

Molecular evolution and functional characterisation of insulin related peptides in molluscs: Contributions of *Crassostrea gigas* genomic and transcriptomic-wide screening

Maëva Cherif-Feildel^{a,b}, Clothilde Heude Berthelin^{a,b}, Beatrice Adeline^{a,b}, Guillaume Rivière^{a,b}, Pascal Favrel^{a,b}, Kristell Kellner^{a,b,*}

^a Normandy University, Caen, France

^b University of Caen Normandie, Unity Biology of Organisms and Aquatic Ecosystems (BOREA), MNHN, Sorbonne University, UCN, CNRS, IRD, Esplanade de la Paix, 14032 Caen, France

ARTICLE INFO

Keywords:

Insulin related peptides
Crassostrea gigas
Molluscs
Phylogeny
Gametogenesis

ABSTRACT

Insulin Related Peptides (IRPs) belong to the insulin superfamily and possess a typical structure with two chains, B and A, linked by disulphide bonds. As the sequence conservation is usually low between members, IRPs are classified according to the number and position of their disulphide bonds. In molluscan species, the first IRPs identified, named Molluscan Insulin-related Peptides (MIPs), exhibit four disulphide bonds. The genomic and transcriptomic data screening in the Pacific oyster *Crassostrea gigas* (Mollusc, Bivalvia) allowed us to identify six IRP sequences belonging to three structural groups. Cg-MIP1 to 4 have the typical structure of MIPs with four disulphide bonds. Cg-ILP has three disulphide bonds like vertebrate Insulin-Like Peptides (ILPs). The last one, Cg-MILP7 has a significant homology with *Drosophila* ILP7 (DILP7) associated with two additional cysteines allowing the formation of a fourth disulphide bond. The phylogenetic analysis points out that ILPs may be the most ancestral form. Moreover, it appears that ILP7 orthologs are probably anterior to lophotrochozoa and ecdysozoa segregation. In order to investigate the diversity of physiological functions of the oyster IRPs, we combine *in silico* expression data, qPCR measurements and *in situ* hybridization. The Cg-ilp transcript, mainly detected in the digestive gland and in the gonadal area, is potentially involved in the control of digestion and gametogenesis. The expression of Cg-mip4 is mainly associated with the larval development. The Cg-mip transcript shared by the Cg-MIP1, 2 and 3, is mainly expressed in visceral ganglia but its expression was also observed in the gonads of mature males. This pattern suggested the key roles of IRPs in the control of sexual reproduction in molluscan species.

1. Introduction

Since their discovering in mammals, the molecules belonging to the insulin family and their potential physiological roles have given rise to a huge interest. The abundant literature dealing with this subject occurs for two main reasons. Firstly, insulin family members are involved in the control of a large variety of physiological processes including reproduction, development, growth, carbohydrate and lipid metabolism (reviewed in Agrogiannis et al., 2014; Dimitriadis et al., 2011; Saltiel and Kahn, 2001; Sliwowska et al., 2014). The involvement of insulin in the control of life span has also been demonstrated (Avogaro et al., 2010). Secondly, insulin belongs to a superfamily including a large number of members. Insulin and Insulin-like Growth Factors (IGFs)

were first described in mammals and other members were later identified as relaxins and a variety of related peptides (Bathgate et al., 2013; Shabanpoor et al., 2009; Wilkinson et al., 2005). Insulin related peptides (IRPs) were also found in almost all animal phyla, so this family of molecules is of particular interest for phylogenetic analysis and study of the evolutionary history. From its large phylogenetic repartition, Jekely (2013) concluded that IRP may be ancestral to all metazoan species.

Invertebrate IRPs sequences have firstly been identified based on biochemical purification or cDNA identification thus producing a large number of IRPs of various taxonomic clades. Another approach relies on genomic and transcriptomic-wide screening. This enabled to complete the panel of IRPs in some species or to identify sequences in new species. These different approaches helped identifying and

* Corresponding author.

E-mail address: kristell.kellner@unicaen.fr (K. Kellner).

<https://doi.org/10.1016/j.ygcen.2018.10.019>

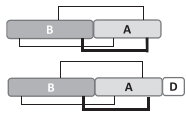
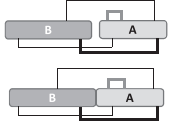
Received 9 July 2018; Received in revised form 25 October 2018; Accepted 26 October 2018

Available online 30 October 2018

0016-6480/ © 2018 Elsevier Inc. All rights reserved.

Table 1

Schematic structure of the main putative mature IRPs sequences, according to the classification of Matsunaga (2017). All IRPs groups have the 2 canonical disulfide bridges connecting the A and B chains. a and b types have an additional inter A and B chains disulfide bridge, only the b type include a fourth intra A chain disulfide bridge. A and B chains may be linked or not depending on the presence and cleavage of a C peptide between them. A supplementary D domain is also present in several members. AG: androgenic gland; IAG: Insulin-like Androgenic gland peptide, ILP: insulin like peptide, IRP: insulin related peptides, RLP: relaxin like peptide. R: Review article, U: Unpublished NCBI data.

			α type		β type	γ type		
			2 canonic disulfide bridges between A and B chains					
			No		1 disulfide intra-A chain bridge			
			1 additional disulfide bridge between A and B chains			No		
Various mature forms								
Species and phylogeny								
			Bibliography					
Chordata	Vertebrata	<i>H. sapiens</i>				Insulin, IGFs, INSL, Relaxin	Shabanpoor et al. (2009) R	
	Tunicata	<i>C. intestinalis</i>				Ci-INS-L1,Ci- INS-L2, Ci-INS-L3	Olinski et al. (2006)	
Ecdysozoa	Nematoda	<i>C. elegans</i>	INS-20 to 31, INS 33 to 36, INS 39	INS 1 to 10, Daf28	INS 11 to 19, INS 32, INS 37, INS 38		Pierce et al. (2001), Li et al (2003), Matsunaga et al. (2017) R	
	Hexapods	<i>B. mori</i>			Bombyxins family A to G		Maruyama et al. (1988) Kondo et al. (1996) Yoshida et al. (1997, 1998), Bayazit (2009) R	
		<i>S. gregaria</i>			Scg-IRP		Badisco et al. (2013)	
		<i>L. migratoria</i>			ILP		Lagueux et al. (1990)	
		<i>A. aegyptii</i>			AaILP 1 to 8		Riehle et al. (2006) R	
		<i>D. melanogaster</i>			DILP 1 to 8		Brogiolo et al. (2001), Grönke et al. (2010) R	
	Crustacea	<i>S. verreauxi</i>			Sv-IAG , Sv-ILP1, Sv-ILP2		Chandler et al. (2017)	
		<i>A. vulgare</i>			Arv AGH (with additional intra B disulfide bridge)		Ventura et al. (2011)	
		<i>M. rosenbergii</i>		Mr IAG	Mr ILP, Mr RLP		Phoungpetchara et al. (2011), Ventura et al. (2009)	
Lophotrochozoa	Mollusca	<i>L. stagnalis</i>		MIP 1, 2, 3, 4, 5, 7			Smit et al. (1988, 1991, 1992, 1993, 1996, 1998), Li et al. (1992)	
		<i>A. californica</i>		AI			Floyd et al., 1999	
		<i>C. geographus</i>		Con-Ins G2	Con-Ins G1		Safavi-Hemami et al. (2015)	
		<i>S. officinalis</i>		So IRP			Zatylny-Gaudin et al. (2016)	
		<i>C. gigas</i>		oIRP			Hamano et al. (2005)	
		<i>L. gigantea</i>		ins 1, ins 2, ins 4	ins 3		Veenstra (2010)	
	Annelida	<i>P. dumerilii</i>		IRP2, IRP3, IRP4, IRP5	IRP1		Conzelmann et al. (2013)	
	Platyhelminths	<i>S. mediterranea</i>			ilp-1		Miller and Newmark (2012)	
		<i>T. solium</i>			ILP-1, ILP-2 (with an additional intra A disulfide bridge)		Wang et al. (2013)	
		<i>S. mansoni</i>			ILP (with an additional intra A disulfide bridge)		Wang et al. (2013)	
Porifera	<i>G. cydonium</i>				sponge insulin		Robitzki et al. (1989)	
Cnidaria	<i>N. vectensis</i>				NV 199266; NV 207484		Anctil (2009)	
	<i>H. magnipapillata</i>				Hv-ILP1; Hv-ILP2; Hv-ILP3		Steele et al. (2009)- U	

(See above-mentioned references for further information.)

characterizing IRPs in a large number of species belonging to all clades of the phylogenetic tree. The most studied members can be found in Table 1.

A comparative analysis of IRP sequences revealed that the percentage of identity among species as well as between insulin related members within one species is relatively low. However, the IRPs display common features. First, they are composed of two chains named B and A which are linked by at least two canonical disulfide bonds (De

Meyts, 2004). Chains B and A originated in the post-translational proteolytic cleavage from a unique pre-propeptide that contained a signal peptide and some facultative additional domains named C peptide, D and E domains. The C peptide is located between the B and A sequences and may be cleaved or not in the mature form. The D and E domains are located at the C-terminal extremity, the latter one being post-translationally cleaved through dibasic residues. One of the most diversified series of IRPs was described in the nematode *Caenorhabditis elegans* and

led to the description of 40 insulins (INS) named INS-1 to INS-39 and DAF28 (Matsunaga, 2017; Pierce et al., 2001). *C. elegans* IRPs can be assigned into three groups named α , β and γ according to their structural properties (Duret et al., 1998; Matsunaga, 2017). The IRPs belonging to α and β types have an additional disulphide bond between the A and B chains. In the β type, a fourth disulphide bond occurs between two residues of the A chain (intra-A disulphide bond). In the γ type, only the two canonical disulphide bonds between A and B are present but the intra-A chain bond also exists (Matsunaga, 2017).

So far, the co-existence of the three types of IRPs in the same species has only been reported in *C. elegans*. Interestingly, the situation regarding IRP types is very different depending on the species considered (Table 1). In deuterostomes, IRPs are only assigned to the γ type, as well as in most of ecdysozoan species, with the exception of the crustacea *Macrobrachium rosenbergii* which has a β type IRP. However, the presence of β type IRPs appears more common in lophotrochozoan species. On the contrary, the γ IRPs were scarcely described in lophotrochozoan and only for species with a sequenced genome. Interestingly, this co-existence of both β and γ types might constitute an evolutionary signature of IRPs in protostomian species. Moreover, distribution of IRPs across the phylogenetic tree led to hypothesise that γ type members constitute the ancestral form.

Investigation of physiological roles of IRPs may improve our understanding of their evolutionary history. Indeed, the multiple biological functions of different members of the insulin family observed in vertebrate species are generally conserved in invertebrates. Indeed, protostomian IRPs assigned to different types present similar functions compared to deuterostome IRPs which are all grouped in the γ type. This functional conservation was clearly demonstrated for *Drosophila melanogaster* and *C. elegans*, since these species benefit from functional approaches mainly based on targeted mutations of insulin-like pathway genes (ligands, receptor or effectors of signal) or RNAi experiments. In both species, IRPs played key roles in the control of growth and development (Böhni et al., 1999; Brogiolo et al., 2001; Garelli et al., 2012; Grönke et al., 2010; Li et al., 2003; Shingleton et al., 2005), energy storage (Ashrafi et al., 2003; Broughton et al., 2005), stress resistance (Liu et al., 2016; Söderberg et al., 2011), response to diet restriction (Broughton et al., 2010; Hibshman et al., 2016), lifespan (Apfeld and Kenyon, 1998; Clancy et al., 2001; Grönke et al., 2010; Partridge et al., 2011) and fecundity (Hsu and Drummond-Barbosa, 2009; Luo et al., 2010; Michaelson et al., 2010; Narbonne et al., 2015). Unlike their vertebrate counterparts, the primary source of IRPs in invertebrates is the nervous system. These assertions rely mainly on immunocytochemical and *in situ* hybridization labelling of the IRPs-producing cells located in the nervous ganglia. However, local expressions in other tissues also exist. Some of them are linked to the role of IRPs in the control of the sexual reproduction. Evidence of this can be seen in crustacean species in which IRPs are expressed in the androgenic gland, an organ involved in the control of sexual differentiation (Manor et al., 2007; Ventura et al., 2011). In the planaria *Schmidtea mediterranea*, ilp-1 expression has also clearly been demonstrated in testis lobes and RNAi inhibition of Insulin-like signal resulted in the reduction of the testis lobe size and in a decrease in the spermatid number in mature animals (Miller and Newmark, 2012).

The pacific oyster *Crassostrea gigas* (e.g. *Magallana gigas*) is a mollusc bivalve belonging to lophotrochozoan species. Previous research in this oyster resulted in the identification of three IRPs transcripts issued from the same precursor (Hamano et al., 2005) and signalling pathway related members (Gricourt et al., 2003; Jouaux et al., 2012). The IRP transcripts are expressed in the visceral ganglia during the annual cycle according to the reproductive and energetic status (Hamano et al., 2005). Moreover, the phylogenetic position of this species and its particular reproduction as an alternative hermaphrodite made *C. gigas* an interesting model for the study of IRPs at molecular and functional levels. In addition, our group has already identified the tyrosine kinase receptor for *C. gigas* IRPs (CIR) as well as three effectors of the

signalling pathway and confirmed the involvement of IRPs in the control of growth, reproduction and nutritional status (Gricourt et al., 2003; Jouaux et al., 2012).

Here, we took advantage of the availability of the genome (Zhang et al., 2012) as well as various transcriptomic data in *C. gigas* (Riviere et al., 2015; Zhang et al., 2012, 2014; Zhao et al., 2012) to better understand the evolutionary history of IRP sequences. For this purpose, we first performed *in silico* analyses to explore the range of IRP sequences and precise their structural characteristics. Next, the phylogenetic analysis of molluscan IRPs was performed. Finally, we investigated the expression level and location of the corresponding transcripts using quantitative PCR and *in situ* hybridization in order to gain insights into their physiological roles. Altogether, our results shed light on structural and functional evolution of the insulin superfamily.

2. Material and method

2.1. Analysis of molecular databases

We used a strategy of sequence similarity to identify sequences encoding protein homologous to model vertebrate species (*Mus musculus* INS and IGFs), ecdysozoan species (*Drosophila melanogaster* DILPs) and lophotrochozoan species (*Lymnaea stagnalis* MIP) insulin superfamily members in the predicted proteome (NCBI), in the transcriptome (GIGATON <http://ngspipelines-Sigenae.toulouse.inra.fr>) and genome (NCBI) of *C. gigas* (Thunberg, 1793, taxid: 29159) using reciprocal BLASTP or TBLASTN. We used the BLOSUM62 matrix for NCBI analysis of the databases SWISSPROT for proteomic search, and Oyster_v9 for genomic search. The sequences used for similarity are listed in the Supplementary data Table S1. Predicted expressions were found using Gigaton databases which allow to retrieve annotations, expression levels in adult tissues and developmental stages (obtained by Zhang et al. (2012)) as well as polymorphism data like described in Riviere et al. (2015).

2.2. Multiple sequence alignment and phylogenetic analysis

The protein sequences of the IRPs family identified in *C. gigas* were aligned by the use of Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) using default parameters (Sievers et al., 2011). The phylogenetic analysis was performed with IRP sequences from four other molluscan species (*Lymnaea stagnalis*, *Lottia gigantea*, *Sepia officinalis* and *Aplysia californica*), one vertebrate species (*Homo sapiens*) and one ecdysozoan species (*Drosophila melanogaster*). The evolutionary analyses were conducted in MEGA X (Kumar et al., 2018) using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992) with the Poisson substitution model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985).

2.3. Analysis of protein sequences

The signal peptide was predicted using PREDISI (<http://www.predisi.de/>). KR potential cleavage sites were identified and compared with those of *L. Stagnalis* (Smit et al., 1988, 1991, 1992, 1993, 1996, 1998). A and B chains as well as C peptide were deduced using the same strategy. Molecular weights were calculated using scansite (http://scansite.mit.edu/calc_mw_pi.html) and verified using ProteinProspector (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct>).

2.4. Genomic organization

The WEBSICIO server was used for the prediction of the genomic organization of scaffolds related to IRP sequences in *C. gigas* (Hatje et al., 2011). The RNA organization was deduced from blastn analysis

linked to genomic organization.

2.5. Animals

Diploid Pacific oysters (*C. gigas*) were provided by a commercial oyster farm in Saint-Vaast la Hougue (Manche, France). Experiments were carried out on animals during their second or third cycle of gametogenesis. Thirty oysters were sampled every two months. The ploidy was individually tested using gill samples as recommended by Jouaux et al. (2010).

2.6. Light microscopy

Visceral ganglia and transverse sections of the gonadal area were fixed for histology in Davidson's fixative at 4 °C (10% glycerol, 20% formaldehyde, 30% ethanol 95°, 30% sterile sea water, 10% acetic acid) for 24 to 72 h. The tissue samples were dehydrated in serial ethanol dilutions, transferred in butanol (Carlo Erba) and embedded in paraffin wax (Roth). Three micrometers sections were stained according to the Prenant-Gabe trichrome protocol (Gabe, 1968). The stages of the gametogenetic cycle were determined according to Heude Berthelin et al. (2001).

2.7. RNA extraction and cDNA synthesis

Four pools of three samples were collected for each tissue sample (adductor muscle, digestive gland, labial palps, inner mantle, edge mantle, gills and heart). Similarly, samples were done for visceral ganglia and gonadal area at each stage of gametogenesis. Each pool was crushed and incubated for 5 min at room temperature in TriReagent® (Sigma Aldrich). One hundred microliters of 1-Bromo-3-chloropropane were added, samples were vortexed and centrifuged (15 min at 4 °C). The aqueous phase was recovered and total RNA extraction was performed using the Nucleospin RNA II kit (Macherey-Nagel). RNA samples were reverse transcribed according to Dheilly et al. (2012).

2.8. Real-time quantitative polymerase chain reaction

The sequences of forward and reverse primers were designed using the Primer3 software and folding was analysed (<http://unafold.rna.albany.edu/?q=mfold>). Quantitative PCR was done using 1X GoTaq SYBR Green Mix (Promega), 8.85 ng cDNA and 900 nM of each primer (Table S2) in a final volume of 15 µL. Each run consisted in one cycle of 5 min at 95 °C followed by 45 cycles of 15 sec at 95 °C and 45 sec at 60 °C. The specific amplification of the target sequences was verified by constructing melting curves (80 cycles of 10 sec, temperature increased by 0.5 °C step from 55 °C to 95 °C). Three reference genes (GAPDH, Arf1, EF1α) were tested for qPCR for each tissue and each condition. They were chosen among the best reference genes as identified by Dheilly et al. (2011). EF1α was found as a reliable normalization gene as no significant difference ($p < 0.05$) of Ct values was observed between the different samples compared. Indeed, EF1α was previously used as reference gene in various tissues of *C. gigas* (Dubos et al., 2018, 2016; Fabioux et al., 2004). Thus, the relative level of each gene expression was calculated for one copy of the EF1α reference gene by using the following formula: $N = 2^{-\Delta Ct}$ (reference to EF1α).

2.9. Normalized cDNA library

Total RNA sample was used for double stranded cDNA synthesis using SMART approach (Zhu et al., 2001). SMART prepared amplified cDNA was then normalized using Duplex Specific Nuclease (DSN) normalization method (Zhulidov et al., 2004). Normalization included cDNA denaturation/reassociation treatment by DNS (Shagin et al., 2002) and amplification of normalized fraction by PCR. Resulted normalized cDNA was used for the “all developmental stages and adult

central nervous system” library preparation of *C. gigas* (Fleury et al., 2009).

2.10. RNA probe preparation

The sequences of interest were amplified by PCR using specific primers (Table S2), with a “all developmental stages and adult central nervous system” directional and normalized cDNA library inserted into the Pal 17.3 plasmid (Evrogen) (Fleury et al., 2009). The samples contained 1X PCR Buffer, 0.2 mM dNTP, primers at 0.2 µM each, 4 mM MgCl₂, 10 ng cDNA, 1U Taq Polymerase. The cycling conditions were as follows: 2 min at 96 °C, 40 cycles at 96 °C, 58 °C and 72 °C during 30 sec each followed by a final step at 72 °C. The PCR products were purified from 1% agarose gel using the Nucleospin® Gel and PCR clean-up kit (Macherey-Nagel). Ligation and transformation were realized according to the protocol of the pGEM®-T Easy Vector Systems kit in JM109 chemocompetent bacteria (Promega). The integration of recombinant plasmids in the transformed bacteria was checked by PCR with T7 and SP6 universal primers. The integrated fragment was then validated by sequencing (GENEWIZ Genomics). For each validated colony, a mini-prep was performed using the Nucleospin® Plasmid kit (Macherey-Nagel). Isolated plasmids were linearized with restriction enzyme either Pst I or Sac II during 2 h at 37 °C and 5 min at 65 °C. After clean up purification (Nucleospin® Gel and PCR Clean-up protocol), the RNA probes were labelled using the protocol of the DIG RNA labelling kit (SP6/T7) (Roche).

2.11. In situ hybridization (ISH)

ISH was performed with 3 µm sections mounted on Superfrost Ultra Plus slides. The sections were heated overnight at 37 °C, 2 min at 56 °C and 30 min at room temperature before dewaxing and rehydration in increasing ethanol dilutions. The sections were treated with Proteinase K (Macherey-Nagel) at 5 µg/mL in Tris-EDTA (Ethylenediaminetetraacetic acid, Sigma-Aldrich) Buffer pH 8 for 10 min between two rinses of 5 min in 1X PBS (Phosphate Buffered Saline, P5493 Sigma-Aldrich). Slides were post fixed with a solution of 1X PBS/PFA 4% (Paraformaldehyde, Fisher Scientific 115–867-11) for 7 min and rinsed for 3 min in PBS and 5 min twice in 2X SSC (Saline Sodium Citrate, 93,017 Sigma-Aldrich). The sections were pre-hybridized for 60 min at 50 °C in hybridization buffer consisting of 4X SSC, 10% dextran sulfate, 1X Denhardt's solution, 2 mM EDTA, 50% deionized formamide (F9037 Sigma-Aldrich) and 500 µg/mL salmon genomic DNA (D1626 Sigma-Aldrich) before the hybridization step with 400 ng/mL of RNA probes overnight at 50 °C. The sections were treated once with 2X SSC and three times with 60% formamide in 0.2X SSC each 5 min at 55 °C before being rinsed twice in 2X SSC and once in 100 mM Tris-HCl pH 7.5 with 150 mM NaCl, 5 min each at room temperature. The slides were treated for 30 min with blocking buffer consisting of 100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% blocking reagent before incubation with alkaline phosphatase-conjugated anti-DIG antibody (polyclonal, Fab fragments) diluted 1:2000 in blocking buffer. Two rinses of 5 min were made with 100 mM Tris-HCl, 150 mM NaCl and one of 10 min with detecting buffer containing 100 mM Tris-HCl pH 9.5, 100 mM NaCl and 50 mM MgCl₂. The sections were incubated overnight with a BCIP (5-bromo-4-chloro-3'-indolylphosphate)/NBT (nitro-blue tetrazolium) solution (11 681 451 001 Roche) diluted at 1:50 and rinsed 5 min with Tris-EDTA buffer. The dehydration process was then carried out with two baths in 100% ethanol and 5 min in Roti-Histol before mounting.

2.12. Statistical analysis

Statistical analyses were performed using R 3.4.3 and RStudio software. A Wilcoxon test and a non-parametric one-way analysis with permutation test were done to determinate significant differences

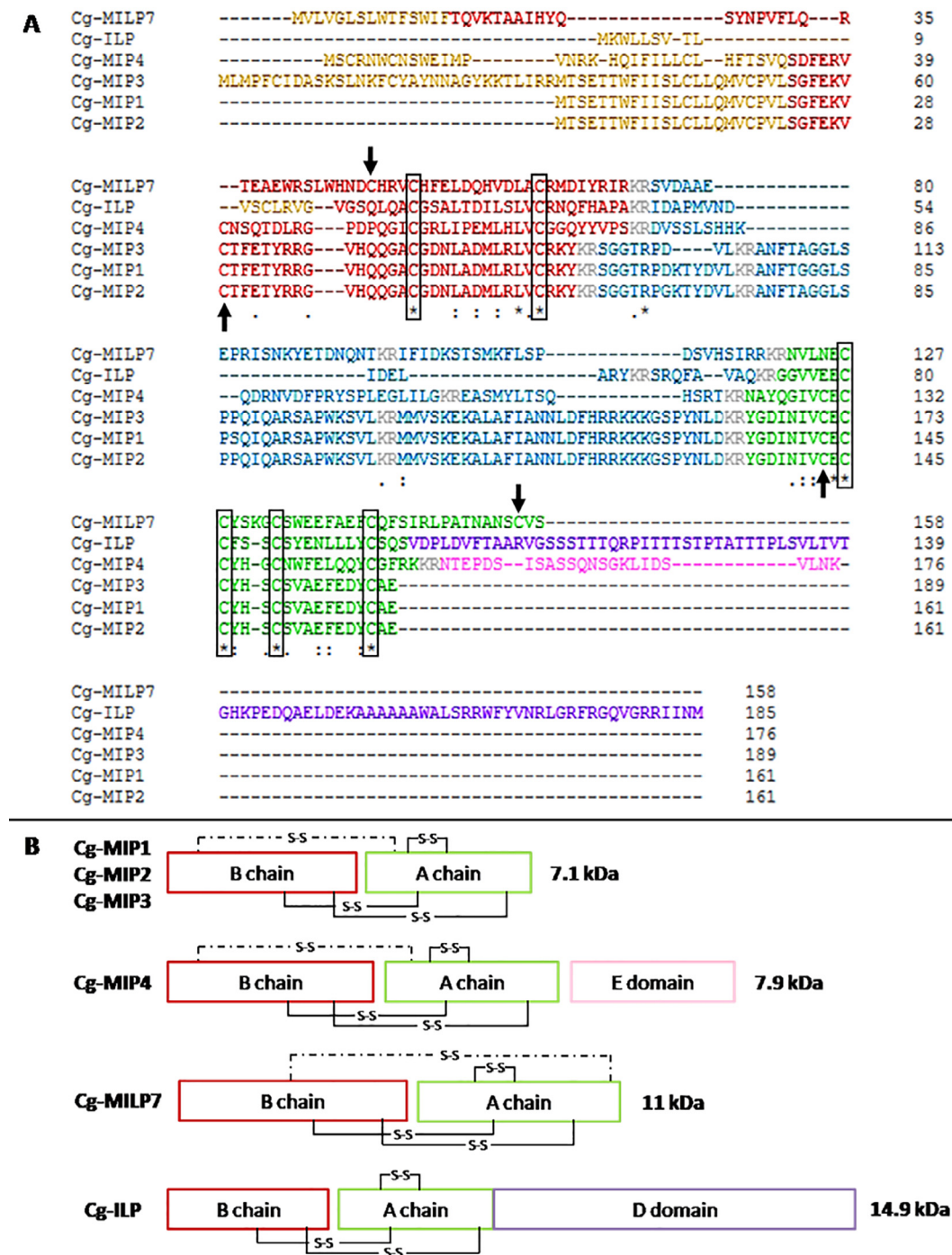


Fig 1. Amino acid alignments (A) and schematic representation of the predicted mature forms (B) for *C. gigas* IRPs. Signal sequences (yellow), B chains (red), C peptides (blue), A chain (green), putative D (purple) and E (pink) domains are specified for each sequence. The putative cleavage sites are either in grey (A) or represented by spaces between chains and domains (B). In the amino acid alignment (A), the six conserved cysteine residues are framed in black and the two additional ones are highlighted by arrows. For the predicted mature forms (B), the three conserved disulphide bonds are annotated by a full line while the putative additional bonds are noted by a dotted line. The estimation of the molecular weight is annotated at the right of each corresponding *C. gigas* IRP.

($\alpha = 0.05$).

3. Results

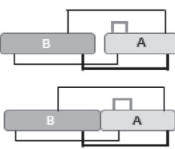
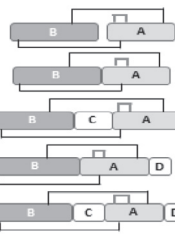
3.1. Identification and classification of IRPs in *C. gigas*

We used sequence similarity and the conserved cysteine residues to identify six sequences of pre-propeptides in the predicted proteome of

C. gigas (NCBI IDs: XP_011455161; BAE48437; XP_011415722; EK18432; EK18433; XP_011435161). The functional analysis with INTERPROSCAN (Quevillon et al., 2005) confirmed that they belong to the insulin superfamily. With these sequences as queries, three mRNA sequences were found in *C. gigas* transcriptome GIGATON (Zhang et al., 2012; Zhao et al., 2012; Riviere et al., 2015) (GIGATON IDs: CHOY-P_ILP.1.1, CHOY-P_MIP5.1.1, and CHOY-P_INS.1.1). In the genome of *C. gigas*, three scaffolds corresponding to these transcripts were detected

Table 2

NCBI accession number, type of IRP with the number of cysteine residues in the B and A chains, signature motif and name for each *C. gigas* IRP.

NCBI IDs	Type	Motif	Name
BAE48437	 β (8 Cys)	C(1X)CC(3X)C(8X)C	Cg-MIP1
XP_011415722			Cg-MIP2
EKC18432			Cg-MIP3
EKC18433			Cg-MIP4
XP_011435161		CC(4X)C(8X)C	Cg-MILP7
XP_011455161	 γ (6 Cys)	CC(3X)C(8X)C	Cg-ILP

(NCBI IDs: NW_011937261.1; NW_011937734.1; NW_011935790.1).

The protein sequences were first aligned to identify conserved domains and structural characteristics of the insulin superfamily (Fig. 1A). The sequences exhibited different signal peptides whose length varies from 15 to 55 amino acids. The B chains, range from 26 to 56 amino acids, always end with a pair of basic amino acids (KR) that allows the cleavage of C peptides which present one or two potential internal cleavage sites. The A chains, ranging from 24 to 37 amino acids, display different signature motifs formed according to the cysteine residues number and position. The identified sequences display 6 or 8 cysteine residues distributed in the B and A chains, which allows the formation of 2 or 3 disulphide bonds between both chains and an additional one intra A chain. Therefore, oyster IRPs belong to both β and γ types (Table 2). Among the six identified sequences, only one (NCBI ID: XP_011455161) has a total of 6 cysteine residues that allows the formation of 3 disulphide bonds. So it belongs to the γ type. Its A chain displays the signature motif of vertebrate Insulin-Like Peptides (ILPs): CC(3X)C(8X)C, this is why it is called Cg-ILP. Moreover, the Cg-ILP sequence presents an uncleaved D domain (88 aa). The other five sequences have 8 cysteine residues allowing the formation of an additional disulphide bonds between A and B chains. For this reason, they belong to the β type. Four of them (NCBI IDs: BAE48437; XP_011415722; EKC18432 and EKC18433) have the signature motif of Molluscan Insulin-related peptides (MIPs) of *Lymnea stagnalis*: C(1X)CC(3X)C(8X)C. Hence, these sequences were called Cg-MIP1 to Cg-MIP4. Cg-MIP1, Cg-MIP2 and Cg-MIP3 exhibit strong identity (> 98%, Table S3). They differ widely from Cg-MIP4 (32.54% identity in average). This last sequence holds an E domain (24 aa) potentially cleaved using the dibasic site in the active form. The last β type sequence (NCBI ID: XP_011435161) presents a different A chain motif CC(4X)C(8X)C and a homology with DILP7 of *Drosophila* stronger than with the other Cg-IRPs (Table S3). The possible existence of four disulphide bonds associated to DILP7 homology led us to name this sequence Cg-MILP7 (Molluscan Insulin-Like Peptide 7). Whereas the identity between sequences is low, the six key cysteine residues are conserved in accordance with insulin superfamily characteristics.

Mature forms were predicted according to the sequence similarity and the identification of cleavage sites (Fig. 1B). The molecular weight of each of these predicted mature forms was calculated. As expected and due to their strong homology, Cg-MIP1, Cg-MIP2 and Cg-MIP3

presented the same molecular weight (7.1 kDa). The Cg-MIP4 differs from the other Cg-MIPs by 16 aa and 24 aa in the A and B chains respectively. The molecular weight of the Cg-MIP4 mature form was estimated at 7.9 kDa after cleavage of the E domain. Cg-MILP7 and Cg-ILP molecular weights were estimated at 11 kDa and 14.9 kDa respectively.

3.2. Genomic organisation of *C. gigas* IRPs

The schematic representation of IRPs genomic location (Fig. S1) showed that IRPs were dispatched on three scaffolds in relation to their structural characteristics. Indeed, all the Cg-MIPs were found on the scaffold 43798. It appeared that Cg-MIP1, Cg-MIP2 and Cg-MIP3 share the same location on this scaffold. Cg-MIP4 was found at a different location, but still on scaffold 43798. Cg-MILP7 and Cg-ILP were localized on scaffold 43,272 and scaffold 737 respectively.

A detailed analysis of the organization of genes, transcripts and peptides confirmed the similarity between Cg-MIP1, Cg-MIP2 and Cg-MIP3 (Fig. 2). Besides sharing the same scaffold and the same transcript, named Cg-mip123 (GIGATON ID: CHOYP_MPI5.1.1), the pre-propeptides structures are close together as well. They all contain B and A chains with three and five cysteine residues respectively and four cleavage sites (KR) in the C peptide and at its extremities. Cg-MIP3 differed from Cg-MIP1 and Cg-MIP2 in that the signal sequence is longer and issued from two different exons (exons 1 and 2) whereas Cg-MIP1 and Cg-MIP2 signal peptides originated from exon 2 only. Both genomic sequences encoding Cg-MIP1 and Cg-MIP2 shared the same predicted organization based on two exons (exons 2 and 3) separated by the intron 2. Interestingly, genomic sequence encoding Cg-MIP4 was located in a different region of scaffold 43798. It displayed three exons and two introns but, while using BLASTN, it appeared to be associated with a different RNA sequence, namely Cg-mip4 (GIGATON ID: CHOYP_INS.1.1). The pre-propeptide sequence Cg-MIP4 appeared to be slightly different from other Cg-MIPs. Its C peptide may be cleaved in two parts instead of three and a potentially cleaved E domain was identified at the C-terminal extremity of the sequence.

The Cg-MILP7 was related to three exons and two introns of scaffold 43272. The associated transcript Cg-milp7 was not present in the databases. Cg-MILP7 presented an amino acid sequence with B and A chains containing three and five cysteine residues respectively and a C peptide potentially cleaved in two parts.

The Cg-ILP was related to its own DNA sequence (scaffold 737) and its own transcript named Cg-ilp (GIGATON ID: CHOYP_ILP.1.1). The genomic sequence possesses three exons and two introns. The amino acid sequence differs due to its six cysteine residues, two of which found on the B chain and four on the A chain. Cg-ILP presents a C peptide, which is potentially cleaved in two parts. A D domain is also present at the C-terminal extremity.

Whatever the IRPs of *C. gigas*, the A and B chains were respectively encoded by two successive exons from the genomic sequences.

3.3. Phylogenetic analysis of molluscan IRPs

The phylogenetic analysis of molluscan IRPs is shown in Fig. 3. For this purpose, we selected IRPs from five molluscan species, *C. gigas* (Cg-MIP1 to 4, Cg-ILP and Cg-MILP7), *S. officinalis* (Insulin), *L. stagnalis* (MIP I to VII and Ls-MILP7), *A. californica* (ILP and Ac-MILP7), *L. gigantea* (Insulin 1 to 4), the vertebrate representative insulin of *Homo sapiens* (Insulin) and the *Drosophila* ILP7 (DILP7). The predicted MILP7 of *A. californica* and *L. stagnalis* were previously described by Veenstra (2010) (ESTs NCBI IDs: EB230021 and ES573590 for *A. californica* and *L. stagnalis* respectively) and the prediction of peptides related to these sequences using EXPASY revealed that they both belong to the MILP7 group. The phylogenetic classification revealed that the Cg-ILP of *C. gigas* and the Insulin 3 of *L. gigantea* are close to each other and grouped with vertebrate insulin. In the same way, the Cg-MIPs are phylogenetically close to the Insulins 1 and 2 of *L. gigantea*, the *S. officinalis* and

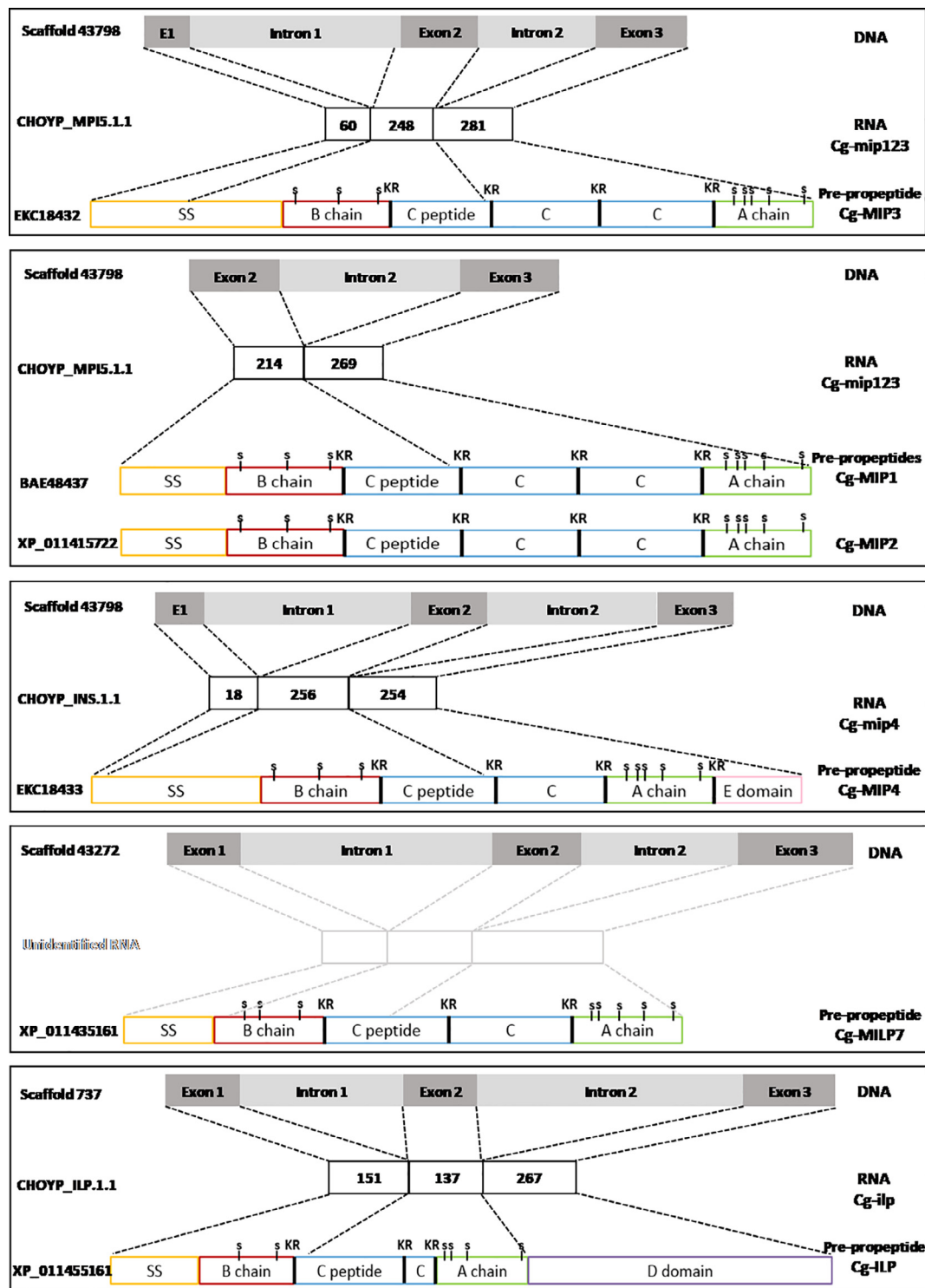


Fig 2. Schematic representation of *C. gigas* scaffolds correlated with transcripts and insulin-like pre-propeptide sequences using WEBSICIO. Cysteine residues (s), cleavage sites (KR) and signal sequences (SS) are annotated. Transcript not found in databases were annotated “Unidentified RNA” and represented in light grey.

A. californica IRPs and all the MIPs of *L. stagnalis*. Cg-MIP1, Cg-MIP2 and Cg-MIP3 were very close with each other but phylogenetically more distant from Cg-MIP4. Likewise, *L. stagnalis* MIP I, MIP II, MIP III, MIP V and MIP VII were grouped together. Cg-MILP7 was grouped with the Insulin 4 of *L. gigantea*, the Ls-MILP7 of *L. stagnalis*, the Ac-MILP7 of *A. californica* and they are phylogenetically close to DILP7.

3.4. Quantification and location of *C. gigas* IRPs expression

The *in silico* expression of transcripts Cg-mip4, Cg-ilp and Cg-mip123 were recorded using GIGATON expression data for the developmental stages of *C. gigas* (Fig. 4) obtained by Zhang et al. (2012a,b). Cg-mip4 and Cg-ilp were found to be expressed in all developmental stages. Both transcripts presented the same patterns but Cg-mip4 expression level is clearly higher. The Cg-mip4 transcript (and to a lesser extent Cg-ilp) increased from the beginning of the trochophore larvae

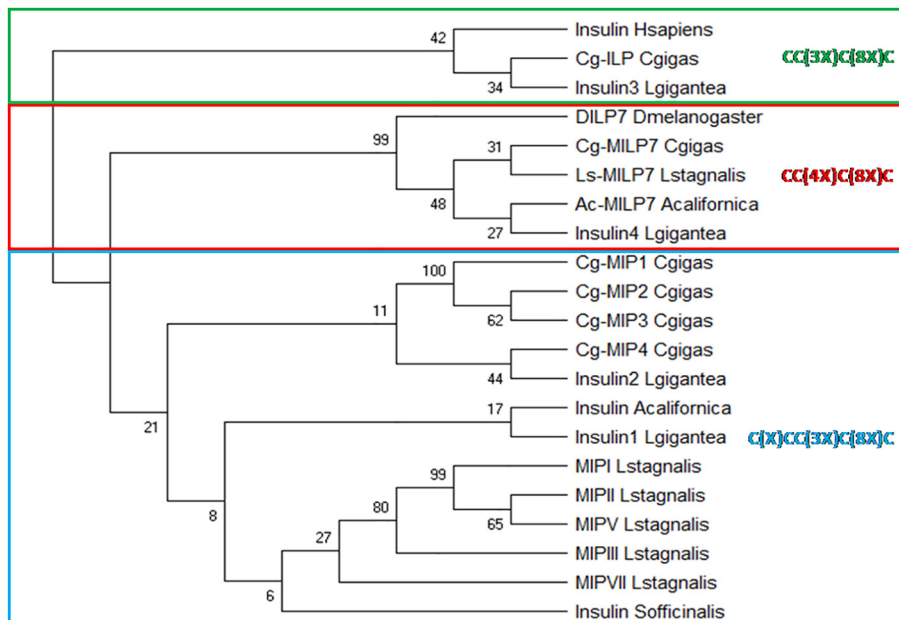


Fig 3. Phylogenetic analysis of some vertebrate, ecdysozoan and molluscan IRPs. *Homo sapiens* (Hsapiens), *Drosophila melanogaster* (Dmelanogaster), *Crassostrea gigas* (Cgigas), *Lymnaea stagnalis* (Lstagnalis), *Sepia officinalis* (Sofficinalis), *Aplysia californica* (Acalifornica) and *Lottia gigantea* (Lgigantea). The phylogenetic tree was constructed using MEGA X software. Bootstraps (500 replicates) are notified next to the branches and signature motifs are annotated for each box.

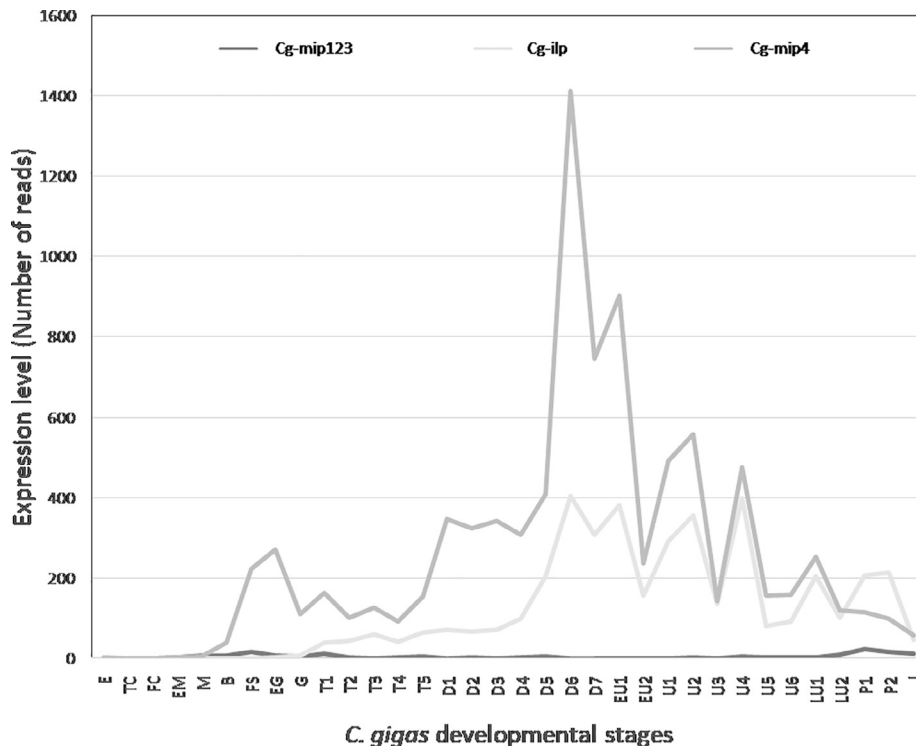


Fig 4. Predicted expressions of Cg-mip123, Cg-ilp and Cg-mip4 transcripts found *in silico* using GIGATON during the developmental stages of *C. gigas*. Unfertilized oocytes (E), two cell embryo (TC), four cell embryo (FC), early morula (EM), Morula (M), Blastula (B), Swimming blastula (FS), early gastrula (EG), gastrula (G), trochophore larvae from 9.5h to 14.5h post-fertilization (T1 to T5), D-shaped larvae from 17.58h to 3.77 days post-fertilization (D1 to D7), early umbo larvae at 5 and 7 days post-fertilization (EU1 and EU2 respectively), umbo larvae from 8 to 14 days post-fertilization (U1 to U6), veliger larvae at 15 and 16 days post-fertilization (LU1 and LU2 respectively), pediveliger larvae at 18.03 and 18.19 days post-fertilization (P1 and P2 respectively), juvenile at 215 days post-fertilization (J) according to Zhang et al. (2012).

stages (T1 to T5) until the juvenile stage (J). The highest expressions of this transcript were observed at the late stages of D-shaped larvae (D6 and D7) and at two umbo larvae stages (U2 and U4). By comparison with the two other transcripts, Cg-mip123 was only weakly expressed during development.

Expressions of Cg-mip123, Cg-ilp and Cg-mip4 transcripts in adult tissues were analysed on the basis of GIGATON transcriptomic data associated to quantitative PCR measurements (Fig. 5). *In silico* expressions showed that all transcripts seemed to be mostly expressed in the digestive gland and in the male mature gonad (Fig. 5A1 to C1). Real time qPCR tends to confirm the overexpression of Cg-mip123 in the digestive gland and the male mature gonadal area, as predicted by *in silico* data. Real time qPCR also revealed that this transcript is expressed

in the adductor muscle and in the visceral ganglia (Fig. 5A1 to A4). The Cg-mip4 is the least transcript expressed in the adult tissues, ten times less than the two others (Fig. 5B2). Concerning Cg-ilp, it appeared that this transcript was significantly overexpressed in the digestive gland as predicted by *in silico* data (Fig. 5C2).

The expression of the three transcripts was measured in the visceral ganglia (Fig. 5A3 to C3) and in the gonadal area (Fig. 5A4 to C4) at each stage of the gametogenetic cycle for both male and female. The results showed that, of the three transcripts, the Cg-mip123 is the one that is mostly expressed in both the visceral ganglia and the gonadal area (Fig. 5A3 and A4). Moreover, Cg-mip123 appeared to be a hundred time more expressed in the visceral ganglia than in the gonadal area. Its expression in the visceral ganglia reaches its highest level at the

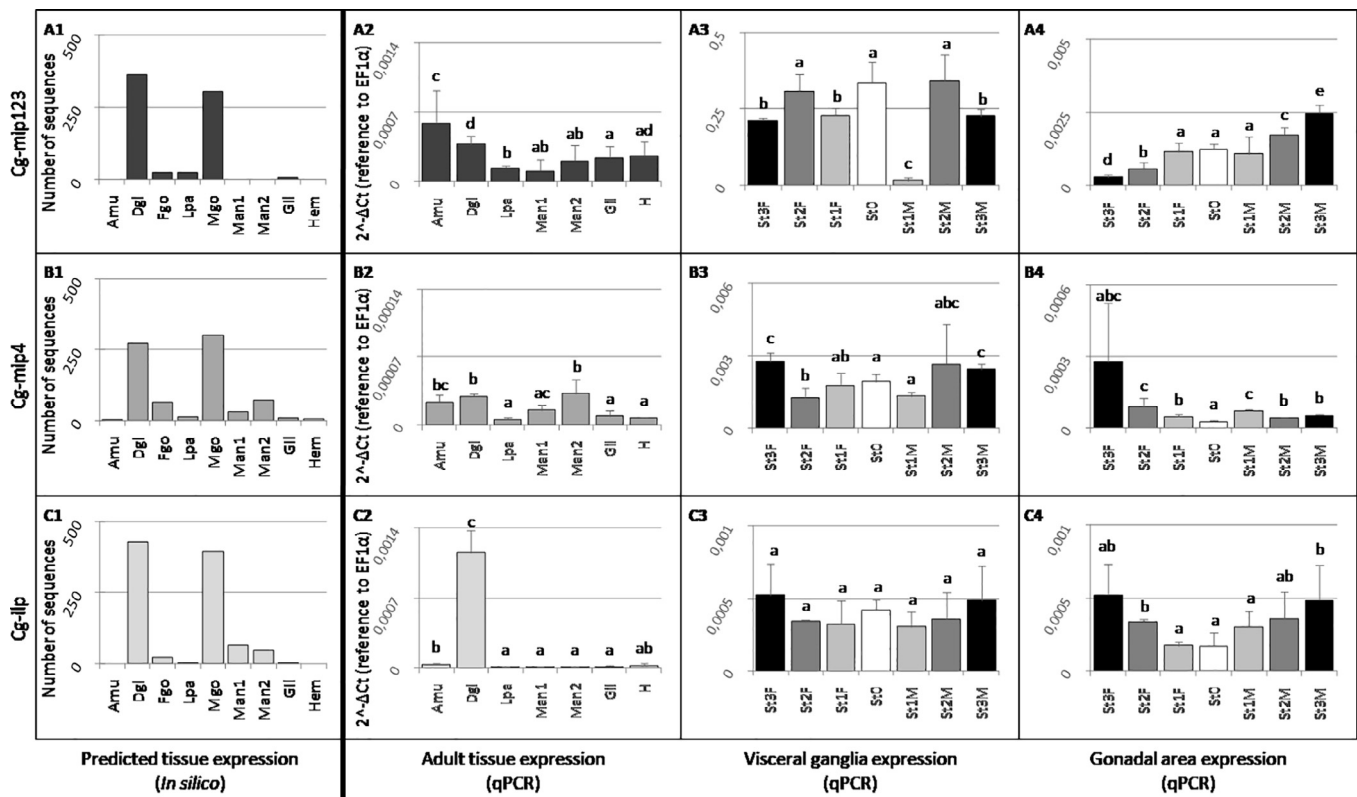


Fig 5. Expressions analyses of Cg-mip123, Cg-mip4 and Cg-ilp transcripts. Predicted expressions found *in silico* using GIGATON (Cg-mip123: A1, Cg-mip4: B1, Cg-ilp: C1) and expressions analyzed by real-time qPCR (Cg-mip123: A2, Cg-mip4: B2, Cg-ilp: C2) in adult tissues of *C. gigas*. Adductor muscle (Amu), Digestive gland (Dgl), female and male gonadal area (Fgo and Mgo respectively), labial palps (Lpa), inner mantle (Man1), mantle edge (Man2), Gill (Gil), hemocytes (Hem), gonadal area (Ago) and heart (H). Expressions measured by real-time qPCR in the visceral ganglia (Cg-mip123: A3, Cg-mip4: B3, Cg-ilp: C3) and in the gonadal area (Cg-mip123: A4, Cg-mip4: B4, Cg-ilp: C4) of animals selected at each reproductive stages. Quiescent stage (St0), male and female gonial mitosis initiating stages (St1M and St1F respectively), male and female germline development stages (St2M and St2F respectively) and male and female sexual maturity stages (St3M and St3F respectively). Each bar is the mean of 4 pools of 3 samples each tested in triplicate. The letters indicate significant differences between bars based on a non-parametric permutation test (p-value < 0.05)

quiescent stage (stage 0) and the male and female germline development stages (stages 2F and 2M). It reaches its lowest level at the initiation of male gonial mitosis (stage 1M). Interestingly, Cg-mip123 expression levels were constant in the gonadal area over the early gametogenetic stages (stage 0 corresponding to the quiescent stage, stages 1M and 1F corresponding to male and female gonial mitosis stages respectively). A significant decrease of Cg-mip123 expression was then observed during the female gametogenesis progression when on the contrary, a significant increase was noted during the male gametogenesis progression. Stage 3M (male at sexual maturity) presented the Cg-mip123 highest expression level in the gonadal area. The Cg-mip4 transcript showed constant levels of expression in the visceral ganglia at the early gametogenetic stages and a tendency to increase at male and female sexual maturity (stages 3F and 3M). In the gonadal area, the Cg-mip4 transcript levels were lower than in the visceral ganglia but tended to increase at the female sexual maturity stage (Fig. 5B3 and B4). Concerning Cg-ilp, no significant difference was shown between the gametogenetic stages in the visceral ganglia. Moreover, the transcript levels were in the same order of magnitude in both the visceral ganglia and the gonadal area (Fig. 5C3 and C4). The Cg-ilp transcript expression in the gonadal area tended to increase as the gametogenesis progressed, in both male and female oysters.

In order to better understand the involvement of Cg-IRPs in reproduction, *in situ* hybridization labelling was only investigated for the transcripts putatively involved in reproduction processes whom expression significantly increase during adult gametogenesis. Subsequently, Cg-mip123 and Cg-ilp expression were explored in the gonadal area and in visceral ganglia. Only the Cg-mip123 transcript

could be detected in the both tissues regardless of the stage of gametogenesis for both males and females. In the visceral ganglia, Cg-mip123 labelled cells appeared to be grouped together in a localised area whatever the stage of gametogenesis (Fig. 6). At the cellular level, Cg-mip123 labelling was localised mainly in the cytoplasm of neurosecretory cells. Variations of labelling intensity may result from the position of the cutting plan in the ganglia. A cartography of visceral ganglia has been established based upon the location of the labelled (Fig. 7). The Cg-mip123 producing cells are located in a lobe of the visceral ganglia, near the epidermal surface. In this area, the positive cells do not constitute a cohesive group but appear to be relatively dispersed. The Cg-mip123 transcript was also clearly detected into the gonadal tubules of mature males at stage 3 (Fig. 8). The labelled cells were located at the periphery of the gonadal tubules in the area containing early germinal cells (spermatogonia) and intra-gonadal somatic cells.

4. Discussion

The present study reports the identification of six IRP sequences in *C. gigas* and the detection of their conserved motif in databases. Moreover, the expression patterns of these IRPs were specified in order to gain insight into their putative biological functions and more precisely their involvement in the adult reproduction.

4.1. Comparison of *C. gigas* IRPs and phylogenetic analysis

Despite the low sequence homology of the members, IRPs

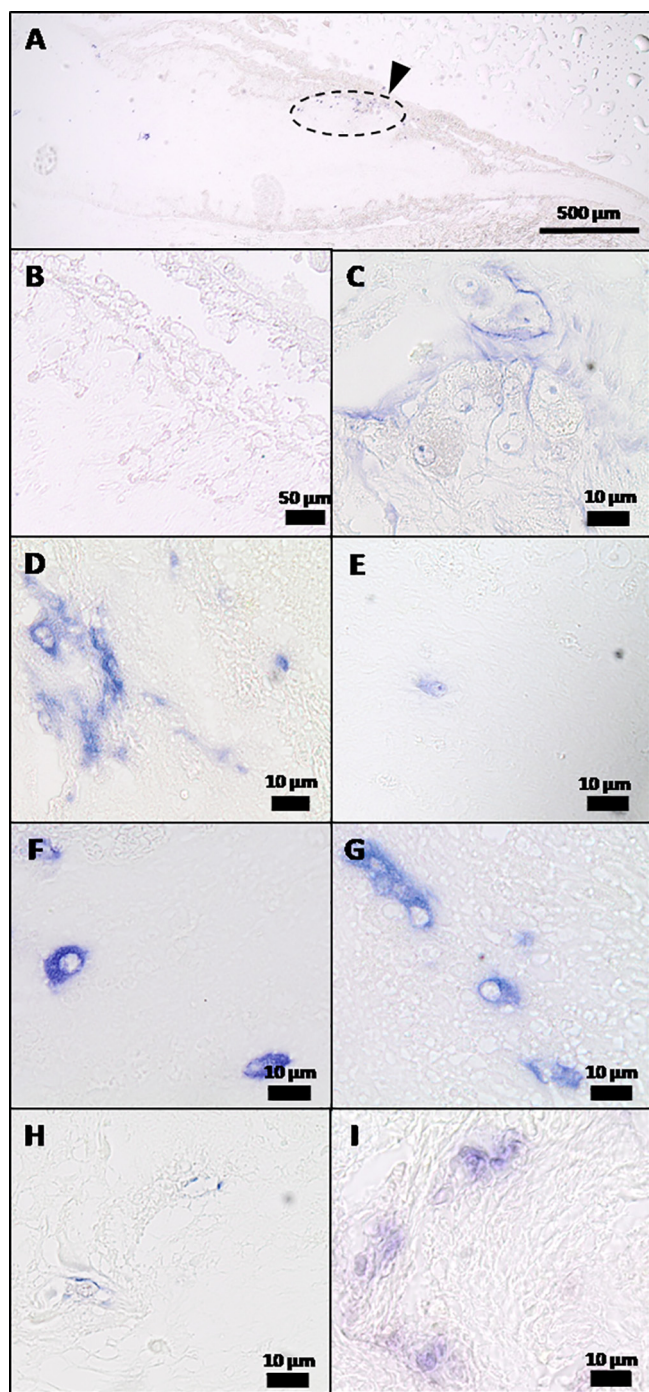


Fig 6. Transcript location of Cg-mip123 in the visceral ganglia by ISH at each stages of gametogenesis. General view of ISH labeling with antisense probe (A) with the labelled area designated by an arrowhead and ISH control with sense probe (B). Quiescent stage (Stage 0, C), gonial mitosis stages female and male (Stage 1, D and E respectively), female and male germline development stages (Stages 2, F and G respectively) and female and male sexual maturity stages (Stages 3, H and I respectively). Observation of neurosecretory cells labelled by Cg-mip123 transcript at magnification 100 for each gametogenic stage (C to I).

configuration led to the identification of conserved signature motifs containing cysteine residues involved in the canonical disulphide bonds which are CC(3X)C(8X)C for Cg-ILP, C(1X)CC(3X)C(8X)C for Cg-MIPs and CC(4X)C(8X)C for Cg-MILP7. The cysteines are responsible for the formation of three or four disulphide bonds depending on the IRPs. They include two canonical bonds between the A and B chains, one intra-A chain and a facultative additional bond between the A and B

chains. Interestingly, the principal segregation in the phylogenetic tree depends first on the IRP signature motif, then on the species. This observation and the fact that all the motifs are found in *L. gigantea* and *C. gigas*, demonstrate that these A chain motifs were present before the segregation of molluscan classes. ILPs were not identified in *L. stagnalis* nor *A. californica*. However, the existence of ILPs cannot be excluded in these species, seen the absence of a whole sequenced genome. Moreover, IRPs identification was targeted in the nervous system. In other gastropod species belonging to the group of the cone snail molluscs, some sequences with ILP motifs were also identified (Table S1). Thus, in *Conus geographus*, sequences with MIP or ILP motifs were identified (Safavi-Hemami et al., 2016). However, these IRPs were produced as venom for predation according to prey preference. They mimic IRPs of their prey, thus interfering with the insulin signal or leading to hypoglycemic shock. However, the authors postulate that the genes of the venom insulins differ from canonical insulin genes and should be issued from a duplication event early in the history of conus. For this reason, we deliberately discarded these molluscan IRPs from our molluscan phylogenetic tree because of their adaptive function.

One of the IRP sequences in *C. gigas*, named Cg-ILP, presented a structural homology with the vertebrates' insulin and a non-cleavable D domain at the C-terminal end of the sequence observed in vertebrates' IGF (Shabanpoor et al., 2009). According to its characteristics, we assigned Cg-ILP in the γ type. The genomic sequence of Cg-ILP presented similarities with the insulin gene structure described in vertebrates. Moreover, as currently observed in other species (Chan and Steiner, 2000), two among the three exons code for the A and B chain sequences and the last one appeared to code for the uncleavable D domain. The identification of Cg-ILP in a lophotrochozoan species demonstrates the ancestral conservation of this IRP gene organization coding for γ type, prior to vertebrate segregation. Indeed, other ILPs with the same structure than the vertebrates' insulins were also found in invertebrate species as *L. gigantea* (Veenstra, 2010), *P. dumerilii* (Conzelmann et al., 2013) and *Nematostella vectensis* (Ancil, 2009).

In this paper, we identified Cg-MIPs belonging to the β type named Cg-MIP 1 to 4. Before *C. gigas* genome availability, Hamano et al. (2005) already identified a cDNA of MIPs, named oIRP, in the visceral ganglia of the oyster. We confirmed here the protein encoded by oIRP corresponds to the Cg-MIP1 sequence in the predicted proteome of *C. gigas*. All Cg-MIPs possess the classical structure of the IRPs of *L. stagnalis* (Smit et al., 1988, 1991, 1992, 1993, 1996, 1998) with an additional disulphide bond between A and B chains. Their designation as MIPs for Molluscan Insulin-related Peptides is due to their first identification in molluscan species but recent data showed that they are rather extended to the lophotrochozoa (Conzelmann et al., 2013). The four Cg-MIP sequences display some differences. Cg-MIP1 and Cg-MIP2 differ only by three amino acids located in the C peptide. These substitutions result from Single Nucleotide Polymorphism (SNPs) as predicted in GIGATON databases (Rivière et al., 2015; Zhang et al., 2012). Indeed, *C. gigas* possesses one of the highest level of genomic polymorphism (Qi et al., 2017; Zhang et al., 2012) and in the Cg-mip123 transcript, SNPs are regularly distributed along the whole sequence. However, amino acid substitutions only occur in the C peptide, cleaved in the mature protein. This situation of higher rate of mutations acceptance in the C peptide was already described by Steiner et al. (1985) for vertebrate IRPs. Cg-MIP3 appeared to be issued from a splicing variant of its transcript. The presence of a cleavable E domain in the Cg-MIP4 structure was also found in *P. dumerilii* IRP-2, IRP-3 and IRP-5 (Conzelmann et al., 2013). The genomic organization presents the classical organization of the IRP family with one exon coding for the B chain and another coding for the A chain (Steiner et al., 1985). Moreover, the Cg-MIPs C peptides are issued from two different exons, as described for MIP II and MIP V genes in *L. stagnalis* (Smit et al., 1998).

The last Cg-IRP sequence identified in our study was called Cg-MILP7. It exhibited three potential A-B disulphide bonds like the Molluscan Insulin-like Peptides but also the A-chain motif of DILP7 of

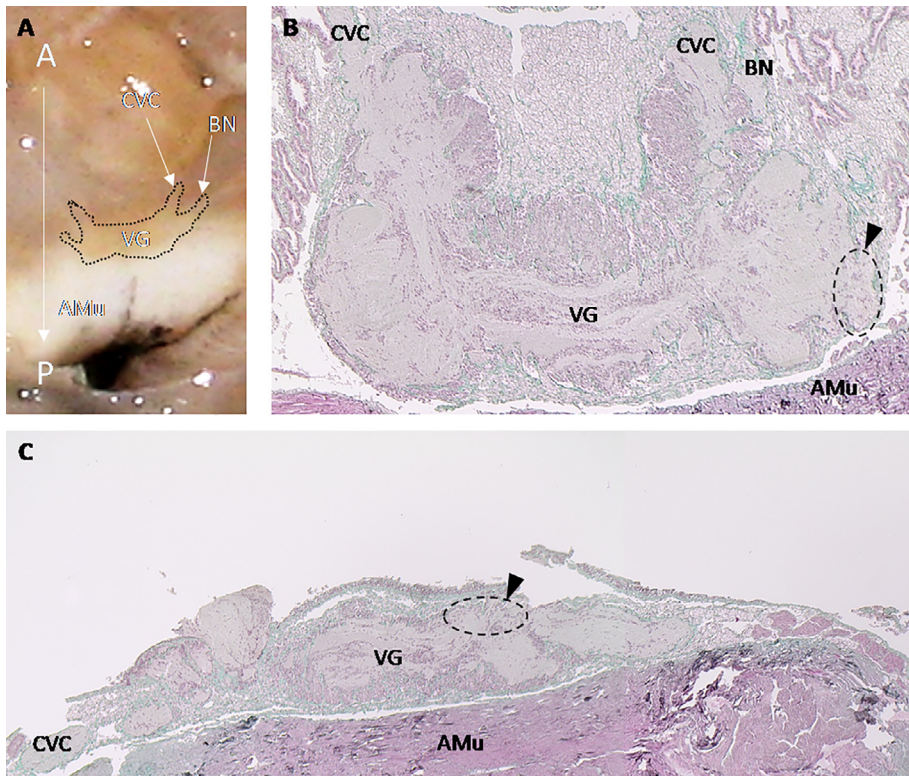


Fig 7. Photography (A) and trichrome Prenant-Gabe coloration of transverse (B) and longitudinal (C) sections of *C. gigas* visceral ganglia (VG). The antero-posterior (A/P) axis, the adductor muscle (AMu), cerebro-visceral connective (CVC) and branchial nerve (BN) are annotated. The arrowheads indicate the location where the labelling is visible.

drosophila [CC(4X)C(8X)C]. This ILP7 ortholog was also identified in the genome of *L. gigantea* (Insulin 4), in the ESTs databases of *L. stagnalis* (ES573590 and ES573220) and *A. californica* (EB230021) (Veenstra, 2010). As expected, all these molluscan MILP7 sequences appeared grouped together as shown in the phylogenetic tree. Moreover, when the phylogenetic analysis includes the DILP7 of *D. melanogaster*, it appears to be grouped with its homologous molluscan sequences, leading to suppose that the motif should be conserved upstream the ecdysozoan specification. The Cg-milp7 transcript could not be found whatever the databases suggesting a very low or targeted expression. Further experiments would consist in the characterization

of its expression pattern of this atypical IRP.

Until recently, most scientific research on IRPs in molluscan species referred to MIP molecules. The existence of ILP and MILP7 was firstly reported in *L. gigantea*, which was the first mollusc to have its genome sequenced. Our work on the oyster *C. gigas* confirmed the existence of three types of molecules belonging to IRPs in molluscs. The three types are present in several molluscan species and considering the phylogenetic segregation, these types would probably be ancestral to molluscs providing new questions concerning the evolution of the insulin family in metazoan. This superfamily of molecules is known to be involved in multiple physiological regulations. Subsequently, the various IRPs we

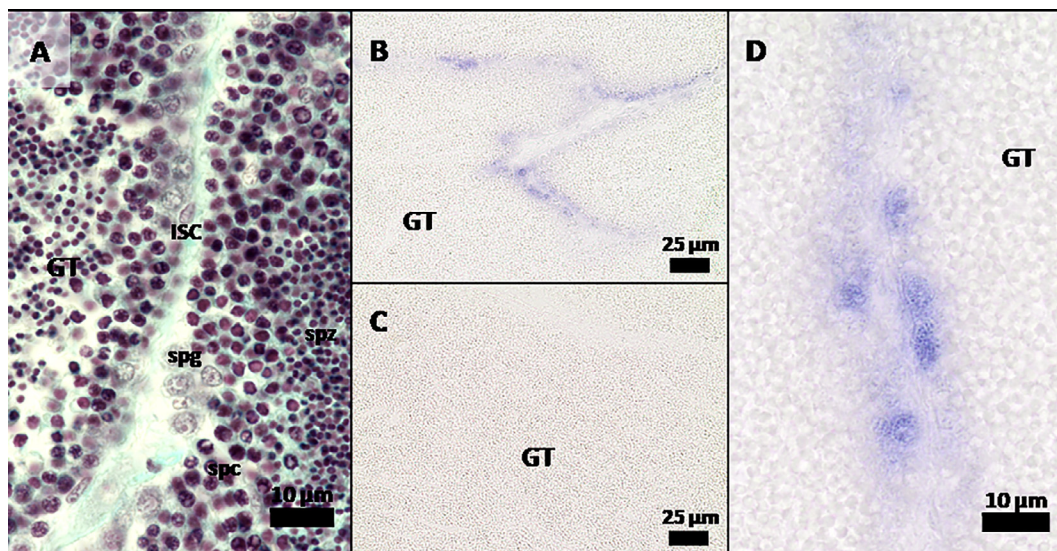


Fig 8. Trichrome Prenant-Gabe coloration of the gonadal tubules (GT) at male sexual maturity (Stage III, A). Spermatogonia (spg), spermatocytes (spc), spermatozoa (spz) and intra-gonadal somatic cell (ISC) are annotated. General view of the transcript location of Cg-mip123 in the gonadal area by ISH at the male sexual maturity stage with antisense probe (B) and control with sense probe (C). Cg-mip123 transcript location of the gonadal tubule at male sexual maturity at magnification 100 (D).

had identified in *C. gigas* probably support specific and conserved physiological roles related to their structural properties, including development, reproduction, sex change and energetic metabolism.

4.2. Evidence of multiple roles of IRPs in *C. gigas*

We also characterized Cg-IRPs expression during development and in various adult tissues in order to investigate their physiological involvement. Predicted expressions were investigated using GIGATON databases and completed by qPCR experiments for adult tissues, as GIGATON databases did not inform about expression in the visceral ganglia and in the gonadal area along the gametogenesis. Indeed, in invertebrate species, IRPs expression was mainly observed in the nervous system. The expression of *C. gigas* IRPs was also mainly detected in the nervous system, with Cg-MIPs being the most expressed IRPs in the visceral ganglia throughout the year. Cg-MIPs should be involved in the central control of multiple physiological processes, in agreement with the results published by Hamano et al. (2005) on the same species. In complement, we also demonstrated local expressions of the other Cg-IRPs in the digestive and gonadal tissues as well as during larval development.

4.2.1. A putative role of Cg-MIP4 during development

According to the databases, Cg-mip4 transcript is mostly expressed during the larval development. The expression of this transcript increased mainly at the end of the D-larvae stage, when larvae start to feed themselves, with minor peaks during umbo stages. The involvement of IRPs in the control of development is highly conserved: in vertebrate species, Insulin-like Growth Factors (IGFs) provide essential signals for the control of embryonic and postnatal development (Dupont and Holzenberger, 2003). In insects and nematodes, the insulin signalling pathway plays essential roles in the control of development and organ growth in relation with the nutritional status (Lin and Smaghe, 2018; Schindler et al., 2014). In the oyster *C. gigas*, Cg-MIP4 might be a key actor in developmental control.

4.2.2. The putative role of Cg-ILP in digestive processes

As expected by *in silico* expression data, Cg-ilp was mainly expressed in the digestive gland and the gonadal area. These results fit with the digestive expression of some IRPs described in various molluscan species as for example *Mytilus edulis* (Fritsch et al., 1976), *Anodonta cygnea*, *Unio pictorum* (Plisetskaya et al., 1978) and *Helisoma duryi* (Abdraba and Saleuddin, 2006). Giard et al. (1998) observed that vertebrate IGF-1 increase amino acid incorporation into proteins in dissociated digestive cells of *Pecten maximus*. More recently, Jouaux et al. (2012) observed the expression of the insulin receptor in the digestive gland of *C. gigas* and suggested a paracrine function of the IRPs produced during digestion. This digestive role of IRPs could also be observed in other invertebrate species. Among the eight DILPs from *D. melanogaster*, DILP3 is expressed in the brain but also in the midgut (Veenstra et al., 2008) and plays a critical role in promoting sugar homeostasis (Kim and Neufeld, 2015). This role in the regulation of nutritional state and carbohydrate metabolism is also crucial and conserved in vertebrate species.

4.2.3. Potential involvement of Cg-ILP and Cg-MIPs in the control of reproduction

In *C. gigas*, Cg-ilp and Cg-mip123 are both expressed in the gonadal area with differential expression related to gametogenetic cycle. Cg-ilp expression significantly increase according to the course of gametogenesis in females as well as in males, allowing to postulate that gonadal Cg-ILP might play a part in the control of gametogenesis and/or associated metabolic pathways. The relatively low Cg-ilp expression did not allow the identification of cells producing Cg-ilp by *in situ* hybridization. However, the involvement of IRPs in the control of gametogenesis is well known in numerous invertebrate species. In the

lophotrochozoan planaria *S. mediterranea*, ilp-1 is expressed in germ cells of the testis lobe and ILP disruption impaired testis growth and spermatogenesis (Miller and Newmark, 2012). In *D. melanogaster* DILP2, DILP3 and DILP5 (presenting the ILPs signature motif [CC(3X)C(8X)C]) are involved in the control of germ stem cell proliferation defining the final reproductive effort (Grönke et al., 2010; Hsu and Drummond-Barbosa, 2009). In the nematode *C. elegans*, Michaelson et al. (2010) indicate that, under replete reproductive conditions, the larval germline responds to insulin signalling (INS-3 and INS-33) to ensure robust germline proliferation and amplify the germline stem cell population. The activation of insulin signalling pathway acts in parallel with Notch signalling, thus affecting the number of cells in the proliferative zone. Moreover, for all these invertebrate species, germline proliferation is closely linked to nutritional status via insulin signalling. If we consider its expression pattern in *C. gigas*, Cg-ILP may be the best candidate to assume the dual control of nutritional status and reproductive effort. Further measurements of Cg-ilp expression based on the nutritional status and germline cells proliferation should be performed to confirm this hypothesis.

The expression of Cg-mip123 was also detected in the gonadal area even though much lower than in visceral ganglia. The Cg-mip123 expression is mainly related to male sexual maturity, both for qPCR and ISH labelling approaches. Interestingly, ISH positive cells were localized near the tubule wall in the location of early germinal cells and associated intragonadal somatic cells (Franco et al., 2008). The small size and low number of labelled cells did not allow us to identify with certainty the nature of the cells but a co-labelling with the germ cells marker Oyvlg (Fabioux et al., 2004) could clarify their belonging to the germinal or somatic lineage. In crustacean species, IRPs were found to be involved in the maintenance of male characteristics (Chandler et al., 2015, 2017; Manor et al., 2007; Sharabi et al., 2016; Ventura et al., 2009). Therefore, the specific detection of Cg-mip123 in the male gonad questions the androgenic role of Cg-mip123 in the oyster, which is an annual alternative hermaphrodite.

5. Conclusion

Regarding the phylogenetic analysis, the most commonly distributed motif is that of ILPs, leading to postulate its ancestry and conservation in metazoan. ILP7 orthologs were found in a diversity of protostomian species, from insects (DILP7) to molluscs (MILP7). ILP7 orthologs may be derived from ILPs resulting from duplication and evolutionary pressure in protostomes. Molluscan MIPs constitutes another group of ILPs, with a different evolutionary history. These MIPs were also found in Annelida (Conzelmann et al., 2013) but were not identified in Platyhelminthes (Wang et al., 2013) at the moment. Our work would lead us gain insight into the evolutionary history of MIPs in lophotrochozoa. So far, our data provided some clues concerning the multiple and conserved roles of IRPs in molluscs. Our expression data suggested the involvement of Cg-ILP and Cg-MIPs in the control of the adult gametogenesis. The roles of IRPs in the control of these physiological processes were rarely mentioned in molluscan species and have now to be explored.

Acknowledgements

PhD of M. Cherif-Feildel was financially supported by the French research ministry (PhD grant). The authors are grateful to M.P. Dubos for providing some visceral ganglia slides used for ISH. Proofreading of the English text has been realised by C. Quint of the centre of translation of UNICAEN.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygcen.2018.10.019>.

References

- Abdraba, A.M., Saleuddin, S., 2006. Insulin-like material in the digestive tract of the freshwater snail *Helisoma duryi* (Mollusca: Pulmonata). *Pure Sci.* 33, 85–92.
- Agrogiannis, G.D., Sifakis, S., Patsouris, E.S., Konstantinidou, A.E., 2014. Insulin-like growth factors in embryonic and fetal growth and skeletal development (Review). *Mol. Med. Rep.* 10, 579–584. <https://doi.org/10.3892/mmr.2014.2258>.
- Ancil, M., 2009. Chemical transmission in the sea anemone *Nematostella vectensis*: a genomic perspective. *Comp. Biochem. Physiol. - Part D Genomics Proteomics* 4, 268–289. <https://doi.org/10.1016/j.cbd.2009.07.001>.
- Apfeld, J., Kenyon, C., 1998. Cell nonautonomy of *C. elegans* daf-2 function in the regulation of diapause and life span. *Cell* 95, 199–210. [https://doi.org/10.1016/S0092-8674\(00\)81751-1](https://doi.org/10.1016/S0092-8674(00)81751-1).
- Ashrafi, K., Chang, F.Y., Watts, J.L., Fraser, A.G., Kamath, R.S., Ahringer, J., Ruvkun, G., 2003. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* 421, 268–272. <https://doi.org/10.1038/nature01279>.
- Avogaro, A., De Kreutzenberg, S.V., Fadini, G.P., 2010. Insulin signaling and life span. *PLoS Arch. Eur. J. Physiol.* 459, 301–314. <https://doi.org/10.1007/s00424-009-0721-8>.
- Bathgate, R.A.D., Halls, M.L., van der Westhuizen, E.T., Callander, G.E., Kocan, M., Summers, R.J., 2013. Relaxin family peptides and their receptors. *Physiol. Rev.* 93, 405–480. <https://doi.org/10.1152/physrev.00001.2012>.
- Bayazit, V., 2009. Evaluation of the Bombyxin Gene and Bombyxin Insulin-like Peptide in Silkworm (*Bombyx Mori*). *Aust. J. of Basic Appl. Sci.* 3, 1032–1042.
- Böhni, R., Riesgo-Escovar, J., Oldham, S., Brogiolo, W., Stocker, H., Andrus, B.F., Beckingham, K., Hafen, E., 1999. Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1–4. *Cell* 97, 865–875. [https://doi.org/10.1016/S0092-8674\(00\)80799-0](https://doi.org/10.1016/S0092-8674(00)80799-0).
- Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., Hafen, E., 2001. An evolutionarily conserved function of the drosophila insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* 11, 213–221. [https://doi.org/10.1016/S0960-9822\(01\)00068-9](https://doi.org/10.1016/S0960-9822(01)00068-9).
- Broughton, S.J., Piper, M.D.W., Ikeya, T., Bass, T.M., Jacobson, J., Driege, Y., Martinez, P., Hafen, E., Withers, D.J., Leivers, S.J., Partridge, L., 2005. Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. *Proc. Natl. Acad. Sci.* 102, 3105–3110. <https://doi.org/10.1073/pnas.0405775102>.
- Broughton, S.J., Slack, C., Alic, N., Metaxakis, A., Bass, T.M., Driege, Y., Partridge, L., 2010. DILP-producing median neurosecretory cells in the *Drosophila* brain mediate the response of lifespan to nutrition. *Aging Cell* 9, 336–346. <https://doi.org/10.1111/j.1474-9726.2010.00558.x>.
- Chan, S.J., Steiner, D.F., 2000. Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *Am. Zool.* 40, 213–222. [https://doi.org/10.1668/0003-1569\(2000\)040\[0213:ITTAPO\]2.0.CO;2](https://doi.org/10.1668/0003-1569(2000)040[0213:ITTAPO]2.0.CO;2).
- Chandler, J.C., Aizen, J., Elizur, A., Hollander-Cohen, L., Battaglene, S.C., Ventura, T., 2015. Discovery of a novel insulin-like peptide and insulin binding proteins in the Eastern rock lobster *Sagmariasus verreauxi*. *Gen. Comp. Endocrinol.* 215, 76–87. <https://doi.org/10.1016/j.ygcen.2014.08.018>.
- Chandler, J.C., Gandhi, N.S., Mancera, R.L., Smith, G., Elizur, A., Ventura, T., 2017. Understanding insulin endocrinology in decapod crustacea: molecular modelling characterization of an insulin-binding protein and insulin-like peptides in the eastern spiny lobster, *Sagmariasus verreauxi*. *Int. J. Mol. Sci.* 18, 1–19. <https://doi.org/10.3390/ijms18091832>.
- Clancy, D.J., Gems, D., Harshman, L.G., Oldham, S., Stocker, H., Hafen, E., Leivers, S.J., Partridge, L., 2001. Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* 292, 104–106. <https://doi.org/10.1126/science.1057991>.
- Conzelmann, M., Williams, E.A., Krug, K., Franz-Wachtel, M., Macek, B., Jékely, G., 2013. The neuropeptide complement of the marine annelid *Platynereis dumerilii*. *BMC Genomics* 14, 906. <https://doi.org/10.1186/1471-2164-14-906>.
- De Meyts, P., 2004. Insulin and its receptor: structure, function and evolution. *BioEssays* 26, 1351–1362. <https://doi.org/10.1002/bies.20151>.
- Dheilly, N.M., Lelong, C., Huvet, A., Favrel, P., 2011. Development of a Pacific oyster (*Crassostrea gigas*) 31,918-feature microarray: identification of reference genes and tissue-enriched expression patterns. *BMC Genomics* 12, 468. <https://doi.org/10.1186/1471-2164-12-468>.
- Dheilly, N.M., Lelong, C., Huvet, A., Kellner, K., Dubos, M.-P., Riviere, G., Boudry, P., Favrel, P., 2012. Gametogenesis in the Pacific oyster *Crassostrea gigas*: a microarrays-based analysis identifies sex and stage specific genes. *PLoS One* 7, e36353. <https://doi.org/10.1371/journal.pone.0036353>.
- Dimitriadis, G., Mitrou, P., Lambadiari, V., Maratou, E., Raptis, S.A., 2011. Insulin effects in muscle and adipose tissue. *Diabetes Res. Clin. Pract.* 93, S52–S59. [https://doi.org/10.1016/S0168-8227\(11\)70014-6](https://doi.org/10.1016/S0168-8227(11)70014-6).
- Dubos, M.P., Bernay, B., Favrel, P., 2016. Molecular characterization of an adipokinetic hormone-related neuropeptide (AKH) from a mollusk. *Gen. Comp. Endocrinol.* 243, 15–21. <https://doi.org/10.1016/j.ygcen.2016.11.002>.
- Dubos, M.P., Zels, S., Schwartz, J., Pasquier, J., Schoofs, L., Favrel, P., 2018. Characterization of a tachykinin signalling system in the bivalve mollusk *Crassostrea gigas*. *Gen. Comp. Endocrinol.* 266, 110–118. <https://doi.org/10.1016/j.ygcen.2018.05.003>.
- Dupont, J., Holzenberger, M., 2003. Biology of insulin-like growth factors in development. *Birth Defects Res. Part C – Embryo Today Rev.* 69, 257–271. <https://doi.org/10.1002/bdrc.10022>.
- Duret, L., Guex, N., Peitsch, M.C., Bairoch, A., 1998. New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. *Genome Res.* 8, 348–353. <https://doi.org/10.1101/gr.8.4.348>.
- Fabioux, C., Pouvreau, S., Le Roux, F., Huvet, A., 2004. Oyster vasa-like gene: a specific marker of the germline in *Crassostrea gigas*. *Biochem. Biophys. Res. Commun.* 315, 897–904. <https://doi.org/10.1016/j.bbrc.2004.01.145>.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y.)* 39, 783. <https://doi.org/10.2307/2408678>.
- Fleury, E., Huvet, A., Lelong, C., de Lorigeril, J., Boulo, V., Gueguen, Y., Bachère, E., Tanguy, A., Moraga, D., Fabioux, C., Lindeque, P., Shaw, J., Reinhardt, R., Prunet, P., Davey, G., Lapègue, S., Sauvage, C., Corporeau, C., Moal, J., Gavory, F., Wincker, P., Moreews, F., Klopp, C., Mathieu, M., Boudry, P., Favrel, P., 2009. Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics* 10, 341. <https://doi.org/10.1186/1471-2164-10-341>.
- Floyd, P., Li, L., 1999. Insulin prohormone processing, distribution, and relation to metabolism in *Aplysia californica*. *J. Neurosci.* 19, 7732–7741.
- Franco, A., Heude Berthelin, C., Goux, D., Sourdaire, P., Mathieu, M., 2008. Fine structure of the early stages of spermatogenesis in the Pacific oyster, *Crassostrea gigas* (Mollusca, Bivalvia). *Tissue Cell* 40, 251–260. <https://doi.org/10.1016/j.tice.2007.12.006>.
- Fritsch, H.A.R., Van Noorden, S., Pearce, A.G.E., 1976. Cytochemical and immunofluorescence investigations on insulin-like producing cells in the intestine of *Mytilus edulis* L. (Bivalvia). *Cell Tissue Res.* 165, 365–369. <https://doi.org/10.1007/BF00222439>.
- Gabe, M., 1968. Techniques histologiques, 6th ed. Paris.
- Garelli, A., Gontijo, A.M., Miguela, V., Caparros, E., Dominguez, M., 2012. Imaginal discs secrete insulin-like peptide 8 to mediate plasticity of growth and maturation. *Science* 336 (80), 579–582. <https://doi.org/10.1126/science.1216735>.
- Giard, W., Lebel, J.M., Boucaud-Camou, E., Favrel, P., 1998. Effects of vertebrate growth factors on digestive gland cells from the mollusc *Pecten maximus* L.: an in vitro study. *J. Comp. Physiol. – B Biochem. Syst. Environ. Physiol.* 168, 81–86. <https://doi.org/10.1007/s003600050123>.
- Gricourt, L., Bonnet, G., Boujard, D., Mathieu, M., Kellner, K., 2003. Insulin-like system and growth regulation in the Pacific oyster *Crassostrea gigas*: hrIGF-1 effect on protein synthesis of mantle edge cells and expression of an homologous insulin receptor-related receptor. *Gen. Comp. Endocrinol.* 134, 44–56. [https://doi.org/10.1016/S0016-6480\(03\)00217-X](https://doi.org/10.1016/S0016-6480(03)00217-X).
- Grönke, S., Clarke, D.-F., Broughton, S., Andrews, T.D., Partridge, L., 2010. Molecular evolution and functional characterization of drosophila insulin-like peptides. *PLoS Genet.* 6, e1000857. <https://doi.org/10.1371/journal.pgen.1000857>.
- Hamano, K., Awaji, M., Usuki, H., 2005. cDNA structure of an insulin-related peptide in the Pacific oyster and seasonal changes in the gene expression. *J. Endocrinol.* 187, 55–67. <https://doi.org/10.1677/joe.1.06284>.
- Hatje, K., Keller, O., Hammesfahr, B., Pillmann, H., Waack, S., Kollmar, M., 2011. Cross-species protein sequence and gene structure prediction with fine-tuned Webscipro 2.0 and Scipio. *BMC Res. Notes* 4, 265. <https://doi.org/10.1186/1756-0500-4-265>.
- Heude Berthelin, C., Laisney, J., Espinosa, J., Martin, O., Hernandez, G., Mathieu, M., Kellner, K., 2001. Storage and reproductive strategy in *Crassostrea gigas* from two different growing areas (Normandy and the Atlantic coast, France). *Invertebr. Reprod. Dev.* 40, 79–86. <https://doi.org/10.1080/07924259.2001.9652500>.
- Hibshman, J.D., Hung, A., Baugh, L.R., 2016. Maternal diet and insulin-like signaling control intergenerational plasticity of progeny size and starvation resistance. *PLOS Genet.* 12, e1006396. <https://doi.org/10.1371/journal.pgen.1006396>.
- Hsu, H.-J., Drummond-Barbosa, D., 2009. Insulin levels control female germline stem cell maintenance via the niche in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1117–1121. <https://doi.org/10.1073/pnas.0809144106>.
- Jékely, G., 2013. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc. Natl. Acad. Sci.* 110, 8702–8707. <https://doi.org/10.1073/pnas.1221833110>.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. <https://doi.org/10.1093/bioinformatics/8.3.275>.
- Jouaux, A., Heude Berthelin, C., Sourdaire, P., Mathieu, M., Kellner, K., 2010. Gametogenic stages in triploid oysters *Crassostrea gigas*: irregular locking of gonial proliferation and subsequent reproductive effort. *J. Exp. Mar. Biol. Ecol.* 395, 162–170. <https://doi.org/10.1016/j.jembe.2010.08.030>.
- Jouaux, A., Franco, A., Heude Berthelin, C., Sourdaire, P., Blin, J.L., Mathieu, M., Kellner, K., 2012. Identification of Ras, Pten and p70S6K homologs in the Pacific oyster *Crassostrea gigas* and diet control of insulin pathway. *Gen. Comp. Endocrinol.* 176, 28–38. <https://doi.org/10.1016/j.ygcen.2011.12.008>.
- Kim, J., Neufeld, T.P., 2015. Dietary sugar promotes systemic TOR activation in *Drosophila* through AKH-dependent selective secretion of Dilp3. *Nat. Commun.* 6, 6846. <https://doi.org/10.1038/ncomms7846>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Kondo, H., Ino, M., Suzuki, A., Ishizaki, H., Iwami, M., 1996. Multiple gene copies for Bombyxin, an insulin-related peptide of the silkworm *Bombyx mori*: structural signs for gene rearrangement and duplication responsible for generation of multiple molecular forms of Bombyxin. *J. Mol. Biol.* 259, 926–937. <https://doi.org/10.1006/jmbi.1996.0370>.
- Lagueux, M., Lwoff, L., Meister, M., Goltzené, F., Hoffmann, J.A., 1990. cDNAs from neurosecretory cells of brains of *Locusta migratoria* (Insecta, Orthoptera) encoding a novel member of the superfamily of insulins. *Eur. J. Biochem.* 187, 249–254. <https://doi.org/10.1111/j.1432-1033.1990.tb15302.x>.
- Li, K.W., Geraerts, W.P.M., Joosse, J., 1992. Purification and sequencing of molluscan

- insulin-related peptide II from the neuroendocrine light green cells in *Lymnaea stagnalis*. *Endocrinology* 130, 3427–3432. <https://doi.org/10.1210/en.130.6.3427>.
- Li, W., Kennedy, S.G., Ruvkun, G., 2003. daf-28 encodes a C. elegans insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes Dev.* 17, 844–858. <https://doi.org/10.1101/gad.1066503>.
- Lin, X., Smaghe, G., 2018. Roles of the insulin signaling pathway in insect development and organ growth. *Peptides* 0–1. <https://doi.org/10.1016/j.peptides.2018.02.001>.
- Liu, Y., Liao, S., Veenstra, J.A., Nässel, D.R., 2016. *Drosophila* insulin-like peptide 1 (DILP1) is transiently expressed during non-feeding stages and reproductive dormancy. *Sci. Rep.* 6, 26620. <https://doi.org/10.1038/srep26620>.
- Luo, S., Kleemann, G.A., Ashraf, J.M., Shaw, W.M., Murphy, C.T., 2010. TGF- β and insulin signaling regulate reproductive aging via oocyte and germline quality maintenance. *Cell* 143, 299–312. <https://doi.org/10.1016/j.cell.2010.09.013>.
- Manor, R., Weil, S., Oren, S., Glazer, L., Aflalo, E.D., Ventura, T., Chalifa-Caspi, V., Lapidot, M., Sagi, A., 2007. Insulin and gender: an insulin-like gene expressed exclusively in the androgenic gland of the male crayfish. *Gen. Comp. Endocrinol.* 150, 326–336. <https://doi.org/10.1016/j.ygcen.2006.09.006>.
- Matsunaga, Y., 2017. Diverse insulin-like peptides in *Caenorhabditis elegans*. *Int. Biol. Rev.* 1, 1–15. <https://doi.org/10.18103/ibr.v1i1.1276>.
- Maruyama, K., Hietter, H., Nagasawa, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., 1988. Isolation and primary structure of bombyxin-iv, a novel molecular species of bombyxin from the silkworm, *bombyx mori*. *Agric. Biol. Chem.* 52, 3035–3041. <https://doi.org/10.1080/00021369.1988.10869178>.
- Michaelson, D., Korta, D.Z., Capua, Y., Hubbard, E.J.A., 2010. Insulin signaling promotes germline proliferation in *C. elegans*. *Development* 137, 671–680. <https://doi.org/10.1242/dev.042523>.
- Miller, C.M., Newmark, P.A., 2012. An insulin-like peptide regulates size and adult stem cells in planarians. *Int. J. Dev. Biol.* 56, 75–82. <https://doi.org/10.1387/ijdb.113443cm>.
- Narbonne, P., Maddox, P.S., Labbe, J.-C., 2015. daf-18/PTEN locally antagonizes insulin signalling to couple germline stem cell proliferation to oocyte needs in *C. elegans*. *Development* 4230–4241. <https://doi.org/10.1242/dev.130252>.
- Olinski, R.P., Dahlberg, C., Thorndyke, M., Hallböök, F., 2006. Three insulin-relaxin-like genes in *Ciona intestinalis*. *Peptides* 27, 2535–2546. <https://doi.org/10.1016/j.peptides.2006.06.008>.
- Partridge, L., Alic, N., Bjedov, I., Piper, M.D.W., 2011. Ageing in *Drosophila*: the role of the insulin/Igf and TOR signalling network. *Exp. Gerontol.* 46, 376–381. <https://doi.org/10.1016/j.exger.2010.09.003>.
- Phoungpetchara, I., Tinikul, Y., Poljaroen, J., Chotwiwatthanakun, C., Vanichviriyakit, R., Sroyraya, M., Hanna, P.J., Sobhon, P., 2011. Cells producing insulin-like androgenic gland hormone of the giant freshwater prawn, *Macrobrachium rosenbergii*, proliferate following bilateral eyestalk-ablation. *Tissue Cell* 43, 165–177. <https://doi.org/10.1016/j.tice.2011.02.001>.
- Pierce, S.B., Costa, M., Wisotzky, R., Devadhar, S., Homburger, S.A., Buchman, A.R., Ferguson, K.C., Heller, J., Platt, D.M., Pasquini, A.A., Liu, L.X., Doberstein, S.K., Ruvkun, G., 2001. Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse C. elegans insulin gene family. *Genes Dev.* 672–686. <https://doi.org/10.1101/gad.867301.4>.
- Plisetskaya, E., Kazakov, V.K., Soltitskaya, L., Leibson, L.G., 1978. Insulin-producing cells in the gut of freshwater bivalve molluscs *Anodonta cygnea* and *Unio pictorum* and the role of insulin in the regulation of their carbohydrate metabolism. *Gen. Comp. Endocrinol.* 35, 133–145.
- Qi, H., Song, K., Li, C., Wang, W., Li, B., Li, L., Zhang, G., 2017. Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). *PLoS One* 12, 1–16. <https://doi.org/10.1371/journal.pone.0174007>.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: protein domains identifier. *Nucl. Acids Res.* 33, 116–120. <https://doi.org/10.1093/nar/gki442>.
- Riehle, M.A., Fan, Y., Cao, C., Brown, M.R., 2006. Molecular characterization of insulin-like peptides in the yellow fever mosquito, *Aedes aegypti*: expression, cellular localization, and phylogeny. *Peptides* 27, 2547–2560. <https://doi.org/10.1016/j.peptides.2006.07.016>.
- Rivière, G., Klopp, C., Ibouniyamine, N., Huvet, A., Boudry, P., Favrel, P., 2015. GigaTON: an extensive publicly searchable database providing a new reference transcriptome in the Pacific oyster *Crassostrea gigas*. *BMC Bioinf.* 16, 401. <https://doi.org/10.1186/s12859-015-0833-4>.
- Robitzki, A., Schröder, H.C., Ugarkovic, D., Pfeifer, K., Uhlenbruck, G., Müller, W.E., 1989. Demonstration of an endocrine signaling circuit for insulin in the sponge *Geodia cydonium*. *EMBO J.* 8, 2905–2909.
- Safavi-Hemami, H., Lu, A., Li, Q., Fedosov, A.E., Biggs, J., Showers Corneli, P., Seger, J., Yandell, M., Olivera, B.M., 2016. Venom insulins of cone snails diversify rapidly and track prey taxa. *Mol. Biol. Evol.* 33, 2924–2934. <https://doi.org/10.1093/molbev/msw174>.
- Saltiel, A.R., Kahn, C.R., 2001. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* 414, 799–806. <https://doi.org/10.1038/414799a>.
- Schindler, A.J., Baugh, L.R., Sherwood, D.R., 2014. Identification of late larval stage developmental checkpoints in *Caenorhabditis elegans* regulated by insulin/IGF and steroid hormone signaling pathways. *PLoS Genet.* 10, 13–16. <https://doi.org/10.1371/journal.pgen.1004426>.
- Shabanpoor, F., Separovic, F., Wade, J.D., 2009. In: Chapter 1 The Human Insulin Superfamily of Polypeptide Hormones, 1st ed. Vitamins and Hormones. Elsevier Inc. [https://doi.org/10.1016/S0083-6729\(08\)00601-8](https://doi.org/10.1016/S0083-6729(08)00601-8).
- Shagin, D.A., Rebrikov, D.V., Kozhemyako, V.B., Altschuler, I.M., Shcheglov, A.S., Zhulidov, P.A., Bogdanova, E.A., Staroverov, D.B., Rasskazov, V.A., Lukanov, S., 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Res.* 12, 1935–1942. <https://doi.org/10.1101/gr.547002>.
- Sharabi, O., Manor, R., Weil, S., Aflalo, E.D., Lezer, Y., Levy, T., Aizen, J., Ventura, T., Mather, P.B., Khalaila, I., Sagi, A., 2016. Identification and characterization of an insulin-like receptor involved in crustacean reproduction. *Endocrinology* 157, 928–941. <https://doi.org/10.1210/en.2015.1391>.
- Shingleton, A.W., Das, J., Vinicius, L., Stern, D.L., 2005. The temporal requirements for insulin signaling during development in *Drosophila*. *PLoS Biol.* 3, 1607–1617. <https://doi.org/10.1371/journal.pbio.0030289>.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. <https://doi.org/10.1038/msb.2011.75>.
- Sliwowska, J.H., Fergani, C., Gawalek, M., Skowronska, B., Fichna, P., Lehman, M.N., 2014. Insulin: Its role in the central control of reproduction. *Physiol. Behav.* 133, 197–206. <https://doi.org/10.1016/j.physbeh.2014.05.021>.
- Smit, A.B., Vreugdenhil, E., Ebberink, R.H., Geraerts, W.P., Klootwijk, J., Joosse, J., 1988. Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. *Nature* 331, 535–538. <https://doi.org/10.1038/331535a0>.
- Smit, A.B., Geraerts, P.M., Meester, I., van Heerikhuizen, H., Joosse, J., 1991. Characterization of a cDNA clone encoding molluscan insulin-related peptide II of *Lymnaea stagnalis*. *Eur. J. Biochem.* 199, 699–703.
- Smit, A.B., Thijsen, S.F., Geraerts, W.P., Meester, I., van Heerikhuizen, H., Joosse, J., 1992. Characterization of a cDNA clone encoding molluscan insulin-related peptide V of *Lymnaea stagnalis*. *Brain Res. Mol. Brain Res.* 14, 7–12.
- Smit, A.B., van Marle, A., van Elk, R., Bogerd, J., van Heerikhuizen, H., Geraerts, W.P., 1993. Evolutionary conservation of the insulin gene structure in invertebrates: cloning of the gene encoding molluscan insulin-related peptide III from *Lymnaea stagnalis*. *J. Mol. Endocrinol.* 11, 103–113.
- Smit, A.B., Spijker, S., Van Minnen, J., Burke, J.F., De Winter, F., Van Elk, R., Geraerts, W.P., 1996. Expression and characterization of molluscan insulin-related peptide VII from the mollusc *Lymnaea stagnalis*. *Neuroscience* 70, 589–596. [https://doi.org/10.1016/0304-4522\(95\)00378-9](https://doi.org/10.1016/0304-4522(95)00378-9).
- Smit, A.B., van Kesteren, R.E., Li, K.W., Van Minnen, J., Spijker, S., Van Heerikhuizen, H., Geraerts, W.P., 1998. Towards understanding the role of insulin in the brain: lessons from insulin-related signaling systems in the invertebrate brain. *Prog. Neurobiol.* 54, 35–54. [https://doi.org/10.1016/S0303-0082\(97\)00063-4](https://doi.org/10.1016/S0303-0082(97)00063-4).
- Söderberg, J.A.E., Birse, R.T., Nässel, D.R., 2011. Insulin production and signaling in renal tubules of *Drosophila* is under control of tachykinin-related peptide and regulates stress resistance. *PLoS One* 6. <https://doi.org/10.1371/journal.pone.0019866>.
- Steiner, D.F., Chan, S.J., Welsh, J.M., Kwok, S.C.M., 1985. Structure and evolution of the insulin gene. *Annu. Rev. Genet.* 19, 463–484. <https://doi.org/10.1146/annurev.ge.19.120185.002335>.
- Veenstra, J.A., 2010. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. *Gen. Comp. Endocrinol.* 167, 86–103. <https://doi.org/10.1016/j.ygcen.2010.02.010>.
- Veenstra, J.A., Agricola, H.-J., Sellami, A., 2008. Regulatory peptides in fruit fly midgut. *Cell Tissue Res.* 334, 499–516. <https://doi.org/10.1007/s00441-008-0708-3>.
- Ventura, T., Manor, R., Aflalo, E.D., Weil, S., Raviv, S., Glazer, L., Sagi, A., 2009. Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypic gender differences and spermatogenesis. *Endocrinology* 150, 1278–1286. <https://doi.org/10.1210/en.2008-0906>.
- Ventura, T., Rosen, O., Sagi, A., 2011. From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades. *Gen. Comp. Endocrinol.* 173, 381–388. <https://doi.org/10.1016/j.ygcen.2011.05.018>.
- Wang, S., Luo, X., Zhang, S., Yin, C., Dou, Y., Cai, X., 2013. Identification of putative insulin-like peptides and components of insulin signaling pathways in parasitic plathelminths by the use of genome-wide screening. *FEBS J.* 281, 877–893. <https://doi.org/10.1111/febs.12655>.
- Wilkinson, T.N., Speed, T.P., Tregear, G.W., Bathgate, R.A., Sherwood, O., Bathgate, R., Samuel, C., Burazin, T., Layfield, S., Claasz, A., Reyntomas, I., Dawson, N., Zhao, C., Bond, C., Summers, R., Parry, L., Wade, J., Tregear, G., Bullesbach, E., Schwabe, C., Bullesbach, E., Schwabe, C., Hansell, D., Bryant-Greenwood, G., Greenwood, F., Winslow, J., Shih, A., Bourell, J., Weiss, G., Reed, B., Stults, J., Goldsmith, L., Sherwood, O., Zimmermann, S., Steding, G., Emmen, J., Brinkmann, A., Nayernia, K., Holstein, A., Engel, W., Adham, I., Nef, S., Parada, L., Bieche, I., Laurent, A., Laurendeau, I., Duret, L., Giovannardi, Y., Frendo, J., Olivieri, M., Fausser, J., Evain-Brion, D., Vidaud, M., Conklin, D., Lofton-Day, C., Haldeman, B., Ching, A., Whitmore, T., Lok, S., Jaspers, S., Lok, S., Johnston, D., Conklin, D., Lofton-Day, C., Adams, R., Jelmsberg, A., Whitmore, T., Schrader, S., Griswold, M., Jaspers, S., Hsu, S., Nakabayashi, K., Nishi, S., Kumagi, J., Kudo, M., Sherwood, O., Hsueh, A., Sudo, S., Kumagai, J., Nishi, S., Layfield, S., Ferraro, T., Bathgate, R., Hsueh, A., Kumagai, J., Hsu, S., Matsumi, H., Roh, J., Fu, P., Wade, J., Bathgate, R., Hsueh, A., Bogatcheva, N., Truong, A., Feng, S., Engel, W., Adham, I., Agoulnik, A., Liu, C., Eriste, E., Sutton, S., Chen, J., Roland, B., Kuei, C., Farmer, N., Jorvall, H., Sillard, R., Lovenberg, T., Liu, C., Chen, J., Sutton, S., Roland, B., Kuei, C., Farmer, N., Sillard, R., Lovenberg, T., Hsu, S., Kudo, M., Chen, T., Nakabayashi, K., Bhalla, A., van der Spek, P., van Duin, M., Hsueh, A., Schwabe, C., Bullesbach, E., Heyn, H., Yoshioka, M., Schwabe, C., LeRoith, D., Thompson, R., Shiloach, J., Roth, J., Georges, D., Viguier-Martinez, M., Poirier, J., Georges, D., Schwabe, C., Roche, P., Crawford, R., Tregear, G., Klönisch, T., Froehlich, C., Tetens, F., Fischer, B., Hombach-Klönisch, S., Hsu, S., Bathgate, R., Scott, D., Chung, S., Ellyard, D., Garreffa, A., Tregear, G., Bathgate, R., Siebel, A., Tovote, P., Claasz, A., Macris, M., Tregear, G., Parry, L., de Rienzo, G., Aniello, F., Branno, M., Minucci, S., Bielawski, J., Yang, Z., Schwabe, C., Bullesbach, E., Schwabe, C., Gowan, L., Reinig, J., Schwabe, C., Warr, G., Dores, R., Rubin, D., Quinn, T., Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S., Ishizaki, H., Brogiolo, W.,

- Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., Hafen, E., Pierce, S., Costa, M., Wisotzky, R., Devadhar, S., Homburger, S., Buchman, A., Ferguson, K., Heller, J., Platt, D., Pasquinelli, A., Liu, L., Doberstein, S., Ruvkan, G., Hombach-Klonisch, S., Abd-Elnaeim, M., Skidmore, J., Leiser, R., Fischer, B., Klonisch, T., Bravo, P., Stewart, D., Lasley, B., Fowler, M., Abd-Elnaeim, M., Saber, A., Hassan, A., Abou-Elmagd, A., Klisch, K., Jones, C., Leiser, R., Brackett, K., Fields, P., Dubois, W., Chang, S., Peer de, Y. Van, Taylor, J., Meyer, A., Bullesbach, E., Schwabe, C., Callard, I., Steinetz, B., Schwabe, C., Callard, I., Goldsmith, L., Bullesbach, E., Schwabe, C., Lacy, E., Reinig, J., Daniel, L., Schwabe, C., Gowan, L., Steinetz, B., O'Byrne, E., Suzuki, Y., Nei, M., Suzuki, Y., Nei, M., Zhang, J., Wyckoff, G., Wang, W., Wu, C., Ting, C., Tsaur, S., Wu, M., Wu, C., Swanson, W., Clark, A., Waldrup-Dail, H., Wolfner, M., Aquadro, C., Rooney, A., Zhang, J., Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Thompson, J., Higgins, D., Gibson, T., Felsenstein, J., Schmidt, H., Strimmer, K., Vingron, M., Haeseler, A., Page, R. von, Page, R., Hedges, S., Yang, Z., Nielsen, R., Yang, Z., 2005. Evolution of the relaxin-like peptide family. *BMC Evol. Biol.* 5, 14. <https://doi.org/10.1186/1471-2148-5-14>.
- Yoshida, I., Tsuzuki, S., Abdel Salam, S.E., Ino, M., Korayem, A.M., Sakurai, S., Iwami, M., 1997. Bombyxin F1 Gene: structure and Expression of a New Bombyxin Family Gene That Forms a Pair with Bombyxin B10 Gene. *Zoolog. Sci.*
- Yoshida, I., Moto, K., Sakurai, S., Iwami, M., 1998. A novel member of the bombyxin gene family: structure and expression of bombyxin G1 gene, an insulin-related peptide gene of the silkworm *Bombyx mori*. *Dev. Genes Evol.* 208, 407–410. <https://doi.org/10.1007/s004270050197>.
- Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Wang, J., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W.H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, S., Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, J., Wang, Q., Steinberg, C.E.W., Wang, H., Li, N., Qian, L., Zhang, G., Li, Y., Yang, H., Liu, X., Yin, Y., Wang, J., 2012a. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54. <https://doi.org/10.1038/nature11413>.
- Zhang, L., Hou, R., Su, H., Hu, X., Wang, S., Bao, Z., 2012b. Network analysis of oyster transcriptome revealed a cascade of cellular responses during recovery after heat shock. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0035484>.
- Zhang, N., Xu, F., Guo, X., 2014. Genomic analysis of the pacific oyster (*Crassostrea gigas*) reveals possible conservation of vertebrate sex determination in a mollusc. *G3 Genes|Genomes|Genetics* 4, 2207–2217. <https://doi.org/10.1534/g3.114.013904>.
- Zhao, X., Yu, H., Kong, L., Li, Q., 2012. Transcriptomic responses to salinity stress in the pacific oyster *Crassostrea gigas*. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0046244>.
- Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., Siebert, P.D., 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30, 892–897. <https://doi.org/10.1126/science.aam8999>.
- Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A., Shagin, D.A., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucl. Acids Res.* 32. <https://doi.org/10.1093/nar/gnh031>. 37e–37.
- Zatylny-Gaudin, C., Cornet, V., Leduc, A., Zanuttini, B., Corre, E., Le Corguillé, G., Bernay, B., Garderes, J., Kraut, A., Couté, Y., Henry, J., 2016. Neuropeptidome of the cephalopod *sepia officinalis*: identification, tissue mapping, and expression pattern of neuropeptides and neurohormones during egg laying. *J. Proteome Res.* 15, 48–67. <https://doi.org/10.1021/acs.jproteome.5b00463>.