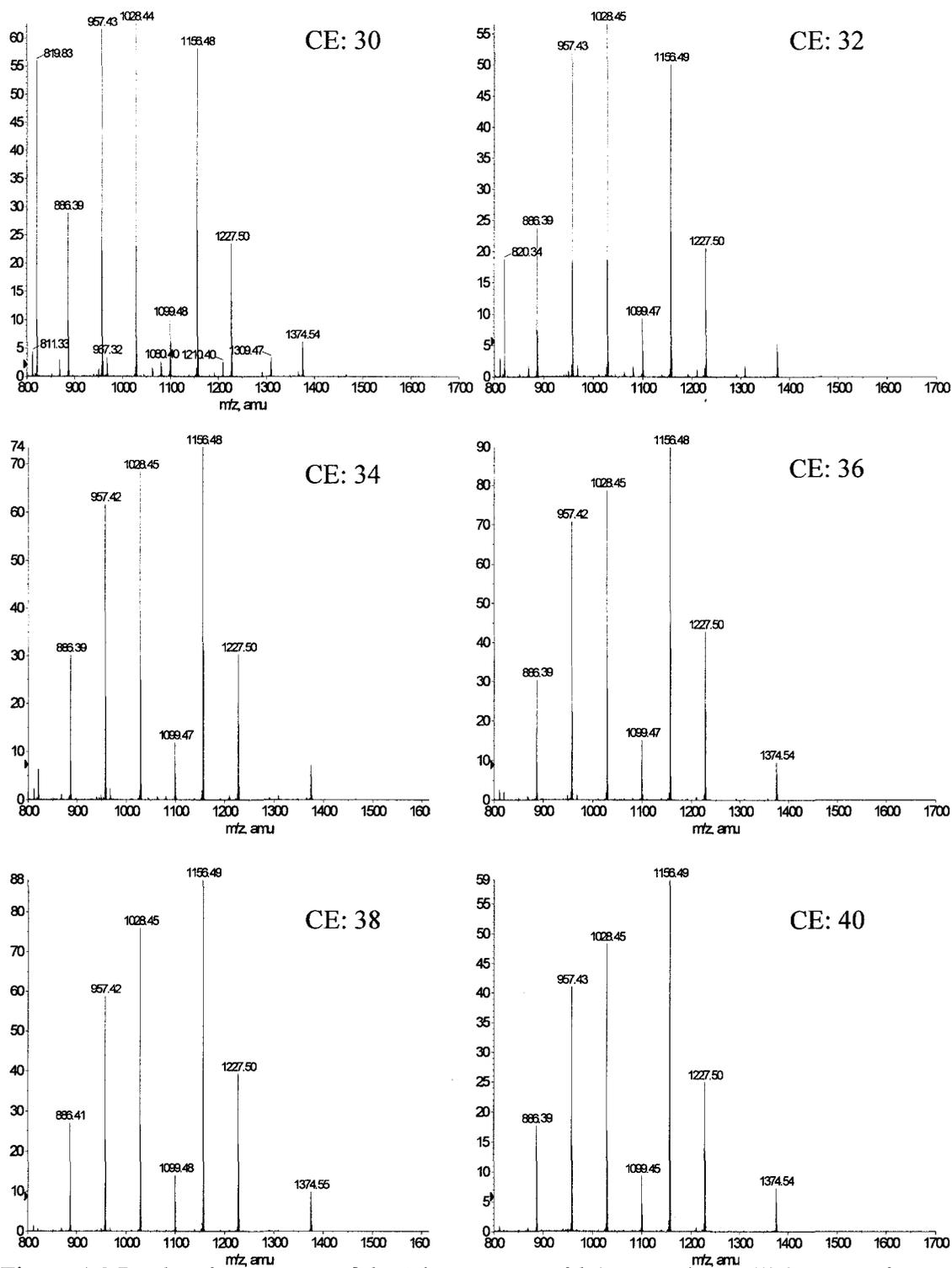


Figure 4.8 Product ion spectra of the ‘signature peptide’ at varying collision energies. The different figures correspond to collision energies (CE) from 30 to 40 V, as indicated in each spectrum.

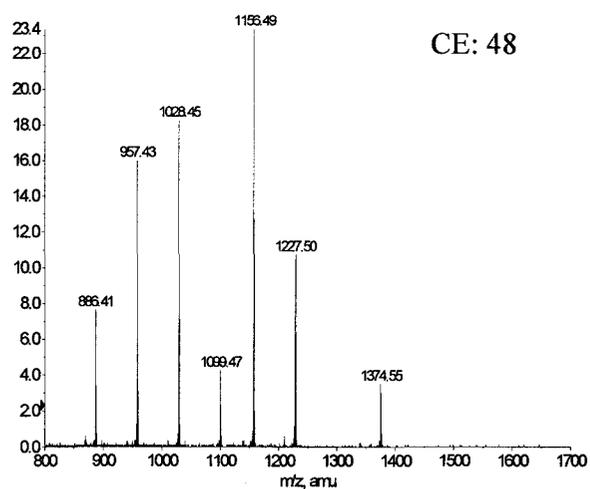
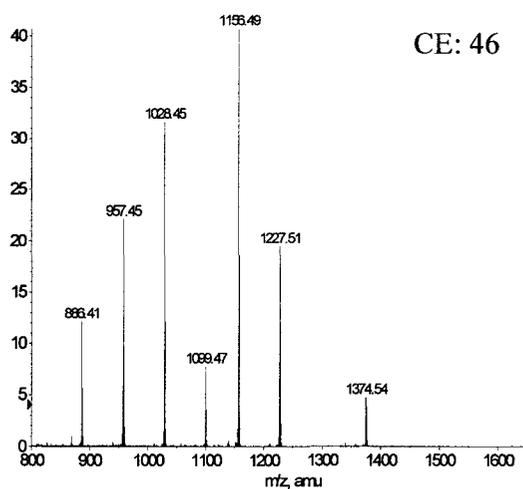
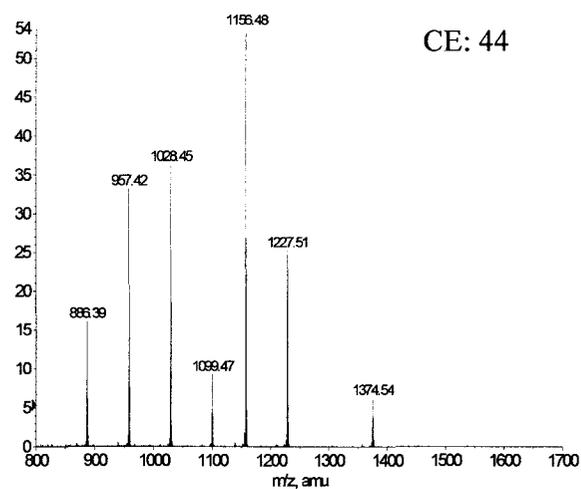
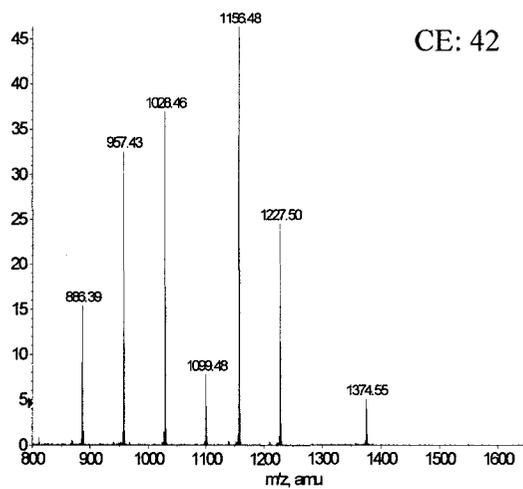


Figure 4.9 Product ion spectra of the 'signature peptide' at varying collision energies. The different figures correspond to collision energies (CE) from 42 to 48 V, as indicated in each spectrum.

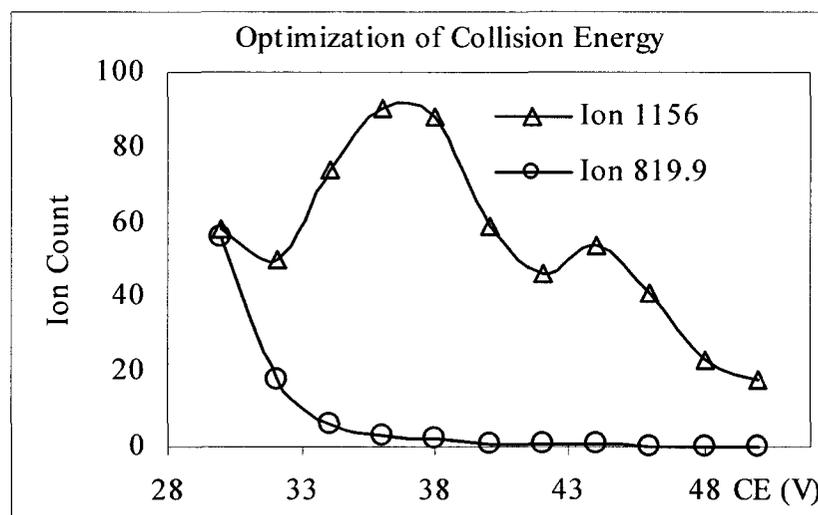


Figure 4.10 Optimization of the collision energy for ‘pseudo’ SRM-MS/MS experiments

However, when the digested plasma sample was analyzed, no signal was produced when monitoring the diagnostic product ions. In fact, when the same sample was analyzed in full scan MS mode, the peaks corresponding to the ‘signature peptide’ and the internal standard were hardly distinguishable from the baseline noise in the spectrum, as shown in Figure 4.12.

The top portion in Figure 4.12 shows the full scan spectrum of digested RT plasma with internal standard, from  $m/z$  400 to 850. The center spectrum shows an expanded scale of the same spectrum from at  $m/z$  815 to 835. The complex mixture present in the plasma digest probably explains this observation, since the ionization competition among the tryptic peptides and other interfering species at the ESI source results in the suppression and/or masking of the ‘signature peptide’ and internal standard signals.

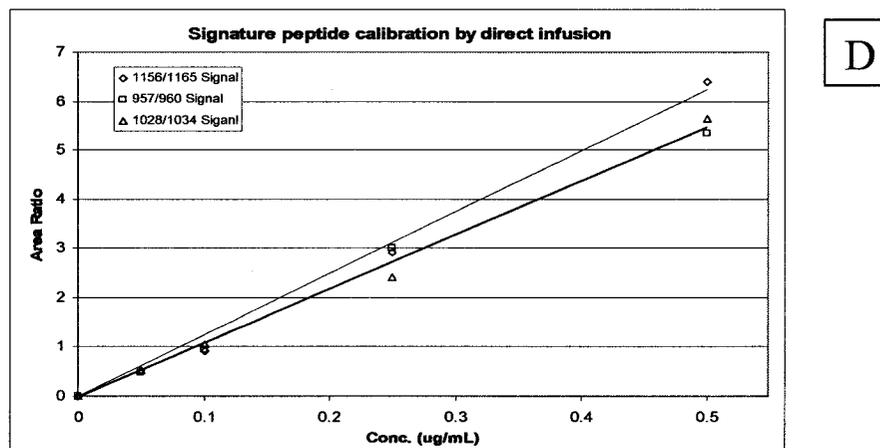
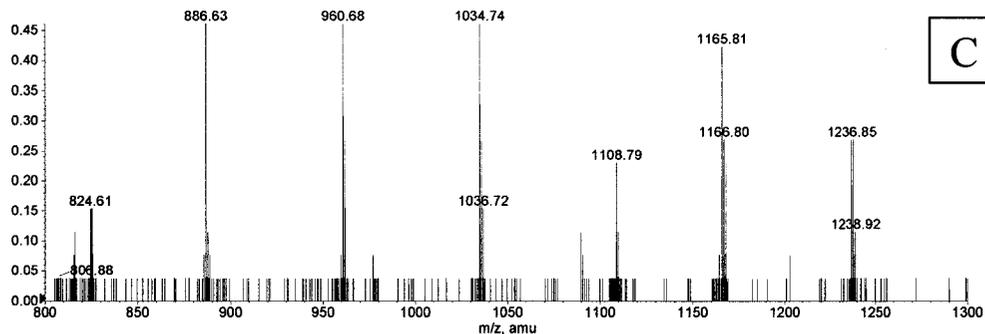
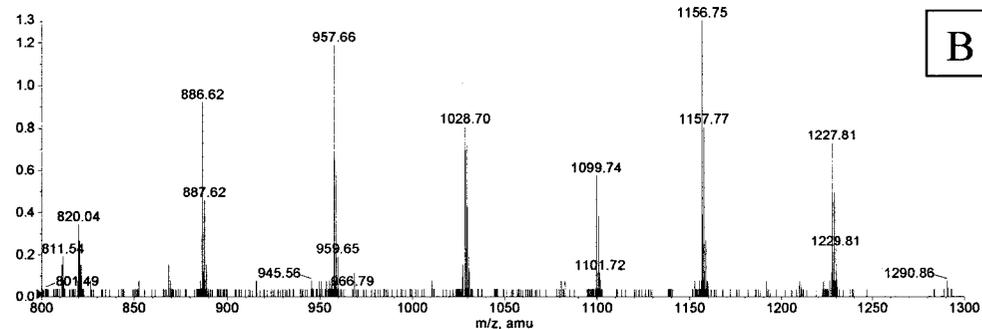
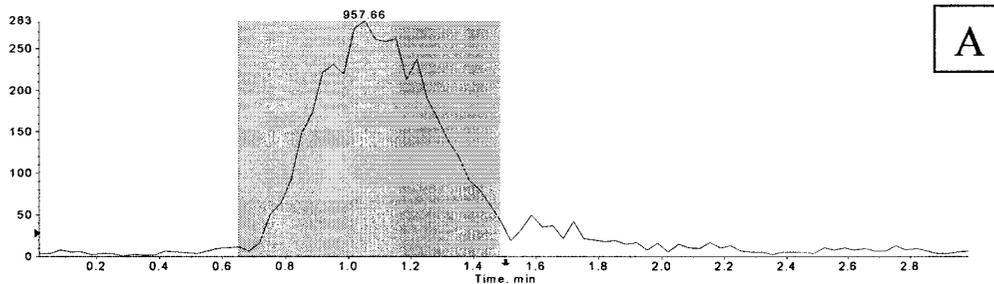
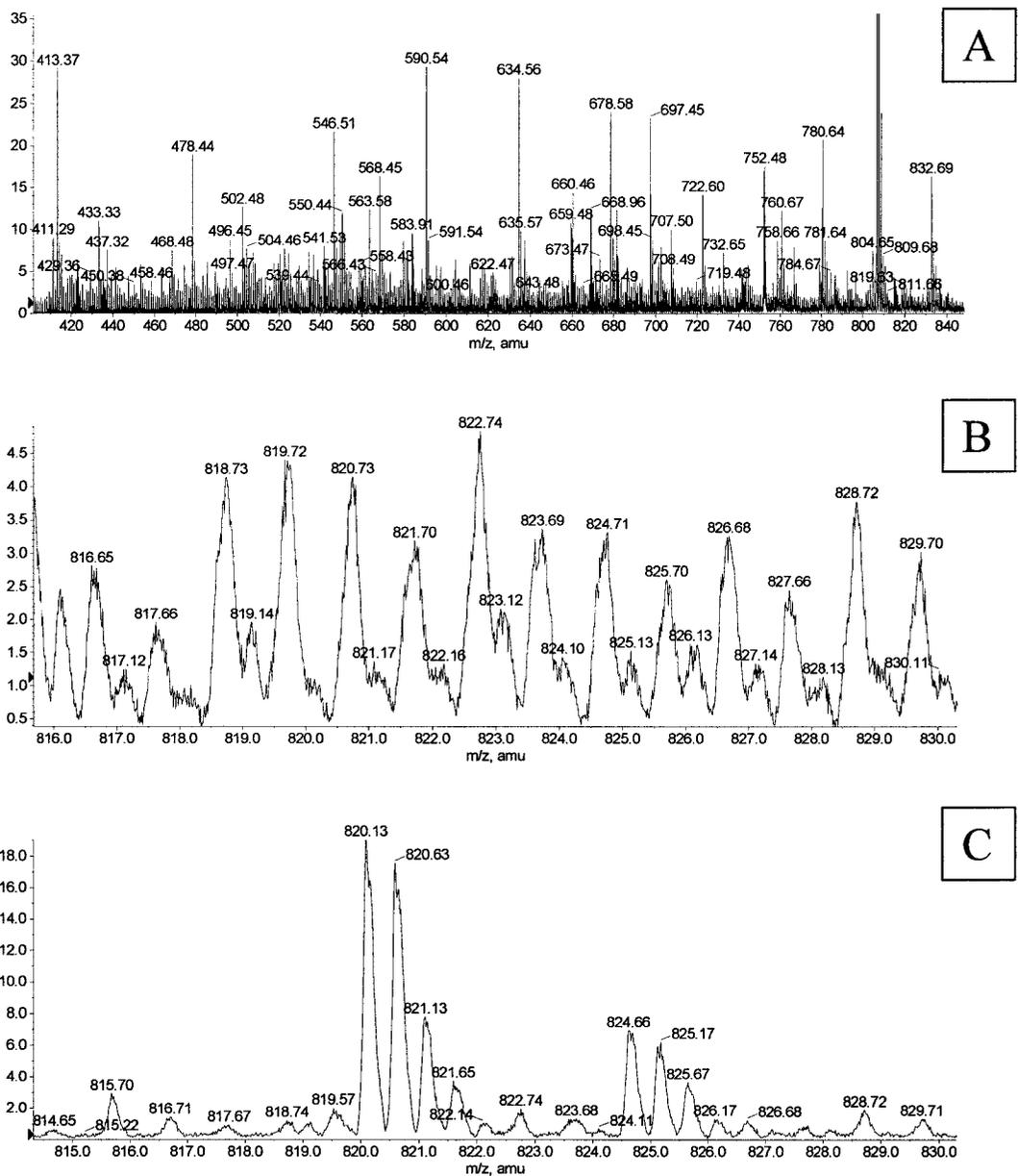


Figure 4.11 Calibration of the 'signature peptide' by direct 'loop' infusion

The top figure (A) shows the total ion count 'chromatogram' after an injection of 10 $\mu$ L of a 0.25 $\mu$ g/mL 'signature peptide' standard. Figures B and C show the product ion spectra obtained for the standard and its labeled isotopic homologue, respectively. The bottom graph (D) shows the calibration curve obtained by direct loop injection, when monitoring the Y<sub>9</sub>, Y<sub>10</sub> and Y<sub>12</sub> diagnostic product ions.



**Figure 4.12** Full scan spectra of samples obtained by direct ‘loop’ infusion  
 The top (A) and middle (B) figure show the full scan spectrum of digested plasma directly infused into the mass spectrometer. The high abundance of background peaks does not permit the visualization of the ‘signature peptide’ or its isotopic homologue at  $m/z$  819.9 and 824.9, respectively. The bottom figure (C) shows, for comparison, an acquisition of a 0.25 $\mu$ g/mL standard spiked with the internal standard under the same conditions.

This spectrum contrasts with that of the standards alone, also acquired in regular MS mode as observed in the bottom spectrum of Figure 4.12. In the latter spectrum, the  $[M+2H]^{2+}$  species at  $m/z$  819.9 and 824.4, corresponding to the 'signature peptide' and the deuterated internal standard, are clearly visible (Note that the calibration of the ToF analyzer is slightly off in these spectra). Furthermore, all the species having the same  $m/z$  would be 'collectively' selected by the first mass resolving analyzer, resulting in fragmentation competition during the CID. This would further affect the appearance of the product ion spectrum. The presence of isobaric species was confirmed, as described later, when the peptides were separated in-line by HPLC prior to MS/MS analyses.

The use of this HPLC-MS/MS hyphenated technique permits the tryptic peptides to be gradually introduced into the mass spectrometer for analysis. Valley-to-valley peak separations of individual peptides typically expected for HPLC are not required when using MS/MS as a detector. In fact, it would be unreasonable, nearly impossible, to expect the many thousands of peptides produced during a complete enzymatic digestion of plasma to be individually separated in a 25 minute separation occurring within a 3 cm chromatographic column. Nevertheless, the chromatographic conditions optimized for this method allowed for the production of strong signals from the product ions derived from the 'signature peptide' and the co-eluting deuterated internal standard.

During the chromatographic separation of a complex plasma tryptic digest, the chances of encountering a peptide with a similar  $m/z$  value to that of the 'signature peptide', with overlapping retention times, can not be ruled out. This situation is illustrated in Figure 4.13. The chromatogram of a plasma tryptic digest was acquired in

both MS and MS/MS modes and the corresponding extracted chromatograms for the precursor and product ions are shown. Figure 4.13A shows the complexity of the protein digest illustrated by the conglomeration of peaks present in the total ion count (TIC) chromatogram acquired in full scan mode. This chromatogram was simplified when the  $[M+2H]^{2+}$  'signature peptide' at  $m/z$  819.9 was extracted, as seen in the extracted ion chromatogram (EIC) shown in Figure 4.13B. However, as expected for such a complex peptide mixture, other peaks at different elution times were also observed for that particular mass.

The same sample was acquired in the 'pseudo' SRM scanning mode (see Figure 4.13C). As expected, this TIC chromatogram is very similar to that found in Figure 4.13B, since both signals are generated or reconstructed from the same selected masses. Figure 4.13D shows the added specificity afforded by tandem mass spectrometry. This figure shows the overlaid EICs corresponding to the  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  diagnostic product ions at  $m/z$  957.5, 1028.57 and 1156.6 obtained from Figure 4.13C.

The traces of these three chromatograms appear as superimposed (an expanded scale view is shown in the inset of Figure 4.13D). None of these diagnostic product ions were observed at other retention times, proving the enhanced specificity of this technique. Furthermore, the EIC (Figure 4.13D) chromatograms show reduced baseline signal and increased signal to noise ratio compared to the TIC chromatograms obtained from the precursor ions (Figure 4.13C).

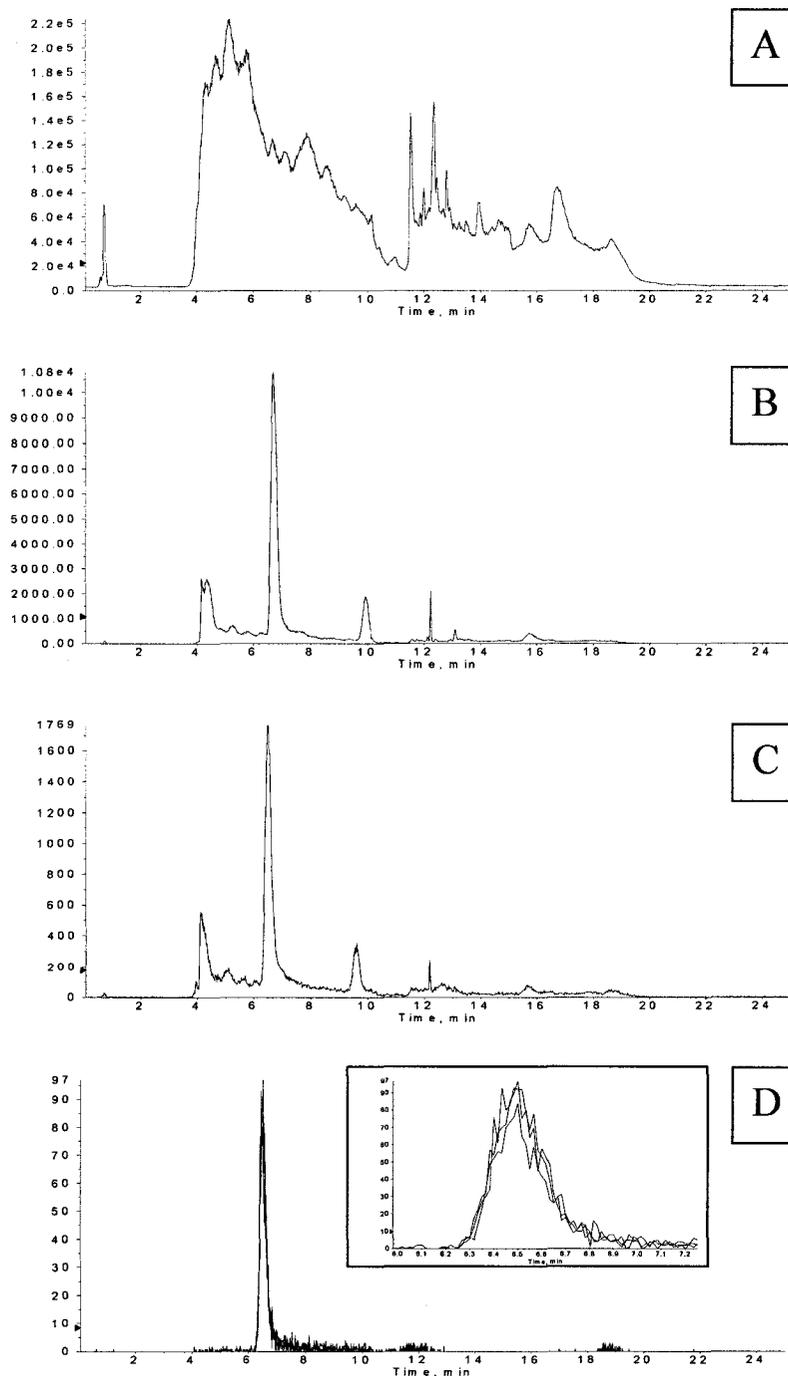


Figure 4.13 Analysis of a digested RT plasma by HPLC-MS and MS/MS

The top figure (A) shows the total ion count (TIC) chromatogram of a digested RT plasma analyzed by full scan MS. Its extracted ion chromatogram (EIC) for  $m/z$  819.9 is shown in figure B. The same sample was acquired in the 'pseudo' SRM-MS/MS mode. The TIC chromatogram is shown in figure C. Finally the three diagnostic product ions,  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$ , were extracted from the spectrum shown in C. The traces of the three overlaid EICs for these ions are shown in figure D. The inset in this last figure shows an expanded scale for the peak eluting at 6.5 minutes.

It should be noted that during the chromatographic separation, the solvent flow rate was decreased to 0.05 mL/min during chromatographic elution of the signature peptide (see Table 4.1) in order for enough ion counts to be accumulated for the product ion spectra. This procedure increased the analytical sensitivity when compared to higher flow rate acquisitions. Also, after the elution of the signature peptide, the flow rate was increased to 0.40 mL/min (see Table 4.1), reducing the chromatographic separation time to 25 minutes, convenient enough to perform reasonable throughput analyses for larger scale sampling.

The extraction of ions from a TIC chromatogram using a TOF detector operated in pseudo-SRM mode is a post-acquisition process. To test whether increased resolution of the QqTOF offers any advantage over the QqQ, the extracted chromatograms were generated using different mass ranges. Figure 4.14A shows the TIC chromatogram of the precursor at  $m/z$  819.9 acquired in the 'pseudo' SRM scanning mode from a plasma digestion sample. Figures 4.14B and 4.14C show the windows of 0.3 and 2.0 Da used around the  $Y_{12}$  product ion at  $m/z$  1156 to obtain the EICs shown in Figures 4.14D and 4.14E, respectively. The 0.3 Da mass window includes the monoisotopic peak exclusively while, the 2.0 Da window is wide enough to include the complete isotopic cluster.

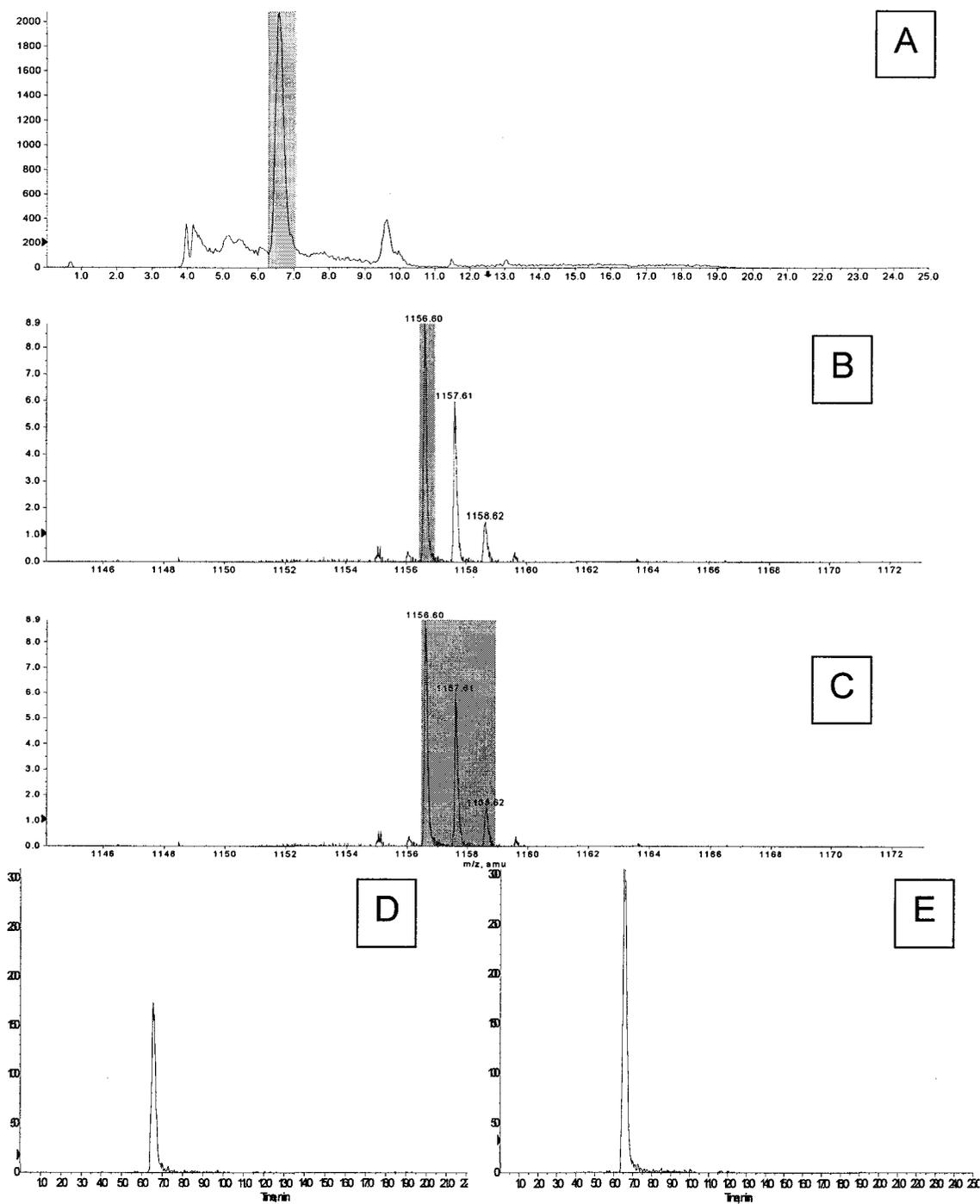


Figure 4.14 Optimization of mass ranges for extracted ion chromatograms

Figure A shows the TIC chromatogram of a digested plasma acquired in the 'pseudo' SRM-MS/MS mode. Figures B and C show the product ion spectra obtained during the elution of the signature peptide. The peak shown in both spectra corresponds to the  $Y_{12}$  product ion at  $m/z$  1156. Figures D and E correspond to the EICs taking into account a mass range of 0.3 and 2.0 Da, respectively. These mass ranges are portrayed as the shaded areas in spectra B and C.

The analysis of Figure 4.14 clearly shows there was no clear increase in specificity by using the higher resolution offered by the QqTOF. None of the extracted chromatograms (Figures. 4.14D and 4.14E) showed a difference in their profile, *i.e.* no interfering peaks were observed in either chromatogram. However, sensitivity was compromised by half with the narrower extraction window (Figure 4.14D), as shown by the less intense peak.

Identical results were obtained from the extracted ion chromatograms of the Y<sub>9</sub> and Y<sub>10</sub> product ions at  $m/z$  957.5 and 1028.57, respectively. Moreover, the use of a wider window would avoid any inconvenience caused by a subtle modification in the ToF calibration parameters, where a slight change would be translated into increased peak integration errors during automated batched data acquisition analysis. Therefore, all subsequent extracted chromatograms were analyzed using a mass range window of 2 Da.

The internal standard was directly infused into the mass spectrometer and analyzed by ToF-MS to confirm that no residual non-deuterated L-alanine had been incorporated instead of L-alanine-3,3,3-d<sub>3</sub> during its synthesis (see Figure 4.15). The peak at  $m/z$  824.5 corresponds to the  $[M+2H]^{2+}$  species. The absence of peaks at  $m/z$  819.9 corresponding to  $[M+2H]^{2+}$  of the non-labeled peptide confirmed the purity of the labeled standard. This requirement was critical, since all the samples analyzed in this study were spiked with the internal standard.

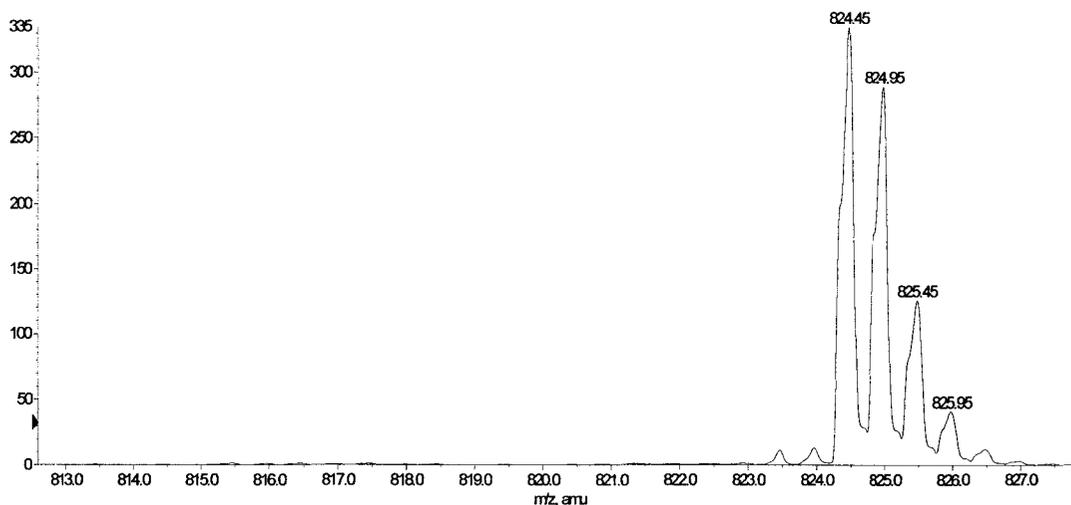


Figure 4.15 Full scan spectrum of an infused labeled ‘signature peptide’ standard  
The absence of a peak at  $m/z$  819.9 confirms the absence of unlabeled residues in the internal standard.

#### 4.4 Sample analysis

After adjustment of the experimental parameters, a calibration curve was constructed for each of the product ions monitored in the ‘pseudo’ SRM experiments acquired on the HPLC-ESI-Qq-ToF MS/MS. The calibration standards were prepared at a concentration of 0.05, 0.10, 0.25 and 0.50  $\mu\text{g}/\text{mL}$  of signature peptide. All standards and unknown plasma samples were spiked with internal standard to produce a final concentration of 0.10  $\mu\text{g}/\text{mL}$  of the deuterated signature peptide. The three calibration curves and the regression parameters obtained for the three monitored product ions are depicted in Figure 4.16. Note that all monitored product ions gave a similar linear

regression equation. The dependent variable was obtained from the area ratios of the  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  monitored product ions to their isotopic homologues.

The values obtained from the regression analysis again confirm the robustness of this method. The limit of quantification (LOQ) for this method was calculated as the interpolated concentration produced by a baseline noise peak multiplied by a factor of ten. This translated into a value of 0.0008  $\mu\text{g/mL}$  of 'signature peptide', equivalent to 0.009 mg/mL of Vtg.

The importance of adding the internal standard to correct for instrumental and processing variations is demonstrated in Figure 4.17. This figure shows the same calibration curves as before, but in these graphs the dependent variable is assigned to the absolute areas of the monitored product ions  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$ , instead of the area ratios. There is an increase in the scattering of the experimental data, statistically indicated by the decrease in the value of the  $R^2$ .

This increase in reproducibility, however, compromised the sensitivity of the method. When single precursor product ion spectra are acquired, the 1<sup>st</sup> stage mass analyzer is fixed on a single mass range value. When operating in SRM mode, the tandem mass spectrometer is programmed to monitor both the products of the 'signature peptide' and its deuterated homologue. This decreased the overall acquisition time for each precursor by a factor of two, since the 1<sup>st</sup> stage analyzer sequentially selected one of the precursors at a time for the product ion spectra to be obtained. However, the increased precision was considered more beneficial than the decrease in analytical sensitivity.

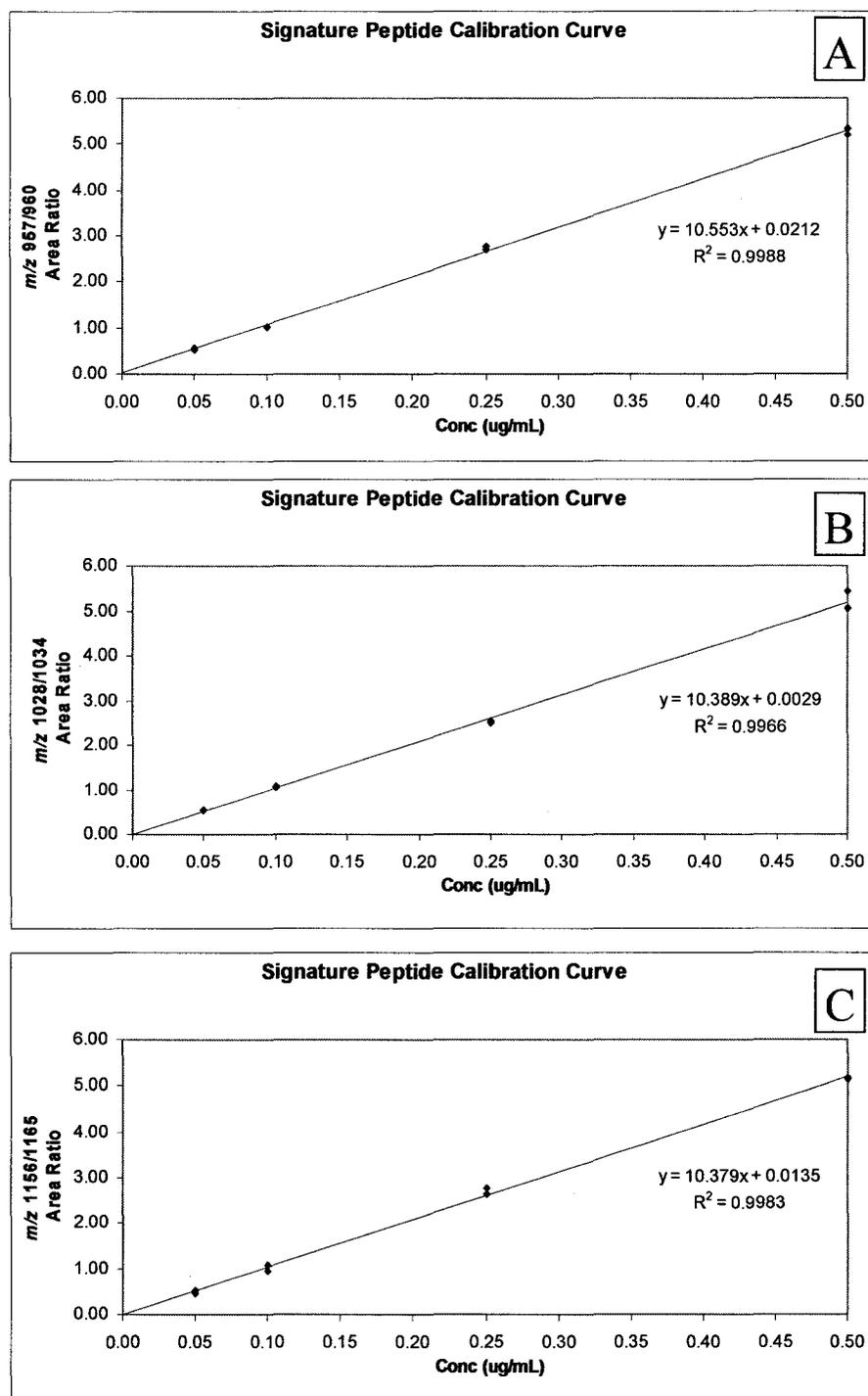
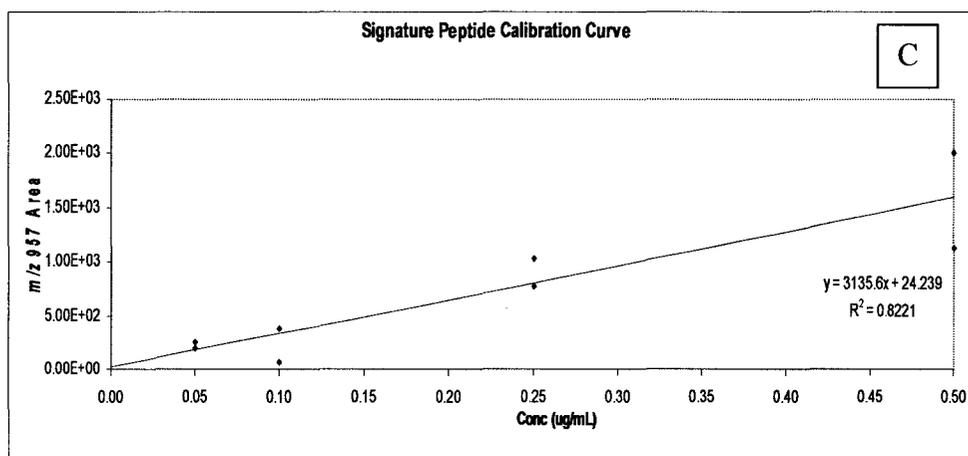
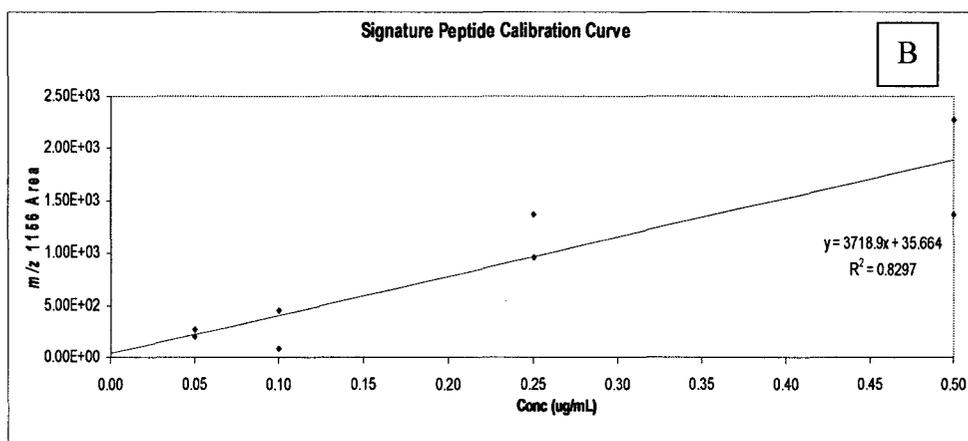
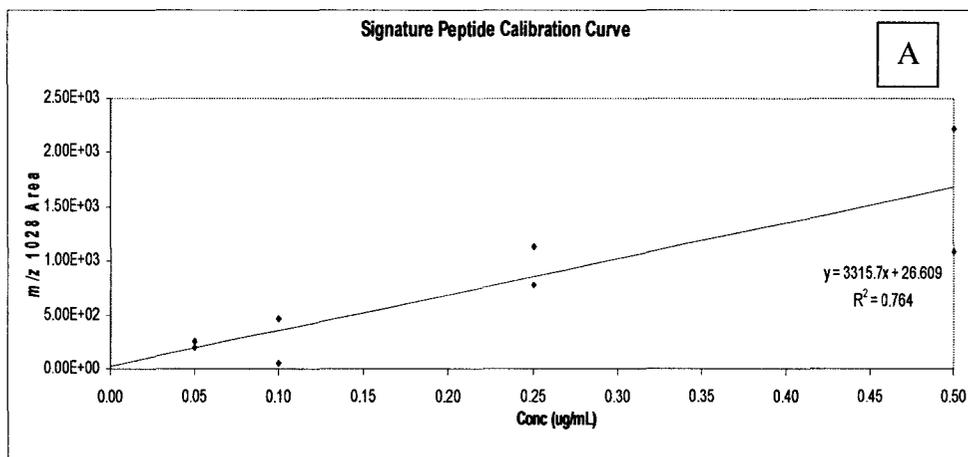


Figure 4.16 Calibration of the 'pseudo' SRM-MS/MS experiments

The calibration curves for the diagnostic  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  product ions is shown in figures A, B and C, respectively. In all cases, the area ratios of the product ions to the corresponding isotopic homologues were taken as the response signals for the dependent variables. All standards were spiked with internal standard to produce a final concentration of  $0.10\mu\text{g/mL}$  of the deuterated signature peptide.



**Figure 4.17 Calibration of classical MS/MS experiments**

The calibration curves for the diagnostic  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  product ions is shown in figures A, B and C, respectively. In all cases, the absolute area values of the product ions were taken as the response signals for the dependent variables. An increased scattering of data is observed.

After injecting the highest concentration standard of 0.50 $\mu$ g/mL, a blank sample (Solvent B) was injected to check for 'carryover' signal. This procedure was repeated after any sample which had a peak corresponding to the 'signature peptide', whose intensity was more intense than that of the highest calibration standard.

The digested plasma samples were run under the same conditions as the standards. Figure 4.18 shows a RT sample prior to (Figure 4.18A) and after 17 $\beta$ -estradiol induction (Figure 4.18B). These figures show the overlaid extracted ion chromatograms of the Y<sub>9</sub>, Y<sub>10</sub> and Y<sub>12</sub> product ions of both 'signature peptide' (SP) and labeled internal standard (IS) at *m/z* 957.5, 1028.57 and 1156.6, and *m/z* 960.5, 1034.61 and 1165.7, respectively. The traces obtained from the 'signature peptide' before 17 $\beta$ -estradiol induction were nearly indistinguishable from the baseline while the traces corresponding to the internal standard formed a clear peak in both samples. The inset shows the expanded scale of these traces, confirming the absence of Vtg before treatment (Figure 4.18A).

The 'signature peptide' and equivalent Vtg plasma concentrations from all the samples analyzed are given in Table 4.2. The identities of the samples analyzed were confirmed by the presence of all three product-ions (Y<sub>9</sub>, Y<sub>10</sub> and Y<sub>12</sub>) in the 'pseudo' SRM analyses at the corresponding elution time of the 'signature peptide'

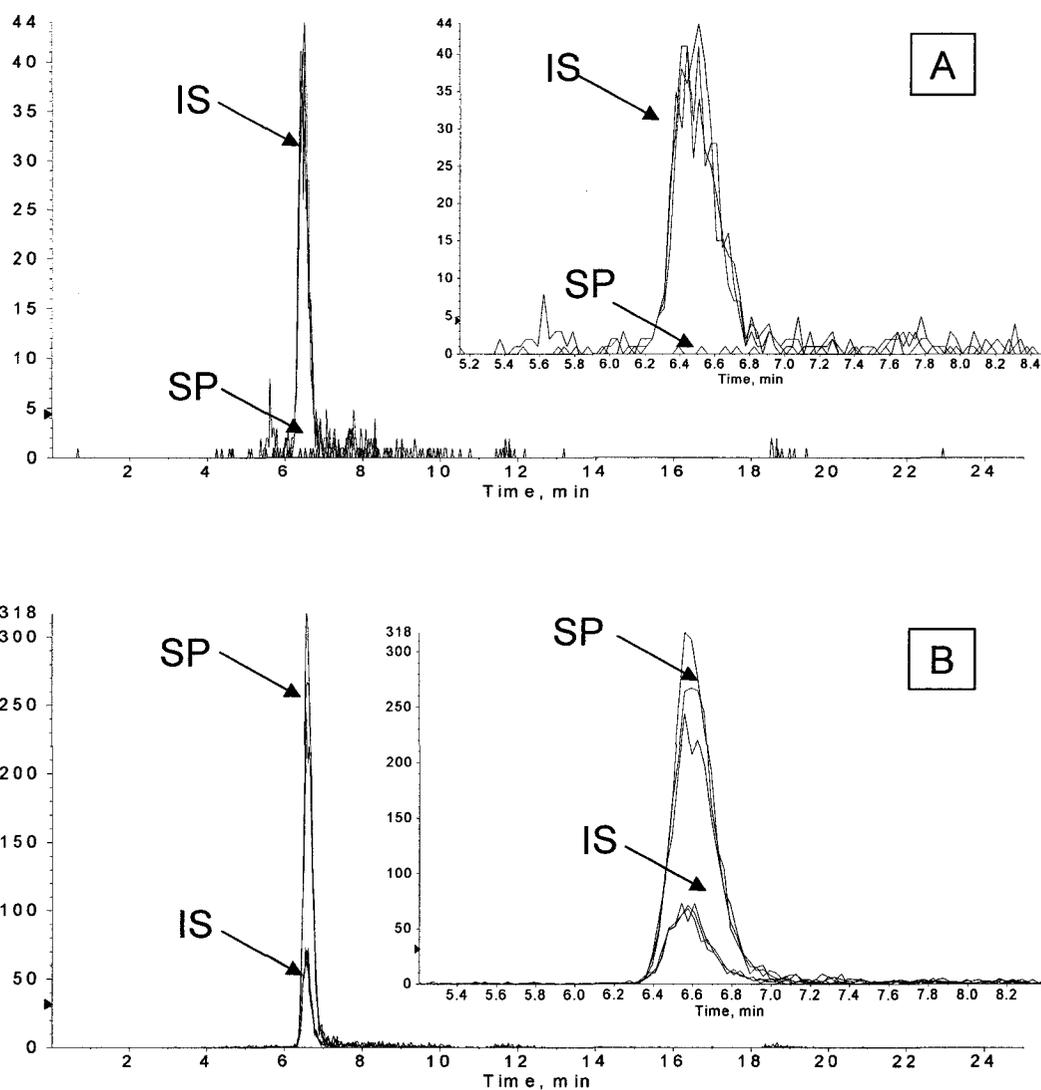


Figure 4.18 Detection of the 'signature peptide' by HPLC-MS/MS

The overlaid EICs obtained from the 'pseudo' SRM-MS/MS analysis for the same fish prior to (Figure A) and following  $\beta$ -estradiol injection (Figure B) are shown. The absence of a signal for the signature peptide (SP) is observed in the control sample (figure A). The trace of the internal standard (IS) is similar in both samples. The inset figures show an expanded scale of the peaks eluting at 6.5 minutes. In these insets, the overlaid signals produced by the  $Y_9$ ,  $Y_{10}$  and  $Y_{12}$  product ions appear as overlapped traces. The concentration of the deuterated signature peptide was  $0.10\mu\text{g/mL}$ .

Table 4.2 Results obtained for the plasma samples analyzed by the ‘signature peptide’ approach

The results for both control (C) and experimental (E) rainbow trout (RT) and Atlantic salmon (AS) are expressed as ‘signature peptide’ (SP) and Vtg concentrations. Only the experimental (E) samples showed significant P values in the t-test for paired samples before and after 17 $\beta$ -estradiol induction.

Sample	Type	Before $\beta$ -estradiol			After $\beta$ -estradiol			t-test P
		SP Conc $\mu\text{g/mL}$	Vtg Conc $\text{mg/mL}$	Vtg Conc $\mu\text{M}$	SP Conc $\mu\text{g/mL}$	Vtg Conc $\text{mg/mL}$	Vtg Conc $\mu\text{M}$	
RT	C	< LOQ	< LOQ	< LOQ	0.096	5.329	0.029	0.11
RT	C	0.041	2.276	0.013	0.085	4.718	0.026	
RT	E	< LOQ	< LOQ	< LOQ	0.576	31.973	0.176	<<0.001
RT	E	< LOQ	< LOQ	< LOQ	0.799	44.352	0.244	
RT	E	0.068	3.775	0.021	0.562	31.196	0.171	
RT	E	< LOQ	< LOQ	< LOQ	0.456	25.312	0.139	
RT	E	< LOQ	< LOQ	< LOQ	0.615	34.138	0.188	
RT	E	0.029	1.610	0.009	0.771	42.798	0.235	
RT	E	< LOQ	< LOQ	< LOQ	0.197	10.935	0.060	
RT	E	< LOQ	< LOQ	< LOQ	0.368	20.428	0.112	
RT	E	< LOQ	< LOQ	< LOQ	0.185	10.269	0.056	
RT	E	< LOQ	< LOQ	< LOQ	0.564	31.307	0.172	
RT	E	0.025	1.388	0.008	0.754	41.854	0.230	
AS	C	0.008	0.444	0.002	0.054	2.998	0.016	
AS	C	0.290	16.098	0.088	0.209	11.601	0.064	
AS	C	< LOQ	< LOQ	< LOQ	0.063	3.497	0.019	
AS	C	< LOQ	< LOQ	< LOQ	0.103	5.717	0.031	
AS	E	< LOQ	< LOQ	< LOQ	0.628	34.860	0.192	<<0.001
AS	E	0.223	12.379	0.068	0.969	53.789	0.296	
AS	E	< LOQ	< LOQ	< LOQ	0.439	24.369	0.134	
AS	E	< LOQ	< LOQ	< LOQ	0.653	36.248	0.199	
AS	E	0.008	0.444	0.002	0.540	29.975	0.165	
AS	E	< LOQ	< LOQ	< LOQ	N/A	N/A	N/A	

For quantification purposes, the concentrations obtained from each of the diagnostic product ions were averaged. The coefficient of variation was typically below 10% among the three concentrations obtained for each sample. The significance of the concentrations should be placed in the order of magnitude. For this reason, none of these results were corrected for recovery. Concentration values of zero analytically mean less than the LOQ. As expected, most of the juvenile RT had non-detectable plasma levels of Vtg before 17 $\beta$ -estradiol induction, except for a couple of larger female trout which were probably at stages of early sexual maturation. As for the AS, two samples were confirmed as females, after the animals were killed and their ovaries were found to be filled with eggs. This would explain the high levels of Vtg found in some of the fish before 17 $\beta$ -estradiol induction, characteristic of the vitellogenic reproductive phase.

A t-test for paired samples was applied to each one of the groups (control RT, experimental RT, control AS and experimental AS). As expected, a statistically significant ( $P < 0.001$ ) increase in Vtg plasma concentration was established for all experimental fish, whereas no significant increase was observed for the control fish.

#### **4.5 Recovery studies**

Four peptide standard solutions of 0.1  $\mu\text{g/mL}$  were prepared and processed under the exact conditions to which the plasma samples were subjected, to test for the recovery

of the entire procedure, starting from the denaturing step before the enzymatic digestion step (see section 4.3.1).

The results from the recovery tests are depicted in Table 4.3. The peptide concentration for each recovery sample was calculated as the average concentration obtained from each of the Y<sub>9</sub>, Y<sub>10</sub> and Y<sub>12</sub> calibration curves. An average of these, the standard deviation (SD) and the coefficient of variation (CV) percent were calculated. Reproducibility of this methodology was within acceptable limits having a coefficient of variation of 3.4%. However, recovery results averaging 67% were not as good as expected. Nevertheless, because of the systematic and reproducible loss of peptide under the described experimental conditions, these recovery results could be incorporated into a correction factor to account for peptide loss during sample handling and experimental procedure.

Table 4.3 Statistical analysis of recovery studies

Four 0.1µg/mL signature peptide solutions were prepared (R<sub>1</sub> to R<sub>4</sub>). After analysis, the average concentration, the standard deviation (SD) and the percent coefficient of variation (CV %) were calculated. The recovery % was based on 0.10µg/mL as 100% recovery.

Sample	Peptide Conc. µg/mL	Average Conc. µg/mL	SD µg/mL	CV%	Recovery %
R <sub>1</sub>	0.070				
R <sub>2</sub>	0.069				
R <sub>3</sub>	0.067	0.067	0.002	3.4	67
R <sub>4</sub>	0.064				

A possible cause for this low recovery could have been the chemical transformation of the peptide. The enzymatic digestion occurs under mild reaction conditions (37°C and pH 7.4), however the previous denaturing step takes place at 60°C in 8M urea. Carbamylation of terminal NH<sub>2</sub> groups in proteins could occur under these conditions as described in the literature (Cejka *et al.*, 1968; Makler, 1980; Lippincott and Apostol, 1999).

To check for this and any other possible chemical modifications of the standard peptide, an Information Dependant Acquisition (IDA) survey scan was performed on one of the recovery sample tubes. The product ion spectra, excluding that corresponding to the 'signature peptide', obtained from all the additional minor peaks present in the chromatogram were investigated to see whether any of these were derived from the 'signature peptide'. A careful analysis of all these spectra showed that none of them contained the carbamylated sequence tags characteristic of the 'signature peptide'.

One of the major concerns with the proposed technique was the possibility of incomplete digestion of the plasma samples. Other authors have also addressed this problem in different ways. For example, some researchers have shown the differences in SDS-PAGE plasma protein profiles prior to and after the digestion step (Zhang *et al.*, 2004a). Alternatively, another study assessed the completion of enzymatic digestion by monitoring the peptides obtained from surrogate synthetic peptides containing the same tryptic cleavage site as the native protein analyzed (Barnidge *et al.*, 2003).

In this study, possible incomplete digestion of Vtg was analyzed using a different approach. A series of hypothetical candidate sequences corresponding to the product of enzymatic missed cleavages of the ‘signature peptide’ were proposed. These can be observed in Table 4.4. Subsequently, digested plasma samples of RT were injected into the HPLC-MS/MS and processed in the ‘pseudo’ SRM mode, except that a different precursor ion from Table 4.4 was selected for each run.

The TIC chromatograms obtained for these precursors at  $m/z$  819.92, 933.49, 1361.22, 1474.79, 1182.65 and 1891.65, revealed some suspicious peaks, which are shown in Figures 4.19 A, B, C, D and E, respectively. However, the product ion spectra obtained from them showed no evidence of any of the fragment ions expected from the candidate peptides. This suggested that digestion was complete, at least in this region of the protein.

Table 4.4 Candidate peptides for the incomplete digestion of RT Vtg

The first peptide in the list (\*) corresponds to the ‘signature peptide’. The slash (-) in the remaining peptides indicates the positions of missed cleavage sites.

Peptide sequence	[M+H] <sup>+</sup> <i>m/z</i>	[M+2H] <sup>2+</sup> <i>m/z</i>
TYFAGAAADVLEVGR*	1638.84	819.92
AR-TYFAGAAADVLEVGR	1865.98	933.49
TYFAGAAADVLEVGR-TEGIQEALLK	2721.44	1361.22
AR-TYFAGAAADVLEVGR-TEGIQEALLK	2948.58	1474.79
TVVAK-AR-TYFAGAAADVLEVGR	2364.30	1182.65
TYFAGAAADVLEVGR-TEGIQEALLK-PPAPENADR	3781.97	1891.49

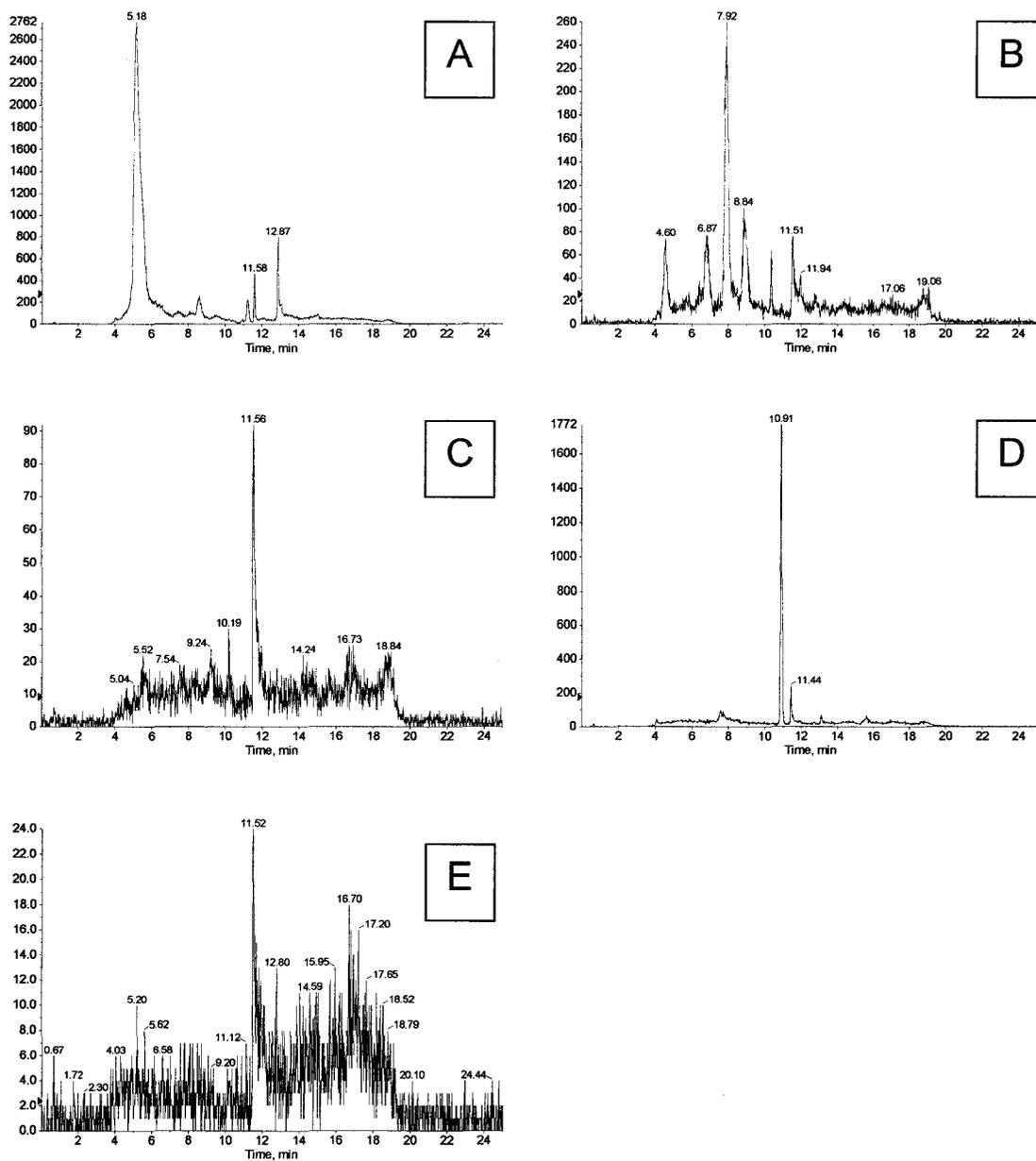


Figure 4.19 Candidate peaks for incomplete digestion peptides of Vtg

A 'pseudo' SRM-MS/MS analysis was performed for all the candidate peptides of the incomplete Vtg digestion listed in Table 4.4. The spectra A, B, C, D and E correspond to the TIC chromatograms obtained for the  $[M+2H]^{2+}$  precursors at  $m/z$  933.49, 1361.22, 1474.79, 1182.65 and 1891.49, respectively.

To further test the completion of the digestion step, a study was performed to follow the production of the 'signature peptide' at timed intervals during enzymatic incubation. A RT plasma was subjected to the digestion step explained with varied incubation times. The results obtained at 2, 6, 14, 18, 24 and 30 hours after the initiation of the digestion are illustrated in Figure 4.20.

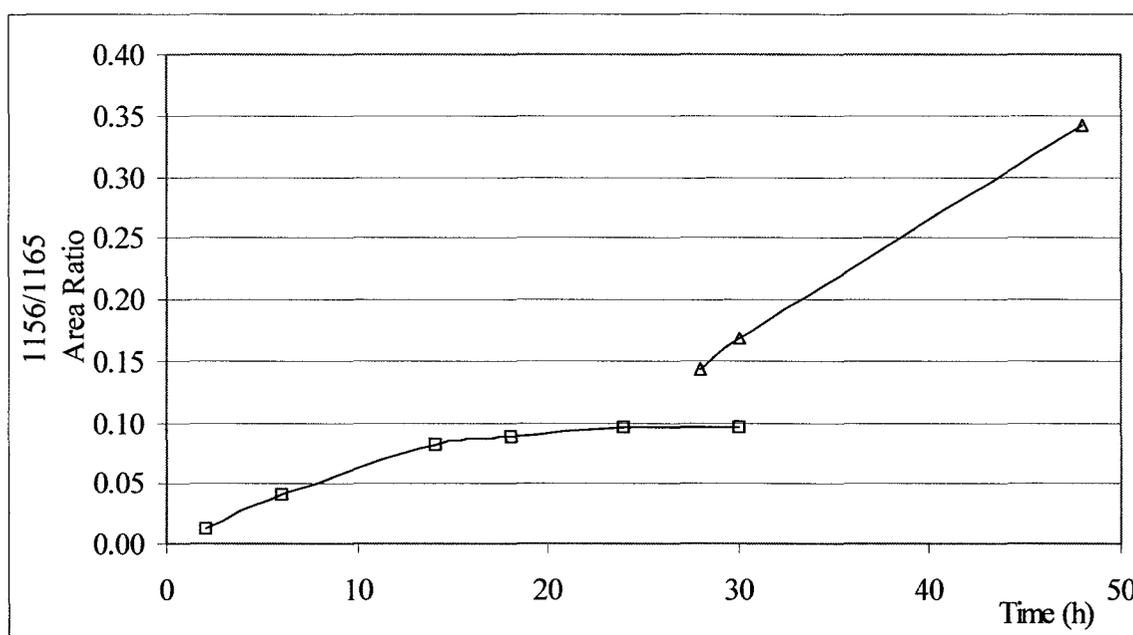


Figure 4.20 'Signature peptide' release during incubation with trypsin

The progress of the enzymatic digestion was monitored during a 30 hour period (□) by measuring the release of the 'signature peptide' analyzed HPLC-MS/MS. After 24 hours, a second aliquot of trypsin was added and monitored for an additional 24 hour period (Δ).

At 24 hours, there is a clear plateau, indicating that the digestion process has stopped. In a parallel experiment, the same sample was subjected to the same enzymatic conditions, with the addition of an extra aliquot of enzyme after the initial 24 hours

(indicated in Figure 4.20 by the 24+24 hour incubation label). The reaction was monitored to a total of 30 and 48 hours from the initiation of the digestion. The results are also portrayed in Figure 4.20. The increase in the signal of the 'signature peptide' indicated that the digestion process did not conclude in the initial 24 hours. Thus, the plateau referred earlier is probably due to inactivation or degradation of the enzyme present in the digestion. This is very important since, some of the published literature usually regard this plateau as the completion of the digestion reaction (Zhang *et al.*, 2004a).

#### **4.6 Vitellogenin stability studies**

Finally, the preservation of samples prior to analysis was studied to observe the effect of freeze-thaw cycles on the final results. This is a critical issue since blood sampling usually occurs during fieldwork, and very rarely under laboratory conditions, and there may be a considerable period of time between sampling and laboratory processing. This lag time increases the possibility of Vtg loss as this protein is known to be highly susceptible to degradation (Silversand *et al.*, 1993). To examine the storage stability of Vtg, unstable conditions were simulated by subjecting the same sample to an increasing number of freeze-thaw cycles prior to analysis.

A fresh plasma sample of Atlantic salmon was separated into 4 aliquots. The first was frozen and stored at -20°C. The remaining three were subjected to 2, 4 and 6 freeze-

thaw 24 hour cycles and finally stored at -20°C. The four aliquots were analyzed as regular plasma samples.

The results of this study are given in Table 4.5 and demonstrate that harsh storage conditions assayed in this test clearly do not affect the results obtained. In addition, the reproducibility of the concentration values is consistent with that obtained from the recovery test.

Table 4.5 Effect of freeze-thaw cycles on final concentrations

After analysis, the average concentration, the standard deviation (SD) and the percent coefficient of variation (CV %) were calculated. The latter was consistent with the values obtained in the Recovery studied shown in Table 4.3.

Freeze-thaw cycles	Peptide conc. (µg/mL)	Average conc. (µg/mL)	SD (µg/mL)	CV %
1	0.431			
3	0.449	0.451	0.019	4.17
5	0.449			
7	0.477			

## **Chapter 5: Discussion and conclusions**

### **5.1 Similarities between Atlantic salmon and rainbow trout vitellogenin**

In the past, Vtgs of different species have been sequenced and stored in protein databases. A BLAST search was performed using the RT Vtg as the query input. Results showed a list of at least fifty vitellogenic related proteins and precursors exhibiting homology. When comparing specifically the published sequences of the RT (See Appendix 2: Q92093) and AS Vtg (Appendix 2: Q8UWG2 and Q800N9), protein homology was evident as seen in the result of the FASTA (Pearson, 1990; Pearson, 1991) output file (Figure 5.1) using the local alignments (Lalign) program (see Section 3.1.7.5). Q92093 exhibited a 91.429% identity in a 175 amino acid overlap with Q8UWG2, and 81.041% identity in a 269 amino acid overlap with Q800N9. Extracted and purified under the same conditions, the molecular mass of AS Vtg seemed to have a Mr only 2.6% smaller than that of the RT Vtg, according to the results of the MALDI-ToF-MS experiments.

Conservative substitutions, that is, aminoacids with similar chemical properties have been replaced (eg. Leu for Ile, Glu for Asp etc) were found along the sequence of both proteins. As observed in Figure 5.1, evidence suggested that further similarities between the Vtg of both species would be found in the study. Indeed, the results shown in the present study seem to support this.



the RT Vtg, two of which had also matched the partial sequence of the published AS Vtg. In spite of the similarities, there is some evidence to postulate amino acid variants. Finally, a remark should be made about the sequenced peptides obtained from the CID-MS/MS of the selected precursor peptide ions at  $m/z$  795.18, 893.28 and 791.05. These are apparently exclusive to the AS Vtg, possibly indicating regions of sequence variability. Further characterization of these and other non-identified fragments will allow post translational modifications and genetic variants to be confirmed.

## **5.2 The ‘signature peptide’ approach: The advantages and disadvantages**

A simple and straightforward HPLC-MS/MS technique has been presented and published for the quantification of Vtg plasma levels in both RT and AS (Cohen *et al.*, 2006). The use of a QqToF operated in the ‘pseudo’ SRM mode allowed us to monitor three diagnostic product ions, enhancing specificity without compromising sensitivity. The specificity was not increased by reducing the mass window of the extracted ion chromatograms from 2 to 0.3 Da. Moreover, sensitivity was reduced by a factor of approximately two.

The use of the common ‘signature peptide’ for both species is a great advantage in regions where both species play an important role in commercial aquaculture ventures. Furthermore, it would not be surprising to find in the future other related species that share this common tryptic peptide. This will almost certainly be corroborated with genomic studies advancing in these other species. Nevertheless, it would be a mistake to

choose this and any other candidate ‘signature peptide’ based only on their presence in the sequences of the protein databases. The experiments carried out in this work clearly show that not all of the predicted tryptic peptides appear and/or are amenable to MS/MS analysis. Different conditions can explain this behavior such as presence of PTM in the ‘signature peptide’, regions or sequences not accessible to enzymatic digestion, or structures that are resilient to CID fragmentation in tandem mass spectrometry. These situations are not always predictable based on the information of the peptide sequence. Therefore, the selection of the ‘signature peptide’ must be supported by prior characterization of the protein by similar techniques as those used during the quantification method. Evidently, the support of the sequence information given by the protein databases facilitates the selection process.

The ‘signature peptide’ approach described herein is a rapid and facile technique that could be considered an alternative to the existing immunoassays, such as ELISA and RIA. Figure 5.2 shows, in summary, the steps required for the development of a classical immunoassay versus the development of the ‘signature peptide’ approach.

The ‘signature peptide’ approach circumvents the intensive work required for the development and production of specific polyclonal and monoclonal antibodies. Moreover, it takes advantage of a laboratory synthesized high quality homogeneous standard peptide, and overcomes the difficulty of purifying this complex phosphoglycolipoprotein from fish blood. A detailed comparison of samples analyzed by the ‘signature peptide’ approach and a commercial kit available for quantification of RT Vtg is given in the next section.

Additionally, there has been a declining trend in the use of RIA due to radioactivity hazards, higher regulatory costs and significant disposal problems caused by radiochemicals.

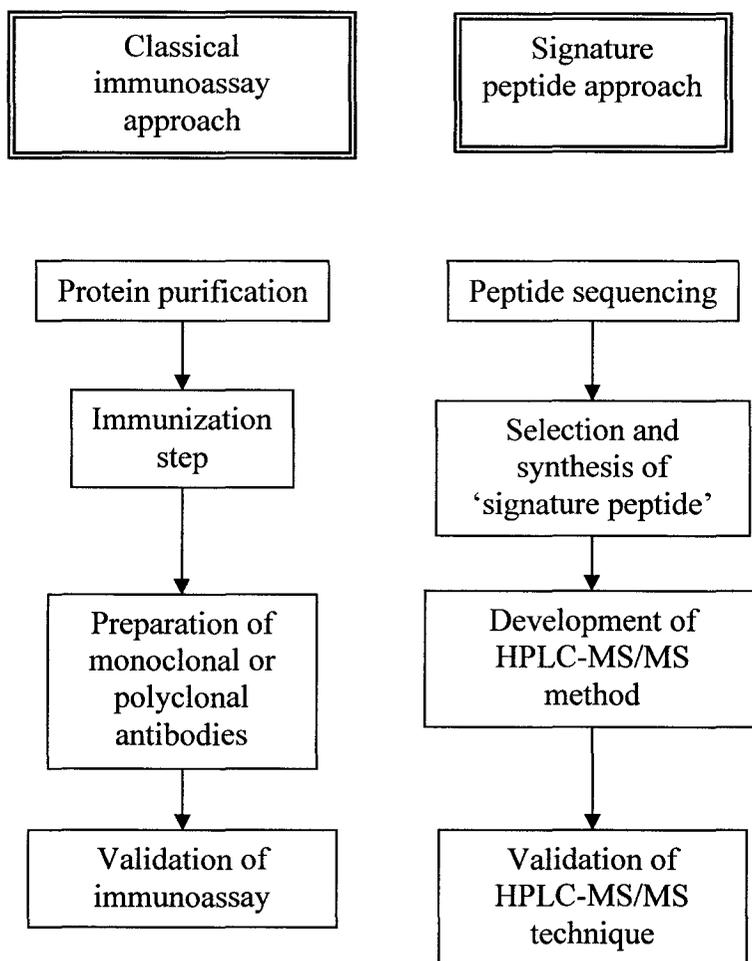


Figure 5.2 Comparison of the 'signature peptide' approach to classical immunoassay method development

In contrast to antibody affinity in immunoassays, the specificity of this technique relies on the distinctiveness of the 'signature peptide' and the power of tandem mass spectrometry to detect its diagnostic product ions. This is definitely one of the strong points of this procedure since this method will probably be less susceptible to the false positive results sometimes encountered in immunological assays. This hypothesis will eventually be tested when large-scale analyses of Vtg are performed and compared to results from both techniques.

As described in the discussion here, the trade off in this technique is the sensitivity, which is acceptable as long as the analytical LOQ is not significantly compromised. The LOQ for the 'signature peptide' method is in the order of ng/mL. However, when expressed in protein concentration, there was a loss in sensitivity by an approximate factor of 150 due to the molecular mass difference between the 'signature peptide' and the native protein. This yields a LOQ (0.009 mg/mL) for Vtg plasma levels that is well below those of the physiological levels found in mature female fish, which according to a study varies between 0.01 to 0.20 mg/mL (Copeland *et al.*, 1986). Thus, for this particular study, the LOQ did not require any improvement. However, when used as a biomarker for ecotoxicological purposes, this limit might not be low enough as those reported by some immunological tests, in the order of ng/mL of Vtg. It is important to point out that the method described in this work uses only 5  $\mu$ L of plasma for the digestion step. Nevertheless, taking into account all the dilution factors, only 20 nL of the original plasma are finally injected for HPLC separation. Furthermore, considering the nebulization and ionization efficiency of the ESI source, the actual amount of plasma entering the mass spectrometer possibly lies in the order of the picoliters.

If required, the limit of detection and sensitivity of this method could be improved in several ways. However, it should be noted that any change to the analytical method would compromise its performance or advantages in some way. These modifications can be introduced at the level of sample processing or further on, during instrumental analyses. Table 5.1 lists a summary of these modifications, together with a brief explanation on the consequences of these changes. The combination of some of these changes might result in an increase in sensitivity of several orders of magnitude. All these issues were considered when optimizing the technique, although they can obviously be changed according to the particular objectives and requirements for different types of experiments. When adapting this technique, it is important that the enzymatic digestion step should be optimized to yield the highest amount of signature peptide, by both adjusting the concentration of enzyme added and the digestion incubation time.

Another benefit of the described technique is the low volume (5 $\mu$ L) of plasma required for analysis. Although blood extraction is definitely an invasive but non-destructive procedure for biomarker analysis (Casin *et al.*, 2002), the small volume used allows the fish to be returned to their habitat with minimal stress impact.

The technique described uses extreme protein denaturing conditions and thorough enzymatic digestion prior to HPLC-QqToF-MS/MS analysis and therefore, partial protein hydrolysis is less likely to affect this method. In addition, as described earlier, because non-reactive hydrophobic amino acids are predominant, the probability of peptide modification within the 'signature peptide' region is minimized. As shown in the stability studies discussed earlier in section 4.6, sample handling and storage should become a less serious issue when compared to other techniques when this stage is critical.

Table 5.1 Modifications that could increase sensitivity and limit of detection of the 'signature peptide' approach.

Modification	Level	Consequences
Increasing volume of plasma	Sample processing	Higher quantities of trypsin needed for digestion. Need for larger HPLC column Increased analysis cost
Reducing dilution factors post digestion	Sample processing	Dirtier samples injected into HPLC-MS/MS Reduced lifetime of HPLC column Higher maintenance of MS interface Increased analysis cost
Increasing injection volumes into HPLC-MS/MS	Instrumental: HPLC	Need for larger HPLC column Increased 'carryover' signal More cleaning steps required Increased analysis time Increased analyses cost
Reducing chromatographic flow rate	Instrumental: HPLC	Increased time of analysis
Aligning tip of ESI capillary to skimmer of MS	Instrumental: MS/MS	Higher maintenance of MS interface Increased analysis cost
Monitoring a product ion with stronger signal at lower <i>m/z</i> .	Instrumental: MS/MS	Reduced specificity

Finally, a good point to mention is the issue as to what units should be used for quantification of Vtg. Usually all reported work in the literature expresses Vtg plasma levels in mass per volume units (eg mg/mL). This brings up an inherent problem if the concentrations are to be converted to molarity, because the molecular mass of Vtg can vary according to the degree of post translational modifications and its lipid content. The approach described herein for this particular 'signature peptide' lacks this inconvenience

since the molar ratio of this peptide to Vtg is known. For this reason, in future studies measuring Vtg with a similar approach, the concentration results expressed as moles (or fractions) per volume as presented here (eg  $\mu\text{M}$ ), is recommended. This would standardize results and make it extremely convenient for comparing results from different species that have Vtg molecules of different molecular masses.

### **5.3 ELISA vs. the ‘signature peptide’ approach: Which is better?**

As described earlier, the ‘signature peptide’ approach described in this work is an alternative to developing immunoassays, such as ELISA. However, several commercial kits are already available for measuring Vtg levels for many species. Such is the case for RT Vtg. In order to compare the performance of both methods, the same samples analyzed by HPLC-MS/MS were also tested by an ELISA developed for RT Vtg (See Appendix 3 for procedure details). The results are given and compared in Table 5.2.

A graphic representation of these results is given in Figure 5.3, in which all the results from the ‘signature peptide’ approach are plotted against the ELISA results. A thorough examination of this table and figure reveals some interesting observations. With respect to RT Vtg, all values below the LOQ of the ‘signature peptide’ approach (0.009 mg/mL) were also below this value for the ELISA. Most of these samples corresponded to the fish prior to  $17\beta$ -estradiol induction. This situation is portrayed in the top graph of Figure 5.3 as overlapping experimental at the intersection of both axes.

Table 5.2 Comparison of Vtg concentrations obtained from ELISA and HPLC-MS/MS using the signature peptide approach

The results for the 'signature peptide' approach taken from Table 4.2. The results for both control (C) and experimental (E) rainbow trout (RT) and Atlantic salmon (AS) are expressed as Vtg concentrations in mg/mL units.

Vtg concentrations (mg/mL)					
Sample	Type	Before $\beta$ -estradiol		After $\beta$ -estradiol	
		HPLC-MS/MS	ELISA	HPLC-MS/MS	ELISA
RT	C	<LOQ	0.001	5.329	2.849
RT	C	2.276	2.188	4.718	13.014
RT	E	<LOQ	0.000	31.973	17.605
RT	E	<LOQ	0.020	44.352	17.760
RT	E	3.775	9.155	31.196	17.643
RT	E	<LOQ	0.001	25.312	17.919
RT	E	<LOQ	0.002	34.138	18.498
RT	E	1.610	0.602	42.798	17.685
RT	E	<LOQ	0.008	10.935	16.914
RT	E	<LOQ	0.000	20.428	18.000
RT	E	<LOQ	0.000	10.269	12.333
RT	E	<LOQ	0.001	31.307	17.248
RT	E	1.388	0.595	41.854	17.949
AS	C	0.444	0.000	2.998	0.000
AS	C	16.098	0.151	11.601	0.000
AS	C	<LOQ	0.000	3.497	0.000
AS	C	<LOQ	0.000	5.717	0.000
AS	E	<LOQ	0.000	34.860	0.425
AS	E	12.379	0.063	53.789	0.717
AS	E	<LOQ	0.000	24.369	0.386
AS	E	<LOQ	0.000	36.248	0.304
AS	E	0.444	0.000	29.975	0.248
AS	E	<LOQ	0.000	N/A	N/A

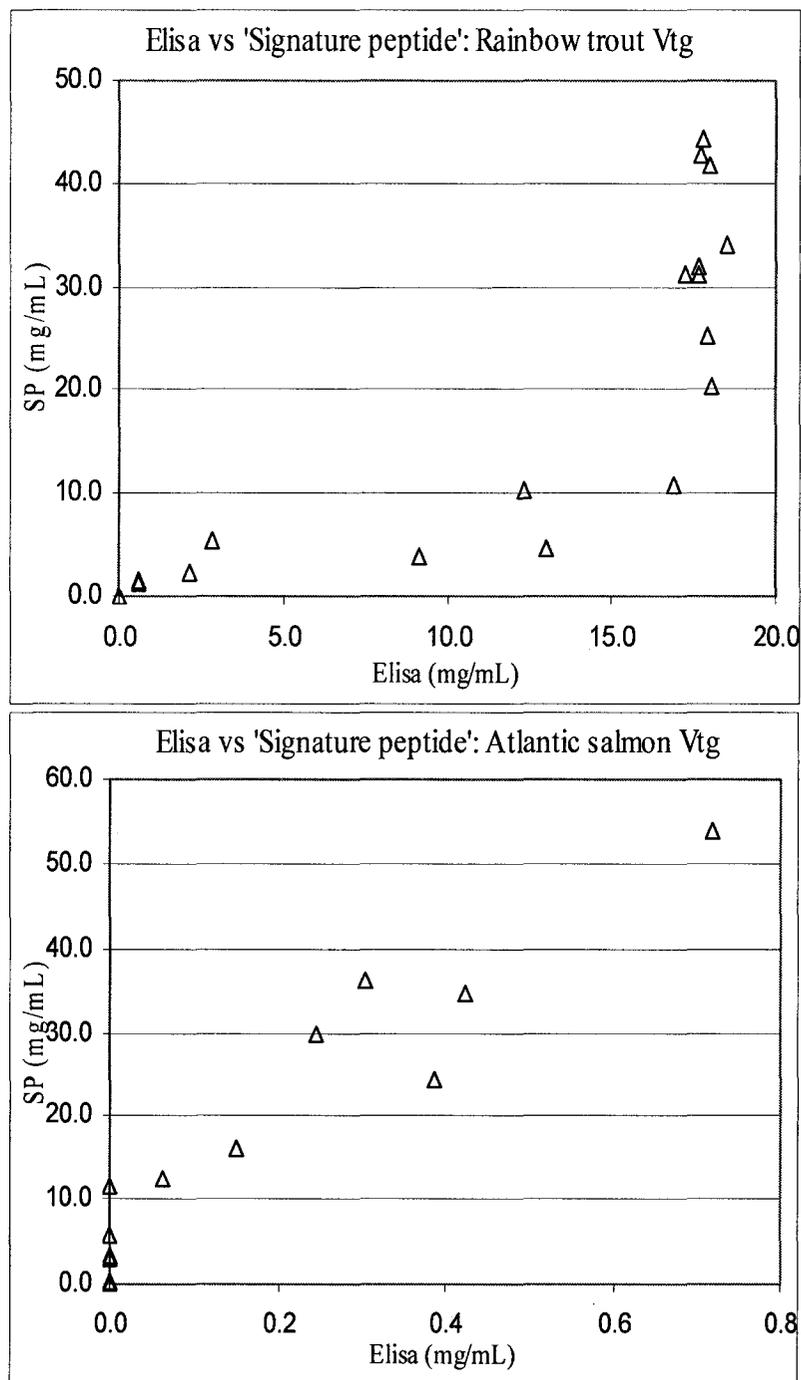


Figure 5.3 Comparative plots for samples analyzed by the 'signature peptide' approach and ELISA

The top and bottom graph correspond to the rainbow trout and Atlantic salmon samples, respectively. In both plots, the results obtained by the 'signature peptide' approach are plotted in the ordinate (SP) and the ELISA results in the abscissa.

A good correlation was observed for the experimental points below 15mg/mL for RT Vtg. Beyond this point, the results from the ELISA kit seem to plateau, unable to detect higher concentrations. This situation is due to the high absorbance values obtained for these samples by the microplate reader, almost at the limits of that obtained for the highest concentration standard. It should be noted that this unusual response in the ELISA method could be corrected by preparing additional dilutions of the plasma samples. This is a common situation encountered in ELISA where, usually, serial dilutions of the sample have to be prepared to insure the samples will fall within the valid response range of the standard curve.

Another advantage of the 'signature peptide' approach lies in the fact that serial dilutions do not seem to be necessary since the analytical response of the HPLC-MS/MS is linear from the LOQ to the high end mg/mL levels. Nevertheless, the comparison of the results obtained for the same samples analyzed by both methods shows them to be within the order of magnitude. In spite of the differences observed for each individual sample, these results obtained are encouraging, considering that, under physiological conditions, the concentrations of plasma Vtg can vary by several orders of magnitude, ranging from ng/mL to the mg/mL levels (Copeland *et al.*, 1986; Wallaert and Babin, 1994; Bon *et al.*, 1997). However, large scale field studies should be carried out to validate this technique. These would provide sufficient data to perform statistical analyses in order to compare the commercial ELISA kits with the technique presented in this thesis.

Another interesting issue is raised when analyzing the data obtained for the AS plasma samples. Although the ELISA kit is intended for use in RT samples, the antibody

used for Vtg recognition exhibits some cross reactivity to AS Vtg. This situation is evident in the bottom graph of Figure 5.3. However, the concentration values differ by one order of magnitude. To correct this difference, this ELISA kit should have been calibrated using AS Vtg as a Standard. This would have considerably affected the sensitivity of the ELISA kit since, clearly, the affinity of the antibody used is much lower for AS Vtg.

Once more, similar results were obtained from both methods when analyzing most of the AS samples prior to, and after  $17\beta$ -estradiol induction, as shown by the overlapping data at the intersection of both axes (Bottom graph of Figure 5.3). However, the ELISA method was unable to detect some low Vtg levels found in some of the lowest concentrations. Above 10mg/mL measured by the signature peptide approach, a reasonable linear relationship was found between the methods.

Is one method better than the other? Unfortunately, the word 'better' is too broad to obtain a generalized answer. Both methods have advantages and disadvantages. Probably, the specific conditions required for a particular objective will determine whether one method should be chosen over the other.

#### **5.4 Atlantic cod vitellogenin sequence: Similarities to haddock and its implications**

As a result of the experiments performed on AC Vtg, the sequences of some diagnostic tryptic peptides have been elucidated. The database search has shown that AC

Vtg shares a series of common peptides with the two different known Vtg sequences of haddock, a closely related species. These findings suggest that Atlantic cod might also co-express at least two distinct forms of Vtg (Cohen *et al.*, 2005). This is further supported by the finding of the tryptic peptide ions at  $m/z$  1458.74 and 1465.82. These two overlapping variant peptides belonging to haddock Vtg B (FFGQEIGFASIDK) and Vtg A (FLGQEIAFANLDK) have been simultaneously extracted from the trypsin digestion of the Atlantic cod Vtg band in the SDS-PAGE.

For quantification of AC Vtg, a similar HPLC-MS/MS approach can be adopted as described earlier for AS and RT Vtg. Unfortunately, none of the peptides sequenced in this work by MS/MS for AS and RT matched those of AC Vtg. Sequence alignment of RT and haddock Vtg have shown that a few common tryptic peptides are present in both species. However, as described previously, this observation is not sufficient to propose a peptide as a candidate 'signature peptide'. Furthermore, its presence in haddock does not imply its presence in AC Vtg.

Nonetheless, the use of a common peptide found in both AC and haddock Vtg could be used with the same advantages as described for AS and RT Vtg. The fact that these species appear to express two distinct forms of Vtg would further exploit the advantages of the 'signature peptide' approach. The selection of a peptide which is present in both forms would allow quantification of total Vtg in a single assay. For example, the tryptic peptide DLGLAYTEK, sequence by MS/MS in AC Vtg was also found in haddock Vtg. Also, since this peptide is present in both forms of haddock Vtg, it may be a perfect candidate for total Vtg quantification. Alternatively, tryptic peptides like

YESFAVAR or VHVDALLAMR could be used to distinctively quantify one of the Vtg forms.

In this sense, the 'signature peptide' approach would offer a great advantage over immunoassays. Because purification and isolation of these two forms of Vtg would be extremely difficult, production of antibodies from the immunization protocols would be greatly impaired. Moreover, even if these two forms of Vtg were able to be individually purified, developing epitope specific antibodies distinct to each would be more than a challenge.

## **5.5 Application of the 'signature peptide' approach to future studies**

As a result of the work described in this thesis, a new horizon has emerged for different research projects in which measurements of Vtg are involved. It is now possible to develop in a relatively short period of time, a simple, specific and accurate technique for quantification of plasma Vtg. Currently, at the Northwest Atlantic Fisheries Center in Newfoundland, Canada, several projects are taking place, in which measurement of Vtg plays an important role.

A research group led by Dr. Atef Mansour is currently studying the causes of gastric dilation and air sacculitis in RT. This disease is of great concern in aquaculture, since it diminishes the commercial value of the cultured fish. Environmental parameters such as temperature and salinity are suspected to be possible causes of this disease. Early maturation has also been proposed for affecting the onset of this sickness. Plasma Vtg

levels will be used in this work, together with other histological techniques to determine the stage of sexual maturation in these fish. If early maturation proves to be a cause of this disease, aquaculture might make use of this information to determine the most convenient time for animal slaughter in order to optimize the commercial return from the caged stocks.

Another project led by Dr. Mansour is studying the effect of the photoperiod on the maturation of Atlantic cod. Once more, aquaculture is particularly interested in this issue, since sexual maturation stops the growth of fish muscular tissue. Thus, the commercial value of the fish, which is mostly dependant on the size of the fillets, is affected. There is evidence suggesting that increasing the hours of light in the light-darkness cycle would slow down the process of sexual maturation. This is done practically by illuminating the fish with an artificial source of light during the night hours. Again, in this study, a quantitative method for measuring Atlantic cod Vtg will be developed using the same approach as described in this thesis. The sexual maturation of the fish under this controlled photoperiod experiment can then be assessed by the Vtg levels in plasma.

Finally, a research proposal supervised by Dr. Joseph Banoub and Dr. Joanne Morgan has recently been approved and funded to investigate the causes of the decrease of the stocks of Greenland halibut (*Reinhardtius hippoglossoides*), also known as Greenland turbot, a species which has been declared endangered. An alteration in the spawning cycle of this species, probably due to an increase in the average temperature of waters in this region, is believed to be one of the causes for this situation. Once more, plasma Vtg levels will be used to assess the reproductive status of this species. A similar

approach as described in this thesis will be taken to develop a quantitative technique for Greenland halibut Vtg. This project is even more challenging, since the Vtg sequence for this particular species is completely unknown.

Finally, I am completely aware, that in the not too distant future (and probably even sooner than I imagine), even most of the sophisticated and advanced techniques used in this work will also be considered a relic. Nevertheless, the ideas expressed in this thesis should build a bridge between the traditional techniques employed for protein quantification and the virtues offered by the power of chromatography coupled to tandem mass spectrometry.

## **5.6 Summary and recommendations**

In this thesis, proteolytic peptides of RT, AS and AC Vtg have been characterized and sequenced using ESI and MALDI MS and MS/MS. A new method for quantification of AS and RT Vtg has been proposed and developed using HPLC coupled to ESI-MS/MS availing of a characteristic ‘signature peptide’ common to both species. This approach could be considered as an alternative to developing immunoassays, especially in cases when no commercial kits are available for a particular protein. The specificity attained by this technique was the most remarkable feature, although the LOD levels, in the order of  $\mu\text{g/mL}$ , were not as low as expected. A critical step for this procedure is the enzymatic digestion, and precaution should be taken when adapting this technique to other conditions.

## Reference List

- Ackermann, G. E., Schwaiger, J., Negele, R. D., and Fent, K. (2002). Effects of long-term nonylphenol exposure on gonadal development and biomarkers of estrogenicity in juvenile rainbow trout *Oncorhynchus mykiss*. *Aquat. Toxicol.* *60*, 203-221.
- Aebersold, R. and Goodlett, D. R. (2001). Mass spectrometry in proteomics. *Chem. Rev.* *101*, 269-295.
- Ahlborg, U. G., Lipworth, L., Titus-Ernstoff, L., Hsieh, C. C., Hanberg, A., Baron, J., Trichopoulos, D., and Adami, H. O. (1995). Organochlorine compounds in relation to breast cancer, endometrial cancer, and endometriosis: an assessment of the biological and epidemiological evidence. *Crit. Rev. Toxicol.* *25*, 463-531.
- Aida, K., Ngan, P. V., and Hibiya, T. (1973). Physiological studies on gonadal maturation of fishes-I Sexual difference in composition of plasma protein of Ayu in relation to gonadal maturation. *Bull. Jpn. Soc. Sci. Fisher.* *39*, 1091-1106.
- Ait-Aissa, S., Ausseil, O., Palluel, O., Vindimian, E., Garnier-Laplace, J., and Porcher, J. M. (2003). Biomarker responses in juvenile rainbow trout (*Oncorhynchus mykiss*) after single and combined exposure to low doses of cadmium, zinc, PCB77 and 17beta-oestradiol. *Biomarkers* *8*, 491-508.
- Allard, A. S., Gunnarsson, M., and Svenson, A. (2004). Estrogenicity in bile of juvenile rainbow trout as measure of exposure and potential effects of endocrine disruptors. *Environ. Toxicol. Chem.* *23*, 1187-1193.
- Alslev, B., Korsgaard, B., and Bjerregaard, P. (2005). Estrogenicity of butylparaben in rainbow trout *Oncorhynchus mykiss* exposed via food and water. *Aquat. Toxicol.* *72*, 295-304.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403-410.
- Anderson, M. J., Olsen H., Matsumura F. and Hinton D.E. (1996). In vivo modulation of 17-beta-estradiol-induced vitellogenin synthesis and estrogen receptor in rainbow trout (*Oncorhynchus mykiss*). *Toxicol. Appl. Pharmacol.* *137*, 210-218.
- Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., and Pearson, T. W. (2004). Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* *3*, 235-244.

- Anderson, T. A., Levitt, D. G., and Banaszak, L. J. (1998). The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure*. 6, 895-909.
- Arukwe, A., Knudsen, F. R., and Goksoyr, A. (1997). Fish zona radiata (eggshell) protein: a sensitive biomarker for environmental estrogens. *Environ. Health Perspect.* 105, 418-422.
- Arukwe, Augustine and Goksoyr, Anders (1998). Xenobiotics, xenoestrogens and reproduction disturbances in fish. *Sarsia* 83, 225-241.
- Arukwe, A. and Goksoyr, A. (2003a). Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. *Comp. Hepatol.* 2, 1-21.
- Babin, P. J. and Vernier, J. M. (1989). Plasma lipoproteins in fish. *J. Lipid Res.* 30, 467-489.
- Babin, P. J., Deryckere, F., and Gannon, F. (1995). Presence of an extended duplication in the putative low-density-lipoprotein receptor-binding domain of apolipoprotein B. Cloning and characterization of the domain in salmon. *Eur. J. Biochem.* 230, 45-51.
- Bailey, R. E. (1957). The effect of estradiol on serum calcium, phosphorus, and protein of goldfish. *J. Exp. Zool.* 136, 455-469.
- Baker, M. E. (1988a). Invertebrate vitellogenin is homologous to human von Willebrand factor. *Biochem. J.* 256, 1059-1061.
- Baker, M. E. (1988b). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem. J.* 255, 1057-1060.
- Banaszak, L. J. and Seelig, J. (1982). Lipid domains in the crystalline lipovitellin/phosvitin complex: a phosphorus-31 and deuterium nuclear magnetic resonance study. *Biochemistry* 21, 2436-2443.
- Banoub, J., Thibault, P., Mansour, A., Cohen, A., Heeley, D. H., and Jackman, D. (2003). Characterisation of the intact rainbow trout vitellogenin protein and analysis of its derived tryptic and cyanogen bromide peptides by matrix-assisted laser desorption/ionisation time-of-flight-mass spectrometry and electrospray ionisation quadrupole/time-of-flight mass spectrometry. *Eur. J. Mass Spectrom.* 9, 509-524.
- Banoub, J., Cohen, A., Mansour, A., and Thibault, P. (2004). Characterization and de novo sequencing of Atlantic salmon vitellogenin protein by electrospray tandem and matrix-assisted laser desorption/ionization mass spectrometry. *Eur. J. Mass Spectrom.* 10, 121-134.

Barnidge, D. R., Dratz, E. A., Martin, T., Bonilla, L. E., Moran, L. B., and Lindall, A. (2003). Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.* *75*, 445-451.

Barr, J. R., Maggio, V. L., Patterson, D. G., Jr., Cooper, G. R., Henderson, L. O., Turner, W. E., Smith, S. J., Hannon, W. H., Needham, L. L., and Sampson, E. J. (1996). Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* *42*, 1676-1682.

Bemanian, V., Male, R., and Goksoyr, A. (2004). The aryl hydrocarbon receptor-mediated disruption of vitellogenin synthesis in the fish liver: Cross-talk between AHR- and ERalpha-signalling pathways. *Comp Hepatol.* *3*, 1-14.

Biemann, K. (1990). Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* *193*, 455-479.

Bjerregaard, P., Korsgaard, B., Christiansen, L. B., Pedersen, K. L., Christensen, L. J., Pedersen, S. N., and Horn, P. (1998). Monitoring and risk assessment for endocrine disruptors in the aquatic environment: a biomarker approach. *Arch. Toxicol. Suppl.* *20*, 97-107.

Bjerregaard, P., Andersen, D. N., Pedersen, K. L., Pedersen, S. N., and Korsgaard, B. (2003). Estrogenic effect of propylparaben (propylhydroxybenzoate) in rainbow trout *Oncorhynchus mykiss* after exposure via food and water. *Comp Biochem. Physiol C. Toxicol. Pharmacol.* *136*, 309-317.

Bjornsson, B. T., Haux, C., Forlin, L., and Deftos, L. J. (1986). The involvement of calcitonin in the reproductive physiology of the rainbow trout. *J. Endocrinol.* *108*, 17-23.

Bogdanov, B. and Smith, R. D. (2005). Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* *24*, 168-200.

Bon, E., Barbe, U., Nunez, Rodriguez J., Cuisset, B., Pelissero, C., Sumpster, J. P., and Le Menn, F. (1997). Plasma vitellogenin levels during the annual reproductive cycle of the female rainbow trout (*Oncorhynchus mykiss*): establishment and validation of an ELISA. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* *117*, 75-84.

Boon, J. P., van Zanden, J. J., Lewis, W. E., Zegers, B. N., Goksoyr, A., and Arukwe, A. (2002). The expression of CYP1A, vitellogenin and zona radiata proteins in Atlantic salmon (*Salmo salar*) after oral dosing with two commercial PBDE flame retardant mixtures: absence of short-term responses. *Mar. Environ. Res.* *54*, 719-724.

Bowman, C. J. and Denslow, N. D. (1999). Development and Validation of a Species- and Gene-Specific Molecular Biomarker: Vitellogenin mRNA in Largemouth Bass (*Micropterus salmoides*). *Ecotoxicology* *88*, 399-416.

Brennan, M. D., Weiner, A. J., Goralski, T. J., and Mahowald, A. P. (1982). The follicle cells are a major site of vitellogenin synthesis in *Drosophila melanogaster*. *Dev. Biol.* *89*, 225-236.

Brion, F., Rogerieux, F., Noury, P., Migeon, B., Flammarion, P., Thybaud, E., and Porcher, J. M. (2000). Two-step purification method of vitellogenin from three teleost fish species: rainbow trout (*Oncorhynchus mykiss*), gudgeon (*Gobio gobio*) and chub (*Leuciscus cephalus*). *J. Chromatogr. B Biomed. Sci. Appl.* *737*, 3-12.

Brooks, J. M. and Wessel, G. M. (2004). The major yolk protein of sea urchins is endocytosed by a dynamin-dependent mechanism. *Biol. Reprod.* *71*, 705-713.

Brown, M. A., Carne, A., and Chambers, G. K. (1997a). Purification, partial characterization and peptide sequences of vitellogenin from a reptile, the tuatara (*Sphenodon punctatus*). *Comp Biochem. Physiol B Biochem. Mol. Biol.* *117*, 159-168.

Brown, M. S., Herz, J., and Goldstein, J. L. (1997b). LDL-receptor structure. Calcium cages, acid baths and recycling receptors. *Nature* *388*, 629-630.

Bruins, A. P. (1991). Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrom. Rev.* *10*, 53-77.

Buckner, G. D., Martin, J. H., and Hull, F. E. (1930). The distribution of blood calcium in the circulation of laying hens. *Am. J. Physiol.* *93*, 86-89.

Buisine, N., Trichet, V., and Wolff, J. (2002). Complex evolution of vitellogenin genes in salmonid fishes. *Mol. Genet. Genomics* *268*, 535-542.

Campbell, C. M. and Idler, D. R. (1980). Characterization of an estradiol-induced protein from rainbow trout serum as vitellogenin by the composition and radioimmunological cross reactivity to ovarian yolk fractions. *Biol. Reprod.* *22*, 605-617.

Casin, S., Fossi, M. C., Cavallaro, K., Marsili, L., and Lorenzani, J. (2002). The use of porphyrins as a non-destructive biomarker of exposure to contaminants in two sea lion populations. *Mar. Environ. Res.* *54*, 769-773.

Cejka, J., Vodrazka, Z., and Salak, J. (1968). Carbamylation of globin in electrophoresis and chromatography in the presence of urea. *Biochim. Biophys. Acta* *154*, 589-591.

Celius, T. and Walther, B. T. (1998). Oogenesis in Atlantic salmon (*Salmo salar* L.) occurs by zonagenesis preceding vitellogenesis in vivo and in vitro. *J. Endocrinol.* *158*, 259-266.

Charles, E. and Hogben, L. (1933). The serum calcium and magnesium level in the ovarian cycle of the laying hen. *J. Exp. Physiol.* *23*, 343-349.

- Cheek, A. O., Brouwer, T. H., Carroll, S., Manning, S., McLachlan, J. A., and Brouwer, M. (2001). Experimental evaluation of vitellogenin as a predictive biomarker for reproductive disruption. *Environ. Health Perspect.* *109*, 681-690.
- Chen, J. S., Sappington, T. W., and Raikhel, A. S. (1997). Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J. Mol. Evol.* *44*, 440-451.
- Chen, T. T., Howard, D. A., Agellon, L. B., Lin, C. M., and Davies, S. L. (1989). Estrogen-controlled gene expression induction of two estrogen-responsive genes in the liver of rainbow trout *salmo-gairdneri*. *Physiol. Zool.* *62*, 25-37.
- Chernushevich, I. V. (2000). Duty cycle improvement for a quadrupole-time-of-flight mass spectrometer and its use for precursor ion scans. *Eur. J. Mass Spectrom.* *6*, 471-479.
- Chernushevich, I. V., Loboda, A. V., and Thomson, B. A. (2001). An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* *36*, 849-865.
- Chrestien, J. P. The search for cod, a delicacy for meatless days. A fishing expedition on the Saint-Andre (1774). <http://www.civilization.ca/hist/lifelines/liron01e.html> . 2005.
- Christiansen, L. B., Pedersen, K. L., Korsgaard, B., and Bjerregaard, P. (1998). Estrogenicity of Xenobiotics in Rainbow Trout (*Oncorhynchus mykiss*) using in vivo Synthesis of Vitellogenin as a Biomarker. *Mar. Environ. Res.* *46*, 137-140.
- Cohen, A. M., Mansour, A. A., and Banoub, J. H. (2005). 'De novo' sequencing of Atlantic cod vitellogenin tryptic peptides by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry: similarities with haddock vitellogenin. *Rapid Commun. Mass Spectrom.* *19*, 2454-2460.
- Cohen, A. M., Mansour, A., and Banoub, J. H. (2006). Absolute quantification of Atlantic salmon and Rainbow trout vitellogenin by the 'signature peptide' approach using electrospray ionization QqToF tandem mass spectrometry. *J. Mass Spectrom.* *41*, 646-658.
- Colborn, T. and Clement, T. (1992). Chemically-induced alterations in sexual and functional development-- the wildlife/human connection. ( Princeton, N.J.: Princeton Scientific Pub. Co.).
- Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolic, L., and Smith, R. D. (2000). Utility of accurate mass tags for proteome-wide protein identification. *Anal. Chem.* *72*, 3349-3354.
- Cooper, C. A., Gasteiger, E., and Packer, N. H. (2001). GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics.* *1*, 340-349.

- Copeland, P. A., Sumpter, J. P., Walker, T. K., and Croft, M. (1986). Vitellogenin levels in male and female rainbow trout (*Salmo gairdneri* Richardson) at various stages of the reproductive cycle. *Comp. Biochem. Physiol. B* 83, 487-493.
- Covens, M., Covens, L., Ollevier, F., and De Loof, A. (1987). A comparative study of some properties of vitellogenin (Vg) and yolk proteins in a number of freshwater and marine teleost fishes. *Comp. Biochem. Physiol. B* 88, 75-80.
- Creighton, T. E. and Charles, I. G. (1987). Sequences of the genes and polypeptide precursors for two bovine protease inhibitors. *J. Mol. Biol.* 194, 11-22.
- Crowther, J. R. (2000). *The Elisa Guidebook*. (Ney Jersey: Humana Press).
- Damstra, T., Barlow, S., Bergman, A., Kavlock, R., and Van Der Kraak, G. (2002a). Global Assessment of the State-of-the-Science of Endocrine Disruptors. Programme on Chemical Safety (IPCS) WHO/PCS/EDC/02.s. World Health Organization, Geneva, Switzerland..
- Damstra, T., Page, S. W., Herrman, J. L., and Meredith, T. (2002b). Persistent organic pollutants: potential health effects? *J. Epidemiol. Community Health* 56, 824-825.
- Darain, F., Park, D. S., Park, J. S., and Shim, Y. B. (2004). Development of an immunosensor for the detection of vitellogenin using impedance spectroscopy. *Biosens. Bioelectron.* 19, 1245-1252.
- Davail, B., Pakdel, F., Bujo, H., Perazzolo, L. M., Waclawek, M., Schneider, W. J., and Le Menn, F. (1998). Evolution of oogenesis: the receptor for vitellogenin from the rainbow trout. *J. Lipid Res.* 39, 1929-1937.
- Dawson, P. H. (1986). Quadrupole mass analyzers: Performance, design and some recent applications. *Mass Spectrom. Rev.* 5, 1-37.
- De Vlaming, V. L., Wiley, H. S., Delahunty, G., and Wallace, R. A. (1980). Goldfish (*Carassius auratus*) vitellogenin: induction, isolation, properties and relationship to yolk proteins. *Comp. Biochem. Physiol.* 67B, 613-623.
- Denslow, N. D., Chow, M. C., Kroll, K. J., and Green, L. (1999). Vitellogenin as a Biomarker of Exposure for Estrogen or Estrogen Mimics. *Ecotoxicology* 8, 385-398.
- DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2005). Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome. Res.* 4, 377-386.

- Ding, J. L., Lim, E. H., Li, H. F., Kumar, J. K., Lee, S. L., and Lam, T. J. (2004). Expression of recombinant vitellogenin in the yeast *Pichia pastoris*. *Biotechnol. Bioeng.* *85*, 330-339.
- Douglas, D. J., Frank, A. J., and Mao, D. (2005). Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* *24*, 1-29.
- Elias, D. A., Monroe, M. E., Marshall, M. J., Romine, M. F., Belieav, A. S., Fredrickson, J. K., Anderson, G. A., Smith, R. D., and Lipton, M. S. (2005). Global detection and characterization of hypothetical proteins in *Shewanella oneidensis* MR-1 using LC-MS based proteomics. *Proteomics* *5*, 3120-3130.
- Eng, Jimmy K., McCormack, Ashley L., and Yates, I. I. I. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* *5*, 976-989.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1990). Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* *9*, 37-70.
- Folmar, L. C., Denslow, N. D., Wallace, R. A., Lafleur, G., Gross, T. S., Bonomelli, S., and Sullivan, C. V. (1995). A highly conserved N-terminal sequence for teleost vitellogenin with potential value to the biochemistry, molecular biology and pathology of vitellogenesis. *J. Fish Biol.* *46*, 255-263.
- Folmar, L. C., Hemmer, M. J., Denslow, N. D., Kroll, K., Chen, J., Cheek, A., Richman, H., Meredith, H., and Grau, E. G. (2002). A comparison of the estrogenic potencies of estradiol, ethynylestradiol, diethylstilbestrol, nonylphenol and methoxychlor in vivo and in vitro. *Aquat. Toxicol.* *60*, 101-110.
- Fremont, L. and Riazi, A. (1988). Biochemical analysis of vitellogenin from rainbow trout (*Salmo gairdneri*): fatty acid composition of phospholipids. *Reprod. Nutr. Dev.* *28*, 939-952.
- Garcia-Reyero, N., Raldua, D., Quiros, L., Llaveria, G., Cerda, J., Barcelo, D., Grimalt, J. O., and Pina, B. (2004). Use of vitellogenin mRNA as a biomarker for endocrine disruption in feral and cultured fish. *Anal. Bioanal. Chem.* *378*, 670-675.
- Gattiker, A., Gasteiger, E., and Bairoch, A. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics* *1*, 107-108.
- Geng, M., Ji, J., and Regnier, F. E. (2000). Signature-peptide approach to detecting proteins in complex mixtures. *J. Chromatogr. A* *870*, 295-313.
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A* *100*, 6940-6945.

- Glückmann, M. and Karas, M. (1999). The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization. *J. Mass Spectrom.* *34*, 467-477.
- Godovac-Zimmermann, J. and Brown, L. R. (2001). Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* *20*, 1-57.
- Goksøyr, A., Arukwe, A., Larsson, J., Cajaraville, MP, Hauser, L, Nilsen, BM, Lowe, D., and Matthiessen, P (2003). Links between the cellular and molecular response to pollution and the impact on reproduction and fecundity including the influence of endocrine disrupters. In *Impacts of Marine Xenobiotics on European Commercial Fish-Molecular Effects and Population Responses*, A. Lawrence, ed. (London, UK: Caldwell Publishers).
- Gottlieb, T. A. and Wallace, R. A. (1981). Intracellular phosphorylation of vitellogenin in the liver of estrogen- stimulated *Xenopus laevis*. *J. Biol. Chem.* *256*, 4116-4123.
- Gottlieb, T. A. and Wallace, R. A. (1982). Intracellular glycosylation of vitellogenin in the liver of estrogen- stimulated *Xenopus laevis*. *J. Biol. Chem.* *257*, 95-103.
- Goulas, A., Triplett, E. L., and Taborsky, G. (1996). Isolation and characterization of a vitellogenin cDNA from rainbow trout (*Oncorhynchus mykiss*) and the complete sequence of a phosvitin coding segment. *DNA Cell Biol.* *15*, 605-616.
- Grandori, R. (2003). Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry. *J Mass Spectrom.* *38*, 11-15.
- Griffin, T. J. and Aebersold, R. (2001). Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* *276*, 45497-45500.
- Griffiths, W. J., Jonsson, A. P., Liu, S., Rai, D. K., and Wang, Y. (2001). Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* *355*, 545-561.
- Guilhaus, M. (1995). Principles and instrumentation in time-of-flight mass spectrometry. *J. Mass Spectrom.* *30*, 1519-1532.
- Gygi, S. P. and Aebersold, R. (1999). Absolute quantitation of 2-D protein spots. *Methods Mol. Biol.* *112*, 417-421.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* *17*, 994-999.
- Gygi, S. P. and Aebersold, R. (2000). Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* *4*, 489-494.

- Gygi, S. P., Rist, B., and Aebersold, R. (2000). Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.* *11*, 396-401.
- Gygi, S. P., Rist, B., Griffin, T. J., Eng, J., and Aebersold, R. (2002). Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.* *1*, 47-54.
- Hagedorn, H. H. and Kunkel, J. G. (1979). Vitellogenin and vitellin in insects. *Ann. Rev. Entomol.* *24*, 475-505.
- Hara, A. (1976). Iron-binding activity of female-specific serum proteins of rainbow trout (*Salmo gairdneri*) and chum salmon (*Oncorhynchus keta*). *Biochim. Biophys. Acta* *427*, 549-557.
- Hara, A. and Hirai, H. (1978). Comparative studies on immunochemical properties of female-specific serum protein and egg yolk proteins in rainbow trout (*Salmo gairdneri*). *Comp. Biochem. Physiol. B* *59*, 339-343.
- Hara, A., Takano, K., and Hirai, H. (1983). Immunochemical identification of female specific serum protein, vitellogenin, in the medaka, *Oryzias latipes* (Teleosts). *Comp. Biochem. Physiol.* *76A*, 135-141.
- Hara, A., Sullivan, C. V., and Dickhoff, W. W. (1993). Isolation and some characterization of vitellogenin and its related egg yolk proteins from coho salmon (*Oncorhynchus kisutch*). *Zool. Sci.* *10*, 245-256.
- Hartling, R. C., Pereira, J. J., and Kunkel, J. G. (1997). Characterization of a heat-stable fraction of lipovitellin and development of an immunoassay for vitellogenin and yolk protein in winter flounder (*Pleuronectes americanus*). *J. Exp. Zool.* *278*, 156-166.
- Heppell, S. A., Denslow, N. D., Folmar, L. C., and Sullivan, C. V. (1995). Universal assay of vitellogenin as a biomarker for environmental estrogens. *Environ. Health Perspect.* *103 Suppl. 7*, 9-15.
- Hermanson, M. A., Ericsson, L. H., Neurath, H., and Walsh, K. A. (1973). Determination of the amino acid sequence of porcine trypsin by sequenator analysis. *Biochemistry* *12*, 3146-3153.
- Hernandez, P., Muller, M., and Appel, R. D. (2006). Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass Spectrom. Rev.* *25*, 235-254.
- Hess, A. F., Bills, C. E., Weinstock, M., and Rivkin, H. (1928). Differences in calcium level of the blood between male and female cod. *Proceedings of the Soc. Exper. Biol. Med.* *25*, 349-350.

- Hillenkamp, F. and Karas, M. (1990). Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol.* 193, 280-295.
- Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 63, 1193A-1203A.
- Hiramatsu, N., Ichikawa, N., Fukada, H., Fujita, T., Sullivan, C. V., and Hara, A. (2002). Identification and characterization of proteases involved in specific proteolysis of vitellogenin and yolk proteins in salmonids. *J. Exp. Zool.* 292, 11-25.
- Hori, S. H., Kodama, T., and Tanahashi, K. (1979). Induction of vitellogenin synthesis in goldfish by massive doses of androgens. *Gen. Comp. Endocrinol.* 37, 306-320.
- Hughes, J. S., Titus, R. W., and Smith, B. L. (1927). The increase in the calcium. of hen's blood accompanying egg production. *Science* 65, 264.
- Hulo, N., Sigrist, C. J., Le, Saux, V, Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Res.* 32, D134-D137.
- Hyllner, S. J., Silversand, C., and Haux, C. (1994). Formation of the vitelline envelope precedes the active uptake of vitellogenin during oocyte development in the rainbow trout, *Oncorhynchus mykiss*. *Mol. Reprod. Dev.* 39, 166-175.
- Idler, D. R., Hwang, S. J., and Crim, L. W. (1979). Quantification of vitellogenin in Atlantic salmon (*Salmo salar*) plasma by radioimmunoassay. *J. Fish. Res. Board Can.* 36, 574-578.
- Jeanville, P. M., Woods, J. H., Baird, T. J., III, and Estape, E. S. (2000). Direct determination of ecgonine methyl ester and cocaine in rat plasma, utilizing on-line sample extraction coupled with rapid chromatography/quadrupole orthogonal acceleration time-of-flight detection. *J. Pharm. Biomed. Anal.* 23, 897-907.
- Jennie, Lill (2003). Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom. Rev.* 22, 182-194.
- Jeppsson, J. O., Kobold, U., Barr, J., Finke, A., Hoelzel, W., Hoshino, T., Miedema, K., Mosca, A., Mauri, P., Paroni, R., Thienpont, L., Umemoto, M., and Weykamp, C. (2002). Approved IFCC reference method for the measurement of HbA1c in human blood. *Clin. Chem. Lab. Med.* 40, 78-89.
- Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., and Regnier, F. (2000). Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J. Chromatogr. B Biomed. Sci. Appl.* 745, 197-210.

- Jobling, S., Casey, D., Rogers-Gray, T., Oehlmann, J., Schulte-Oehlmann, U., Pawlowski, S., Baunbeck, T., Turner, A. P., and Tyler, C. R. (2004). Comparative responses of molluscs and fish to environmental estrogens and an estrogenic effluent. *Aquat. Toxicol.* *66*, 207-222.
- Johnson, A. C., Aerni, H. R., Gerritsen, A., Gibert, M., Giger, W., Hylland, K., Jurgens, M., Nakari, T., Pickering, A., Suter, M. J., Svenson, A., and Wettstein, F. E. (2005). Comparing steroid estrogen, and nonylphenol content across a range of European sewage plants with different treatment and management practices. *Water Res.* *39*, 47-58.
- Jonsson, A. P., Aissouni, Y., Palmberg, C., Percipalle, P., Nordling, E., Daneholt, B., Jornvall, H., and Bergman, T. (2001). Recovery of gel-separated proteins for in-solution digestion and mass spectrometry. *Anal. Chem.* *73*, 5370-5377.
- Jowett, T. and Postlethwait, J. H. (1980). The Regulation of yolk polypeptide synthesis in *Drosophila* ovaries and fat body by 20-hydroxyecdysone and a juvenile hormone analog. *Dev. Biol.* *80*, 225-234.
- Kanungo, J., Petrino, T. R., and Wallace, R. A. (1990). Oogenesis in *Fundulus heteroclitus*. VI. Establishment and verification of conditions for vitellogenin incorporation by oocytes in vitro. *J. Exp. Zool.* *254*, 313-321.
- Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* *60*, 2299-2301.
- Karas, M., Bahr, U., and Giebmann, U. (1991). Matrix-assisted laser desorption ionization mass spectrometry. *Mass Spectrom. Rev.* *10*, 335-357.
- Karas, M., Gluckmann, M., and Schafer, J. (2000). Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* *35*, 1-12.
- Khalaila, I., Peter-Katalinic, J., Tsang, C., Radcliffe, C. M., Aflalo, E. D., Harvey, D. J., Dwek, R. A., Rudd, P. M., and Sagi, A. (2004). Structural characterization of the N-glycan moiety and site of glycosylation in vitellogenin from the decapod crustacean *Cherax quadricarinatus*. *Glycobiology* *14*, 767-774.
- Kime, D. E. (1995). The effects of pollution on reproduction in fish. *Rev. Fish Biol. Fish.* *5*, 52-95.
- King, J. L. and Jukes, T. H. (1969). Non-Darwinian evolution. *Science* *164*, 788-798.
- Kinter, M. and Sherman, N. E. (2000). Protein sequencing and identification using tandem mass spectrometry., D.M.Desiderio and N.M.Nibbering, eds. (New York, NY: John Wiley & Sons).

Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2005). The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35, 265-273.

Knott, T. J., Pease, R. J., Powell, L. M., Wallis, S. C., Rall, S. C., Jr., Innerarity, T. L., Blackhart, B., Taylor, W. H., Marcel, Y., Milne, R., and . (1986). Complete protein sequence and identification of structural domains of human apolipoprotein B. *Nature* 323, 734-738.

Knudsen, F. R., Schou, A. E., Wiborg, M. L., Mona, E., Tollefsen, K., Stenersen, J., and Sumpter, J. P. (1997). Increase of plasma vitellogenin concentration in rainbow trout (*Oncorhynchus mykiss*) exposed to effluents from oil refinery treatment works and municipal sewage. *Bull. Environ. Contam. Toxicol.* 59, 802-806.

Köster, C., Kahr, M. S., Castoro, J. A., and Wilkins, C. L. (1992). Fourier transform mass spectrometry. *Mass Spectrom. Rev.* 11, 495-512.

Kunkel, J. G. and Pan, M. L. (1976). Selectivity of yolk protein uptake: comparison of vitellogenins of two insects. *J. Insect Physiol.* 22, 809-818.

Kunkel, J. G. and Nordin, J. H. (1986). Yolk Proteins. In *Comprehensive Insect Physiology Biochemistry Pharmacology*, L.I.Gilbert and T.A.Miller, eds. (London, UK: Pergamon Press), pp. 83-111.

Kwon, H. C., Hayashi, S., and Mugiya, Y. (1993). Vitellogenin induction by estradiol-17-beta in primary hepatocyte culture in the rainbow trout, *Oncorhynchus mykiss*. *Comp. Biochem. Physiol. B* 104, 381-386.

Kwon, J. Y., Prat F., Randall C. and Tyler, C. R. (2001). Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout (*Oncorhynchus mykiss*). *Biol. Reprod.* 65, 1701-1709.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lam, T.J. (1983). Environmental influences on gonadal activity in fish. W.S.Hoar and D.J.Randall, eds. (Ney York, NY: Academic Press), pp. 65-116.

Larkin, P., Knoebel, I., and Denslow, N. D. (2003). Differential gene expression analysis in fish exposed to endocrine disrupting compounds. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 136, 149-161.

Laskowski, M. (1935). Über die Phosphoverbindungen. im. Blutplasma der Leg-. ehene. *Biochem. Zeitschr.* 275, 293-300.

- Latonnelle, K., Fostier, A., Le Menn, F., and Bennetau-Pelissero, C. (2002a). Binding affinities of hepatic nuclear estrogen receptors for phytoestrogens in rainbow trout (*Oncorhynchus mykiss*) and Siberian sturgeon (*Acipenser baeri*). *Gen. Comp. Endocrinol.* *129*, 69-79.
- Latonnelle, K., Le Menn, F., Kaushik, S. J., and Bennetau-Pelissero, C. (2002b). Effects of dietary phytoestrogens in vivo and in vitro in rainbow trout and Siberian sturgeon: interests and limits of the in vitro studies of interspecies differences. *Gen. Comp. Endocrinol.* *126*, 39-51.
- Le Bail, P. Y. and Breton, B. (1981). Rapid determination of the sex of puberal salmonid fish by a technique of immunoagglutination. *Aquaculture* *22*, 367-375.
- Le Guellec, K., Lawless, K., Valotaire, Y., Kress, M., and Tenniswood, M. (1988). Vitellogenin gene expression in male rainbow trout (*Salmo gairdneri*). *Gen. Comp. Endocrinol.* *71*, 359-371.
- Le Menn, F., Davali, B., Pelissero, C., Ndiaye, P., Bon, E., Perazzolo, L., and Rodriguez, J. (2000). A new approach to fish vitellogenesis. In *Proceedings of the 6th International Symposium on the Reproductive Physiology of Fish: 1999 July 4–9*, B.Norberg, O.S.Kjesbu, G.L.Taranger, E.Andersson, and S.O.Stefansson, eds. (Bergen, Norway: Bergen. John Grieg A/S), pp. 281-284.
- Li, A., Sadasivam, M., and Ding, J. L. (2003). Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. *J. Biol. Chem.* *278*, 2799-2806.
- Lindholst, C., Pedersen, K. L., and Pedersen, S. N. (2000). Estrogenic response of bisphenol A in rainbow trout (*Oncorhynchus mykiss*). *Aquatic. Toxicol.* *48*, 87-94.
- Lindholst, C., Pedersen, S. N., and Bjerregaard, P. (2001). Uptake, metabolism and excretion of bisphenol A in the rainbow trout (*Oncorhynchus mykiss*). *Aquat. Toxicol.* *55*, 75-84.
- Lindholst, C., Wynne, P. M., Marriott, P., Pedersen, S. N., and Bjerregaard, P. (2003). Metabolism of bisphenol A in zebrafish (*Danio rerio*) and rainbow trout (*Oncorhynchus mykiss*) in relation to estrogenic response. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* *135*, 169-177.
- Lippincott, J. and Apostol, I. (1999). Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. *Anal. Biochem.* *267*, 57-64.
- Lubzens, E., Lissauer, L., Levavi-Sivan, B., Avarre, J. C., and Sammar, M. (2003). Carotenoid and retinoid transport to fish oocytes and eggs: what is the role of retinol binding protein? *Mol. Aspects Med.* *24*, 441-457.

- Luo, Y., Rudy, J. A., Uboh, C. E., Soma, L. R., Guan, F., Enright, J. M., and Tsang, D. S. (2004). Quantification and confirmation of flunixin in equine plasma by liquid chromatography-quadrupole time-of-flight tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* *801*, 173-184.
- Mackay, M. E. and Lazier, C. B. (1993). Estrogen responsiveness of vitellogenin gene expression in rainbow trout (*Oncorhynchus mykiss*) kept at different temperatures. *Gen. Comp. Endocrinol.* *89*, 255-266.
- Madsen, S. S., Skovbolling, S., Nielsen, C., and Korsgaard, B. (2004). 17-Beta estradiol and 4-nonylphenol delay smolt development and downstream migration in Atlantic salmon, *Salmo salar*. *Aquat. Toxicol.* *68*, 109-120.
- Makler, M. T. (1980). Carbamylation of human HbF on carboxy methyl cellulose column in 8 M urea pH 6.7. *Biochem. Biophys. Res. Commun.* *94*, 967-973.
- Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., Kochl, S., Ritchie, P. J., Bradbury, P., Hussain, F. S., Amey, J., Vanloo, B., Rosseneu, M., Infante, R., Hancock, J. M., Levitt, D. G., Banaszak, L. J., Scott, J., and Shoulders, C. C. (1999). The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J. Mol. Biol.* *285*, 391-408.
- March R.E. (2000). Quadrupole ion trap mass spectrometer. In *Encyclopedia of Analytical Chemistry*, Robert A. Meyers, ed. (Chichester, UK: John Wiley & Sons Ltd), pp. 1-24.
- Marshall A.G., Hendrickson C.L., and Jackson G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* *17*, 1-35.
- Marvin, L. F., Delatour, T., Tavazzi, I., Fay, L. B., Cupp, C., and Guy, P. A. (2003). Quantification of o,o'-dityrosine, o-nitrotyrosine, and o-tyrosine in cat urine samples by LC/ electrospray ionization-MS/MS using isotope dilution. *Anal. Chem.* *75*, 261-267.
- Masselon, C., Anderson, G. A., Harkewicz, R., Bruce, J. E., Pasa-Tolic, L., and Smith, R. D. (2000). Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal. Chem.* *72*, 1918-1924.
- Matsubara, T., Nagaie, M., Ohkubo, N., Andoh, T., Sawaguchi, S., Hiramatsu, N., Sullivan, C. V., and Hara, A. (2003). Multiple vitellogenins and their unique roles in marine teleosts. *Fish Physiol. Biochem.* *28*, 295-299.
- Mayadas, T. N. and Wagner, D. D. (1992). Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc. Natl. Acad. Sci USA* *89*, 3531-3535.

- McClain, J. S., Oris, J. T., Burton, G. A., Jr., and Lattier, D. (2003). Laboratory and field validation of multiple molecular biomarkers of contaminant exposure in rainbow trout (*Oncorhynchus mykiss*). *Environ. Toxicol. Chem.* *22*, 361-370.
- McDonald, M. R. and Riddle, O. (1945). The effect of reproduction and estrogen administration on the partition of calcium, phosphorus, and nitrogen in pigeon plasma. *J. Biol. Chem.* *159*, 445-464.
- McGinnis, S. and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* *32*, W20-W25.
- Meininger, T., Raag, R., Roderick, S., and Banaszak, L. J. (1984). Preparation of single crystals of a yolk lipoprotein. *J. Mol. Biol.* *179*, 759-764.
- Mellanen, P., Petanen, T., Lehtimaki, J., Makela, S., Bylund, G., Holmbom, B., Mannila, E., Oikari, A., and Santti, R. (1996). Wood-derived estrogens: studies in vitro with breast cancer cell lines and in vivo in trout. *Toxicol. Appl. Pharmacol.* *136*, 381-388.
- Meucci, V. and Arukwe, A. (2005). Detection of vitellogenin and zona radiata protein expressions in surface mucus of immature juvenile Atlantic salmon (*Salmo salar*) exposed to waterborne nonylphenol. *Aquat. Toxicol.* *73*, 1-10.
- Michalet, S., Favreau, P., and Stocklin, R. (2003). Profiling and in vivo quantification of proteins by high resolution mass spectrometry: the example of goserelin, an analogue of luteinizing hormone-releasing hormone. *Clin. Chem. Lab. Med.* *41*, 1589-1598.
- Miescher, F. (1897). *Histochemische und physiologische Arbeiten*. Leipzig: Vogel.
- Mills, L. J. and Chichester, C. (2005). Review of evidence: are endocrine-disrupting chemicals in the aquatic environment impacting fish populations? *Sci. Total Environ.* *343*, 1-34.
- Milne, R., Theolis, R., Jr., Maurice, R., Pease, R. J., Weech, P. K., Rassart, E., Fruchart, J. C., Scott, J., and Marcel, Y. L. (1989). The use of monoclonal antibodies to localize the low density lipoprotein receptor-binding domain of apolipoprotein B. *J. Biol. Chem.* *264*, 19754-19760.
- Mommsen, T. P. and Walsh, P. J. (1988). Vitellogenesis and oocyte assembly. In *The physiology of developing fish: Part A: Eggs and larvae*, W.S.Hoar, D.J.Randall, and E.M.Donaldsson, eds. (New York, NY: Academic press), pp. 347-406.
- Mouchel, N., Trichet, V., Betz, A., Le Pennec, J. P., and Wolff, J. (1996). Characterization of vitellogenin from rainbow trout (*Oncorhynchus mykiss*). *Gene* *174*, 59-64.

- Muddiman, D. C., Gusev, A. I., and Hercules, D. M. (1995). Application of secondary ion and matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for the quantitative analysis of biological molecules. *Mass Spectrom. Rev.* *14*, 383-429.
- Nagler, J. J., Ruby, S. M., Idler, D. R., and So, Y. P. (1987). Serum phosphoprotein phosphorus and calcium levels as reproductive indicators of vitellogenin in highly vitellogenic mature female and estradiol-injected immature rainbow trout *salmo-gairdneri*. *Can. J. Zool.* *65*, 2421-2425.
- Nakari, T. (2004). Estrogenicity of municipal effluents assessed in vivo and in vitro. *Environ. Toxicol.* *19*, 207-215.
- Nelson, J. S. (1984). *Fishes of the World*. (New York, NY: John Wiley & Sons).
- Ng, S. S. and Dain, J. A. (1976). The natural occurrence of sialic acids. In *Biological roles of sialic acid*, A. Rosenberg and C. Schengrund, eds. (New York, NY: Plenum Press), pp. 59-102.
- Ng, T. B. and Idler, D. R. (1983). Yolk formation and differentiation in teleost fishes. In *Fish physiology*, W.S. Hoar and J.L. Specker, eds. (New York, NY: Academic Press), pp. 373-404.
- Nilsen, B. M., Berg, K., Eidem, J. K., Kristiansen, S. I., Brion, F., Porcher, J. M., and Goksoyr, A. (2004). Development of quantitative vitellogenin-ELISAs for fish test species used in endocrine disruptor screening. *Anal. Bioanal. Chem.* *378*, 621-633.
- Norbeck, A. D., Monroe, M. E., Adkins, J. N., Anderson, K. K., Daly, D. S., and Smith, R. D. (2005). The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J. Am. Soc. Mass Spectrom.* *16*, 1239-1249.
- Norberg, B. and Haux, C. (1985). Induction, isolation and a characterization of the lipid content of plasma vitellogenin from two *Salmo* species: rainbow trout (*Salmo gairdneri*) and sea trout (*Salmo trutta*). *Comp. Biochem. Physiol B* *81*, 869-876.
- Norberg, B. and Haux, C. (1988). An homologous radioimmunoassay for brown trout (*Salmo trutta*) vitellogenin. *Fish Physiol. Biochem.* *5*, 59-68.
- Okoumassoun, L. E., Averill-Bates, D., Gagne, F., Marion, M., and Denizeau, F. (2002). Assessing the estrogenic potential of organochlorine pesticides in primary cultures of male rainbow trout (*Oncorhynchus mykiss*) hepatocytes using vitellogenin as a biomarker. *Toxicology* *178*, 193-207.
- Olsson, P. E., Kling, P., Petterson, C., and Silversand, C. (1995). Interaction of cadmium and oestradiol-17 beta on metallothionein and vitellogenin synthesis in rainbow trout (*Oncorhynchus mykiss*). *Biochem. J.* *307 (Pt 1)*, 197-203.

- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* *1*, 376-386.
- Palmer, B.D., Huth L.K., Pioto D.L. and Selcer K.W. (1998). Vitellogenin as a biomarker for xenobiotic estrogens in an amphibian model system. *Environ. Toxicol. Chem.* *17*, 30-36.
- Pan, M. L., Bell, W. J., and Telfer, W. H. (1969). Vitellogenic blood protein synthesis by insect fat body. *Science* *165*, 393-394.
- Pasa-Tolic, L., Masselon, C., Barry, R. C., Shen, Y., and Smith, R. D. (2004). Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* *37*, 621-33, 636.
- Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* *183*, 63-98.
- Pearson, W. R. (1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* *11*, 635-650.
- Pedersen, S. N., Christiansen, L. B., Pedersen, K. L., Korsgaard, B., and Bjerregaard, P. (1999). In vivo estrogenic activity of branched and linear alkylphenols in rainbow trout (*Oncorhynchus mykiss*). *Sci. Total Environ.* *233*, 89-96.
- Pedersen, K. L., Pedersen, S. N., Christiansen, L. B., Korsgaard, B., and Bjerregaard, P. (2000). The preservatives ethyl-, propyl- and butylparaben are oestrogenic in an in vivo fish assay. *Pharmacol. Toxicol.* *86*, 110-113.
- Pedersen, K. H., Pedersen, S. N., Pedersen, K. L., Korsgaard, B., and Bjerregaard, P. (2003). Estrogenic effect of dietary 4-tert-octylphenol in rainbow trout (*Oncorhynchus mykiss*). *Aquat. Toxicol.* *62*, 295-303.
- Peng, J. and Gygi, S. P. (2001). Proteomics: the move to mixtures. *J. Mass Spectrom.* *36*, 1083-1091.
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* *20*, 3551-3567.
- Qian, W. J., Monroe, M. E., Liu, T., Jacobs, J. M., Anderson, G. A., Shen, Y., Moore, R. J., Anderson, D. J., Zhang, R., Calvano, S. E., Lowry, S. F., Xiao, W., Moldawer, L. L., Davis, R. W., Tompkins, R. G., Camp, D. G., and Smith, R. D. (2005). Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. *Mol. Cell Proteomics* *4*, 700-709.

- Radice, S., Ferraris, M., Marabini, L., and Chiesara, E. (2002). Estrogenic activity of procymidone in primary cultured rainbow trout hepatocytes (*Oncorhynchus mykiss*). *Toxicol. In Vitro* *16*, 475-480.
- Rehulka, J., Minarik, B., Adamec, V., and Rehulkova, E. (2005). Investigations of physiological and pathological levels of total plasma protein in rainbow trout, *Oncorhynchus mykiss* (Walbaum). *Aquacul. Res.* *36*, 22-32.
- Reis-Henriques, M. A., Cruz, M. M., and Pereira, J. O. (1997). The modulating effect of vitellogenin on the synthesis of 17beta-estradiol by rainbow trout (*Oncorhynchus mykiss*) ovary. *Fish Physiol. Biochem.* *16*, 181-186.
- Reith, M., Munholland, J., Kelly, J., Finn, R. N., and Fyhn, H. J. (2001). Lipovitellins derived from two forms of vitellogenin are differentially processed during oocyte maturation in haddock (*Melanogrammus aeglefinus*). *J. Exp. Zool.* *291*, 58-67.
- Ren, L., Lewis, S. K., and Lech, J. J. (1996). Effects of estrogen and nonylphenol on the post-transcriptional regulation of vitellogenin gene expression. *Chem. Biol. Interact.* *100*, 67-76.
- Riddle, O. and Reinhart, W. H. (1926). Studies on the physiology of reproduction in birds: xxi. Blood calcium changes in the reproductive cycle. *Am. J. Physiol.* *76*, 660-676.
- Riggs, L., Sioma, C., and Regnier, F. E. (2001). Automated signature peptide approach for proteomics. *J. Chromatogr. A* *924*, 359-368.
- Rime, H., Guitton, N., Pineau, C., Bonnet, E., Bobe, J., and Jalabert, B. (2004). Post-ovulatory ageing and egg quality: a proteomic analysis of rainbow trout coelomic fluid. *Reprod. Biol. Endocrinol.* *2*, 26.
- Roubal, W. T., Lomax, D. P., Willis, M. L., and Johnson, L. L. (1997). Purification and partial characterization of English sole (*Pleuronectes vetulus*) vitellogenin. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* *118*, 613-622.
- Routledge, E. J., Sheahan, D., Desbrow, C., Brighty, G. C., Waldock, M., and Sumpter, J. P. (1998). Identification of estrogenic chemicals in STW effluent: 2. In vivo responses in trout and roach. *Environ. Sci. Technol.* *32*, 1559-1565.
- Russell, D. H. (1997). High-resolution Mass Spectrometry and Accurate Mass Measurements with Emphasis on the Characterization of Peptides and Proteins by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry. *J. Mass Spectrom.* *32*, 263-276
- Safe, S. and Krishnan, V. (1995). Cellular and molecular biology of aryl hydrocarbon (Ah) receptor-mediated gene expression. *Arch. Toxicol. Suppl* *17*, 99-115.

Safe, S. H. (1995). Environmental and dietary estrogens and human health: is there a problem? *Environ. Health Perspect.* *103*, 346-351.

Sammar, M., Levi, L., Hurvitz, A., and Lubzens, E. (2005). Studies on retinol-binding protein during vitellogenesis in the Rainbow Trout (*Oncorhynchus mykiss*). *Gen. Comp. Endocrinol.* *141*, 141-151.

Schmieder, P., Tapper, M., Linnun, A., Denny, J., Kolanczyk, R., and Johnson, R. (2000). Optimization of a precision-cut trout liver tissue slice assay as a screen for vitellogenin induction: comparison of slice incubation techniques. *Aquatic. Toxicol.* *49*, 251-268.

Schneider, W. J. (1996). Vitellogenin receptors: oocyte-specific members of the low-density lipoprotein receptor supergene family. *Int. Rev. Cytol.* *166*, 103-137.

Schultz, I. R., Orner, G., Merdink, J. L., and Skillman, A. (2001). Dose-response relationships and pharmacokinetics of vitellogenin in rainbow trout after intravascular administration of 17 $\alpha$ -ethynylestradiol. *Aquat. Toxicol.* *51*, 305-318.

Schwaiger, J., Mallow, U., Ferling, H., Knoerr, S., Braunbeck, T., Kalbfus, W., and Negele, R. D. (2002). How estrogenic is nonylphenol? A transgenerational study using rainbow trout (*Oncorhynchus mykiss*) as a test organism. *Aquat. Toxicol.* *59*, 177-189.

Shoulders, C. C., Brett, D. J., Bayliss, J. D., Narcisi, T. M., Jarmuz, A., Grantham, T. T., Leoni, P. R., Bhattacharya, S., Pease, R. J., Cullen, P. M., and . (1993). Abetalipoproteinemia is caused by defects of the gene encoding the 97 kDa subunit of a microsomal triglyceride transfer protein. *Hum. Mol. Genet.* *2*, 2109-2116.

Shoulders, C. C., Narcisi, T. M., Read, J., Chester, A., Brett, D. J., Scott, J., Anderson, T. A., Levitt, D. G., and Banaszak, L. J. (1994). The abetalipoproteinemia gene is a member of the vitellogenin family and encodes an alpha-helical domain. *Nat. Struct. Biol.* *1*, 285-286.

Shukla, A. K. and Futrell, J. H. (2000). Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom.* *35*, 1069-1090.

Shyu, A. B., Blumenthal, T., and Raff, R. A. (1987). A single gene encoding vitellogenin in the sea urchin *Strongylocentrotus purpuratus*: sequence at the 5' end. *Nucleic Acids Res.* *15*, 10405-10417.

Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., and Bucher, P. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinform.* *3*, 265-274.

- Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005). Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* *77*, 2187-2200.
- Silversand, Christer, Hyllner, Sven J., and Haux, Carl (1993). Isolation, immunochemical detection, and observations of the instability of vitellogenin from four teleosts. *J. Exp. Zool.* *267*, 587-597.
- Smith, R. D., Loo, J. A., Ogorzalek, R. R., Busman, M., and Udseth, H. R. (1991). Principles and practice of electrospray ionization - mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* *10*, 359-452.
- Smith, R. D., Anderson, G. A., Lipton, M. S., Pasa-Tolic, L., Shen, Y., Conrads, T. P., Veenstra, T. D., and Udseth, H. R. (2002). An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* *2*, 513-523.
- So, Y. P., Idler, D. R., and Hwang, S. J. (1985). Plasma vitellogenin in landlocked Atlantic salmon (*Salmo salar* Ouananiche): isolation, homologous radioimmunoassay and immunological cross-reactivity with vitellogenin from other teleosts. *Comp. Biochem. Physiol. B* *81*, 63-71.
- Specker, J. L. and Sullivan, C. V. (1994). Vitellogenesis in fish: Status and perspectives. In *Perspectives in comparative endocrinology*, K.G.Davey, R.G.Peter, and S.S.Tobe, eds. Ottawa: National Research Council of Canada, pp. 304-315.
- Spieth, J., Nettleton, M., Zucker-Aprison, E., Lea, K., and Blumenthal, T. (1991). Vitellogenin motifs conserved in nematodes and vertebrates. *J. Mol. Evol.* *32*, 429-438.
- Srdic, Z., Reinhardt, C., Beck, H., and Gloor, H. (1979). Autonomous yolk protein synthesis in ovaries of *Drosophila* cultured in-vivo. *W. R. Arch. Dev. Biol.* *187*, 255-266.
- Strittmatter, E. F., Ferguson, P. L., Tang, K., and Smith, R. D. (2003). Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* *14*, 980-991.
- Sumpter, J. P. and Jobling, S. (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environ. Health Perspect.* *103 Suppl 7*, 173-178.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* *2*, 151-153.
- Tao, Y., Hara, A., Hodson, R. G., Woods, LC I., and Sullivan, C. V. (1993). Purification, characterization and immunoassay of striped bass (*Morone saxatilis*) vitellogenin. *Fish Physiol. Biochem.* *12*, 31-46.

- Tata, J. R. and Smith, D. F. (1979). Vitellogenesis: a versatile model for hormonal regulation of gene expression. *Recent Prog. Horm. Res.* 35, 47-95.
- Taylor, J. A. and Johnson, R. S. (1997). Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 11, 1067-1075.
- Taylor, J. A. and Johnson, R. S. (2001). Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal. Chem.* 73, 2594-2604.
- Telfer, W. H. and Telfer, W. H. (1954). Immunological studies of insect metamorphosis. II. The role of a sex-limited blood protein in egg formation by the *Cecropia* silkworm. *J. Gen. Physiol.* 37, 539-558.
- Thomas-Jones, E., Thorpe, K., Harrison, N., Thomas, G., Morris, C., Hutchinson, T., Woodhead, S., and Tyler, C. (2003). Dynamics of estrogen biomarker responses in rainbow trout exposed to 17beta-estradiol and 17alpha-ethinylestradiol. *Environ. Toxicol. Chem.* 22, 3001-3008.
- Thorpe, K. L., Hutchinson, T. H., Hetheridge, M. J., Scholze, M., Sumpter, J. P., and Tyler, C. R. (2001). Assessing the biological potency of binary mixtures of environmental estrogens using vitellogenin induction in juvenile rainbow trout (*Oncorhynchus mykiss*). *Environ. Sci. Technol.* 35, 2476-2481.
- Thorpe, K. L., Cummings, R. I., Hutchinson, T. H., Scholze, M., Brighty, G., Sumpter, J. P., and Tyler, C. R. (2003). Relative potencies and combination effects of steroidal estrogens in fish. *Environ. Sci. Technol.* 37, 1142-1149.
- Thurston, R. V. (1967). Electrophoretic patterns of blood serum proteins from rainbow trout (*Salmo gairdneri*). *J. Fish. Res. Board Can.* 24, 2169-2188.
- Todd, J. F. (1991). Ion trap mass spectrometer - past, present, and future (?). *Mass Spectrom. Rev.* 10, 3-52.
- Tollefsen, K. E., Mathisen, R., and Stenersen, J. (2002). Estrogen mimics bind with similar affinity and specificity to the hepatic estrogen receptor in Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). *Gen. Comp. Endocrinol.* 126, 14-22.
- Tollefsen, K. E., Mathisen, R., and Stenersen, J. (2003). Induction of vitellogenin synthesis in an Atlantic salmon (*Salmo salar*) hepatocyte culture: a sensitive in vitro bioassay for the oestrogenic and anti-oestrogenic activity of chemicals. *Biomarkers* 8, 394-407.
- Toppari, J., Larsen, J. C., Christiansen, P., Giwercman, A., Grandjean, P., Guillette, L. J., Jr., Jegou, B., Jensen, T. K., Jouannet, P., Keiding, N., Leffers, H., McLachlan, J. A., Meyer, O., Muller, J., Rajpert-De Meyts, E., Scheike, T., Sharpe, R., Sumpter, J., and

- Skakkebaek, N. E. (1996). Male reproductive health and environmental xenoestrogens. *Environ. Health Perspect.* *104 Suppl 4*, 741-803.
- Trichet, V., Buisine, N., Mouchel, N., Moran, P., Pendas, A. M., Le Pennec, J. P., and Wolff, J. (2000). Genomic analysis of the vitellogenin locus in rainbow trout (*Oncorhynchus mykiss*) reveals a complex history of gene amplification and retroposon activity. *Mol. Gen. Genet.* *263*, 828-837.
- Tyler, C. R., Sumpter, J. P., and Bromage, N. R. (1988). In-vivo ovarian uptake and processing of vitellogenin in the rainbow trout *salmo-gairdneri*. *J. Exp. Zool.* *246*, 171-179.
- Tyler, C. R., Sumpter, J. P., and Campbell, P. M. (1991). Uptake of vitellogenin into oocytes during early vitellogenic development in the rainbow trout *oncorhynchus-mykiss walbaum*. *J. Fish Biol.* *38*, 681-690.
- Tyler, C. R. (1993). Electrophoretic patterns of yolk proteins throughout ovarian development and their relationship to vitellogenin in the rainbow trout, *Oncorhynchus mykiss*. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* *106*, 321-331.
- Tyler, C. R. and Lancaster, P. (1993). Isolation and characterization of the receptor for vitellogenin from follicles of the rainbow trout, *Oncorhynchus mykiss*. *J. Comp. Physiol. B Biochem. Systemic Environ. Physiol.* *163*, 225-233.
- Tyler, C. R. and Sumpter, J. P. (1996). Oocyte growth and development in teleosts. *Rev. Fish Biol. Fisher.* *6*, 287-318.
- Tyler, C. R., Van Aerle, Ronny, Hutchinson, Tom H., Maddix, Sue, and Trip, Heleanne (1999). An in vivo testing system for endocrine disruptors in fish early life stages using induction of vitellogenin. *Environ. Toxicol. Chem.* *18*, 337-347.
- Tyler, C. R., van Aerle, R., Nilsen, M. V., Blackwell, R., Maddix, S., Nilsen, B. M., Berg, K., Hutchinson, T. H., and Goksoyr, A. (2002). Monoclonal antibody enzyme-linked immunosorbent assay to quantify vitellogenin for studies on environmental estrogens in the rainbow trout (*Oncorhynchus mykiss*). *Environ. Toxicol. Chem.* *21*, 47-54.
- Tyler, C. R., Spary, C., Gibson, R., Santos, E. M., Shears, J., and Hill, E. M. (2005). Accounting for differences in estrogenic responses in rainbow trout (*Oncorhynchus mykiss*: Salmonidae) and roach (*Rutilus rutilus*: Cyprinidae) exposed to effluents from wastewater treatment works. *Environ. Sci. Technol.* *39*, 2599-2607.
- Utter, F. M. and Ridgway, G. J. (1965). A serologically detected serum factor associated with maturity in English sole, *Paropherys vetulus*, and Pacific halibut, *Hippoglossus stenolepis*. *Fish. Bull.* *66*, 47-58.

- van Bohemen, C. G. and Lambert, J. G. (1981). Estrogen synthesis in relation to estrone, estradiol, and vitellogenin plasma levels during the reproductive cycle of the female rainbow trout, *Salmo gairdneri*. *Gen. Comp. Endocrinol.* *45*, 105-114.
- van Bohemen, C. G., Lambert, J. G., and Peute, J. (1981). Annual changes in plasma and liver in relation to vitellogenesis in the female rainbow trout, *Salmo gairdneri*. *Gen. Comp. Endocrinol.* *44*, 94-107.
- van den Heuvel, M. R. and Ellis, R. J. (2002). Timing of exposure to a pulp and paper effluent influences the manifestation of reproductive effects in rainbow trout. *Environ. Toxicol. Chem.* *21*, 2338-2347.
- van den Heuvel, M. R., Ellis, R. J., Tremblay, L. A., and Stuthridge, T. R. (2002). Exposure of reproductively maturing rainbow trout to a New Zealand pulp and paper mill effluent. *Ecotoxicol. Environ. Saf.* *51*, 65-75.
- Van den, Belt K., Verheyen, R., and Witters, H. (2003). Comparison of vitellogenin responses in zebrafish and rainbow trout following exposure to environmental estrogens. *Ecotoxicol. Environ. Saf.* *56*, 271-281.
- van Montfort, B. A., Canas, B., Duurkens, R., Godovac-Zimmermann, J., and Robillard, G. T. (2002). Improved in-gel approaches to generate peptide maps of integral membrane proteins with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* *37*, 322-330.
- Verslycke, T., Vandenberg, G. F., Versonnen, B., Arijs, K., and Janssen, C. R. (2002). Induction of vitellogenesis in 17 $\alpha$ -ethinylestradiol-exposed rainbow trout (*Oncorhynchus mykiss*): a method comparison. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* *132*, 483-492.
- Vetillard, A. and Bailhache, T. (2005). Cadmium: an endocrine disrupter that affects gene expression in the liver and brain of juvenile rainbow trout. *Biol. Reprod.* *72*, 119-126.
- von der Decken, A. and Waters, S. (1993). Modulation of hepatic chromatin structure in response to 17-beta estradiol induced activation of the vitellogenin gene regions in Atlantic salmon, *Salmo salar*. *Acta Biochim. Pol.* *40*, 23-28.
- Wallace R.A. and Jared D.W. (1968). Studies on amphibian yolk. VII. Serum phosphoprotein synthesis by vitellogenic females and estrogen-treated males of *Xenopus laevis*. *Can. J. Biochem.* *46*, 953-959.
- Wallace, R. A. (1985). Vitellogenesis and oocyte growth in nonmammalian vertebrates. *Dev. Biol.* *1*, 127-177.
- Wallaert, C. and Babin, P. J. (1994). Age-related, sex-related, and seasonal changes of plasma lipoprotein concentrations in trout. *J. Lipid Res.* *35*, 1619-1633.

- Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* *75*, 4818-4826.
- Watts, M., Pankhurst, N. W., Pryce, A., and Sun, B. (2003). Vitellogenin isolation, purification and antigenic cross-reactivity in three teleost species. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* *134*, 467-476.
- Wiegand, M. D. (1982). Vitellogenesis in fishes., C.J.J.Richter and H.J.Goos, eds. Pudoc, Wageningen, The Netherlands.
- Wiley, H. S., Opresko, L., and Wallace, R. A. (1979). New methods for the purification of vertebrate vitellogenin. *Anal. Biochem.* *97*, 145-152.
- Wiley, H. S. and Wallace, R. A. (1981). The structure of vitellogenin. Multiple vitellogenins in *Xenopus laevis* give rise to multiple forms of the yolk proteins. *J. Biol. Chem.* *256*, 8626-8634.
- Wollnik, H. (1993). Time-of-flight mass analyzers. *Mass Spectrom. Rev.* *12*, 89-114.
- Wunschel, D., Schultz, I., Skillman, A., and Wahl, K. (2005). Method for detection and quantitation of fathead minnow vitellogenin (Vtg) by liquid chromatography and matrix-assisted laser desorption/ionization mass spectrometry. *Aquat. Toxicol.* *73*, 256-267.
- Yadete, F., Arukwe, A., Goksoyr, A., and Male, R. (1999). Induction of hepatic estrogen receptor in juvenile Atlantic salmon in vivo by the environmental estrogen, 4-nonylphenol. *Sci. Total Environ.* *233*, 201-210.
- Yadete, F. and Male, R. (2002). Effects of 4-nonylphenol on gene expression of pituitary hormones in juvenile Atlantic salmon (*Salmo salar*). *Aquat. Toxicol.* *58*, 113-129.
- Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001). Proteolytic <sup>18</sup>O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* *73*, 2836-2842.
- Yao, X., Afonso, C., and Fenselau, C. (2003). Dissection of proteolytic <sup>18</sup>O labeling: endoprotease-catalyzed <sup>16</sup>O-to-<sup>18</sup>O exchange of truncated peptide substrates. *J. Proteome. Res.* *2*, 147-152.
- Yao, Z. and Crim, Laurence W. (1996). A biochemical characterization of vitellogenins isolated from the marine fish ocean pout (*Macrozoarces americanus* L.), lumpfish (*Cyclopterus lumpus*) and Atlantic cod (*Gadus morhua*). *Comp. Biochem. Physiol. B* *113*, 247-253.

- Yates, J. R., III (1998). Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* *33*, 1-19.
- Yeo, I.K. and Mugiya Y. (1997). Effects of extracellular calcium concentrations and calcium antagonists on vitellogenin induction by estradiol-17-beta in primary hepatocyte culture in the rainbow trout *Oncorhynchus mykiss*. *General and Comparative Endocrinology* *105*, 294-301.
- Yokota, Y. and Kato, K. H. (1988). Degradation of yolk proteins in sea urchin eggs and embryos. *Cell Differ.* *23*, 191-200.
- Yu, L. R., Conrads, T. P., Uo, T., Issaq, H. J., Morrison, R. S., and Veenstra, T. D. (2004). Evaluation of the acid-cleavable isotope-coded affinity tag reagents: application to camptothecin-treated cortical neurons. *J. Proteome Res.* *3*, 469-477.
- Zhang, F., Bartels, M. J., Brodeur, J. C., and Woodburn, K. B. (2004a). Quantitative measurement of fathead minnow vitellogenin by liquid chromatography combined with tandem mass spectrometry using a signature peptide of vitellogenin. *Environ. Toxicol. Chem.* *23*, 1408-1415.
- Zhang, H., Yan, W., and Aebersold, R. (2004b). Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* *8*, 66-75.
- Zimmer, J. S. D., Monroe, M. E., Qian, W. J., and Smith, R. D. (2006). Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev.* *25*, 450-482.
- Zwarenstein, H. and Shapiro, H. A. (1933). Metabolic changes associated with endocrine activity and the reproductive cycle in *Xenopus laevis*. III. Changes in the calcium content of the serum associated with captivity and the normal reproductive cycle. *J. Exp. Biol.* *10*, 372-378.

### **Appendix 1. Single and three letter codes for the protein amino acids**

Alanine	A	Ala
Cysteine	C	Cys
Aspartic Acid	D	Asp
Glutamic Acid	E	Glu
Phenylalanine	F	Phe
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Lysine	K	Lys
Leucine	L	Leu
Methionine	M	Met
Asparagine	N	Asn
Proline	P	Pro
Glutamine	Q	Gln
Arginine	R	Arg
Serine	S	Ser
Threonine	T	Thr
Valine	V	Val
Tryptophan	W	Trp
Tyrosine	Y	Tyr

**Appendix 2. Amino acid sequences of rainbow trout and Atlantic salmon vitellogenin**

Rainbow trout vitellogenin sequence: ENTRY Q92093

10	20	30	40	50	60
MRAVVLALTL	ALVASQSVNF	APDFAASKTY	VYKYEALLLG	GLPEEGLARA	GVKVISKVL
70	80	90	100	110	120
SAVAENTYLL	KLVNPEIFEY	SGVWPKDPFV	PAAKLTSALA	AQFSIPIKFE	YAKGVVGKVL
130	140	150	160	170	180
APTAVSETVL	NVHRGILNIL	QLNIKKTQNV	YELQEAGAQQ	VCKTHYVIRE	DAKAERIHLT
190	200	210	220	230	240
KSKDLNNCQQ	RIMKDFGLAY	TEKCVETRQR	GEALMGAATY	NYLMKPADNG	ALILEATVTE
250	260	270	280	290	300
LHQFTPFNEM	SGAAQMEAKQ	MLTFVEIKKD	PIIVPDNNYV	HRGSIRYEFA	TEILQMPIQL
310	320	330	340	350	360
LKISNARAQA	VKILNHLVTY	NTAPVHEDAP	LKFLQFIQLL	RMASSETINA	IWAEFKAKPA
370	380	390	400	410	420
YRHWILDVAV	SIGSSVAVRF	IKEKFLAGDI	TIFEAAQALV	AAVHMVAADL	ETVKLVESLA
430	440	450	460	470	480
FNHKIQTHPV	LRELTMLGYG	TMVSKYCVEH	PNCPAELVKP	IHELAVQAVA	NSKFEELSMV
490	500	510	520	530	540
LKALGNAGHP	ASIKPITKLL	PVFGTAAAAAL	PLRVQADAVL	ALRNIAKREP	RMVQEVAVQL
550	560	570	580	590	600
FMDKALHPEL	RMLACIVLFE	TKPPMGLVIT	LASILKTEKN	MQVASFTYSH	MMSLTRSTAP
610	620	630	640	650	660
DFASVAAACN	VAVKMLSNKF	RRLSCHFSQA	IHLDAYSNPL	RIGAAASAFY	INDAATLFPR
670	680	690	700	710	720
TVVAKARTYF	AGAAAADVLEV	GVRTEGIEEA	LLKLPPAPEN	ADRITKMRRV	IKALSDWRSL
730	740	750	760	770	780
ATSKPLASIY	VKFFGQEIAF	ANIDKSIIDQ	ALQLANSPSA	HALGRNALKKA	LLAGATFQYV
790	800	810	820	830	840
KPLLAEEVRR	IFPTAVGLPM	ELSYTAAVA	KAYVNVRTL	TPALPETFHA	AQLLKTNIEL
850	860	870	880	890	900
HAEVRPSIVM	HTFAVMGVNT	AFIQAAIMAR	AKVRTIVPAK	FAAQLDIANG	NFKFEAFPVS
910	920	930	940	950	960
PPEHIAAAHI	ETFAVARNVE	DVPAERITPL	IPAQGVARST	QQSRDKLTSM	IADSAASFAG
970	980	990	1000	1010	1020
SLRSSEILY	SDLPSNFKPI	IKAIVVHLEE	TICVERLGVK	ACFEFTSESA	AFIRNTLFYN
1030	1040	1050	1060	1070	1080
MIGKHSVLIS	VKPSASEPAI	ERLEFEVQVG	PKAAEKIKV	ITMNEEEEP	EGKTVLLKLK
1090	1100	1110	1120	1130	1140
KILLPDLKNG	TRASSSSSSS	SSSSSRSSSS	RSRSRKSESS	SSSSSSSSRI	SKRDGPDQPY
1150	1160	1170	1180	1190	1200
NPNDRKFKKN	HKDSQSTSNV	ISRSKSSASS	FHAIYKQDKF	LGNKLAPMVI	ILFRLVRADH
1210	1220	1230	1240	1250	1260
KIEGYQVTAY	LNKATSRLQI	IMAALDENDN	WKLCADGVLL	SKHKVTAKIA	WGAECKDYNT
1270	1280	1290	1300	1310	1320
FITAETGLVG	PSPAVRLRLS	WDKLPKVPKA	VWRYVRIVSE	FIPGYIPYYL	ADLVPMQKDK
1330	1340	1350	1360	1370	1380
NNEKQIQFTV	VATSERLDV	ILKTPKMTLY	KLGVNLPCL	PFESMTDLSP	FDDNIVNKIH
1390	1400	1410	1420	1430	1440
YLFSEVNAVK	CSMVRDTLTT	FNNKKYKINM	PLSCYQVLAQ	DCTTELKFMV	LLKKDHASEQ
1450	1460	1470	1480	1490	1500

NHINVKISDI	DVDLYTEDHG	VIVKVNEMEI	SNDNLPYKDP	SGSIKIDRKG	KGVS LYAPSH
1510	1520	1530	1540	1550	1560
GLQEVYFDKY	SWKIKVVDWM	KGQTCGLCGK	ADGENRQEYR	TPSGRLTKSS	VSFAHSWVLP
1570	1580	1590	1600	1610	1620
SDSCRDASEC	LMKLESVKLE	KQVIVDDRES	KCYSVEPVLR	CLPGCLPVRT	TPITIGFHCL
1630	1640	1650			
PVDSNLNRSE	GLSSIEYKSV	DLMEKAEAHV	ACRCSEQCM		

**Atlantic salmon vitellogenin sequence: ENTRY Q800N9**

	10	20	30	40	50	60
MKAVVFALTL	ALVASQHVNF	APEFEASKTY	VYKYEALLLG	GLPEEGLSRA	GIKVISKVL I	
70	80	90	100	110	120	
SAIAPTIYML	KLVDPEIFEY	SGVWPKDPFV	PASKLTSALA	AQLLTPIKFE	YANGVVGKVF	
130	140	150	160	170	180	
APTGVSETVL	NIHRGILNIL	QLTFKKTQNV	YEMQEAGAQG	VCKTHYVISE	DAKTGYIH LT	
190	200	210	220	230	240	
KTKDLNHCQE	RIIKEFGLAY	TENCVECQOK	GKNLRGAAAY	NYIMKPTATG	AIIMEATVTE	
250	260	270				
LHQFSPFNEM	NGAAQMEAKQ	TFAFVEIKKT				

**Atlantic salmon vitellogenin sequence: ENTRY Q8UWG2**

	10	20	30	40	50	60
KIKQKGEVGS	LYAPSHGLQE	VYFDKNSWKI	RVVDWMKGQT	CGLCGKADGE	VRQEYSTPSG	
70	80	90	100	110	120	
RLTKSSVSFA	HSWVLPDSC	RDASECLMTF	ESVKLEKQVI	VDDKESKCYS	VEPVLRCLPG	
130	140	150	160	170		
CLPVRTTPIT	IGFHCLPVDS	NLSRSEGLSS	FYEKSVDLRE	KAEAHVACRC	SQQCI	

### **Appendix 3: ELISA test for rainbow trout and Atlantic salmon vitellogenin**

Rainbow trout and Atlantic salmon plasma samples were analyzed by an ELISA kit for RT Vtg (Biosense Laboratories, Bergen, Norway), according to its instruction manual. Although this kit is specific for RT Vtg, the suppliers of this kit claim it can be used for AS Vtg, using a correction factor of 20. Briefly, all plasma samples were diluted 1:100, 1:3000 and 1:90000 according to their concentration estimated in the HPLC-QqTof-MS/MS experiments. Simultaneously, a calibration curve was prepared by dilution of the RT Vtg stock solution to produce concentrations ranging from 200 to 0.39ng/mL. All dilutions were kept on ice until use.

One hundred  $\mu\text{L}$  of standards and diluted samples were loaded in duplicate to the ELISA plate and incubated at  $4^{\circ}\text{C}$  overnight. Subsequently, the ELISA plate was washed three times with  $300\mu\text{L}$  of the provided washing buffer per well. Next, all wells were incubated at room temperature with the detecting antibody for 1 hour on an orbital shaker. After 5 washing steps with  $300\mu\text{L}$  of washing buffer per well,  $100\mu\text{L}$  of the Ellman's reagent substrate solution were added to each well. The plate was placed on the orbital shaker in the dark for one hour. Finally, the plate was placed in a microplate reader (Synergy HT, Bio-tek Instruments. Winooski VT, USA), and the absorbance of each well was read at 405 nm. A 4-parameter transformation of the data was used for regression analysis of the standard curve, and for extrapolating the concentration of the unknown samples. The latter were then corrected by the corresponding dilution factors.





