Paper II

Genomic and proteomic analyses reveal a single form of vitellogenin in the basal clupeocephalan Atlantic herring, with partial degradation of derivative yolk proteins during oocyte hydration

Børge A Kristoffersen[1], Audun Nerland[2], Frank Nilsen[1,2], Jelena Kolarevic[1], Roderick Nigel Finn[1*]

[1] Department of Biology, University of Bergen, Bergen High Technology Center, Thormøhlensgate 55, N-5008 Bergen, Norway
[2] Institute of Marine Research, Post box 1870 Nordnes, N-5817 Bergen, Norway

Abstract

Atlantic herring is a comparatively primitive marine teleost spawning benthic eggs that undergo moderate oocyte hydration during meiotic maturation. This group of teleosts was amongst the first to radiate in the oceans during the early Cretaceous. We have previously shown that a small pool of free amino acids (FAA) significantly contributes to the osmolarity of the ovulated egg. To determine whether yolk proteolysis represents the underlying mechanism in this species, we cloned hepatically expressed vitellogenin (*vtg*) and examined its deduced structure in relation to deposited oocyte and egg yolk proteins (Yp). Unlike all species of teleost studied to date we found that only a single *vtg* transcript (*chvtgAc*) is expressed. Genomic analyses of exon-intron structures conserved between *vtgs* of Atlantic herring and zebrafish revealed a small polymorphic intron between exons 9 and 10 in Atlantic herring, and a small variation (~5%) in exons 6-10. This suggest that Atlantic herring has at least two *vtg* genes, but only one form of Vtg. The data also suggest that duplication of ostariophysan *vtgAo1* and *vtgAo2* genes occurred in a lineage-specific manner after separation of the Ostariophysi from the Clupeiformes. The deduced amino acid sequence shows ChvtgAc to conform to the pentapartite NH_2-(LvH-Pv-LvL-ß´-CT)-COO^- structure of complete teleost Vtgs. Multiple phylogenetic analyses consistently clustered the *chvtgAa* transcript and ChvtgAa protein as the basal sister group to the Ostariophysi in full congruence with the Clupeocephalan rank. Three-dimensional modeling of ChvtgAa against lamprey lipovitellin revealed that the tertiary structure is highly conserved. Identification of the oocyte and egg Yp by N-terminal microsequencing, mass spectrometry, Western immunoblotting, and gel electrophoresis showed that some proteolytic processing occurs during oocyte hydration. The data indicate that the phosvitin domain, the smallest yet reported for teleosts, and an N-terminal fragment of the lipovitellin light chain contribute to the FAA pool. We also provide the first molecular evidence for the presence of the C-terminal coding region of Vtg in the yolk protein pool of any teleost. The data suggest that yolk proteolysis and the generation of an organic osmolyte pool of FAA was an adaptive response to spawning in seawater also for the Clupeiformes, but was not evolutionarily successful in terms of biodiversity until *vtg* gene duplication and neofunctionalization occurred in the Acanthomorpha.

Key words

Vitellogenin, gene duplication, oocyte hydration, meiotic maturation, osmolality, osmoregulation, yolk proteolysis, yolk proteins, endosome, lysosome, herring, zebrafish

Introduction

In teleosts, the differential expression of multiple forms of vitellogenin (*vtg*) and subsequent processing of the Vtg proteins following clathrin-mediated endocytosis has been shown to result in a complex suite of deposited yolk proteins (Yp) in the growing oocyte (Babin et al., 2007; Finn 2007a). Owing to a convoluted heritage of several rounds of genomic- and lineage-specific gene duplications with dependent and independent gene losses it has proved challenging to understand the functional precursor-product relations of this important family of high-density lipoproteins (Finn & Kristoffersen 2007). Members of the Acanthomorpha have been the best studied with up to three forms of *vtg* (*vtgAa*, *vtgAb* and *VtgC*) reported for this group (Ding *et al*. 1989, LaFleur *et al*. 1995, 2005, Matsubara *et al*. 1999, 2003, Reith *et al*. 2001, Shimizu *et al*. 2002, Hiramatsu *et al.* 2003, Ohkubo *et al.* 2004, Fujiwara *et al.* 2005, Sawaguchi *et al.* 2005, 2006a, Amano *et al.* 2007, Davis *et al.* 2007, Finn, 2007b, Finn & Kristoffersen, 2007, Kolarevic *et al.* 2008). Among the more primitve members of Salmonidae (Protacanthopterygii) multiple copies of two forms (*vtgAsa* and *vtgAsb*) are known, while in the genus *Oncorhynchus*, the *vtgAsb* form has been lost (Mouchel *et al.* 1996, 1997, Trichet *et al.* 2000, Buisine *et al.* 2002). Up to three forms (*vtgAo1*, *vtgAo2* and *vtgC*) are recognized in Ostariophysi (Wang *et al.* 2000, 2005, Mikawa *et al.* 2006, Miracle *et al.* 2006, Kang *et al.* 2007) and two forms (three genes: *vtgAe1-3*) are known for the Elopomorpha (Okumura *et al.* 2002, Wang *et al.* 2006, Finn and Kristoffersen, 2007). It is only among the more ancient Chondrostei (Sharrock *et al.* 1992) and Hyperoartia (Anderson *et al.* 1998) that single forms of *vtg* are known.

Atlantic herring is a comparatively primitive marine teleost from the order Clupeiformes that spawns benthic adhesive eggs. Recent morphological and molecular phylogenetic analyses have placed Clupeiformes as the basal sister group to the Ostariophysi under the higher rank of Clupeocephala (Inoue *et al.* 2003, 2005, Nelson, 2006, Li & Ortí, 2007, Mabee *et al.* 2007). Our previous analysis of the evolution of vertebrate *vtg* genes highlighted the lack of molecular data for this group of teleosts (Finn & Kristoffersen, 2007). In the present study we have cloned a single *vtg* transcript and investigated the structural relationships of the deduced peptide to deposited Yps in the oocyte and ovulated egg (OV egg) of Atlantic herring. This study is the first to document the full *vtg* transcript and Yps in this class of teleost. As in all oviparous species of marine teleosts, the oocyte of Atlantic herring undergoes a pre-adaptive hydration prior to spawning in the hyperosmotic oceanic environment (Kristoffersen & Finn, 2008). In this latter study we have shown that inorganic

ions are the major osmolytes that drive the maturational influx of water in this species. Intriguingly, a small pool of free amino acids (FAA) that appear during meiotic maturation. This FAA pool due to the relatively low water content of the ovulated egg (OV egg) in Atlantic herring, contribute 29% to the ovoplasmic osmolarity (Kristoffersen & Finn, 2008). Since it is known that acanthomorph species rely on depolymerization of mainly the lipovitellin heavy chain of the VtgAa form to generate the osmotic driving force for oocyte hydration, we wanted to investigate whether a similar process occurs in more ancestral marine teleosts as exemplified by Atlantic herring.

Materials and Methods

Fish and samples

Atlantic herring (*Clupea harengus*) were collected using gill nets during the autumn and spring spawning seasons in coastal waters near Bergen, Norway. Mature females were euthanized (blow to the head) and transported on ice to the Institute of Biology, University of Bergen where biometric data (fork length, ± 0.1 cm), and gravimetric data (female whole body, liver and ovary wet masses, ± 0.1 g) were determined. Gonadosomatic index (GSI) was calculated as the ovarian fraction of the female body mass, and livers from females with an intermediate GSI indicating vitellogenesic livers were dissected out and transferred to -80°C as described by Kristoffersen *et al*. (2007).

cDNA isolation and cloning

Total RNA was isolated using Trizol Reagent (Life-Technologies) followed by mRNA purification (Oligotex™ mRNA kit, Qiagen) in accordance with the guidelines of the kit protocols. A cDNA library was made using an Invitrogen kit: "SuperScript™ Plasmid System with Gateway® Technology for cDNA Synthesis and Cloning" with a pBluescript (r) II SK (+) vector and XL10-Gold®Ultracompetent cells from Stratagene.

Based on sequences from Atlantic salmon *vtg* (Yadetie *et al*. 1999) degenerate primers were designed to obtain sequences from Atlantic herring *vtg*: degenerate primer sense #1: (5'-CMA GMA GMA CCC AGG ART G-3') and degenerate primer antisense #2: (5'-CAG GTC

TTG GCT CAR CAY TGY-3'). Both primers gave a strong single band of approximately 400 base pairs (bp) in length. XL10-Gold colonies were grown on LB media containing ampicillin (100 μL mL^{-1}) and a digoxigenin (DIG)-labelling kit (Invitrogen) was used to identify positive clones from Southern blots. Positive colonies were further grown and plasmid DNA purified using a Miniprep kit (Qiaprep Miniprep Kit, Qiagen).

Plasmid DNA was sequenced using M13 vector primers with a Big-Dye 3.1 Sequencing kit on a ABI Sequencer 3730xl (AME Bioscience). Initial sequencing produced ESTs that yielded nucleotide (nt) sequences from the 3´-region of *vtg*, including the stop codon. Atlantic herring gene specific primers (GSP: sense 5'-ATT CGT CGC ATC TTC CCC A-3', antisense 5'TGG CTC TAG CCC ATG ACA GC-3') were designed from this region and 5'-rapid amplification of cDNA ends (5'-RACE, Invitrogen GeneRacer™ Kit and pCR4-TOPO vector using a TOPO TA cloning kit) was used to generate a new 5-RACE library. To screen this library Southern blots using a new DIG probe (~300 bp) based on the GSP primers was performed as described above. Positive colonies were selected, grown and sequenced as above.

The remaining 5´-region including the start codon was obtained using a new set of GSP primers (antisense # 3: 5'- GAG CCA AAT CCA GGG AGG ATC TTC A-3' and antisense # 4: 5'-GAG AGT AGA CAC CAG TGC GAC ATT A-3') using SMART™ RACE kit (Clonetech) and TOPO TA cloning kit (Invitrogen) as above. Plasmid DNA purification and sequencing also performed as above. Since it is known that up to three forms of *vtg* are expressed in acanthopterygian teleosts, degenerate antisense primers were specifically designed from red seabream *vtgAb* (5'-CAT CTG GGC AGC GCC ATT CAA GAT GTT GA-3') and *vtgC* (5'-TGG TAC TGG GTC ATC CAG GTT CTG CAT CAT-3') and used to screen the Clonetech cDNA library.

Since only a single *vtg* transcript was identified using this approach, a genomic exon-intron strategy was employed to search for other silent variants. Forward (s = sense) and reverse (a = antisense) primer pairs were designed against conserved exons flanking heterogeneous introns based upon an alignment of zebrafish genomic DNA (gDNA) (*vtgAo1*: ENSDARG00000033770 and *vtgAo2*: ENSDARG00000055809) and our putative Atlantic herring *vtgAc*.

Exon 4-5

*1) vtgAo1*_s3 (bp 573-601) 5'- CCT CAG CAC TGG CTG CTC AGC TTC AGA T -3'
*2) vtgAo1*_a3 (bp 914-940) 5'- TCA TGA TTC TCT CCT GGC AGT GGC TCA -3'

Exon 4-5

*3) vtgAo2*_s3 (bp 616-643) 5'- CTC AGC ACT GGC TGC TCA GCT TCA GA CT -3'
*4) vtgAo2*_a3 (bp 916-942) 5'- GCG ATC TGC CTT TGG ATC CTC ACT GAT -3'

Exon 6-7

*5) chvtgAc*_s3 (bp 622-649) 5'- AGC CTG ACT GGA GCA GCA ACC TAC AGC T -3'
*6) chvtgAc*_a3 (bp 818-842) 5'- TGC AGA GCT CCA CGG GCC AGA TAA T -3'

Exon 9-10

*7) chvtgAc*_s1 (bp 1097-1124) 5'- TGG TTG GAA CAC CAG TTG CTG TGA GGT T -3'
*8) vhvtgAc*_a1a (bp 1281-1308) 5'- GGT GCC ATA ACC CAG CAT AGC AAT CTC A -3'
*9) chvtgAc*_a1b (bp 1324-1349) 5-' GCT GGG CAA GTA GGA ACT GCA GCA CA -3'

Exon 6-10

*5) chvtgAc*_s3 (bp 622-649) 5'- AGC CTG ACT GGA GCA GCA ACC TAC AGC T -3'
*9) chvtgAc*_a1b (bp 1324-1349) 5-' GCT GGG CAA GTA GGA ACT GCA GCA CA -3'

Where exon numbers refer to the zebrafish genes (www.ensembl.org, Zv7), and bp numbers refer to zebrafish gDNA and Atlantic herring cDNA, respectively. Atlantic herring gDNA was isolated from a single female liver by homogenisation in phenol/chloroform, precipitated in ethanol and resolubilised in ddH_2O. PCR, cloning and sequencing of products were performed as described above.

Verification of purity and/or size of PCR products was performed using 1% agarose gels visualized using BioRad Gel Doc 2000. Determination of RNA or DNA content was conducted using a NanoDrop ND-1000 Spectrophotometer. PCR amplification of products was performed on a Applied Biosystems GeneAmp PCR System 9700 or a Eppendorf Mastercycler ep gradient S PCR thermal cycler.

Sequence Analyses

Multiple phylogenetic analyses of the Atlantic herring *vtg* codons (Bayesian, maximum likelihood, maximum parsimony) and deduced amino acid (aa) sequence

(Bayesian, maximum likelihood, maximum parsimony, neighbour joining) were performed on full-length sequences and subdomains as described by (Finn & Kristoffersen, 2007). Three-dimensional-modeling was achieved using the CBS service for prediction of 3D homology models (Lund 2002). Identification of the putative signal peptide, phosphorylation and glycosylation sites were conducted according to the recommendations of Blom et al (1999, 2004) Julenius *et al*. (2004) and Emanuelsson *et al*. (2007) using the CBS services.

Yolk protein analyses

Yolk proteins (Yp) from oocytes (ooc) and ovulated eggs (OV egg) were solubilized as described by Finn *et al.* (Finn *et al.* 2002a) followed by electrophoresis and visualization as described by Finn (2007a) using Coomassie brilliant blue (CBB) G-250, silver stains (Wako) and phosvitin visualization using the Gelcode phosphoprotein staining kit (Pierce).

Western immunoblotting was performed as described by Matsubara *et al.* (Matsubara *et al.* 1999) and Finn (2007a) using polyvalent antisera raised against barfin flounder (*Verasper moseri*) yolk proteins: a_{Vm}-LvH-Aa, 107 kDa, a_{Vm}-LvH-Ab, 94 kDa, a_{Vm}-LvL-Aa, 30 kDa, a_{Vm}-LvL-Ab, 28 kDa, and a_{Vm}-ß´, 17 kDa.

In order to further identify the oocyte and egg Yps, bands were excised and submitted to either N-terminal microsequencing (Applied BioSystems ProCise cLC 492) or analyses by mass spectrometry (MS). N-terminal microsequencing was performed following SDS-PAGE and semi-dry transblotting to PVDF membranes as described by Finn (2007a). For mass spectrometry, Yps were prepared according to Kolarevic et al (2008) and identified using q-TRAP and q-TOF fragment analysis (q-TOF FA) at the Norwegian Proteomics Unit (PROBE). Tryptic digests were analyzed with a hybrid triple quadrupole linear ion trap (4000 q-TRAP, Applied Biosystems) coupled to a nanoflow chromatography system. The liquid chromatography (LC) separation was achieved using a LC Packings Integrated System (Dionex, Camberley, UK) consisting of a FAMOS microautosampler, Switchos microcolumn switching device and an Ultimate nanopump. Samples were loaded with 0.05% trifluoroacetic acid into a C18 µ-precolumn cartridge. Peptides were eluted from a homemade C18 column (i.d. 75 µm, 15 cm) using a column loading device from Proxeon with packing material reprosil-Pur 120 (3.5 µm) (Dr Maisch GmbH, Germany) using a 45 min gradient with a flow

of 0.2 µL min^{-1} and a mobile phase of acetonitrile in ddH_2O and 0.1% formic acid. qTOF analyses were performed as described by Kolarevic et al. (2008).

Results

A single *vtg* sequence was obtained from cDNA libraries of vitellogenic livers of female Atlantic herring. The *vtg* sequence was constructed from 56 ESTs ranging in size from ~400 - 1000 bp, with a coverage overlap of between 2 – 20 for all regions. No discrepancies existed in these ESTs that might suggest the presence of more than one expressed form of *vtg*. The full-length cDNA sequence contained an open reading frame of 4821 nt encoding 1607 aa including the predicted signal peptide of 15 aa with a total calculated molecular mass of 174 kDa. No clones were obtained using the GSPs specific for red seabream *vtgAb* and *vtgC*.

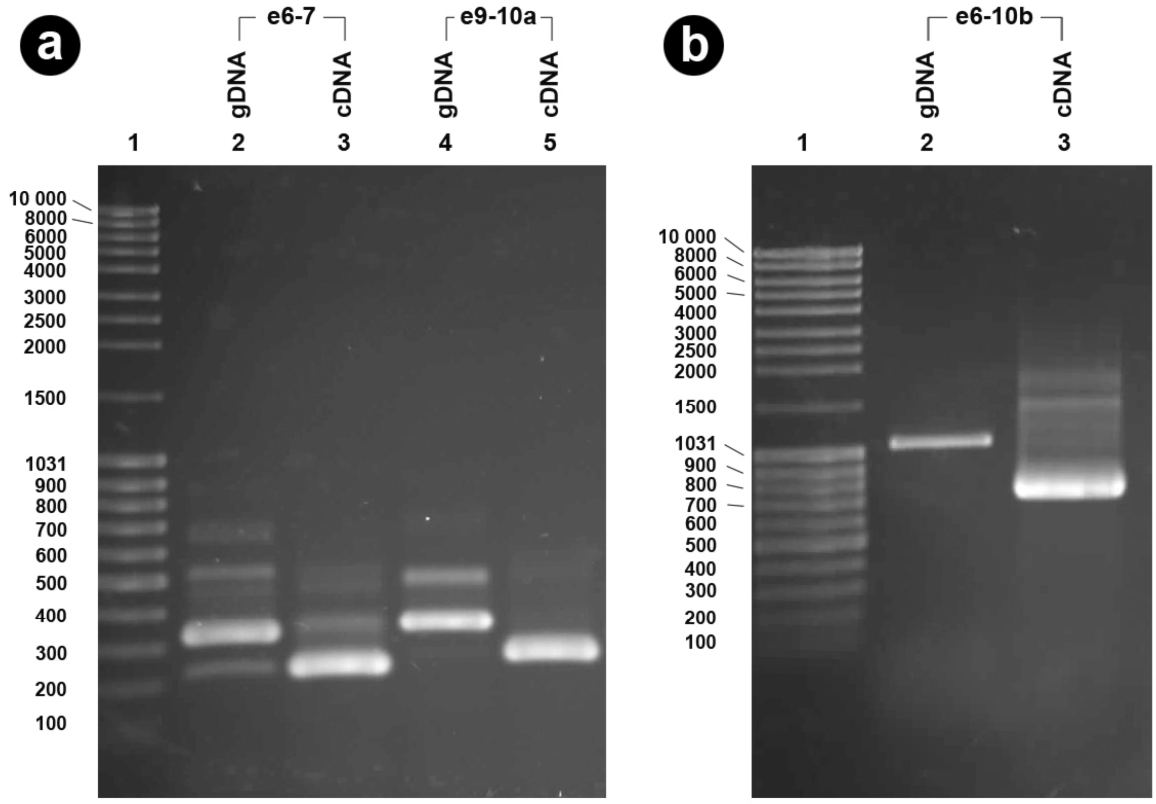

Figure 1. PCR products amplified from genomic extracts of Atlantic herring (*Clupea harengus*) liver. Exon primer pairs (panel (a) e6-7; e9-10a; panel (b) e6-10b) conserved between Atlantic herring and zebrafish *vtgAo1* and *vtgAo2* genes were used to amplify both genomic (gDNA) and transcript (cDNA) products. Lane 1 in both panels shows numbers of bases pairs in the DNA ladder.

To further investigate that only a single form of *vtg* exists in the Atlantic herring genome, we isolated gDNA and examined its structure in relation to *vtgAo1* and *vtgAo2* in zebrafish. Using the exon primer pairs 6-7, 9-10a, and 6-10b to amplify Atlantic herring

gDNA and cDNA, single major bands were visualized from the PCR products (Fig. 1). The size difference between each of the major gDNA and cDNA bands closely matched the size of the introns (Table 1). No PCR products were obtained using the zebrafish-specific exon primer pairs for exons 4-5 (data not shown). For exon primer pair 9-10a, a shadow band was noted approximately 110 bp above the major gDNA band, an observation repeated with exon primer pair 9-10b (data not shown). However, no such shadow bands were observed for the PCR products generated from exon primer pair 6-10b. To test whether the shadow bands represent a novel form of *vtg*, all bands were submitted to sequencing. The data revealed that a small polymorphic intron (83 vs 98 bp) exists between exons 9 to 10 (Table 1), while the shadow band had the same sequence length and identity as the major bands, despite the size difference. Analyses of >100 clones further revealed two variants differing in sequence identity by ~5%. These data thus show that at least two closely related gene variants exist in the genome, but that only a single form of complete-type *vtg* is expressed in Atlantic herring.

Table 1: Intron identity scores and lengths for Atlantic herring genomic *chvtgAc* between exons 6-9. Two polymorphic introns are shown for *chvtgAc* intron $9^{1,2}$. Intron sequences are compared to zebrafish *vtgAo1* and *vtgAo2* genes.

intron	6	7	8	9^1	9^2
ID to *vtgAo1* (%)	36	38	48	34	37
ID to *vtgAo2* (%)	47	38	44	7	9
chvtgAc length (bp)	103	91	88	83	98

Alignment of the deduced aa sequence of the Atlantic herring Vtg against full length Atlantic halibut VtgAa and VtgAb and common carp VtgAo2 revealed that the Atlantic herring Vtg represents a complete type Vtg containing the LvH, Pv, LvL, ß´ and CT subdomains (Fig. 2). In accordance with the recently proposed classification of vertebrate Vtgs, we named the Atlantic herring protein sequence ChvtgAc, where the final "c" represents Clupeiformes type Vtg.

Figure 2. Multiple sequence alignment of the deduced amino acid sequence of Atlantic herring (*Clupea harengus*) ChvtgAc, Atlantic halibut (HhvtgAa, ABQ58113, HhvtgAb, ABQ58114) and common carp (VtgAo2, BAD51933). Identical residues are boxed in dark grey while similar residues have light grey background. The sequences are annotated with known and predicted cleavage sites based on acanthomorph data (Finn, 2007a, b). The Vtg receptor-minimal interaction domain (Vtgr-MID) is identified in accordance with the findings of Li *et al*. (2003). Mass spectrometry and N-terminal sequence results are annotated as follows: black solid bars for N-terminal sequences, light grey bars for mass spectrometry hits (Q-Trap and Q-Tof) for Yp1 and Yp2, dark grey bars for Yp7 and Yp9. Predicted N-linked glycosylation and mucin-type O-linked glycosylation sites were identified after Blom *et al*. (2004) and Julenius *et al*. (2004), respectively.


Signal peptide
LvH
HhvtgAa
HhvtgAb
CcvtgAo2
ChvtgAc
q-TRAP Yp1 & Yp2
Vtgr MID
q-TRAP Yp11
q-TRAP Yp1&2
Missing C-C bridge
OGlyc
N-term Yp12
N-term Yp10 & Yp17
Pv
Ooc LvL-CS
LvL-MCS
N-term Yp8 & Yp16
q-TOF Yp6 & Yp16
ß´
q-TOF FA Yp6
"QEY"
q-TOF FA Yp4
NGlyc

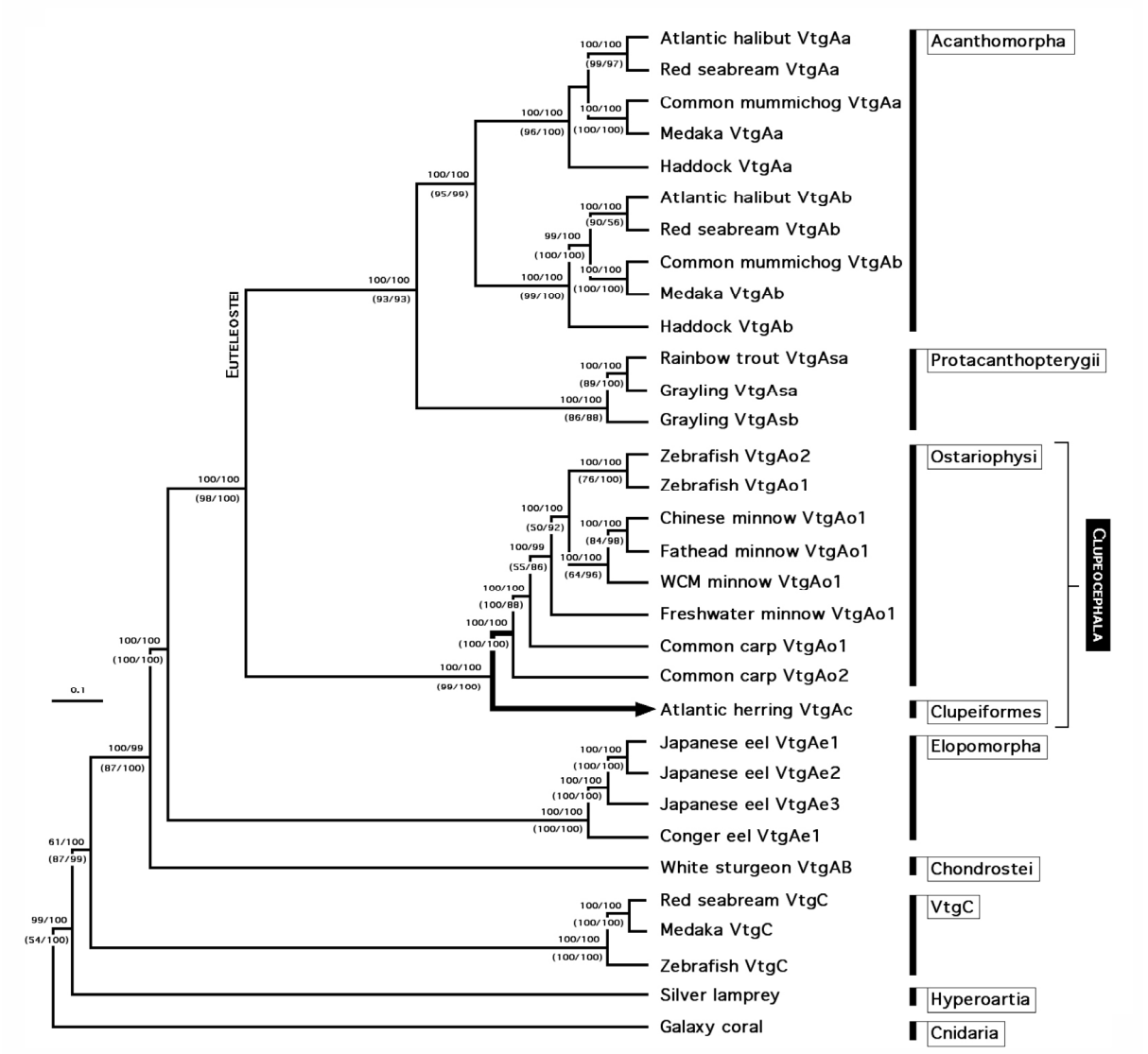

Figure 3. Combined Bayesian majority rule consensus (amino acid (aa) and codons), maximum likelihood (codons), and maximum parsimony (aa and codons) phylogenetic tree for Atlantic herring (*Clupea harengus*) vitellogenin in relation to other fishes. Galaxy coral is used as outgroup. Numbers above nodes indicate protein/codon Bayesian posterior probabilities (consensus of 6,500 trees from 650,000 generations; burnin 3,500; total generations = 1,000,000). Numbers below nodes in parentheses are parsimony bootstrap values (1000 bootstraps) of aa/codon alignments. See methods for further details of phylogenetic analyses. Accession numbers for taxa: Atlantic halilbut VtgAa: ABQ58113, Red seabream VtgAa: AB181838, Common mummichog VtgAa: U07055, Medaka VtgAa: AB064320, Haddock VtgAa: AF284035, Atlantic halibut VtgAb: ABQ58114, Red seabream VtgAb: AB181839, Common mummichog VtgAb: FHU70826, Medaka VtgAb: AB0744891, Haddock VtgAb: AF284034, Rainbow trout VtgAsa: X92804, Grayling VtgAsa: AF545753, Grayling VtgAsb: Zebrafish VtgAo2: ENSDARP00000061164, Zebrafish VtgAo1: AF406784, AF545754, Chinese minnow VtgAo1: EF639846, Fathead minnow VtgAo1: AF130354, White cloud mountain (WCM) minnow VtgAo1: ABN13867, Freshwater minnow VtgAo1: EU100019, Common carp VtgAo1: AF414432, Common carp VtgAo2: AB106873, Japanese eel VtgAe1: AY775788, Japanese eel VtgAe2: AY423445, Japanese eel VtgAe3: AY423444, Conger eel: AB185334; White sturgeon VtgAB: U00455, Red seabream VtgC: AB181840, Medaka VtgC: ENSORLP00000008173, Zebrafish VtgC: AF254638, Silver lamprey Vtg: M88749, Galaxy coral Vtg: BAD74020.

Multiple phylogenetic analyses of the codons and aa alignments consistently clustered the Atlantic herring sequence as the basal sister group to the Ostariophysi with high posterior probabilities for the Bayesian analyses, and high bootstrap values for the maximum

parsimony analyses (Fig. 3). The topology of the tree was further corroborated by maximum likelihood analysis of the codon alignment (data not shown). Neighbour-joining methods proved unable to resolve the phylogenetic relationships and gave both polytomies and clustering incongruent with current understanding of teleost phylogeny (data not shown).

The close relationship between the Atlantic herring Vtg and ostariophysan Vtgs was supported by highest identity and similarity scores between the groups (Fig. 4). Truncated ostariophysian Vtg had the highest similarity (73%) and identity (57%) values among extant teleost taxa. These similarity and identity scores increased to 78% and 62% for the LvH domain and 84% and 69% for the Vwfd domain, respectively. To compare the homology of the Atlantic herring *chvtgAc* and zebrafish *vtgAo1* and *vtgAo2* transcripts, we removed the Vwfd region from *vtgAo2*. This revealed that *chvtgAc* was equally (66.5%) homologous to both zebrafish genes. Similarly, the *chvtgAc* intron identities did not distinguish between the *vtgAo1* and *vtgAo2* forms (Table 1).

	Acanthomorpha VtgAa	Acanthomorpha VtgAb	Protacanthopterygii VtgAsa	Ostariophysi	Elopomorpha	VtgC	Chondrostei	Hyperoartia	Cnidaria
Identity									
Truncated	50	50	56	57	54	25	38	26	15
LvH	55	55	62	61	60	27	45	31	17
LvL	42	42	45	53	51	22	33	25	9
Vwfd	62	60	68	69	65		26	30	11
Similarity									
Truncated	67	68	71	73	69	46	56	44	30
LvH	72	74	76	78	76	51	65	52	32
LvL	64	66	66	72	67	41	52	43	26
Vwfd	79	78	80	84	78		40	48	23

Figure 4. Mean similarity and identity scores for the putative Atlantic herring (*Clupea harengus*) ChvtgAc against truncated (LvH-Pv-LvL), lipovitellin heavy chains (LvH), lipovitelin light chains (LvL) and Von Willebrand type D domains (Vwfd) of piscine and cnidarian vitellogenins (Vtg) organized according to phylogenetic rank. See legend to Fig. 3 for accession numbers.

In order to understand the structural relationships between the parent Vtg precursor and the deposited Yps (Fig. 5a), the putative Atlantic herring Vtg protein was modelled using the CBS prediction 3D server. A positive hit that gave lower than threshold ($E < 0.05$) was only found for the 3D structure of silver lamprey lipovitellin (PBD identifier 1LSH, with a score of 558, E: $1e^{-158}$ for 1LSH chain A and a score of 81, E: $6e^{-15}$ for 1LSH chain B). Aligning the ChvtgAc to the lamprey structure mask in Cn3D revealed that, despite a low identity between the LvH domains (31%, Fig. 4), the conformational structure of the herring Vtg appears to be highly conserved (Fig. 5b). The 338 identical residues that were identified

between the two modelled subdomains were not clustered in any particlar region, but were distributed primarily (73%) in secondary structures throughout the LvH.

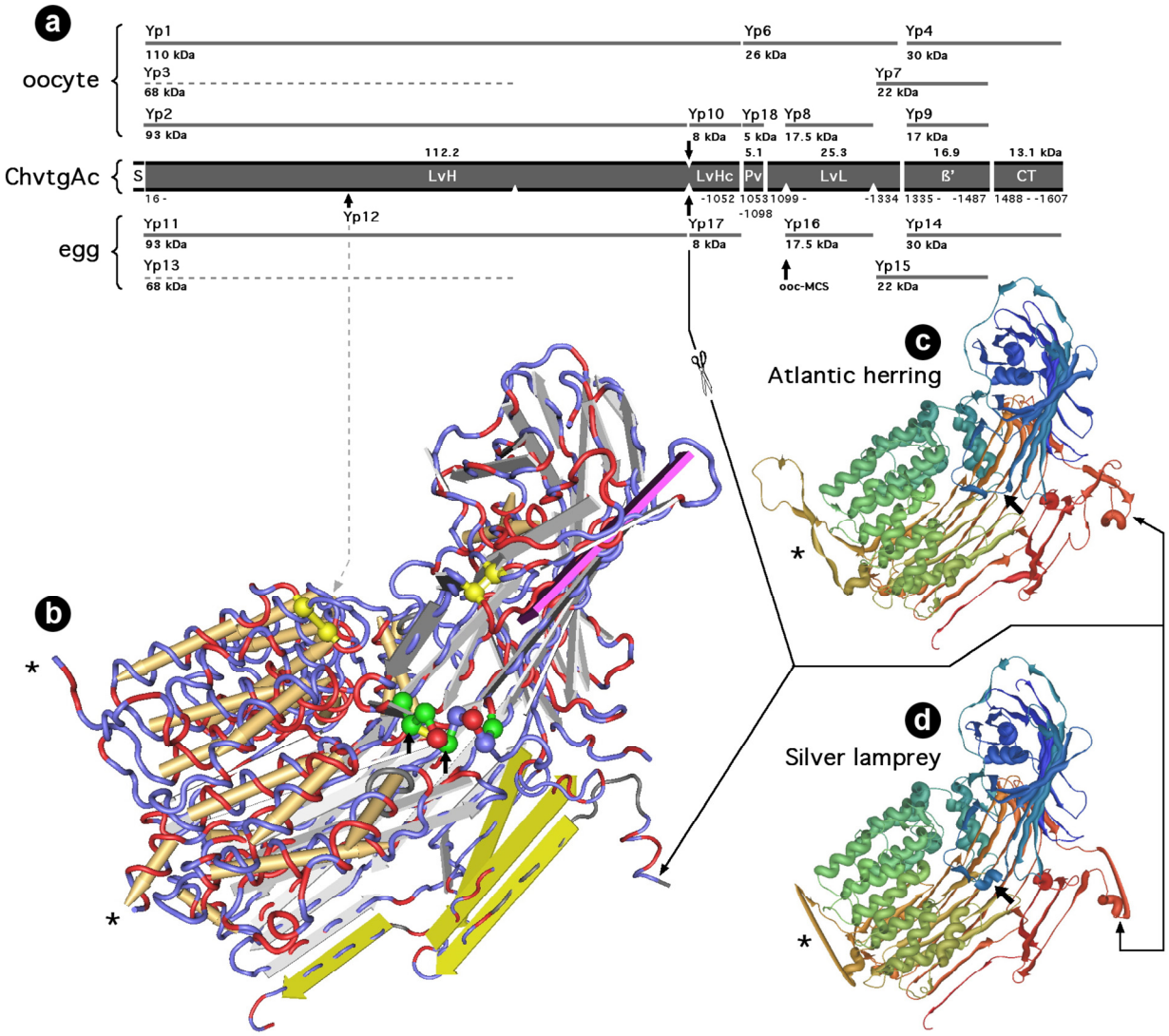

Figure 5. Linear-scale and three-dimensional models of Atlantic herring (*Clupea harengus*) vitellogenin in relation to oocyte and egg yolk proteins and resolved crystal structure of lamprey lipovitellin. (**a**) Linear-scale reconstruction of the deduced domain structure of Atlantic herring vitellogenin (ChvtgAc). Major yolk proteins (Yp) identified by N-terminal microsequencing, mass spectrometry, Western immunoblotting and electrophoresis are shown for the oocyte and egg. (**b**) Three-dimensional map showing the salient features of the ChvtgAc heavy chain. The molecule is folded in relation to the lamprey lipovitellin structure mask, and rendered as worms in Cn3D. Red indicates identical residues, blue indicates non-identical residues. Cys residues are shown in yellow rendered as ball and stick, and linked where known disulfide bridges exist. Missing residues in the ChvtgAc sequence are rendered in green ball and stick, and vertical arrows highlight the missing Cys residues. Asterisks highlight the second region of conformational disparity described in the text. β-sheets at the bottom of the structure are colored yellow to illustrate the N-terminus of the cleaved LvHc peptide (Thr_{959}: Yp10 in the oocyte and Yp17 in the ovulated egg). (**c**) Cartoon render of the lipovitellin heavy chain of ChvtgAc orientated as in panel B to illustrate regions of conformational disparity (asterisk and arrowed LvHc cleavage site). (**d**) Cartoon render of lamprey lipovitellin orientated as in panel B to illustrate the regions of conformational disparity (asterisk and arrowed ChvtgAc LvHc cleavage site).

Interestingly the majority (63%) of these fully conserved residues were internal rather than at the aqueous interface. Three regions that showed conformational disparity were noted. The first was due to the absence of two highly conserved cystines (1LSHA: Cys_{198} and Cys_{201}

from the signal Met) and three other residues in the ChvtgAc sequence (Fig. 5b). These cystines form a disulfide bridge that stabilizes a small α-helix at the base of the N-sheet in lamprey lipovitellin (Fig. 5d). This α-helix is lacking in the same region of the putative ChvtgAc molecule (Fig. 5c). The second region of conformational disparity forms an external β-sheet that loops out between β-sheets 4 and 5 at the base of the A-sheet (Figs. 5b-d). The third region that differed between the two molecules represented an extended β-sheet leading to a loop and short α-helix that traverses the outside of the lipid pocket between the C-sheet and A-sheet (not modelled in Fig. 5b, and obscured at the back of the molecules in Figs 5c and 5d). Based on the CBS prediction services, only a single, non-conserved, N-linked glycosylation site (Asn_{1582}) was identified in the CT region of ChvtgAc.

Separation of the oocyte and egg proteins by SDS-PAGE revealed several major bands and many minor bands, but comparatively few band shifts between the two stages (Fig. 6a). In the OV egg, however, the 110 kDa (Yp1) and 15 kDa bands that were present in the oocyte had disappeared, while a stronger signal for the 8 kDa (Yp17) protein was observed. Despite repeated attempts to identify the 15 kDa band by N-terminal microsequencing, MS and Western immunoblotting, the identity of this peptide remains elusive. Separate analyses of the oocyte and egg Yp profiles by silver staining corroborated the CBB stains (Fig. 6b). In the silver stains the oocyte 8 kDa and 26 kDa (Yp6) bands were better visualized compared to the CBB stained gels. In order to identify which proteins were Yps, each major and minor band was submitted to N-terminal microsequencing in duplicate. Only Yp8 (17.5 kDa) and Yp10 (8 kDa) in the oocyte, and Yp12 (88 kDa), Yp16 (17.5 kDa) and Yp17 (8 kDa) in the OV egg yielded N-terminal data (Fig. 2, Fig. 6a). These data revealed that Yp8 in the oocyte and Yp16 in the egg had identical N-termini located at the conserved acanthomorph oocyte maturational cleavage site (ooc-MCS) of the LvL (Ala_{1140} in ChvtgAc). Similary Yp10 and Yp17 were found to have identical N-termini that mapped to Thr_{959} in the C-terminal region of the ChvtgAc LvH. This cleavage site occurs between an extended β-sheet and short α-helix that leads to the final four β-sheets of the LvH A-sheet prior to the Pv domain (Fig. 5b-d). A minor protein in the OV egg (Yp12) was mapped to Pro_{370} in the α-helical domain of the LvH. This site occurs 1 aa downstream from Thr_{369} which is a predicted mucin-type O-linked glycosylation site.

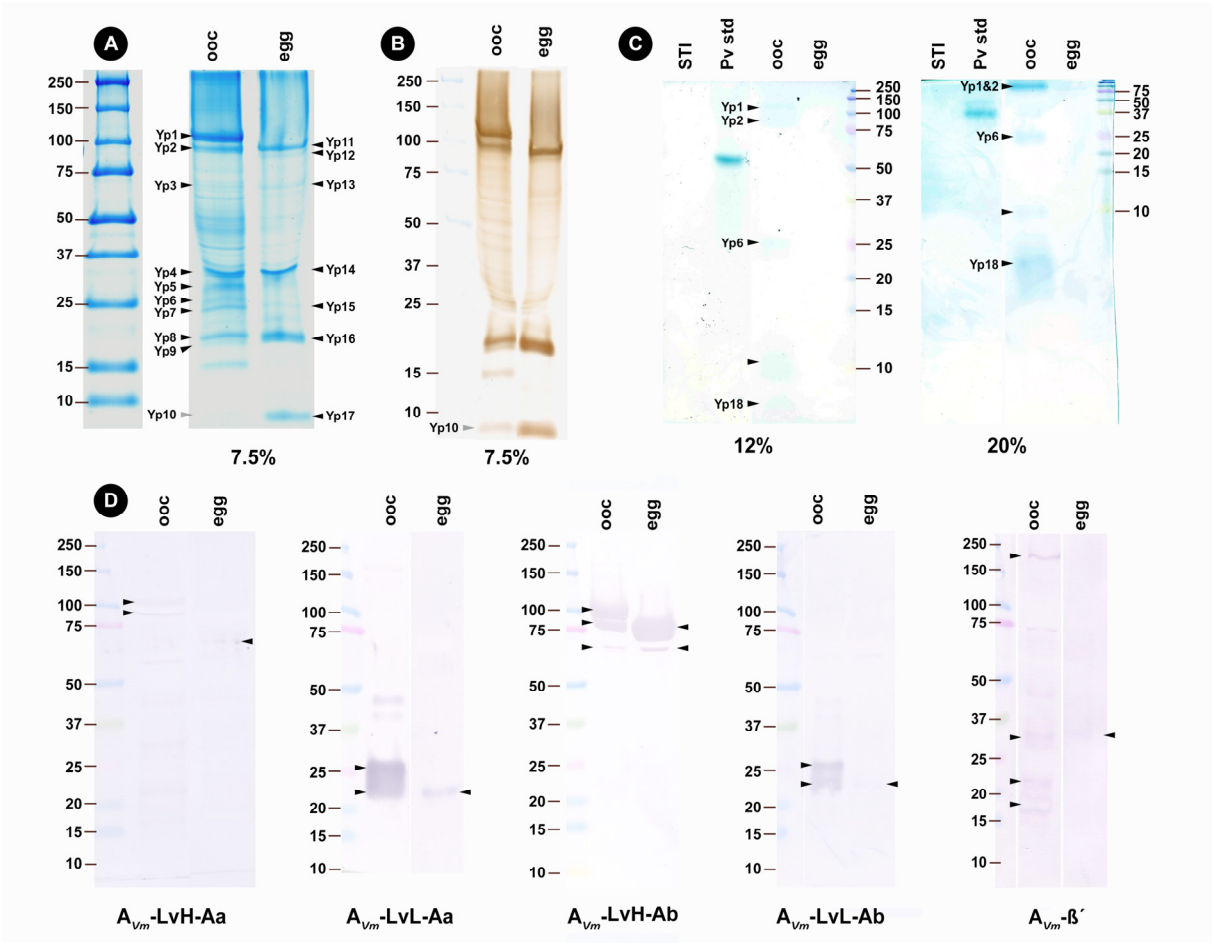

Figure 6. Tris-tricine electrophoretic profiles and Western immunoblots of oocyte and egg yolk proteins of Atlantic herring (*Clupea harengus*). A) Coomassie Brilliant Blue stained gels (7.5% homogeneous); molecular weight markers (kDa) are shown to the left. B) Silver stains of oocyte and egg yolk proteins (7.5% homogeneous). C) Phosphoprotein stains of oocyte and egg yolk proteins (12% and 20% homogeneous: Pv: phosvitin; STI: serine trypsin inhibitor). D) Western immunoblots (7.5% homogeneous) using antibodies raised against purified barfin flounder (*Verasper moseri*) yolk proteins. ooc: oocyte; egg: ovulated egg; See materials and methods for further details.

To further identify Yps, tryptic digests of excised CBB stained bands were studied by MS (q-TRAP and q-TOF) and by western immunoblotting using antibodies raised against purified barfin flounder Yps. The MS data yielded several peptides from Yp1, Yp2, Yp4, Yp6 and Yp8 in the oocyte, and Yp11 and Yp16 in the OV egg. Peptides from Yp1, Yp2 and Yp11 were precisely mapped to regions of the ChvtgAc LvH, while peptides from Yp6, Yp8 and Yp16 were precisely mapped to regions of the ChvtgAc LvL (Fig. 2). To verify that Yp6 in the oocyte is the LvL, the second peptide (TVDVIVK) was verified by q-TOF FA sequencing (Fig. 7a). These data are fully congruent with the N-terminal analyses and also suggest that Yp6 contains the full LvL with a predicted mass of 25.3 kDa, while Yp8 in the oocyte and Yp16 in the OV egg are processed variants. Yp4 (30 kDa) contained two peptides that mapped to the CT domain of ChvtgAc. Since this is the first molecular evidence for the

presence of the CT domain in the deposited Yp pool of any teleost, we verified the peptides by q-TOF FA sequencing (Fig 7b).

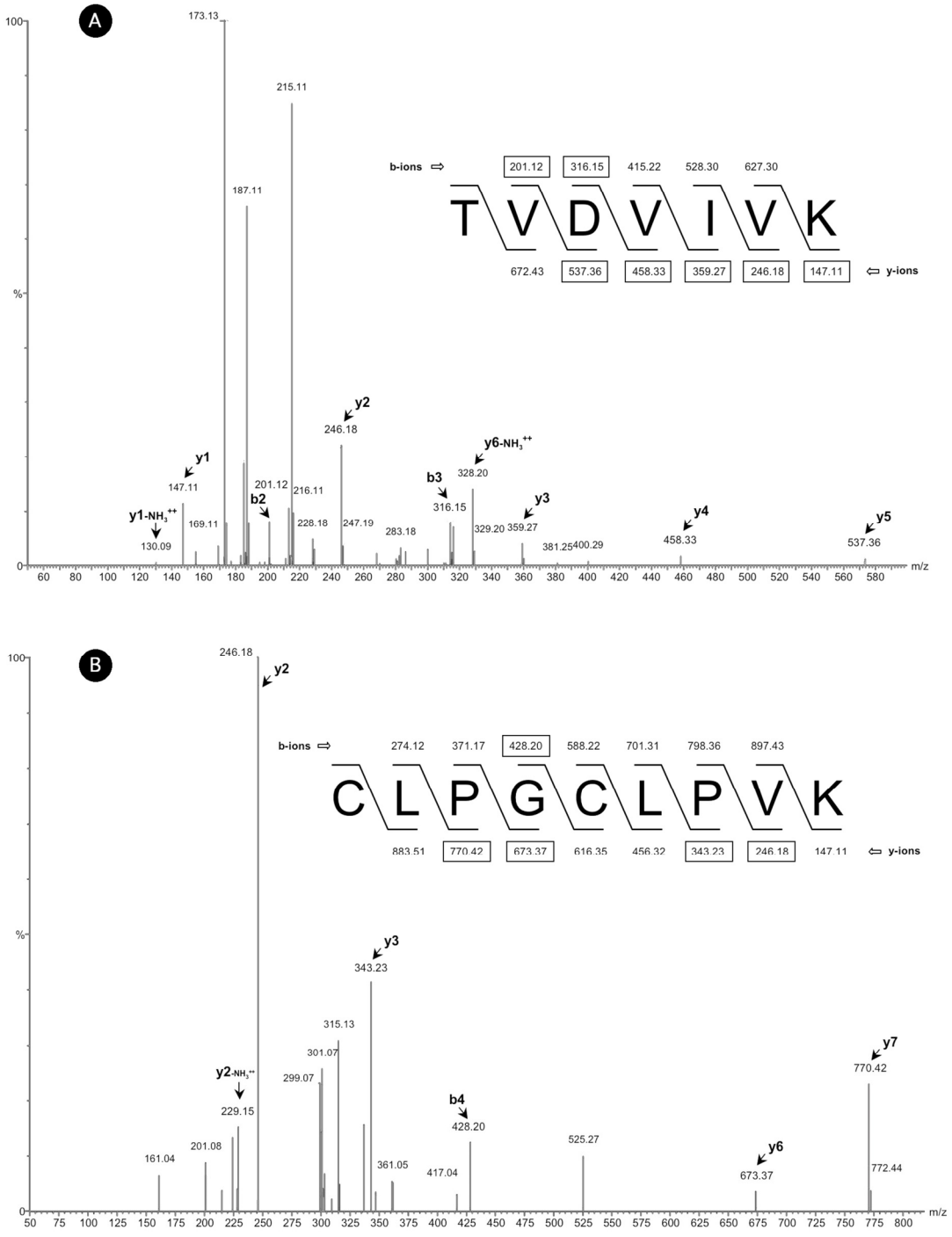

Figure 7. Mass spectrometry (MS/MS) fragment analyses collected from a hybrid triple quadrupole linear ion trap of A) Yp6: lipovitellin light chain and B) Yp4: containing the C-terminal coding region of Atlantic herring (*Clupea harengus*). The expected fragmentation patterns are shown inset with verified b and y ions outlined in boxes. See materials and methods for further details.

Two bands (Yp1 and Yp2) in the oocyte were weakly immunoreactive to A_{Vm}-LvH-Aa after 3 hr incubations, while one band (Yp13; 68 kDa) in the OV egg was weakly immunoreactive to this antibody (Fig. 6d). These reactions were much stronger to A_{Vm}-LvH-

Ab (bands were visualized within 5 min) with the addition of the strongest reaction to Yp11 in the OV egg. A further reaction to this antibody was noted in the oocyte (Yp3, 68 kDa) indicating that Yp3 and Yp13 are minor cleavage variants of the LvH. Immunoblots against the LvL antibodies (A_{Vm}-LvL-Aa and A_{Vm}-LvL-Ab) revealed strong reactions to Yp6 and Yp8 in the oocyte, and weaker reactions to Yp 16 in the OV egg and thus validated the MS and N-terminal sequence data. In some blots, a reaction was also noted to Yp5 (28 kDa) (data not shown) indicating that Yp5 is an LvL cleavage variant. In contrast to the data for LvH antibodies, reactions to A_{Vm}-LvL-Aa (22 min incubation) were stronger than A_{Vm}-LvL-Ab (100 min). Detection of the ß´ was achieved after 3 hr reactions to A_{Vm}-ß´. Four bands were visualized in the oocyte: a 180 kDa band that corresponded to the full ChvtgAc, a 30 kDa band that corresponded to Yp4, a 22 kDa band that corresponded to Yp7, and a 17 kDa band that corresponded to Yp9. Only the 30 kDa (Yp 13) band was visualized in the OV egg (Fig. 6d).

The deduced primary structure of ChvtgAc shows it to have a fully conserved KKIL site and FxK LvL cleavage site (ooc-LvL-CS) that flank a small Pv domain. This polyserine region contains 56% serine residues (23 mol mol^{-1} Pv), of which 83% are predicted to be phosphorylated. This Pv domain is the smallest recorded to date with a deduced M_r of 5.1 kDa. A band of this size was recorded in 12% and 20% tris-tricine gels following phosphoprotein staining (Fig. 6c), but could not be seen in 7.5% gels using the same buffer system (data not shown). Four other bands were observed in the 12% and 20% gels: 110 and 93 kDa bands that corresponded to Yp1 and Yp2; 26 kDa band that corresponded to Yp6, and an 11 kDa band that was not seen in the CBB or silver stained gels. No bands could be visualized for OV eggs.

Discussion

Here we report for the first time, a full-length sequence of *vtg* from a basal clupeocephalan. We found that unlike other teleosts, Atlantic herring expresses only a single form of *vtg*, with a deduced aa sequence conforming to the linear NH_2-(LvH-Pv-LvL-ß´-CT)-COO^- pentapartite structure of teleost complete Vtgs. A single *vtg* form in this group of teleosts is consistent with preliminary reports of only a single *vtg* transcript or a single eluted

peak of Vtg in Pacific herring (*Clupea pallasii*) (Koya *et al.* 2003; Matsubara *et al.* 2003). Since it is known that the Ostariophysi and the more recently evolved Acanthomorpha express multiple forms of *vtg*, including *vtgC* (Wang *et al.* 2000, 2005, Mikawa *et al.* 2006, Miracle *et al.* 2006, Kang *et al.* 2007), we used a GSP and exon primer-pair based strategy to search for further *vtg* transcripts or genomic variants in Atlantic herring. Two genomic variants were found that differed in exon identity by ~ 5%, and contained a small polymorphic intron between exons 9-10 (Table 1). Conversely the zebrafish *vtgAo1* and *vtgAo2* genes showed 10% variability in the same exon regions, and contains polymorphic introns between each of the exons 6-10, including a highly polymorphic variant between exons 9-10 (103 bp: *vtgAo1*; 574 bp: *vtgAo2*). Taken together, our data suggest that Atlantic herring may have cis-duplicated genes, or allelic variants, but has only a single form of *vtg*. In the context of vertebrate *vtg* evolution, the loss of *vtgC* from Atlantic herring would fit the recently proposed scheme of Finn & Kristoffersen (2007). Further, comparison of the introns and truncated (without the Vwfd region) transcripts with zebrafish *vtgAo1* and *vtgAo2* revealed that *chvtgAc* is equally homologous to both zebrafish genes. We thus propose that duplication of ostariophysan *vtgAo1* and *vtgAo2* genes occurred in a lineage-specific manner after separation of the Ostariophysi from the Clupeiformes.

In further support of this notion, the phylogenetic analyses of the Atlantic herring *chvtgAc* and ChtgAc show that it clusters as a basal sister group to the Ostariophysi in full congruence with current understanding of teleost phylogeny and clupeocephalan rank (Inoue *et al.* 2003; 2005; Nelson 2006; Mabee *et al.* 2007).

Despite a comparatively low identity score with silver lamprey lipovitellin, the conformational structure of ChvtgAc appears to be highly conserved. As found previously for Atlantic halibut (Finn 2007b), the fully conserved residues are primarily located in secondary structures rather than loops. The observation that the majority of the fully conserved residues are also located internally suggests that the outside of the protein is evolving at a faster rate. Given that Vtgs are secreted to the extracellular fluids for systemic distribution, the driving force for such adaptation could be the extracellular solute environment of each species. Such an environment would support hydrophilic aa substitution, while inner residues are conserved to stabilise the tertiary structure and provide lipophilic binding sites within the lipid pocket. It is further noteworthy that few glycosylation sites are conserved between species (Finn 2007b). In the present context, only a single N-linked glycosylation site (Asn_{1582}) was predicted for the CT region and two mucin-type O-glycosylation sites (Thr_{369} and Thr_{786}) in

the LvH domain of ChvtgAc. This contrasts the semi-conserved sites identified in the Pv and LvL domains (Finn 2007b). Few studies have identified which residues are glycosylated in Vtgs, but both vertebrate and invertebrate Vtgs are glycosylated (Gottlieb & Wallace 1982; Don-Wheeler & Engelmann 1997; Khalaila *et al.* 2004), and the major function is thought to enhance solubility of the lipoprotein following secretion. Since we have not been able to demonstrate the presence of the CT domain in the deposited Yps (see below), it remains to be established whether the secreted ChvtgAa lipoprotein is glycosylated at this Asn_{1582}-Ile_{1583}-Ser_{1584} sequon. Our observation that the N-terminus of Yp8 occurs 1 aa downstream of a predicted mucin-type O-glycosylation site (Julenius *et al.* 2005) is intriguing, since the N-terminal Pro_{370}, which occurs on the outside of the upper part of the α-helical domain, could be glycosylated and explain the high M_r of this peptide in the gels. In order to explain the size of this OV egg band based on the linear structure of Vtg, it would extend beyond the Pv domain into the LvL. However neither of the LvL antibodies reacted to this band, nor was it visualized in the Pv stains. Dephosphorylation of the Yps, which is known to occur during oocyte maturation (Sawaguchi *et al.* 2006b; Finn 2007a), could explain the latter observation, but not the former. Owing to the surprisingly low number of predicted glycosylation sites, however, it remains to be established whether Atlantic herring Vtg is glycosylated.

In recent studies of the evolution of vertebrate *vtg* genes, it has been shown that Cys residues are particularly highly conserved (Finn & Kristoffersen, 2007). The current finding that the second pair of fully conserved Cys residues (Cys_{198} and Cys_{201}) are lacking in ChvtgAc is novel for a complete type Vtg. These residues form a disulfide bridge that stabilizes a small α-helix at the base of the N-sheet (Thompson & Banaszak 2002) that leads to the β-sheets flanking the Vtg receptor minimal interaction domain (Vtgr-MID) identified by Li *et al*. (2003). While the lack of this Cys-doublet may have important implications for Vtg-Vtgr interactions in the Atlantic herring, the putative 3D model suggests that conformational adjustments are minor. The second region of conformational disparity in the ChvtgAc protein forms an external β-sheet that loops out between β-sheets 4 and 5 at the base of the A-sheet. This region represents the cleavage site that separates the C-terminal fraction of the LvH (Yp6 and Yp10). This site in Atlantic herring occurs 4 aa downstream from the same locus in Atlantic halibut VtgAb (Finn 2007b), and is thus a homologous LvHc domain. The four β-sheets that remain (coloured yellow in Fig. 5b) are at the centre of the Vtg dimer at the back of the A-sheet, and their separation from the rest of the LvHn domain (Yp2) may be important for the unlinking of the Vtg homodimer, and/or release of lipid cargo following

deposition or during oocyte maturation. Since Yp2 (LvHn) and Yp6 (LvHc) are derivatives of Yp1, the full LvH domain, some processing of this region occurs during vitellogenesis. However, the absence of Yp1 and intensifying of Yp7 and Yp10 in the OV egg demonstrates that the majority of processing into the LvHn and LvHc subdomains occurs during oocyte maturation. The amphipathic nature of these four β-sheets comprising the LvHc does not preclude their continued function in maintaining solubility of the deposited or released lipids.

In the Atlantic halibut, the Pv domain of the HhvtgAb paralogue remains attached to the LvHc-Ab subdomain, but to the LvL-Aa domain of the HhvtgAa paralogue (Finn 2007b). This seems not to be the case for the LvHc in Atlantic herring as no band of a size corresponding to the LvHc-Pv (~13 kDa) was observed in the Pv stains. Due to the modular structure of vertebrate Vtgs, it was possible to identify the Pv domain between the conserved cleavage sites, KKIL and FxK ooc-CS (Finn 2007b). The Pv domain of the Atlantic herring is the smallest reported to date, with a calculated size of only 5.1 kDa. This finding is consistent with an earlier prediction of a 4.2 kDa phosphoprotein in Pacific herring, based on sedimentation equilibrium data (Inoue *et al.* 1971). Only the fathead minnow, a member of the Clupeacephala: Ostariophysi, with a Pv of 5.3 kDa has a similarly small domain (Korte *et al.* 1999). In addition to the smallest Pv domain, the Atlantic herring also has the smallest LvH domain (112.2 kDa) of any complete type Vtg sequenced to date. These findings support the hypothesis of a functional relationship between the two domains (Finn 2007b).

Several bands were visualized using phosphoprotein specific stains, all of which were visualized in the oocyte and none in the egg (see Fig. 6c). The absence of any positive bands in the egg is consistent with maturational de-phosphorylation and an increase in PO_4^{3-} during oocyte hydration in Atlantic herring (Kristoffersen & Finn 2008). To separate and visualize the bands both 12% and 20% gels were run. The 26 kDa band (Yp6) in the oocyte also reacted to both of the acanthomorph LvL antibodies suggesting that the Atlantic herring Pv is attached to the LvL as found for other species (Matsubara *et al.* 1999; Sawaguchi *et al.* 2006a; Amano *et al.* 2007; Finn 2007a). LvH had several predicted phosphorylated sites, and two bands matching Yp1 and Yp2 in size could be seen in the oocyte. The observed 10 kDa and 5 kDa bands does match a temporal cleavega variant of Yp18 that still retains part of the LvL from the ooc-LvL to the ooc-MCS, while the 5 kDa matches the PV domain; Yp18. The identities of Yp8 and Yp16 were established by N-terminal microsequencing and MS. These data demonstrate that both proteins are identical fragments of the LvL cleaved at the

conserved ooc-MCS. The intensification of Yp16 in the OV egg, but the depletion of Yp6 from all of the gels, including the western immunoblots and phosphoprotein stains suggests a causal relationship. It has been shown for all pelagophils and several benthophils studied to date (see Finn 2007b), that the LvL is processed during oocyte maturation, this also appears to be true for Atlantic herring.

The ß´ of teleost Vtgs is a conserved Cys-rich domain that is homologous to Vwfd. Under non-reducing conditions it has shown to be present in dimeric form (~30 kDa) that resolves into monomers (~17 kDa) under reducing conditions (Hiramatsu *et al.* 2002). Owing to the comparatively high conservation of this domain, the use of non species-specific antibodies have proved successful in identifying this Yp (Hiramatsu *et al.* 2002; Amano *et al.* 2007; Kolarevic *et al.* 2008). The ß´ is clearly present in the deduced primary structure of ChvtgAa and is the most highly conserved of all of the domains among the teleost taxa. It was therefore surprising not to find any bands immunoreactive to the A_{Vm}-ß´. It presently remains unclear whether the ß´ is incorporated into the oocytes of Atlantic herring. As co-members of the Clupeocephala, it is interesting to note that the Vwfd has been lost in the major *vtg* gene (*vtgAo1*), but not the minor gene (*vtgAo2*) of the Ostariophysi (Wang *et al.* 2005, Finn & Kristoffersen 2007, Kang *et al.* 2007). The loss of this domain would have occurred after separation from the last common ancestor of the Clupeiformes and Ostariophysi and after the cis-duplication of the *vtgAo1* and *vtgAo2* genes.

It is now well established that the maturational proteolysis of primarily the LvH-Aa domain in marine pelagophils generates an organic osmolyte pool of FAA that drives hydration of the oocyte prior to ovulation and oviposition. We have previously argued that this mechanism was a key adaptation of acanthomorph marine pelagophils to the hyperosmotic spawning environment of the oceans (Finn & Kristoffersen 2007). With the exception of the common mummichog (Greeley *et al.* 1986; 1991; McPherson *et al.* 1989; LaFleur *et al.* 2005) no study has examined the combined ionic and proteolytic events during oocyte hydration of a marine benthophil. Our separate study of the major osmolyte changes during oocyte hydration in the more ancestral Atlantic herring showed that inorganic ions are the major driving force for water influx (Kristoffersen & Finn 2008). However, a small pool of FAA (Δ1.8% of dry mass) was also found to contribute significantly to the osmotic pressure of the OV egg. This increase is only fractional compared to pelagic eggs, where Δ15-16% of dry mass is typical (Finn *et al.* 2002a,b). Due to the lower water content of the Atlantic herring OV egg (70% of wet mass) compared to a typical pelagic egg (>90% of wet

mass), the small pool of FAA contributed 29% to the total osmolarity. The present finding of maturational processing of the major Yps suggests that Yp depolymerization is the source of the small pool of FAA. The region that cannot be accounted for in the OV egg is the Pv domain, together with an N-terminal fragment of the LvL. These regions are also known to be proteolysed to FAA in pelagophils (Finn 2007a). Further evidence in support of this fractional degradation comes from the significant increase in free P_i and free Ser in the OV egg (Kristoffersen & Finn 2008).

We thus argue that yolk proteolysis and the generation of an organic osmolyte pool of FAA was an adaptive response to spawning in seawater prior to the rise of Acanthomorpha. In support of our previous study (Finn & Kristoffersen 2007), however, this mechanism was not evolutionarily successful in terms of biodiversity until gene duplication and neofunctionalization occurred in the Acanthomorpha.

Acknowledgments

The authors thank Rune Male for providing degenerate primers from salmonid *vtg*, Takahiro Matsubara (Hokkaido National Fisheries Research Institute, Japan) for providing the barfin flounder antibodies, Stig Mæhle and Heidi Kongshaug (Institute of marine Research, Bergen) for technical assistance, Mathew Betts and David Liberles (The Computational Biology Unit, Bergen) for technical assistance and Olav Mjaavatten and Magnus Arntzen (PROBE, Department of Biomedicine, Bergen) for help with the mass spectromic analyses.

References

Amano H, Fujita T, Hiramatsu N, Shimizu M, Sawaguchi S, Matsubara T, Kagawa H, Nagae M, Sullivan CV, Hara A. 2007. Egg Yolk Proteins in Gray Mullet (*Mugil cephalus*): Purification and classification of multiple lipovitellins and other vitellogenin-derived yolk proteins and molecular cloning of the parent vitellogenin genes. J. Exp. Zool. 307A:324–341

Anderson TA, Levitt DG, Banaszak LJ. 1998. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. Structure 6:895-909.

Babin PJ, Carnevali O, Lubzens E, Schneider WJ. 2007. Molecular aspects of oocyte vitellogenesis in fish. In: Babin PJ, Cerdà J, Lubzens E (eds) The Fish Oocyte: From Basic Studies to Biotechnological Applications. Springer, Dordrecht, The Netherlands, p 39-76.

Blom, N., Gammeltoft, S., and Brunak, S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294:1351-1362.

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 4:1633-1649.

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

Craik JCA, Harvey SM. 1984. Biochemical changes occuring during final maturation of eggs of some marine and freshwater teleosts. J. Fish Biol. 24(5):599-610.

Davis LK, Hiramatsu N, Hiramatsu K, Reading BJ, Matsubara T, Hara A, Sullivan CV, Pierce AL, Hirano T, Grau EG. 2007. Induction of three vitellogenins by 17beta-estradiol with concurrent inhibition of the growth hormone-insulin-like growth factor 1 axis in a euryhaline teleost, the tilapia (*Oreochromis mossambicus*). Biol. Reprod. 77:614–625.

Ding JL, Hee PL, Lam TJ. 1989. Two forms of vitellogenin in the plasma and gonads of male *Oreochromis aureus*. Comp. Biochem. Physiol. 93B:363-370.

Don-Wheeler G, Engelmann W. 1997. The Biosynthesis and Processing of Vitellogenin in the Fat Bodies of Females and Males of the Cockroach *Leucophaea maderae*. Insect Biochem. Mol. Biol. 27: 901-918.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2:953 - 971.

Fabra M, Raldúa D, Bozzo MG, Deen PMT, Lubzens E, Cerdà J. 2006. Yolk proteolysis and aquaporin-1o play essential roles to regulate fish oocyte hydration during meiosis resumption. Dev. Biol. 295:250-262.

Finn RN. 2007a. Vertebrate yolk complexes and the functional implications of phosvitins and other subdomains in vitellogenins. Biol. Reprod. 76, 926–935.

Finn RN. 2007b. The maturational disassembly and differential proteolysis of paralogous Vitellogenins in a marine pelagophil teleost: A conserved mechanism of oocyte hydration. Biol. Reprod. 76, 936–948.

Finn RN, Kristoffersen BA. 2007. Vitellogenin gene duplication in relation to the “3R hypothesis”: Correlation to the pelagic egg and the oceanic radiation of teleosts. PLoS ONE PLoS ONE 2(1): e169. doi:10.1371/journal.pone.0000169.

Finn RN, Østby GC, Norberg B, Fyhn HJ. 2002a. *In vivo* oocyte hydration in Atlantic halibut (*Hippoglossus hippoglossus*), proteolytic liberation of free amino acids, and ion transport, are driving forces for osmotic water influx. J. Exp. Biol. 205:211-224.

Finn RN, Wamboldt M, Fyhn HJ. 2002b. Differential processing of yolk proteins during oocyte hydration in marine fishes (Labridae) that spawn benthic and pelagic eggs. Mar. Ecol-Prog. Ser. 237:217-226.

Fujiwara Y, Fukada H, Shimizu M, Hara A. 2005. Purification of two lipovitellins and development of immunoassays for two forms of their precursors (vitellogenins) in medaka (*Oryzias latipes*). Gen. Comp. Endocrinol. 143:267-277.

Gottlieb TA, Wallace RA. 1982. Intracellular glycosylation of vitellogenin in the liver of estrogen-stimulated *Xenopus laevis*. J. Biol. Chem. 257:95-103.

Greeley MS Jr, Calder DR, Wallace RA (1986) Changes in teleost yolk proteins during oocyte maturation: correlation of yolk proteolysis with oocyte hydration. Comp. Biochem. Physiol. B 84: 1-9

Greeley MS Jr, Calder DR, Wallace RA. 1991. Changes in size, hydration and low molecular weight osmotic effectors during meiotic maturation of *Fundulus* oocytes *in vivo*. Comp. Biochem. Physiol. 100A:639-674.

Hiramatsu N, Hara A, Hiramatsu K, Fukada H, Gregory M. Weber GM, Denslow ND, Sullivan CV. 2002. Vitellogenin-derived yolk proteins of white perch, *Morone americana*: purification, characterization, and vitellogenin-receptor binding. Biol. Reprod. 67:655-667.

Inoue S, Kaneda-Hayashi T, Sugiyama H, Ando T. 1971. Studies of phosphoproteins from fish eggs. J. Biochem. 69:1003-1011.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2005. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol. Phylogen. Evol. 26:110-120.

Inoue JG, Miya M, Venkatesh B, Nishida M. 2005. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene S349:227-235.

Julenius K, Mølgaard A, Gupta R, Brunak S. 2004. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 15:153–164.

Kang BJ, J.-H. Jung JH, Lee JM, Lim SG, Saito H, Kim MH, Kim YJ, Saigusa M, Han CH. 2007. Structural and expression analyses of two vitellogenin genes in the carp, *Cyprinus carpio*. Comp. Biochem. Physiol. (in press) doi:10.1016/j.cbpb.2007.07.088

Khalaila I, Peter-Katalinic J, Tsang C, Radcliffe CM, Aflalo ED, Harvey DJ, Dwek RA, Rudd PM, 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.

Kolarevic J, Nerland A, Nilsen F, Finn RN. 2008. Goldsinny wrasse (*Ctenolabrus rupestris*) is an extreme *vtgAa*-type pelagophil teleost. Mol Reprod Dev DOI 10.1002/mrd.20845.

Korte JJ, Kahl MD, Jensen KM, Pasha MS, Parks LG, LeBlanc GA, Ankley GT. 1999. Fathead minnow vitellogenin: Complementary DNA sequence and messenger RNA and protein expression after 17β-estradiol treatment. Environ. Toxicol. Chem. 19:972–981.

Koya Y, Soyano K, Yamamoto K, Obana H, Matsubara T. 2003. Oocyte development and serum profiles of vitellogenin and steroid hormone levels in captive female Pacific herring *Clupea pallasii* during their first maturational cycle. Fish. Sci. 69:137-145.

Kristoffersen BA, Finn RN. 2008. Major osmolyte changes during oocyte hydration of a clupeocephalan marine benthophil: Atlantic herring (*Clupea harengus*). Mar. Biol. (in press) DOI:10.1007/s00227-008-0961-8.

LaFleur GJ Jr, Byrne BM, Kanungo J, Nelson LD, Greenberg RM, Wallace RA. 1995. *Fundulus heteroclitus* vitellogenin: the deduced primary structure of a piscine precursor to noncrystalline, liquid-phase yolk protein. J. Mol. Evol. 41:505-521.

LaFleur GJ Jr, Raldúa D, Fabra M, Carnevali O, Denslow N, Wallace RA, Cerdà J. 2005. Derivation of major yolk proteins from parental vitellogenins and alternative processing during oocyte maturation in *Fundulus heteroclitus*. Biol. Reprod. 73:815-824.

Li C, Ortí G. 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. Mol. Phylogen. Evol. 44:386–398.

Li A, Sadasivam M, Ding JL. 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.

Lund O, Nielsen M, Lundegaard, C. Worning P. 2002. X3M a Computer Program to Extract 3D Models. CASP5: A102.

Mabee PM, Arratia G, Coburn M, Haendel M, Hilton EJ, Lundberg JG, Mayden RL, Rios N, Westerfield M. 2007. Connecting evolutionary morphology to genomics using ontologies: a case study from Cypriniformes including zebrafish. J. Exp. Zool. (Mol Dev Evol) 308B: DOI 10.1002/jez.b

Maissey J. 1996. Discovering fossil fishes. New York: Westview Press.

Matsubara T, Ohkubo N, Andoh T, Sullivan CV, Hara A. 1999. Two forms of vitellogenin, yielding two distinct lipovitellins, play different roles during oocyte maturation and early development of barfin flounder, *Verasper moseri*, a marine teleost that spawns pelagic eggs. Dev. Biol. 213:18-32.

Matsubara T, Nagae M, Ohkubo N, Andoh T, Sawaguchi S, Hiramatsu N, Sullivan C, Hara A. 2003. Multiple vitellogenins and their unique roles in marine teleosts. Fish. Physiol. Biochem. 28:295-299.

McPherson R, Greeley MS, Wallace RA. 1989. The influence of yolk protein proteolysis on hydration in the oocytes of *Fundulus heteroclitus*. Dev. Growth. Differ. 31:475-483.

Mouchel N, Trichet V, Betz A, Le Pennec JP, Wolff J. 1996. Characterization of vitellogenin from rainbow trout (*Oncorhynchus mykiss*). Gene 174:59-64.

Mouchel N, Trichet V, Naimi BY, Le Pennec JP, Wolff J. 1996. Structure of a fish (*Oncorhynchus mykiss*) vitellogenin gene and its evolutionary implication. Gene 197:147-142.

Nelson JS. 2006. Fishes of the world: John Wiley and Sons, Inc: New York.

Ohkubo N, Andoh T, Mochida K, Adachi S, Hara A, Matsubara T. 2004. Deduced primary structure of two forms of vitellogenin in Japanese common goby (*Acanthogobius flavimanus*). Gen. Comp. Endocrinol. 137:19-28.

Reith M, Munholland J, Kelly J, Finn RN, Fyhn HJ. 2001. Lipovitellins derived from two forms of vitellogenin are differentially processed during oocyte maturation in haddock (*Melanogrammus aeglefinus*). J. Exp. Zool. 291B:58-67.

Sawaguchi S, Koya Y, Yoshizaki N, Ohkubo N, Andoh T, Hiramatsu N, Sullivan CV, Hara A, Matsubara T. 2005. Multiple vitellogenins (Vgs) in mosquitofish (*Gambusia affinis*): Identification and characterization of three functional Vg genes and their circulating and yolk protein products. Biol. Reprod. 72:1045-1060.

Sawaguchi S, Kagawa H, Ohkubo N, Hiramatsu N, Sullivan CV, Matsubara T. 2006a. Molecular characterization of three forms of vitellogenin and their yolk protein products during oocyte growth and maturation in red seabream (*Pagrus major*), a marine teleost spawning pelagic eggs. Mol. Reprod. Dev. 73(6):719-736.

Sawaguchi S, Ohkubo N, Matsubara T. 2006b. Identification of two forms of vitellogenin-derived phosvitin and elucidation of their fate and roles during oocyte maturation in the barfin flounder, *Verasper moseri*. Zool. Sci. 23:1021-1029.

Selman K, Wallace RA, Cerdà J. 2001. Bafilomycin A1 inhibits proteolytic cleavage and hydration but not yolk crystal disassembly or meiosis during maturation of sea bass oocytes. J. Exp. Zool. 290:265-278.

Sharrock WJ, Rosenwasser TA, Gould J, Knott J, Hussey D, Gordon JI, Banaszak L. 1992. Sequence of lamprey vitellogenin. Implications for the lipovitellin crystal structure. J. Mol. Biol. 226:903-907.

Shimizu M, Fujiwara Y, Fukada H, Hara A. Purification and identification of a second form of vitellogenin from ascites of medaka (*Oryzias latipes*) treated with estrogen. J. Exp. Zool. 293:726–735.

Thompson JR, Banaszak LJ. 2002. Lipid-protein interactions in lipovitellin. Biochemistry 41:9398-9409.

Trichet V, Buisine N, Mouchel N, Moran P, Pendas AM, Le Pennec JP, 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. S263:828-837.

Yadetie F, Arukwe A, Anders Goksoyr A, 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.