Molecular Cloning and Expression Analysis of cDNAs Encoding an Insulin-like Androgenic Gland Factor from Three Palaemonid Species, *Macrobrachium lar*, *Palaemon paucidens* and *P. pacificus*

Kota BANZAI[1], Susumu IZUMI[1,2] and Tsuyoshi OHIRA[1,2]*

[1] Course of Biological Sciences, Graduate School of Science, Kanagawa University (Hiratsuka, Kanagawa 259-1293, Japan)

[2] Department of Biological Sciences, Faculty of Science, Kanagawa University (Hiratsuka, Kanagawa 259-1293, Japan)

Abstract

An androgenic gland hormone (AGH) is synthesized in the male-specific androgenic gland (AG) and is responsible for sex differentiation in crustaceans. To date, AGH cDNAs have been cloned from only three isopod species, but not decapod crustaceans. Recently, however, cDNAs encoding the AG specific polypeptide, which was designated as an insulin-like AG factor (IAG), have been cloned from two decapods, the red-claw crayfish *Cherax quadricarinatus* and the giant freshwater prawn *Macrobrachium rosenbergii*. These studies suggested that IAG may be involved in controlling sex differentiation in decapod crustaceans, but this point remains unclear. In order to accumulate knowledge on IAG, here we cloned IAG cDNAs from three palaemonid species (Crustacea, Decapoda), *M. lar*, *Palaemon paucidens* and *P. pacificus*. The three IAG precursors consisted of a signal peptide, a B chain, a C peptide and an A chain. This organization corresponds to that of the known isopod AGH and decapod IAG precursors. The three IAGs showed higher sequence identities to *M. rosenbergii* IAG (49~72%) than to *C. quadricarinatus* IAG (29~32%). These results are consistent with their phylogenetic relationship. In the analysis of the tissue-specific gene expression of the cloned three IAGs by RT-PCR, each IAG was detected only in the AG. Our data obtained in this study will facilitate further investigation of sex differentiation controlled by the AG in decapods.

Introduction

The androgenic gland (AG) is specific to male crustaceans and responsible for regulating sex differentiation. The AG was first discovered in the amphipod *Orchestia gammarella*[3]. Since then, the role and function of the AG have been examined by the following *in vivo* bioassays. In the terrestrial isopod *Armadillidium vulgare*, the implantation of the AGs into young females induced not only the formation of a male-specific outer morphology, but also gonadal masculinization[9]. Conversely, the ablation of the AGs caused young males to become feminized[24]. Since the injection of an extract from AGs into juvenile females showed the same phenomena as the implantation of the AG[9], it has been thought that male sex

differentiation is regulated by the androgenic gland hormone (AGH) synthesized in and secreted from the AG.

Purification of AGH from *A. vulgare* has been attempted by some research groups[7,12,17]. Purified *A. vulgare* AGH (Arvul-AGH) was found to be inactivated by treatment with trypsin or reductive carboxymethylation, indicating that Arvul-AGH is a protein with disulfide bond(s)[17]. Subsequently, the primary structure of Arvul-AGH was determined to be a heterodimeric glycoprotein[13] and a cDNA encoding Arvul-AGH precursor was cloned[18]. Similarly to pro-insulin, Arvul-AGH is expressed as a single peptide chain consisting of a B chain, a C peptide and an A chain (pro-AGH). Removal of the C peptide by processing at two dibasic sites, Lys-Arg, converted pro-AGH into a mature heterodimeric form comprising the B and A chains. The mature Arvul-AGH

had an *N*-linked glycan moiety in the A chain, which was essential for conferring biological activity[19]. The mature Arvul-AGH possessed eight Cys residues, four of which formed two inter-chain disulfide bonds between the B and A chains, while the remaining four residues formed two intra-chain disulfide bonds within the B and A chains, respectively[19].

In addition to Arvul-AGH, to date, cDNAs encoding AGH precursors have been cloned from two terrestrial isopods *Porcellio scaber* and *P. dilatatus*[20]. The two AGH precursors consisted of a signal peptide, a B chain, a C peptide and an A chain, which are organized similarly to that of the Arvul-AGH precursor. The amino acid sequences of the A and B chains were highly conserved among the three isopod species, while those of the C peptides exhibited low sequence similarity. Both the AGHs contained a single *N*-glycosylation motif (Asn-Xaa-Ser/Thr) in the A chain, the position of which was completely conserved with that of Arvul-AGH.

In decapod crustaceans, the role of AG has also been studied. In the giant freshwater prawn *Macrobrachium rosenbergii*, implantation of AGs into young female prawns caused these genetic females to become phenotypic reproductive males[15]. In the red-claw crayfish *Cherax quadricarinatus*, the implantation of AGs, which were hypertrophied due to having been destalked bilaterally, into immature female animals induced male-specific morphologies and inhibited female secondary sex characteristics in recipient females[10]. These results indicate that AGs of decapod species also control sex differentiation. Recently, cDNA libraries were constructed via suppression subtractive hybridization of the AGs of *C. quadricarinatus* and *M. rosenbergii*[11,26]. Expressed sequence tags from these libraries were sequenced whereupon two AG-specific genes of *C. quadricarinatus* and *M. rosenbergii*, respectively, were detected. Both gene products consisted of a signal peptide, a B chain, a C peptide and an A chain. The organizations of these gene products were the same as that of insulin precursor; therefore, they were named as an insulin-like AG factor (IAG). *C. quadricarinatus* IAG (Chqua-IAG) and *M. rosenbergii* IAG (Maros-IAG) had six and eight cysteine residues, respectively, in their predicted mature peptides (A and B chains). The positions of six cysteine residues of Chqua-IAG were the same as six of the eight cysteine residues in Maros-IAG[11,26]. Furthermore, the six cysteine residues were well-conserved within an insulin family, including isopod AGHs. Therefore, the two IAG molecules were thought to be putative AGHs in decapod species. However, Chqua-IAG and Maros-IAG showed low sequence similarities with the three isopod AGHs and moreover, shared little similarity. Moreover, the biological activi-

ties of the two IAGs have not yet been examined.

Both the marine and freshwater shrimp aquacultures are significant global industries. To further develop shrimp aquaculture, more efficient seed production techniques are required. Monosex culture has the potential to improve the production efficiency of shrimp aquaculture based on dramatic increases in yield and profit due to observed sexual dimorphism in growth. Therefore, it is desirable to gain a better understanding of shrimp IAG, because mating sex-reversed shrimp males (neofemales) induced by the administration of AG[13], with normal males may result in all-male populations. In order to obtain information on the amino acid sequences of shrimp IAG, here we describe the cloning of cDNAs encoding IAG precursors from three palaemonid species (Crustacea, Decapoda), *Macrobrachium lar*, *Palaemon paucidens* and *P. pacificus*, and the respective AG-specific gene expression in each case.

Materials and methods

1. Animals

Mature males of *M. lar*, *P. paucidens* and *P. pacificus* were purchased from Internet shops and identified by their external morphology as described previously[14]. The muscle, hepatopancreas, eyestalk, nerve cord (thoracic ganglia and abdominal ganglia), testis with sperm duct, sperm duct, and terminal ampulla, including the androgenic gland, were dissected out under a microscope from each species, flash-frozen in liquid nitrogen and stored at −80°C until use.

2. Isolation of total RNA

Total RNAs were extracted from the above six tissues of the three palaemonid species by the acid guanidium thiocyanate-phenol-chloroform extraction method[4].

3. Cloning of *M. lar* IAG (Malar-IAG) cDNA

First strand cDNA for RT-PCR and 3′-rapid amplification of cDNA ends (RACE) was synthesized by reverse-transcription of 500 ng of total RNA obtained from the sperm duct and terminal ampulla, including the androgenic gland of *M. lar*, with the adaptor oligo-dT primer (Table 1) using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA). Using the first strand cDNA as a template, PCR was performed with two degenerate oligonucleotide primers (Malar-F and Malar-R in Table 1 and Fig. 1A), which were designed based on the amino acid sequence of Maros-IAG[26]. The reaction mixture (20 μl) containing 1/200 aliquot of the first strand cDNA, 0.5 μM of each primer and TaKaRa *LA Taq* (Takara Bio, Otsu, Shiga, Japan) was prepared according to the

Table 1. Nucleotide sequences of oligonucleotide primers used for RT-PCR, 5′-RACE and 3′-RACE

Primer name	Sequence (5′-3′)
Adaptor oligo-dT primer	AACTGGAAGAATTCGCGGCCGCAGGAA(T)$_{18}$
Malar-F	CACAAYAAYTAYATHAAYCC
Malar-R	GCGTTRTTRCARCAYTCYTC
Papau-F	GTTGACTTYGAYTGYGGNGA
Papau-R	GCGACYTCYTCTAARTTRCA
Papac-F	CTCTCCGTNGANTTYGAYTGY
Papac-R	ATGCAATAYTCNGCNACYTA
Malar 5′RACE-R1	GTCCCGTCGCAGGCGACGCTTGGACTGC
Malar 5′RACE-R2	TTCTTCGCGGCTCAGCGTCAGGTGCTGC
Papau 5′RACE-R1	GTGTCGGCGATTCGAGTGCAGCATCGTG
Papau 5′RACE-R2	GTTCGGTGAAGTTACGCACAGAACGCCG
Papac 5′RACE-R1	GTCGGAGTCTGCGGTTCGTGTGGAGCAT
Papac 5′RACE-R2	CCATGGTGGTCGGCACATCCCCACAATC
Malar 3′RACE-F1	CCAAGCGTCGCCTGCGACGG
Malar 3′RACE-F2	GGAAGACTCCGAGGGAGGAG
Papau 3′RACE-F1	TATGACGACCACCTTTGCGT
Papau 3′RACE-F2	TCCAAAGAACGGCGTTCTGT
Papac 3′RACE-F1	AGGGTATACAAACCATTCGT
Papac 3′RACE-F2	ATCCGAGAGATTTCGCTCAG
3AP-R1	AACTGGAAGAATTCGCGGCCG
3AP-R2	AAGAATTCGCGGCCGCAGGAA
Malar-cF	CCTCTTCCTTCTATTTCCGC
Malar-cR	CGGGATGATCTTTCAGACCT
Papau-cF	AGTTTAAAATGGCATACAGG
Papau-cR	CTAGGCTGAGCTGCAGGTGTTC
Papac-cF	AGTGAGGAGTCAGGGGTGCT
Papac-cR	GAAACTGACGGCCACAGACA
Actin-F	TGTAYGCCTCTGGYCGYACC
Actin-R	CVACRTCRCACTTCATGATGS

manufacturer's protocol. The cycling conditions were 94°C for 3 min, followed by 35 cycles of 94°C for 30 sec - 52°C for 30 sec - 72°C for 1 min, and 72°C for 5 min respectively.

For 5′-RACE, first-strand cDNA was newly synthesized with 500 ng of total RNA from the sperm duct and terminal ampulla, including the androgenic gland of *M. lar*, by using a SMART RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA) with the exception of MMLV Reverse Transcriptase (Clontech), which was substituted by SuperScript II Reverse Transcriptase (Invitrogen). Two specific primers (Malar 5′RACE-R1 and Malar 5′RACE-R2 in Table 1) were designed based on the nucleotide sequence of the Malar-IAG cDNA fragment obtained from RT-PCR. A cDNA fragment encoding the 5′-region of Malar-IAG was amplified by two rounds of PCR. In the first PCR, the first strand cDNA was used as a template, and amplification was primed by a pair of primers, a universal primer mix (UPM, Clontech) and Malar 5′RACE-R1, as shown in Fig. 1A. In the second PCR, the diluted first PCR product was used as a template, and amplification was primed by a pair of nested primers, a nested universal primer (NUP, Clontech) and Malar 5′RACE-R2, as shown in Fig. 1A. Each PCR mixture was prepared as described above. The cycling conditions for each PCR amplification were 94°C for 3 min, 5 cycles of 94°C for 30 sec - 72°C for 3 min, 5 cycles of 94°C for 30 sec - 70°C for 30 sec - 72°C for 3 min, 25 cycles of 94°C for 30 sec - 68°C for 30 sec - 72°C for 3 min, and 72°C for 5 min respectively.
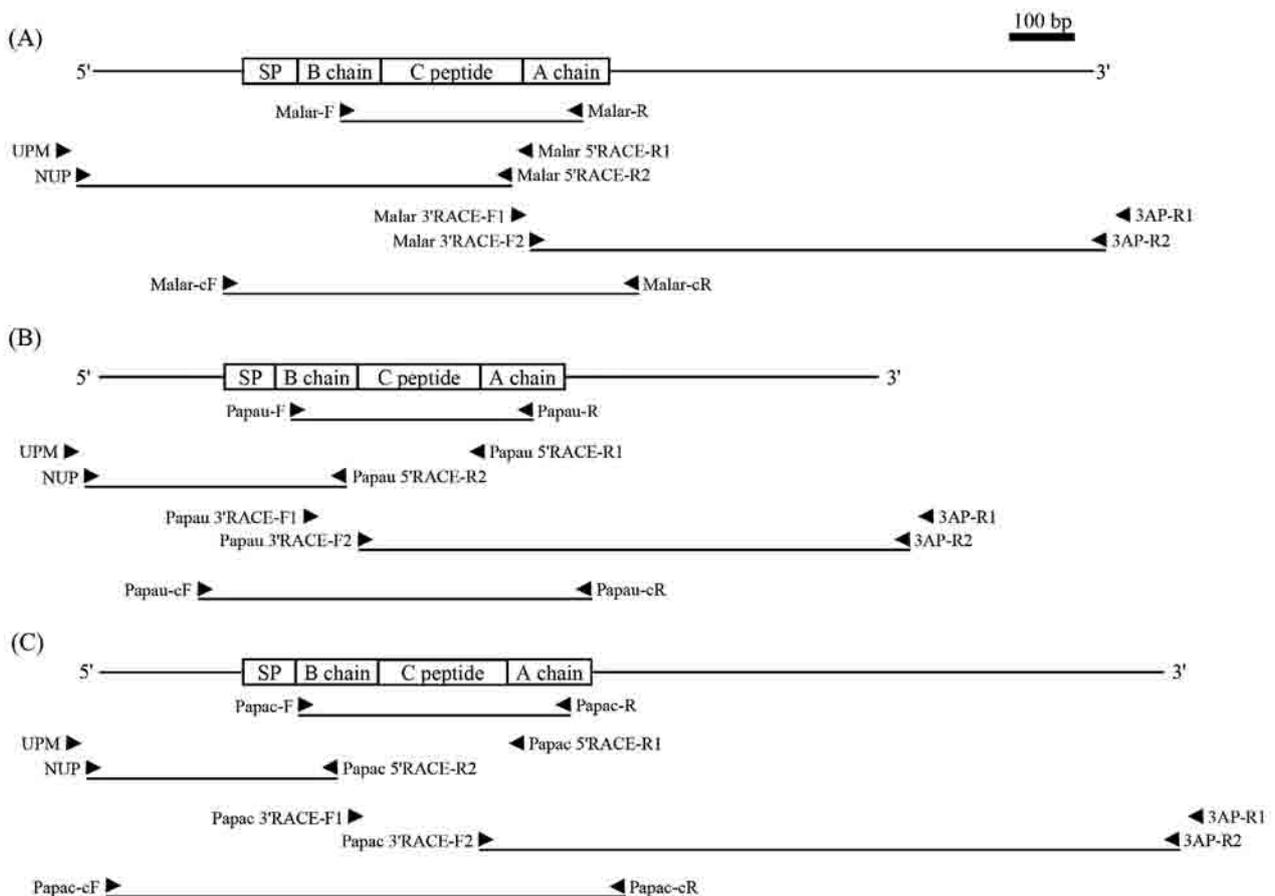
Fig. 1. **Schematic representation of the cDNA structures of (A)** *Macrobrachium lar* **IAG (Malar-IAG), (B)** *Palaemon*
paucidens **IAG (Papau-IAG) and (C)** *P. pacificus* **IAG (Papac-IAG), and the locations of the oligonucleotide primers**
Arrowheads represent the primers, and the lines under the arrowheads indicate the amplified cDNA fragments. Boxes
represent the open reading frame. Lines on both sides of the boxes indicate the 5′- and 3′-untranslated regions. SP is the
abbreviation for signal peptide.

For 3′-RACE, two specific primers (Malar 3′RACE-F1 and Malar 3′RACE-F2 in Table 1) were designed based on the nucleotide sequence of the Malar-IAG cDNA fragments obtained from RT-PCR. A cDNA fragment encoding the 3′-region of Malar-IAG was amplified by two rounds of PCR. In the first PCR, the first strand cDNA as described above was used as a template and amplification was primed by a pair of primers, Malar 3′RACE-F1 and 3AP-R1, as shown in Fig. 1A. In the second PCR, the diluted first PCR product was used as a template and amplification was primed by a pair of nested primers, Malar 3′RACE-F2 and 3AP-R2, as shown in Fig. 1A. Each PCR mixture was prepared as described above. The cycling conditions for each PCR amplification were 94°C for 3 min, followed by 35 cycles of 94°C for 30 sec - 53°C for 30 sec - 72°C for 3 min, and 72°C for 5 min respectively. The nucleotide sequence of Malar-IAG cDNA, including the full-length open reading frame (ORF), was confirmed by PCR amplification using spe-

cific primers, Malar-cF and Malar-cR shown in Table 1 and Fig. 1A.

**4. Cloning of *P. paucidens IAG* (Papau-IAG) and *P.
pacificus* IAG (Papac-IAG) cDNAs**

First strand cDNAs used in RT-PCR, 5′-RACE, and 3′-RACE for Papau-IAG and Papac-IAG were synthesized by the same methods as described above except for the use of the sperm ducts and terminal ampullae, including the androgenic glands from *P. paucidens* and *P. pacificus*, respectively, instead of *M. lar*.

RT-PCRs were performed with degenerate primers (Table 1, Figs. 1B and 1C), which were designed based on the amino acid sequences of Maros-IAG and Malar-IAG. The cycling conditions for Papau-IAG and Papac-IAG were 94°C for 3 min, followed by 40 cycles of 94°C for 30 sec - 52 °C for 30 sec - 72°C for 1 min, and 72°C for 5 min respectively.

5′-RACE and 3′-RACE were performed with gene

specific primers (Table 1, Figs. 1B and 1C), which were designed based on the nucleotide sequences of Papau-IAG and Papac-IAG cDNA fragments obtained from RT-PCR. The cycling conditions for 5′RACE and 3′RACE were the same as described above.

5. Nucleotide sequence analysis

All PCR products were subcloned into a pMD-20T vector using a Mighty TA-cloning kit (Takara Bio). Both strands of the plasmid DNAs were sequenced on an ABI PRISM 3130 genetic analyzer (Applied Biosystems, Foster City, CA, USA) using a Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

6. Sequence data analysis

Cleavage sites of the signal peptide of the three IAG precursors were predicted using the SignalP 3.0 program (http://www.cbs.dtu.dk/services/SignalP/). Propeptide cleavage sites and *N*-linked glycosylation sites of the three IAG precursors were predicted by the ProP 1.0 program (http://www.cbs.dtu.dk/services/ProP/) and the NetNGlyc 1.0 program (http://www.cbs.dtu.dk/services/NetNGlyc/), respectively. The Clustal W program[25] was used with default settings to align the amino acid sequences of IAG precursors from *C. quadricarinatus* (GenBank accession no. DQ851163), *M. rosenbergii* (FJ409645), *M. lar* (AB579012), *P. paucidens* (AB588013) and *P. pacificus* (AB588014).

7. Tissue-specific gene expression of Malar-IAG, Papau-IAG and Papac-IAG

The total RNA from six tissues; the muscle, hepatopancreas, eyestalk, nerve cord, testis with the sperm duct, and sperm duct and terminal ampulla including the androgenic gland from each species, were treated with DNase I (Takara bio) at 37°C for 30 min. The first strand cDNA of each tissue was synthesized by the same method as RT-PCR in the section of "Cloning of *M. lar* IAG (Malar-IAG) cDNA". The first strand cDNAs were used as templates and each amplification was primed by a pair of specific primers, Malar-cF and Malar-cR for Malar-IAG, Papau-cF and Papau-cR for Papau-IAG, or Papac-cF and Papac-cR for Papac-IAG, respectively (Table 1). Each PCR mixture was prepared as described above. As a positive control, an actin cDNA fragment in each tissue was amplified using two degenerate primers, actin-F and actin-R (Table 1), which were designed based on the amino acid sequences of two conserved regions of the vertebrate and invertebrate actins. The PCR products were separated on a 2.0% agarose gel and stained with ethidium bromide.

Results

1. Cloning of IAG cDNAs from *M. lar*, *P. paucidens* and *P. pacificus*

Amplification of IAG cDNA fragments from *M. lar*, *P. paucidens* and *P. pacificus* were performed by RT-PCR using degenerate oligonucleotide primers. Subsequently, the 5′- and 3′-regions of the three cDNAs were amplified by 5′- and 3′-RACE using specific primers based on the nucleotide sequences obtained by RT-PCR. In order to confirm each nucleotide sequence, each cDNA fragment, including the full-length open reading frame (ORF), was amplified. The three IAG cDNAs of *M. lar*, *P. paucidens* and *P. pacificus* were designated as Malar-IAG, Papau-IAG, and Papac-IAG, respectively, and deposited in the DNA Data Bank of Japan (DDBJ) database (accession nos. AB579012, AB588013 and AB588014, respectively).

2. Characterization of Malar-IAG cDNA

Nucleotide and deduced amino acid sequences of Malar-IAG cDNA are shown in Fig. 2A. The cDNA was 1,529 nucleotides in length, comprising a 5′-untranslated region (231 bp), an ORF (558 bp), and a 3′-untranslated region (740 bp). The 3′-untranslated region contained a canonical polyadenylation signal (AATAAA). ORF of Malar-IAG cDNA encoded a putative polypeptide consisting of 185 amino acid residues. The N-terminal segment included a high proportion of hydrophobic amino acid residues, and therefore the first 28 amino acid residues were predicted to be a signal peptide by the SignalP program. Four potential *N*-linked glycosylation motifs (Asn-Xaa-Ser/Thr) were observed, and the Asn residues at amino acid positions 42, 74, 80 and 156 were likely to be glycosylated. Putative cleavage sites between Arg[70] and Ser[71], and Arg[142] and Asp[143] were predicted using the ProP 1.0 program. After cleavage, two dibasic processing sites (Arg[69]-Arg[70] and Arg[141]-Arg[142]) were probably removed similarly to those of *A. vulgare* AGH[13] and *Homo sapiens* insulin[21].

3. Characterization of Papau-IAG cDNA

The nucleotide and deduced amino acid sequences of Papau-IAG cDNA are shown in Fig. 2B. The cDNA was 1,169 nucleotides in length, comprising a 5′-untranslated region (192 bp), an ORF (522 bp), and a 3′-untranslated region (455 bp). The 3′-untranslated region did not contain a canonical polyadenylation signal. The ORF of Papau-IAG cDNA encoded a putative polypeptide consisting of 173 amino acid residues. The 26 N-terminal amino acid residues were predicted to be a signal peptide

(A)

```
nt 0001 TTATTCCAAGAGGGGCCAAGACTCTGGGATCATACCTCGAAAGGCTCTTTGTGTCGCTTCCCCTCGTCCG
nt 0071 GTTAACCGGTGTTTTCTACCAGCGTTCTCGGCACCTAAAAATTCGCTTTCTGGCCAACCTTGCACTAATC
nt 0141 CTTCCAGTTCCCTCTTCCTTCTATTTCCGCACATCAACTTCTTCTACTCTCCAGATTTTCTACATCGAAG
nt 0211 TGAAACAAAATAAGCTACAAAATGGGATACTGGAATACCGCGGTGAAGAGTCTTTTCTTGTGGTCACTCG
aa 0001                                    M  G  Y  W  N  T  A  V  K  S  L  F  L  W  S  L
nt 0281 TAGCGCTACTGCATCTGCCCCAACCTTCCTCGAGCTACGAGATCGAATGTCTCACCGTCGAATTCGACTG
aa 0017 V  A  L  L  H  L  P  Q  P  S  S  S  Y  E  I  E  C  L  T  V  E  F  D  C
nt 0351 CAGCAACATTACTTATACCTTAAGCCGTGTCTGCCTGAGGCACAACAACTACATCGACCCAGGACCCTCC
aa 0041 S  N  I  T  Y  T  L  S  R  V  C  L  R  H  N  N  Y  I  D  P  G  P  S
nt 0421 TACATTTCCAGAGGGCGACGATCTGTTGACAACTATACCCAGGGTCCTAACTATACCGTACCTTCTACGA
aa 0064 Y  I  S  R  G  R  R  S  V  D  N  Y  T  Q  G  P  N  Y  T  V  P  S  T
nt 0491 AGTCTCCATCGCTCGTCACCCGAGAGCTACCCACTTGACCATGGCTGACGTAAAACCTCTGAAGGAATC
aa 0087 K  S  P  S  L  V  H  P  R  A  T  H  L  T  M  A  D  V  K  P  L  K  E  S
nt 0561 CAAGGTGGAGGAAGAAGAGGAGGCAAGGATGCAGCACCTGACGCTGAGCCGCGAAGAAGCGAGCAAGATG
aa 0111 K  V  E  E  E  E  A  R  M  Q  H  L  T  L  S  R  E  E  A  S  K  M
nt 0631 CTGCAGTCCAAGCGTCGCCTGCGACGGGACAATGTGAGGAAGACTCCGAGGGAAGAATGTTGCAATAACG
aa 0134 L  Q  S  K  R  R  L  R  R  D  N  V  R  K  T  P  R  E  E  C  C  N  N
nt 0701 ACTCCTTCAGACGTTGCACCTTCGAGGAAGTCGCCGAATATTGCATTGAACTGAATCCCGGCATCAACAC
aa 0157 D  S  F  R  R  C  T  F  E  E  V  A  E  Y  C  I  E  L  N  P  G  I  N  T
nt 0771 CTGCGGAAGCTCCAGGTAGGAGGTCTGAAAGATCATCCCGTCCCGGTCTTATATGCTTGACACAAGATGT
aa 0181 C  G  S  S  R  *
nt 0841 TTAAAGTCAATTCACAGTCTTCGAAATTGCAACATTACTTTTGTGAAGTAGAATGACCAACTAAAGCTGT
nt 0911 CTCTTTGTCTTTCCTCACAATCAACTAAAATATGATTTGTCTTACCCTTACCTTCAGCTAAATGACTCTC
nt 0981 TTGTCTCTGCTTTGAATTCAACATTCGTCTTCTTTGCCCTTAGTTCAGCTTATGGCTTCTGTTTGTCTTA
nt 1051 TCTTTAACATCCACAAAGGTTTGGATTGTCCCACCTTTACCTTCCGCTAAATCTTTCTTTTGTCTTCCCG
nt 1121 CTACTTTAGTCACACTTACTGTCGGCTTGGGATTTTTTATTGGCCACTTTTACCTTCAGCTAAAGGTTC
nt 1191 CTTTTGTCTCAGCCTTGCCTTCCACTAATGGTTCCTCTTGTCTCAACCTTGCCTTCAGCTAATGCTTCCT
nt 1261 CTTGGCTCAACCTTGCCCTCAGCCAAAGATTCCTCGTGTCTAACCCTTGCTTTCAGCTAAACTCTCATTT
nt 1331 ACCTTTGGCTAAAGATTTCCTTTGCTGCACCCTTGCCTAAACGTTCTAATGTCATACCTTTACACGCAAC
nt 1401 AGCATCTAGACATTTCCGAACAATAAACATATTGCATTATTATTGGTGATTCTTGTCTTATGGTGATTCT
nt 1471 TGTCAATGTCTCCGAAAAATTGTTTGATACATCAGTTATTCGTCAAATAAAGGCTCTTG (A)ₙ
```

(B)

```
nt 0001 TTATTCAGTGAGAGGTTAGGAGTTTCGCTCAGACTCAGAGTGGCTCGTTTCCCGCCTATCGCTAGTTTAA
nt 0071 CCCAAACCTATCTCTCTCTCTCTCTATCTCTCGCTCTCTCGATATATATATATACATATATTACTT
nt 0141 CCCACATTCTGATATTTAATTTGTGAAGTCTACAATCAAATTACAGTTTAAAATGGCATACAGGAATATC
aa 0001                                                              M  A  Y  R  N  I
nt 0211 TCAGTCAAGGTGTTCTTTTGTATGGTGATGGCATTCTACTTGGCACAACCTGCTTCCAGTTATACGATTG
aa 0007 S  V  K  V  F  F  C  M  V  M  A  F  Y  L  A  Q  P  A  S  S  Y  T  I
nt 0281 AATGTCTTTCCGTTGATTTTGACTGTGGCGATATGACGACCACCTTTGCGTCCATTTGCAAAAGATATAA
aa 0030 E  C  L  S  V  D  F  D  C  G  D  M  T  T  T  F  A  S  I  C  K  R  Y  N
nt 0351 TGGCTACATCAACCCTGGACCAGCATACATTTCCAAAGAACGGCGTTCTGTGCGTAACTTCACCGAACAT
aa 0054 G  Y  I  N  P  G  P  A  Y  I  S  K  E  R  R  S  V  R  N  F  T  E  H
nt 0421 CCAACCAAGTCTTCCCCCACCTTCTTTCACCCTCGAGCCACCCACTTGACAAAGCGTGACGAAGAAACTC
aa 0077 P  T  K  S  S  P  T  F  F  H  P  R  A  T  H  L  T  K  A  D  E  E  T
nt 0491 TGAAGACGTTGGAGATGCCATGGGAGATTCGTGACATGTCGCTCGGTCGTAAGGAAGCAAACATGATGCT
aa 0100 L  K  T  L  E  M  P  W  E  I  R  D  M  S  L  G  R  K  E  A  N  M  M  L
nt 0561 GCACTCGAATCGCCGACACCGACGGGACAGCGTGAGGAAGACTCCGAGGGAGGAGTGCTGTAGCAACACC
aa 0124 H  S  N  R  R  H  R  R  D  S  V  R  K  T  P  R  E  E  C  C  S  N  T
nt 0631 ACTTTCAGACGCTGCACTTACGAGGAAGTCGCAAAATATTGCATTGAAGTGCGTCCCGGAGTGAACACCT
aa 0147 T  F  R  R  C  T  Y  E  E  V  A  K  Y  C  I  E  V  R  P  G  V  N  T
nt 0701 GCAGCTCAGCCTAGGACGGATAATGGATGTTCCCATGGCCGTTTCACGCCTGCCTTTTTGAAAGCCAAAT
aa 0170 C  S  S  A  *
nt 0771 GAAAATCCGCGAAAACGTTACTAATTCATCAACCTTATATTCAAGATCTTCAGAGAACTAAGTCTGACCT
nt 0841 TCATCATAAGCTTTCATTGATTTAAGTTAACCTTCGTTGAAAGCTTTCATTATCTTCGGCTCTTCTTGAC
nt 0911 TTTTCTTTTCGGGTAACCATCAGCGAAAACTTACTTCAGTTTAGGCTCACCTTCGGCGAATGTTTTCTTG
nt 0981 ATTTATGCTAACCTGCAATGGAAAATTTCTTTGATTTCAGAGTACTCTCTGCTAAAATTCTATTTTATCA
nt 1051 GACACTTTCACTTCCTAACACGCAAACATTACTGAACTCTTAGCAAATTGTATATCATCGGTTCCATAAA
nt 1121 ATTGTTTTATGCTTTAGTAATTGCAAAATAATTATATTTGAGAATATTG (A)ₙ
```

Fig. 2. The nucleotide and deduced amino acid sequences of cDNAs encoding (A) Malar-IAG, (B) Papau-IAG and (C) Papac-IAG

The putative signal peptide sequences are marked in bold, the putative B chains are underlined, the putative C peptides are in italics, and the putative A chains are dotted-underlined. The gray colored boxes and open boxes indicate the putative *N*-linked glycosylation motifs (N-X-S/T) and the putative dibasic processing sites. The asterisks indicate the putative stop codons. Polyadenylation signals are shown by wavy lines. The nucleotide sequences of the cDNAs encoding Malar-IAG, Papau-IAG and Papac-IAG have been deposited in the DNA Data Bank of Japan (accession nos. AB579012, AB588013 and AB588014, respectively).

(C)

```
nt 0001 TTATTCAGTGAGGAGTCAGGGGTGCTCGCTCGGACTCGGTGGATCATTTCCCACGAATCGCTTGTTTAAC
nt 0071 CCAAAACTCTCTCGCTTTCTCTCTCTCTCTTTCTCTCTCTCTTTCAGTCTCGCTCTCTCTCGGCTTGTTT
nt 0141 TTTCTCGTTCCTAAAAGCTTATATAACTTCCTGCTTTCTCATACTTACTGTGTGGAGTCGACAATCAAAT
nt 0211 TACTGCAGAAAATGGCAGACTGGAATATCGGAGTCAAGAATGTGTTCTTTTGTGTCTTGCTGGCGTTATG
aa 0001                 M  A  D  W  N  I  G  V  K  N  V  F  F  C  V  L  L  A  L  C
nt 0281 CTGGATGCAGCCTGCTTCCAGCTACAGCATTGAATGTCTTACGGTTGATATGGATTGTGGGGATGTGCCG
aa 0021 W  M  Q  P  A  S  S  Y  S  I  E  C  L  T  V  D  M  D  C  G  D  V  P
nt 0351 ACCACCATGGCATCAGTTTGTAGGGTATACAAACCATTCGTCCCCCTTAATCATCCCTACCCTTCCAAAA
aa 0044 T  T  M  A  S  V  C  R  V  Y  K  P  F  V  P  L  N  H  P  Y  P  S  K
nt 0421 AGCGTCGTTCTGTGAGCAACTCCACCGAGCACAATCTAGTCGAGTCTTCCTCGTCTTCCTTCCACCCACG
aa 0067 K  R  R  S  V  S  N  S  T  E  H  N  L  V  E  S  S  S  S  F  H  P  R
nt 0491 AGCCACCCACCTGACGAAGGCAAAGGCTGACCAAGACGTCGGACGGCGTTGGAGATGGCAAGCAAGATC
aa 0091 A  T  H  L  T  K  A  K  A  D  Q  D  S  R  T  A  L  E  M  A  S  K  I
nt 0561 CGAGAGATTTCGCTCAGTCGGGAGGAAGCAAACACAATGCTCCACACGAACCGCAGACTCCGACGCCAGG
aa 0114 R  E  I  S  L  S  R  E  E  A  N  T  M  L  H  T  N  R  R  L  R  R  Q
nt 0631 TGGGGAGGAAGACGCTGAGGGAGGAGTGCTGTGGCAACACGCCCTTCAGACCATGCACCTACGAGGAAGT
aa 0137 V  G  R  K  T  L  R  E  E  C  C  G  N  T  P  F  R  P  C  T  Y  E  E  V
nt 0701 CGCAGAATATTGCGAGGTGTTGCATGACGAAGCCCTCACCTGCCCCAGACCTTAGGACTGTTTATAAACG
aa 0161 A  E  Y  C  E  V  L  H  D  E  A  L  T  C  P  R  P  *
nt 0771 TTCCCTGTCTGTGGCCGTCAGTTTCACACCTGTCAGTCAGTCAAATGGAAATCCCTGGAAACGTCACTAA
nt 0841 TCTATCAACCGCATCTTCAAGACCTTCAGATGACCAAGTCTGGCCTTCATTATAAGCTTTTCTTTGGATC
nt 0911 CAAGTTGACATTCTGTGAAAGCTTTCTTTTTCTTGTGCTTTCTTTCGGTGAAATAACCTTTGGTTTACGA
nt 0981 ACTGACCTCCAGATAAAGATTTTTAGACTTACAATCTTCAGATTAGGCTGACTTCCAGCTTATGTTTTCT
nt 1051 TTATTTACGCTAATCTTTTTGGTTTAGACCTACTCTCTGTTAAAGCTTTATTTATCATTGATTTAGACCT
nt 1121 ACTCTCAGCTAAAGCTTTGAATCTTTGATTTAGACCTACTCTCAGCTAAAGTTTCGACTCTTTGATTTAA
nt 1191 ACCTACTCTCAGCTAAAGCTTTGAATCTTTGATTTAGACCTACTCTCAGCTAAAGCTTTGAATCTTTGAT
nt 1261 TTAGACCTACTTTCAGCTAAAGTTTCGACTCTTTGATTTAAACCTACTTTCAGCTAAAGCTTTGAATCTT
nt 1331 TGATTTAGACATACTTTCAGCTAAAGTTTCGACTCTTTGATTTAAACCTACTCTCAGCTAAAGCTTTGAA
nt 1401 TATTTGATTTAGACTACTCTCAGATAAAGCTTTGAATCTTTGATTTAGACCTACTCTCAGCTAAAGCTTT
nt 1471 GAATCTTTATTTCAAGTGACTACTTCACTTACTACCACAAATACGTTACCGAACTCTTACCAGATTGTAT
nt 1541 TTCACCGCTTCCAGAAAATTATCTTGTTTCACTAATTCGCAAAATGAATAAATTTGAGAATGTAG (A)n
```


(A)

	1	2	3	4	5	6	
IAG							— 659 bp
Actin							— 450 bp

(B)

IAG							— 529 bp
Actin							— 450 bp

(C)

IAG							— 788 bp
Actin							— 450 bp

Fig. 3. **RT-PCR analyses for tissue-specific gene expression of (A) Malar-IAG, (B) Papau-IAG and (C) Papac-IAG**

First strand cDNAs from total RNAs of the muscle (lane 1), hepatopancreas (lane 2), eyestalk (lane 3), nerve cord (lane 4), testis with the sperm duct (lane 5), and sperm duct and terminal ampulla, including the androgenic gland (lane 6), were used as templates. Upper and lower electrophoresis images are of PCR products using the IAG-specific primers for each species and the actin degenerate primers, respectively (see Materials and Methods).

by the SignalP program. Two potential N-linked glycosylation motifs were observed. Putative cleavage sites between Arg^{68} and Ser^{69}, and Arg^{131} and Asp^{132} were predicted using the ProP 1.0 program.

4. Characterization of Papac-IAG cDNA

The nucleotide and deduced amino acid sequences of Papac-IAG cDNA are shown in Fig. 2C. The cDNA was 1,605 nucleotides in length, comprising a 5′-untranslated region (221 bp), an ORF (534 bp), and a 3′-untranslated region (850 bp). The 3′-untranslated region contained a canonical polyadenylation signal (AATAAA). The ORF of Papac-IAG cDNA encoded a putative polypeptide consisting of 177 amino acid residues. The 27 N-terminal amino acid residues were predicted to be a signal peptide by the SignalP program. A potential N-linked glycosylation motif was observed. Putative cleavage sites between Arg^{69} and Ser^{70}, and Arg^{135} and Gln^{136} were predicted using the ProP 1.0 program.

5. Tissue-specific gene expression of Malar-IAG, Papau-IAG and Papac-IAG

In RT-PCR, using specific primer pairs for Malar-IAG, Papau-IAG and Papac-IAG, cDNA fragments with corresponding sizes of 659, 529 and 788 bp, respectively, were detected only in cDNA synthesized from the sperm

duct and terminal ampulla, including the androgenic gland, but not from the muscle, hepatopancreas, eyestalk, nerve cord, or testis with the sperm duct (Figs. 3A-C). An actin cDNA fragment of each species as a positive control was detected in all tissues (Figs. 3A-C).

Discussion

In this study, three cDNAs were isolated from three palaemonid species, *M. lar*, *P. paucidens* and *P. pacificus*. All three cDNAs encoded polypeptides consisted of a signal peptide, a B chain, a C peptide, and an A chain. The organization of the three molecules corresponds to that of the known two IAGs, Chqua-IAG and Maros-IAG[11,26]. Moreover, the three transcripts were expressed only in the AG. Collectively, in this study, the three polypeptides from *M. lar*, *P. paucidens*, *P. pacificus* were designated as Malar-IAG, Papau-IAG and Papac-IAG, respectively.

The amino acid sequences of Malar-IAG, Papau-IAG and Papac-IAG precursors were compared with those of two known molecules, Maros-IAG and Chqua-IAG precursors (Fig. 4). Malar-IAG, Papau-IAG and Papac-IAG showed higher sequence identities to Maros-IAG (72, 69 and 49%, respectively) than to Chqua-IAG (32, 29 and 32%, respectively). These results are consistent with the following phylogenetic relationship. *M. lar*, *P. paucidens* and *P. pacificus* as well as *M. rosenbergii* belong to the same family (Palaemonidae), whereas *C. quadricarinatus* is a freshwater crayfish of the family

Parastacidae. In isopods, however, sequence identities of AGH precursors between different families, the Armadilliidae and the Porcellionidae, are clearly higher (63% between Arvul-AGH and *Porcellio scaber* AGH, and 65% between Arvul-AGH and *P. dilatatus* AGH)[20] than those of IAG between the Palaemonidae, and Parastacidae (around 30%). This may be due to differences in the molecular evolutionary pathway between IAG and AGH. In order to clarify this point, further studies characterizing the amino acid sequences of additional IAGs and AGHs are required.

Insulin family peptides have six conserved cysteine residues, four of which are harbored in the A chain and the remaining two contained in the B chain[1,2,5,8,16,22,27]. The six cysteine residues in insulin family peptides form two inter-chain disulfide bonds between the B and A chains, and an intra-chain disulfide bond in the A chain. All three isopod AGHs also possess the six conserved cysteine residues, and are thus considered members of the insulin family. The isopod AGHs have two additional conserved cysteine residues in the B chain[13,18,20]. In *A. vulgare*, it has already been demonstrated that the two additional cysteine residues form an intra-chain disulfide bond in the B chain[13,19]. The four palaemonid IAGs (Malar-IAG, Papau-IAG, Papac-IAG and Maros-IAG) also have six conserved cysteine residues (open boxes in Fig. 4) and two additional cysteine residues (gray boxes in Fig. 4). The additional two cysteine residues, one of which is in the A chain and the other in the B chain, are conserved among palaemonid IAGs, but not found in
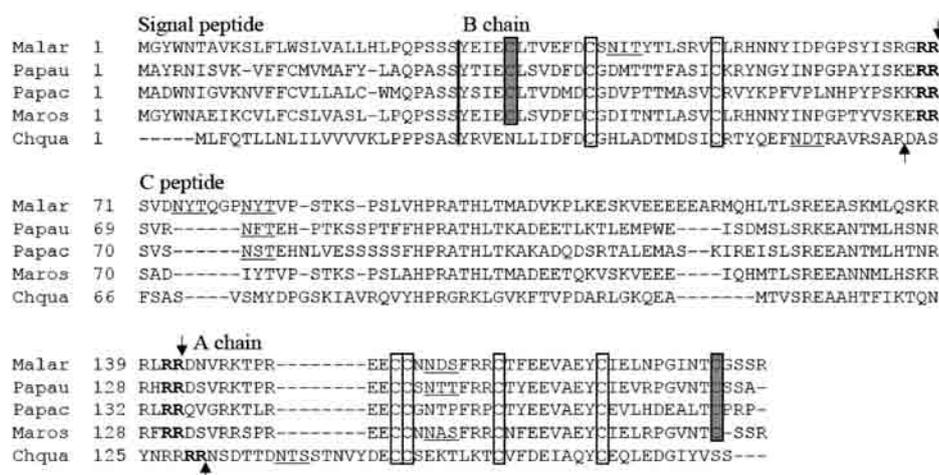

Fig. 4. Amino acid sequence alignment of Malar-IAG, Papau-IAG, Papac-IAG, *Macrobrachium rosenbergii* IAG (Maros-IAG)[11], and *Cherax quadricarinatus* IAG (Chqua-IAG)[26]

Open boxes and gray open boxes indicate the positions of conserved Cys residues in all species and only in palaemonid species, respectively. The putative cleavage sites predicted by the ProP 1.0 program are shown by arrows. The boundary between the signal peptide and the B chain is represented by a vertical line. Bold letters and underlines represent the di-basic cleavage sites and the putative *N*-linked glycosylation motifs (N-X-S/T), respectively.

parastacid IAG, Chqua-IAG (Fig. 4). In addition, the positions of the two additional cysteine residues in palaemonid IAGs differ from those of isopod AGHs, both of which are in the B chain. There is no evidence that the additional two cysteine residues in palaemonid IAGs form inter-chain disulfide bonds between the B and A chains, but *Caenorhabditis elegans* insulin-like peptide and *Lymnaea stagnalis* molluscan insulin-related peptide VII are known as insulin family peptides, which include two additional cysteine residues besides the six conserved cysteine residues[6,23].

In isopods, each of the three AGH has a single conserved *N*-glycosylation motif (Asn-Xaa-Ser/Thr) in the A chain[18,20]. Conversely, the positions and number of *N*-glycosylation motifs are not conserved among decapod IAGs (Fig. 4). For instance, Malar-IAG has one *N*-glycosylation motif in the B chain, two in the C peptide, and one in the A chain. Papac-IAG has only one motif in the C peptide, which indicates a mature molecule comprising the B chain and the A chain, does not possess an *N*-linked glycan moiety. Although the *N*-linked glycan in the A chain of Arvul-AGH is indispensable for biological activity[19], *N*-glycosylation of the A chain of decapod IAG is not likely to be essential for the conferring function. Further experiments on the purification and structural determination of native IAG are needed to clarify this issue.

IAG has been thought of as a putative AGH in decapod species. In *M. rosenbergii*, silencing of the IAG gene using RNA interference methods temporally prevented the regeneration of male-specific sexual characteristics, and led to the arrest of spermatogenesis[26]. This correlates well with the above hypothesis; however, there has been no direct evidence for IAG having AGH activity until now. To overcome this obstacle, we are now able to easily produce a heterodimeric mature Papac-IAG with selective disulfide bond formation in its synthetic A and B chains. Because Papac-IAG is the first example of a non-glycosylated IAG in its mature form, the synthetic Papac-IAG will be useful for *in vivo* and/or *in vitro* biological studies, which provide important information for understanding the IAG function in decapod crustaceans.

Acknowledgments

References

1. Adham, I. M. et al. (1993) Cloning of a cDNA for a novel insulin-like peptide of the testicular Leydig cells. *J. Biol. Chem.*, **268**, 26668-26672.
2. Bell, G. I. et al. (1979) Nucleotide sequence of a cDNA clone encoding human preproinsulin. *Nature*, **282**, 525-527.
3. Charniaux-Cotton, H. (1954) Découverte chez un Crustacé Amphipode (*Orchestia gammarella*) d′une glande endocrine responsable de la différenciation des caractères sexuels primaires et secondaires mâles. *C. R. Acad. Sci. Paris*, **239**, 780-782 [In French].
4. Chomczynski, P. & Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156-159.
5. Fotsis, T., Murphy, C. & Gannon, F. (1990) Nucleotide sequence of the bovine insulin-like growth factor 1 (IGF-1) and its IGF-1A precursor. *Nucleic Acids Res.*, **18**, 676.
6. Gregoire, F. M. et al. (1998) Cloning and developmental regulation of a novel member of the insulin-like gene family in *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.*, **249**, 385-390.
7. Hasegawa, Y., Haino-Fukushima, K. & Katakura, Y. (1987) Isolation and properties of androgenic gland hormone from the terrestrial isopod, *Armadillidium vulgare*. *Gen. Comp. Endocrinol.*, **67**, 101-110.
8. Hudson, P. et al. (1984) Relaxin gene expression in human ovaries and the predicted structure of a human preprorelaxin by analysis of cDNA clones. *EMBO J.*, **3**, 2333-2339.
9. Katakura, Y. (1984) Sex differentiation and androgenic gland hormone in the terrestrial isopod *Armadillidium vulgare*. *In* The Biology of Terrestrial Isopods, eds. Sutton, S. L. & Holdich D. M., Oxford Univ. Press, Oxford, UK, 127-142.
10. Khalaila, I. et al. (2001) Effects of implantation of hypertrophied androgenic glands on sexual characters and physiology of the reproductive system in the female red claw crayfish, *Cherax quadricarinatus*. *Gen. Comp. Endocrinol.*, **121**, 242-249.
11. Manor, R. et al. (2007) Insulin and gender: an insulin-like gene expressed exclusively in the androgenic gland of the male crayfish. *Gen. Comp. Endocrinol.*, **150**, 326-336.
12. Martin, G. et al. (1990) Purification and characterization of androgenic hormone from the terrestrial isopod *Armadillidium vulgare* Latr. (Crustacea, Oniscidea). *Gen. Comp. Endocrinol.*, **80**, 349-354.
13. Martin, G. et al. (1999) The structure of a glycosylated protein hormone responsible for sex determination in the isopod, *Armadillidium vulgare*. *Eur. J. Biochem.*, **262**, 727-736.
14. Miyake, S. (1982) *Japanese crustacean decapods and stomatopods in color Vol. I.* Hoikusha, Osaka, Japan, 26-29 [In Japanese].
15. Nagamine, C. et al. (1980) Masculinization of female *Macrobrachium rosenbergii* (de Man) (Decapoda, Palaemonidae) by androgenic gland implantation. *Gen. Comp. Endocrinol.*, **41**, 442-457.
16. Nagasawa, H. et al. (1984) Amino-terminal amino acid sequence of the silkworm prothoracicotropic hormone: homology with insulin. *Science*, **226**, 1344-1345.

17. Okuno, A., Hasegawa, Y. & Nagasawa, H. (1997) Purification and properties of androgenic gland hormone from the terrestrial isopod *Armadillidium vulgare. Zool. Sci.*, **14**, 837-842.

18. Okuno, A. et al. (1999) Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod *Armadillidium vulgare. Biochem. Biophys. Res. Commun.*, **264**, 419-423.

19. Okuno, A. et al. (2002) Preparation of an active recombinant peptide of crustacean androgenic gland hormone. *Peptides*, **23**, 567-572.

20. Ohira, T. et al. (2003) Molecular cloning and expression analysis of cDNAs encoding androgenic gland hormone precursors from two Porcellionidae species, *Porcellio scaber* and *P. dilatatus. Zool. Sci.*, **20**, 75-81.

21. Short, D. K. et al. (1998) Adenovirus-mediated transfer of a modified human proinsulin gene reverses hyperglycemia in diabetic mice. *Am. J. Physiol. Endocrinol. Metab.*, **275**, 748-756.

22. Smit, A. B. et al. (1991) Characterization of a cDNA clone encoding molluscan insulin-related peptide II of *Lymnaea stagnalis. Eur. J. Biochem.*, **199**, 699-703.

23. Smit, A. B. et al. (1996) Expression and characterization of molluscan insulin-related peptide VII from the mollusc *Lymnaea stagnalis. Neuroscience*, **70**, 589-596.

24. Suzuki, S., Yamasaki, K. & Katakura, Y. (1990) Vitellogenin synthesis in andrectomized males of the terrestrial isopod, *Armadillidium vulgare* (malacostracan Crustacea). *Gen. Comp. Endocrinol.*, **77**, 283-291.

25. Thompson, J., D., Higgins, D. G. & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673-4680.

26. Ventura, T. et al. (2009) Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. *Endocrinology*, **150**, 1278-1286.

27. Wentworth, B. M. et al. (1986) Characterization of the two nonallelic genes encoding mouse preproinsulin. *J. Mol. Evol.*, **23**, 305-312.