Molecular Structure and Organization of Crustacean Hyperglycemic Hormone Genes of *Penaeus monodon*

Amporn Wiwegweaw, Apinunt Udomkit* and Sakol Panyim

Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Nakhon Pathom 73170, Thailand

The Crustacean hyperglycemic hormone (CHH) has been shown to exist as multiple molecular forms in several crustacean species. In *Penaeus monodon*, a gene encoding CHH (so-called *Pem-CHH1*) was recently described. In this study, the molecular structures of two other *CHH* genes (*Pem-CHH2* and *Pem-CHH3*) are reported. Both the *Pem-CHH2* and *Pem-CHH3* genes contain three exons that are separated by two introns that are similar to the structure of other genes in the same family. An analysis of the upstream nucleotide sequences of each *Pem-CHH* gene has identified the putative promoter element (TATA box) and putative binding sites for several transcription factors. The binding sites for CREB, Pit-1, and AP-1 were found upstream of all three *Pem-CHH* genes. A Southern blot analysis showed that at least one copy of each *Pem-CHH* gene was located within the same 10 kb genomic DNA fragment. These results suggest that the *CHH* genes are arranged in a cluster in the genome of *P. monodon*, and that their expression may be modulated by similar mechanisms.

Keywords: CHH, Gene organization, Gene structure, *Penaeus monodon*, Regulatory elements

Introduction

The crustacean hyperglycemic hormone (CHH) is the most abundant neuropeptide that is produced by a group of peptidergic neurons in the medulla terminalis X-organ of decapod crustacean (Charmantier *et al.*, 1997). The primary function of CHH is to regulate the glucose level in the hemolymph. CHH belongs to a peptide hormone family that is mainly comprised of the CHH, molt-inhibiting (MIH), gonad-inhibiting, (GIH) and mandibular organ-inhibiting hormones (MOIH) (Keller, 1992). Peptides in this family share unique characteristics: all are composed of 72-78 amino acids and possess six cysteine residues in identical positions (Chang, 1997).

The amino acid sequences of CHH from several species have been determined (Kegel *et al.*, 1989; Kegel *et al.*, 1991; Tensen *et al.*, 1991b; Huberman *et al.*, 1993; Martin *et al.*, 1993; Yang *et al.*, 1995; Gu and Chang, 1998). The presence of multiple molecular forms of CHH in a single species was observed. For example, two structurally different CHHs that were 92% identical in their amino acid sequences (so-called CHH-A and CHH-B) were reported in the lobster *Homarus americanus* (Tensen *et al.*, 1989). The cDNAs encoding these two forms of CHH in *H. americanus* were subsequently obtained (Tensen *et al.*, 1991a). Two distinctive forms of CHH were also found in the crayfish *Procambarus bouvieri* (Huberman and Aguilar, 1988) and shrimp *Metapenaeus ensis* (Gu *et al.*, 2000). Moreover, in the shrimp *Penaeus japonicus*, at least five neuropeptides possessed hyperglycemic activity (Yang *et al.*, 1997). In addition, two isoforms, resulting from the configurational difference of a single amino acid residue (L- or D-Phe at position 3), were observed for each CHH-A and CHH-B in *H. americanus*. These isoforms had the same molecular mass and showed similar hyperglycemic activity (Soyez *et al.*, 1994). Similar sterioinversion was observed in CHH from crayfish *P. clarkii* (Yasuda *et al.*, 1994).

Recently, a gene encoding CHH from the black tiger shrimp *P. monodon* was characterized (Udomkit *et al.*, 2000). We report here the identification and characterization of the genes encoding another two CHH peptides from *P. monodon*. The structure and putative promoter region of these three *Pem-CHH* genes were analyzed and compared. Preliminary data on the organization of these *CHH* genes was also described.

Materials and Methods

Amplification of *Pem-CHH2* and *Pem-CHH3* genes Genomic

*To whom correspondence should be addressed.
Tel: 66-02-441-9003-7-1240; Fax: 66-02-441-9906
E-mail: staud@mahidol.ac.th

DNA was prepared from abdominal muscle tissue of *P. monodon* by using the Qiagen Genomic-Tip and Genomic DNA buffer set (Qiagen, Chatsworth, USA). The PCR reaction contained 150 ng of the genomic DNA template, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM MgCl$_2$, 200 nM each primer, and 200 μM each of dATP, dCTP, dGTP, and dTTP in a total volume of 50 μl. The reaction was heated to 95ºC for 5 min then 2.5 units of *rTth* DNA polymerase (Biotools, Edmonton, Canada) were added. Amplification was achieved by 35 successive cycles of denaturation at 94ºC for 1 min, annealing at 50ºC for 1 min and extension at 72ºC for 2 min, followed by a 10 min-final extension at 72ºC. The primers that were used to amplify the genes were designed from the 5' and 3' ends of the corresponding cDNA. The nucleotide sequence of these primers were CHH-F: 5' CGGAATTCTCAGTGCAGAGGGAGAGCC 3' and CHH2-R: 5' GCGGATCCCTGCTTTATGAAGACACTG 3' for the *Pem-CHH2* gene and CHH-F and CHH3-R: 5' ATGCTTTATGAAG ACATTAC 3' for the *Pem-CHH3* gene.

Genomic walking into the 5' upstream sequences of *Pem-CHH* genes   The unknown 5' DNA sequences (upstream of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes) were accessed by the strategy of the Universal Genome Walker™ Kit (CLONTECH, Palo Alto, USA). Genomic DNA of *P. monodon* was digested with either *Dra* I or *Sna*B I to generate blunt-ended genomic fragments. The *Dra* I and *Sna*B I digested genomic DNA was then ligated to the Genome Walker Adapter in separate reactions, which resulted in pools of the adapter-ligated genomic fragments, which were referred to as *Dra* I and *Sna*B I Genome Walker libraries, respectively. The unknown flanking sequences of the individual *Pem-CHH* genes were amplified from the genome walker libraries with outer adapter (AP1) and outer gene specific primers (GSP1) that were designed in order to extend the amplification from the 5' terminus of each Pem-CHH gene into their upstream sequences. Secondary PCR was performed with nested adapter (AP2) and nested gene specific primers (GSP2) in order to obtain specifically amplified products. The PCR reactions were carried out according to the manufacturer's instruction. The gene specific primers for the *Pem-CHH2* gene were as follows: CHH2-GSP1: 5 TAAGGAAAA AATCGGACCTTGCTGAAA 3' and CHH2-GSP2: 5' CAAATGA ATAATGCATACAGCGTCAGA 3' and for *Pem-CHH3* gene were CHH3-GSP1: 5' AGGGGATATACAGGGCTTTACATTTCT 3' and CHH3-GSP2: 5' CAGGTTAAGCTCTGCAATTTACAGAGC 3'.

DNA sequence analysis   The PCR products were cloned into the pGEM®-T Easy vector (Promega, Madison, USA). Double-stranded DNA sequencing was performed by the ABI PRISM™ BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, USA). Searches for the transcription factor-binding site were performed using Tfsitescan software on the MIRAGE WWW server (http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl) and Signal Scan software on the BIMAS WWW server (http://bimas.dcrt.nih.gov/molbio/signal).

Southern blot analysis   About 10 μg of *P. monodon*'s genomic DNA was digested with selected restriction enzymes. The genomic fragments were separated on 0.7% agarose gel at 30 V for 21 h. They were subsequently transferred to the GeneScreen Plus® hybridization transfer membrane (DuPont, Boston, USA) with 0.4

N NaOH using a vacuum blotter (Model 785, Bio-RAD, Heracules, USA).

Three DNA probes that are specific to the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes were synthesized from the intronic region of each gene. The fourth probe was synthesized from the coding region of the mature Pem-CHH1, which is almost identical to the same region of the other two *Pem-CHH* genes. The primers that were used for amplification of each probe were as follows: forward CHH1-IF (5' AACCCAACCCAGTTGTGTTGT 3') and reverse CHH1-IR (5' CCACTGGTAGGTCGTACAGGA 3') for the CHH1-specific probe, forward CHH2-IF (5' TTCGTATAGCG GATCCGA 3') and reverse CHH2-IR (5' TCAGTTTTGTTCGGA GGG 3') for CHH2-specific probe, forward CHH-F (5' TCAGTGCAGAGGGAGAGCC 3') and reverse CHH2-IR followed by *Acc* I digestion for the CHH3-specific probe and forward CHH-mat (5' GGAATTCCATATGAGCCTATCCTTCAG 3') and reverse CD-R (5' CGGGATCCCTACTTGCCGAGCCTC TG 3') for the mature CHH1 probe.

The DNA probes were synthesized by the incorporation of Fluorescein-11-dUTP with the Gene Image random prime-labeling module (Amersham Biosciences, Uppsala, Sweden) following the manufacturer's protocol. Hybridization was performed in a hybridization solution (5 × SSC, 0.1% (w/v) SDS, 5% (w/v), dextran sulphate, 20-fold dilution of liquid block, and 100 μg/ml denatured salmon sperm DNA) at 65°C overnight. After hybridization, the first wash was performed in 1 × SSC, 0.1%(w/v) SDS at 65°C for 15 min, followed by the second wash in 0.1 × SSC, 0.1% (w/v) SDS at the same temperature for 15 min. The detection step was conducted according to the protocol of the Gene Images CDP-Star detection module (Amersham Pharmacia Biotech).

Results

Nucleotide sequence analysis of *Pem-CHH2* and *Pem-CHH3* genes   Specific DNA fragments of about 1.2 kb were amplified from *P. monodon*'s genomic DNA using primers that were designed from both ends of the 5' and 3' cDNA encoding Pem-CHH2 and Pem-CHH3 (unpublished data). Figures 1A and 1B show the nucleotide sequences of these *Pem-CHH2* and *Pem-CHH3* genes, as well as the primers that were used for amplification. A sequence analysis using the Gene Finder Program of Baylor College of Medicine (BCM) via the worldwide web revealed that both genes contained three exons and two introns. The donor and acceptor consensus sequences of the splice site (GT-AG) were detected at all exon-intron boundaries. The first and second introns of *Pem-CHH2* were 406 and 268 bp long, whereas those of *Pem-CHH3* comprised 370 and 282 bp, respectively. The first intron interrupted the sequence encoding the signal peptide, while the second intron separated the mature peptide coding sequence in both the *Pem-CHH2* and *Pem-CHH3* genes (Figs. 1A and B). Figure 2 shows the schematic diagram comparing the structure of *Pem-CHH2* and *Pem-CHH3* genes to that of the previously characterized *Pem-CHH1* gene.

(A) Pem-CHH2

CHH-F

```
TCAGTGCAGAGGGAGAGCCTCGAAGTCGGTGGTCCTTGTTCCTTCTGCTCTCTGTCGAGT
CCAGCGAAATGGTTGCGGTCCGATTGGTAAAGGAGGTTTAAGTTCAGGCGGTTTTAACTC
       M  V  A  V  R  L -----------------------------------
TTGTGTAAAGTGCAGAGCTTGTCCTGTGCAAATATGGTTTTTTTTTCGTATAGCGGATCC
------------------------------------------------------------
GATTTTTTTTTATTTTCTTTATTAATTTTCCAATGCTTCTGACGCTGTATGCATTATTCA
------------------------------------------------------------
TTTGTCTTGTGTGTCCCCTCCGAACAAAACTGATTGATTCGAGAGTAGTAGTAAAGCGAA
---------------------- Intron 1 ----------------------
CATAGGAAGTTTTCTAAACATCGAAGTCATAATCATTACATCAATACGGCCACAAAATCC
------------------------------------------------------------
TGTAAGGGTCTCAGGACTGACTTCGGACATGAATAATTTCAGCAAGGTCCGATTTTTTCC
------------------------------------------------------------
TTAGTTACAGATGTTGATACTTGTTGGAAACAAACCGTCCTCTGAAGTTAACTCCTCTCC
------------------------------------------------------------
CTCTGCCAACAGGTGCAGTCAGCTGTTCTTGTGTCCTTGCTGGTGGCACTTCCGGCCTGT
------------ V  Q  S  A  V  L  V  S  L  L  V  A  L  P  A  C
GTCACAACTTCTGAAAACACGAATGAAATACCGGCGTCCATTCTTTCCTCCCCTGGGGAT
 V  T  T  S  E  N  T  N  E  I  P  A  S  I  L  S  S  P  G  D
TCCCTTTCAGAAGACCAAAGCATAAGCAAACGCTCTATATGGTTCAATTCTTGCACGGGC
 S  L  S  E  D  Q  S  I  S  K  R  S  I  W  F  N  S  C  T  G
GTCTACGACCGCGAACTTCTTGTAAGGCTCGACCGCGTGTGCGAAGACTGCTACAACCTC
 V  Y  D  R  E  L  L  V  R  L  D  R  V  C  E  D  C  Y  N  L
TACCGCGACGTCGGAGTGGCGGCCGAATGCAGGTATGTTATTCTAATTCTTACATAAAAA
 Y  R  D  V  G  V  A  A  E  C  R-----------------------------
TGTCAACAACCACTTTATACAGCAGCGAGTATCTGTAATCGCCTGTCAAGATCAGATGTG
------------------------------------------------------------
ATAGTATGCTGGGTAAATACTGTTATCATTTAATAGGTCGATTTTTTAATCACTATAAAA
---------------------- Intron 2 ----------------------
CCTTTATTTAGATTCCAAGGATCTGAGTAAAATCAACTTTACTTCCCTTTTCTCTTGCAT
------------------------------------------------------------
TCTATTTTCATTCCTTCTTTACCTACTTACATGGTTACATAGAACCTTTACATATTGCAG
------------------------------------------------------------
GAGTAACTGTTTCCACAACGAGGTGTTCCTGTACTGCGTGGACTACATGTTCCGGCCTCG
 S  N  C  F  H  N  E  V  F  L  Y  C  V  D  Y  M  F  R  P  R
CCAGAGGAACCAGTACCGGGCCGCCCTCCAGAGGCTCGGCAAGTAGGCGGCTCGTCTTCG
 Q  R  N  Q  Y  R  A  A  L  Q  R  L  G  K  *
GTCAACCCTTCCTGCGGGGACGCTCGCCGTGAGATTTTTTCTAGGCTCTCCTTCGGTTGC
TTCTGTATCCTAAGTTAGAAACGGATAACATATTGACGTCGAGGTGTTTCATGTACAGTG
```

TCTTCATAAAGCAG

CHH2-R

(B) Pem-CHH3

CHH-F

```
TCAGTGCAGGGGGAGAGCCTCGAAGTCGGTGGTCCTTGTTCCCGCTGATCTGTTGCGAGT
CCAGCGAAATGATTGCGGTCCGACTGGTGAGGGAGGTTGGAGAACAGGCGGTTGTAACGG
        M  I  A  V  R  L ----------------------------------
CTCTGTAAATTGCAGAGCTTAACCTGTGCAAATATTCTTTTCGTACAGCGGTTCCGAATG
------------------------------------------------------------
TTTTTATTTGTGTTTATTGATTTTCTGCTGCTGTATACATTATTTATAGAAATGTAAAGC
------------------------------------------------------------
CCTGTATATCCCCTGCGAACAAAAGTGACCGATTCGATAATAGTAGTAAAGCGAATATAG
---------------------- Intron 1 ----------------------
GGAAGAAATCAGTTTCTCTCTCTAAACATTGAAATCACAATCATTATATCAATACGGCCA
------------------------------------------------------------
CAAAAATCCTCAGGACTGAGGGCGAAAAATTTTCTTAGTTACAGATGTTGATACTCAAAC
------------------------------------------------------------
ACTTCCTAAAGTTAACTACTCTCCCTCTGCCGACAGGTGCAGTCAGCTGTTCTGGTGTCC
------------------------------------- V  Q  S  A  V  L  V  S
CTGCTGGTGGCATTTCCGGCCTGTGTCACGACTTCTGAAAACACGAATGAAATACTGGCG
 L  L  V  A  F  P  A  C  V  T  T  S  E  N  T  N  E  I  L  A
TCCATTCTTTCCTCCCCTGGGGATTCCCTTACACGAGAACAAAGCATAAGCAAACGCACC
 S  I  L  S  S  P  G  D  S  L  T  R  E  Q  S  I  S  K  R  T
ATGTCGTTCAGTTCTTGCACGGGTGTCTACGACCGCGAACTCCTTGTAAGGCTCGACCGC
 M  S  F  S  S  C  T  G  V  Y  D  R  E  L  L  V  R  L  D  R
GTGTGCGAAGACTGCTACAACCTCTACCGCGACGTTGGAGTGGCGGCCGAATGTAGGTAT
 V  C  E  D  C  Y  N  L  Y  R  D  V  G  V  A  A  E  C  R----
GTTCTTCCTATTCTTACATAAAATTTGTCAACAACCAATACAAACAACCAGTTTATTATA
------------------------------------------------------------
CAGCAGCCAGTATCTGTGCCAGATGGTCAGTCAAGGGCAGATGTAATGGTATACTTTCAG
------------------------------------------------------------
AATGTCGATTTTTTTTCTTTTTATAATTACTAGAAAACCTTACATTTAGATTTCAAGGAT
---------------------- Intron 2 ----------------------
TTGAGTAAAATCAAATTTGCTACCCTTTACTCTTGATTTCTATTTTCATTCCTTCTTTAC
------------------------------------------------------------
TTATCGGCATGGGTACATAAAAGTTTTGCATTTTACAGGAGTAACTGTTTCCACAACGAG
------------------------------------- S  N  C  F  H  N  E
GTGTTCCTGTACTGCGTGGACTACATGTTTCGGCCTCGCCAAAGGAATCAGTACCGGGCC
 V  F  L  Y  C  V  D  Y  M  F  R  P  R  Q  R  N  Q  Y  R  A
GCCCTGCAGAGGCTCGGCAAGTAGGCGGCTCCTATCCGGCCAGCTCTTCCGCACGGGCGC
 A  L  Q  R  L  G  K  *
TCGCTGTGAGACTTTCTGAATGACCAGACTCTAGGCTCTCTTCGGTTGCCTCTGTATCCT
AAGTTAGAAGCTGACACCATATTGACGCCGAGGTGTTTCATGTGTAATGTCTTCATAAAG
```
CHH3-R

CAT

Fig. 1. Nucleotide and deduced amino acid sequences of the *Pem-CHH2* (A) and *Pem-CHH3* (B) genes of *P. monodon*. Introns are indicated by dashed lines. Amino acids are shown in a one-letter symbol and the asterisk marks the stop codon. Arrows indicate the nucleotide sequences corresponding to the primers used for amplification of each gene.
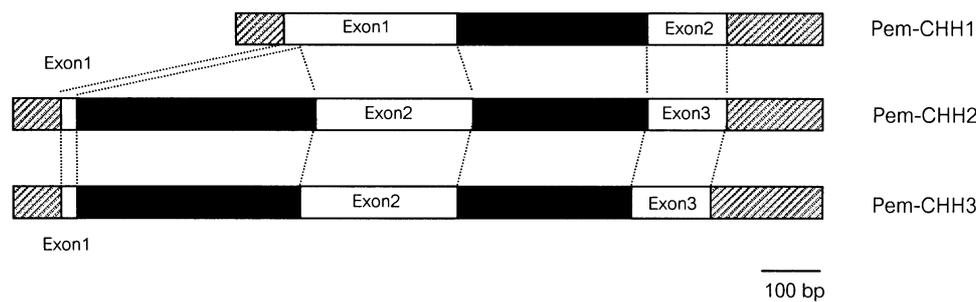

Fig. 2. Schematic diagram comparing the structure of the *Pem-CHH1* (Udomkit *et al.*, 2000), *Pem-CHH2*, and *Pem-CHH3* genes. The hatched boxes represent the 5' and 3' untranslated regions. Exons and introns are shown in white and black boxes, respectively.

Analysis of the putative promoter and probable regulatory elements The 5' upstream sequences (about 500 nucleotides from the transcription start site (+1) of each *Pem-CHH* genes) were analyzed. Sequence similarities were observed between the 200 nucleotides upstream from the +1 position of the *Pem-CHH2* and *Pem-CHH3* genes, where the same region of *Pem-CHH1* showed a higher degree of variation (Fig. 3). In addition, three types of dinucleotide repeats (i.e. AC, AT, and

GT) were found in the further upstream nucleotide sequence of the *Pem-CHH1* gene. No similar repeats were detected in the 5' upstream sequence of the *Pem-CHH2* and *Pem-CHH3* genes that were analyzed in this study.

A prediction for the potential upstream regulatory elements within 500 nucleotides of the upstream sequences, which was carried out using MIRAGE Tfsitescan software and the BIMAS Signal Scan software, revealed a TATA box-like
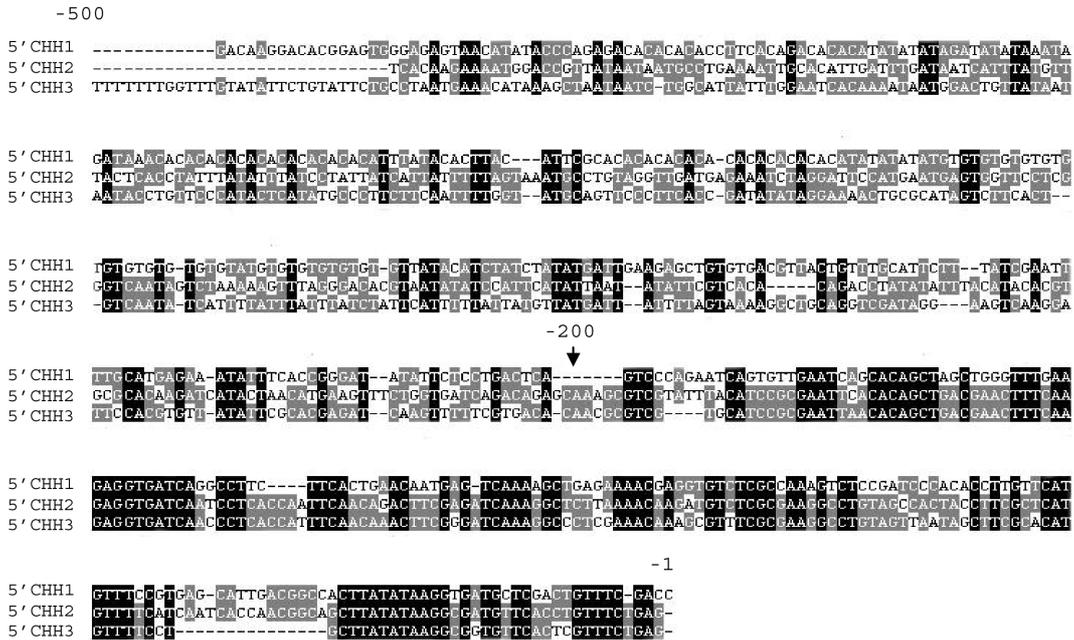
-500

```
5'CHH1   ------------GCACAAGGACACGGAGTCGGAGAGTAAACATATACCCAGAGACACACACACCTTCACAGACACACATATATATAGATATATAAATA
5'CHH2   -----------------------TCACAAGAAAATGGACCGTTATAATAAATGCCTGAAAATTGCACATTGCATTTGATAATCATTTATGTT
5'CHH3   TTTTTTTGGTTTGTATGATTCTGTATTCTGCCTAATGAAACATAAAAGCTAATAATCGTCGGCATTTATTTGGAATCACAAAATAATGCACTGTTATAAT
```

```
5'CHH1   GATAAACACACACACACACACACACACATTTATACACTTAC---ATTCGGCACACACACACA-CACACACACACTATATATATCTGTGTGTGTGTC
5'CHH2   TACTCACCTATTTATATTTATCCTATTATTATTATTTTTAGTAAATCCCTGTAGGTTGATGACAAATCTACCGATTCCATGAATGAGTCGTTCCTCG
5'CHH3   AATACCTGTTCCCATACTCATATGCCCTTCTTCAATTTTGGTT-ATGCAGTTCCCTTCACCCGATAATATAGGAAAACTGCGCATAGTCTTCACTT-
```

```
5'CHH1   TGTGTGTG-TGTGTATGTCTGGTGTGTGT-CTTATACATCTATCTATATGATTGAAGAGCTGTGTGACGTTACTGTTTGCATTTCTT--TATCGAATT
5'CHH2   GGTCAATAGTCTAAAAAGTTTACGGCACACCTAATATATCCCATTCCATATTAATT--ATATTCGTCACA------CAGACCTATATATTTACATACACGT
5'CHH3   -GTTCAATATTCATTTTTATTTATTTATCCTATTCATTTTTATTATTGTTTATGATTT--ATTTTAGTAAAAGCCTGCCACGGTCGATAGG---AAGTTCAAGGAA
```

-200

```
5'CHH1   TTGCATGAGAA-ATATTTCACCGGGAT--ATATTCTCCTGACTCA-------GTCCCAGAATCAGTGTTGAATCAGCACAGCTAGCTGGGTTTGAA
5'CHH2   GCGCACAGATCTAATACTAACATGAAGTTTCTGGTGATCAGACGAGCAAAGCGTCCTATTTACATCCGCGAATTCACACAGCTGACGAACTTTCAA
5'CHH3   TTCCACGTGTT-ATATTCGCACGAGAT--CAAGTTTTTCGTGACA-CAACGCGTCG----TGCATCCGCGAATTAACACAGCTGACGAACTTTCAA
```

```
5'CHH1   GAGGTGATCAGGCCTTC----TTCACTGAACAATGAG-TCAAAAGCTGAGAAAACGAGGTGTCTCGGCAAAGTCTCCGATCCGACACCTTGTTCAT
5'CHH2   GAGGTGATCAATCCTCACCAATTCAACAGACTTCGAGATCAAAGGCTCTTAAAACAAGATGTGTCGCGAAGGCCTGTAGCCACTACCTTCGCTCAT
5'CHH3   GAGGTGATCAACCCTCACCATTTCAACAAACTTCGGGATCAAAGGCCCTCGAAACAAACGGTTTCGCGAAGGCCTGTAGTTAATAGCTTCGCACAT
```

-1

```
5'CHH1   GTTTTCCGGTGAG-CATTGACGGCCACTTATATAAGGTGATGCTCGACTGTTTC-GACC
5'CHH2   GTTTTCATCAATCACCAACGGCAGCTTATATAAGGCGATGTTCACCTGTTTCTGAG-
5'CHH3   GTTTTTGCCT-------------GCTTATATAAGGCGGTGTTCACTCGTTTCTGAC-
```

Fig. 3. Comparison of the 500 nucleotides upstream from the transcription start site of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes. The nucleotides that are identical in all three sequences are highlighted in black; whereas, those that are identical in two of the three sequences are gray. The numbers are positions relative to the transcription start site. The arrow indicates the position 200, downstream of which the sequences are highly conserved among all of the *Pem-CHH* genes.

element (TATAA) that was located at 23 nucleotides upstream of the transcription start site of each gene. In addition, several potential recognition sites for transcription factors were identified (Fig. 4). The putative binding sites for the cAMP responsive element binding protein (CREB), pituitary-specific transcription factor-1 (Pit-1), and activator protein-1 (AP-1) were found in the upstream region of all three CHH-like genes. Moreover, recognition sequences for other transcription factors (i.e. NF-1, CP1, and CP2) have also been identified.

Analysis of the copy numbers and organization of *Pem-CHH* genes in the genome of *P. monodon* A Southern blot analysis was performed in order to gain preliminary data on the copy numbers and organization of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes in *P. monodon*'s genome. The nucleotide sequences that are specific to each *Pem-CHH* gene were used as a probe to detect individual *Pem-CHH* genes in the genome. These probes were unable to hybridize to one another (Fig. 5A). *P. monodon*'s genomic DNA was separately digested with three restriction enzymes that do not cut within any of the three *Pem-CHH* genes (i.e. *Bgl* I, *Eco*R I, and *Hind* III). The result in Fig. 5B shows that there were at least 3 copies of the Pem-CHH1 gene, one copy of the Pem-CHH2 genes, and 2 copies of the Pem-CHH3 gene in the genome. Some common DNA fragments were detected by more than one probe. A fragment of about 7 kb in the *Eco*R I-digested DNA was detected by both of the CHH2 and CHH3 probes, and the 9 kb band in *Eco*R I-digested DNA was detected by both of the CHH1 and CHH3 probes. In addition,

the band above 10 kb in the *Hind* III-digested DNA was detected by all of the probes. All of the DNA fragments that were detected by each gene-specific probe, together with several additional bands, were hybridized to the mature CHH 1 probe (Fig. 5C).

Discussion

Despite the similarity in their amino acid sequences, Pem-CHH2 and Pem-CHH3 were shown to be encoded from genes that are structurally different from the one encoding Pem-CHH1. While Pem-CHH1 contains information that is encoded from two exons (Udomkit *et al.*, 2000), Pem-CHH2 and Pem-CHH3 are both the translated products from three exons. The gene structure of CHH, MIH, and MOIH in other species is similar to that of Pem-CHH2 and Pem-CHH3 (i.e. containing three exons and two introns) (Gu and Chang 1998; Gu *et al.*, 2000; Lu *et al.*, 2000). Compared to the genes for other CHHs, the Pem-CHH1 gene lacks the first intron that separates the coding region for the signal peptide in the others. The only intron of the *Pem-CHH1* gene is located at the same position as intron 2 of *Pem-CHH2* and *Pem-CHH3* genes.

In this study, about 500 nucleotides of the upstream regions from the transcription start site (+1) of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes were characterized. A similarity in the nucleotide sequence within about 200 nucleotides upstream from the +1 position among the three *Pem-CHH* genes suggests that they occurred through gene
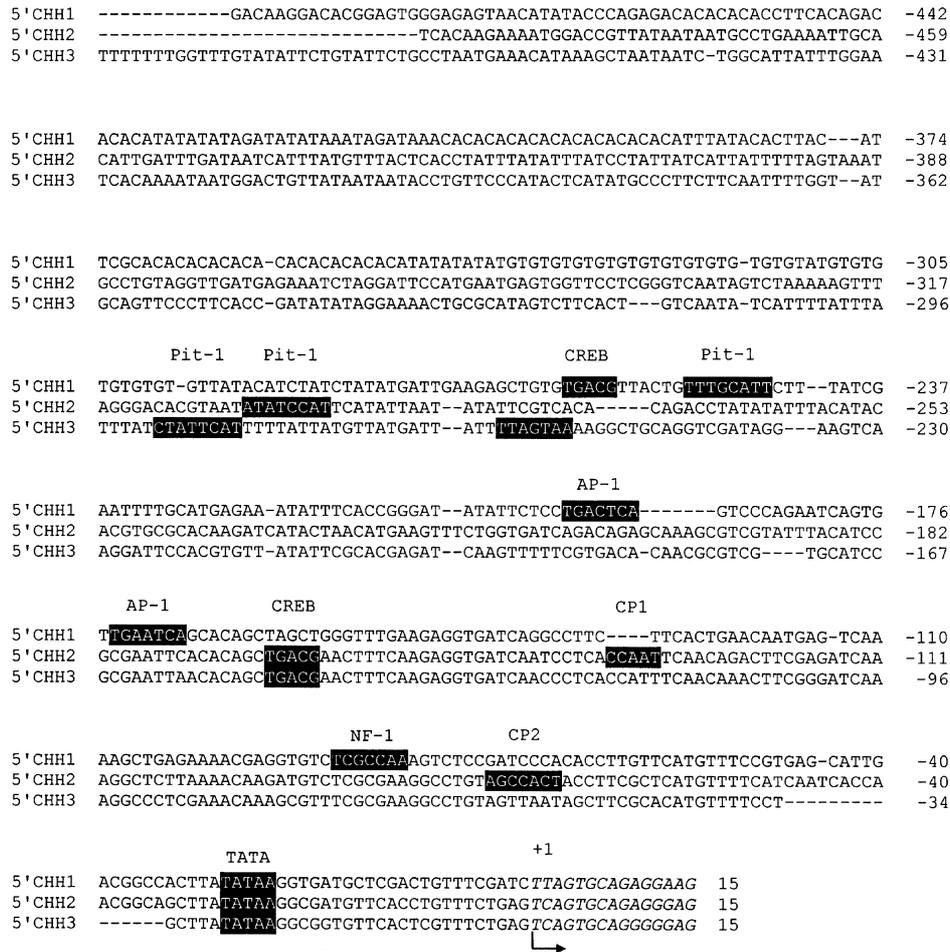
```
5'CHH1  ------------GACAAGGACACGGAGTGGGAGAGTAACATATACCCAGAGACACACACACCTTCACAGAC   -442
5'CHH2  --------------------------TCACAAGAAAATGGACCGTTATAATAATGCCTGAAAATTGCA   -459
5'CHH3  TTTTTTTGGTTTGTATATTCTGTATTCTGCCTAATGAAACATAAAGCTAATAATC-TGGCATTATTTGGAA   -431


5'CHH1  ACACATATATATAGATATATAAATAGATAAACACACACACACACACACACACATTTATACACTTAC---AT   -374
5'CHH2  CATTGATTTGATAATCATTTATGTTTACTCACCTATTTATATTTATCCTATTATCATTATTTTTAGTAAAT   -388
5'CHH3  TCACAAAATAATGGACTGTTATAATAATACCTGTTCCCATACTCATATGCCCTTCTTCAATTTTGGT--AT   -362


5'CHH1  TCGCACACACACACA-CACACACACACATATATATATGTGTGTGTGTGTGTGTGTGTG-TGTGTATGTGTG   -305
5'CHH2  GCCTGTAGGTTGATGAGAAATCTAGGATTCCATGAATGAGTGGTTCCTCGGGTCAATAGTCTAAAAAGTTT   -317
5'CHH3  GCAGTTCCCTTCACC-GATATATAGGAAAACTGCGCATAGTCTTCACT---GTCAATA-TCATTTTATTTA   -296

                 Pit-1    Pit-1                       CREB        Pit-1
5'CHH1  TGTGTGT-GTTATACATCTATCTATATGATTGAAGAGCTGTGTGACGTTACTGTTTGCATTCTT--TATCG   -237
5'CHH2  AGGGACACGTAATATATCCATTCATATTAAT--ATATTCGTCACA-----CAGACCTATATATTTACATAC   -253
5'CHH3  TTTATCTATTCATTTTTATTATGTTATGATT--ATTTTAGTAAAAGGCTGCAGGTCGATAGG---AAGTCA   -230

                                                AP-1
5'CHH1  AATTTTGCATGAGAA-ATATTTCACCGGGAT--ATATTCTCCTGACTCA-------GTCCCAGAATCAGTG   -176
5'CHH2  ACGTGCGCACAAGATCATACTAACATGAAGTTTCTGGTGATCAGACAGAGCAAAGCGTCGTATTTACATCC   -182
5'CHH3  AGGATTCCACGTGTT-ATATTCGCACGAGAT--CAAGTTTTTCGTGACA-CAACGCGTCG----TGCATCC   -167

           AP-1            CREB                         CP1
5'CHH1  TTGAATCAGCACAGCTAGCTGGGTTTGAAGAGGTGATCAGGCCTTC----TTCACTGAACAATGAG-TCAA   -110
5'CHH2  GCGAATTCACACAGCTGACGAACTTTCAAGAGGTGATCAATCCTCACCAATTCAACAGACTTCGAGATCAA   -111
5'CHH3  GCGAATTAACACAGCTGACGAACTTTCAAGAGGTGATCAACCCTCACCATTTCAACAAACTTCGGGATCAA   -96

                     NF-1          CP2
5'CHH1  AAGCTGAGAAAACGAGGTGTCTCGCCAAAGTCTCCGATCCCACACCTTGTTCATGTTTCCGTGAG-CATTG   -40
5'CHH2  AGGCTCTTAAAACAAGATGTCTCGCGAAGGCCTGTAGCCACTACCTTCGCTCATGTTTTCATCAATCACCA   -40
5'CHH3  AGGCCCTCGAAACAAAGCGTTTCGCGAAGGCCTGTAGTTAATAGCTTCGCACATGTTTTCCT---------   -34

                 TATA                      +1
5'CHH1  ACGGCCACTTATATAAGGTGATGCTCGACTGTTTCGATCTTAGTGCAGAGGAAG   15
5'CHH2  ACGGCAGCTTATATAAGGCGATGTTCACCTGTTTCTGAGTCAGTGCAGAGGGAG   15
5'CHH3  ------GCTTATATAAGGCGGTGTTCACTCGTTTCTGAGTCAGTGCAGGGGGAG   15
```

Fig. 4. The putative promoter sequence and regulatory elements identified in the upstream region of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes. Putative TATA box and binding sites for the transcription factors are highlighted. Arrows indicate the transcription start site (+1) of each gene (Udomkit *et al.*, 2000; unpublished data). Numbers on the right show the positions relative to the transcription start site.

duplication events. Likewise, a high degree of similarity was also found in the 400 nucleotides 5' upstream regions of three CHH-like genes in *Metapenaeus ensis* (Gu and Chan, 1998). Although the coding regions are highly conserved among the three *Pem-CHH* genes, the 5' upstream region and the introns of *Pem-CHH2* and *Pem-CHH3* are more similar to each other than to that of *Pem-CHH1*. A comparison of the downstream sequences of these three genes also showed that the nucleotide sequence in this region of *Pem-CHH1* was less conserved than that of *Pem-CHH2* and *Pem-CHH3* (data not shown). Since *Pem-CHH1* contains only a single intron (both *Pem-CHH2* and *Pem-CHH3* contain two introns), *Pem-CHH1* seems to have gone through an independent evolutionary path from its related copies, *Pem-CHH2* and *Pem-CHH3*.

The conserved 200 nucleotides upstream from +1 of each *Pem-CHH* gene may reflect the presence of common regulatory elements among the three *Pem-CHH* genes. On the other hand, the divergence in the more distal region (upstream from the 200 position) suggests that the nucleotides in this region may not be required for the regulatory function. It is,

nevertheless, possible that these sequences may have undergone a divergence in order to acquire a specific regulatory function to each gene. However, no possible recognition sequences for any distinctive regulatory factors have been found within these distal regions.

The presence of a putative site for CREB (cAMP responsive element binding protein), a ubiquitous transcription factor that modulates the gene transcription via the cAMP/PKA signaling pathway (Silva *et al.*, 1998) in the upstream region of *Pem-CHH* genes, suggests the role of cAMP in mediating the expression of these genes. This is supported by previous studies that show that the secretion of CHH in the crayfish *Procambarus clarkii* is modulated by serotonin concentration (Escamilla-Chimal *et al.*, 2002), and that the interaction between serotonin and its receptors can modulate the cAMP level (Tierney, 2001). Pit-1 is one of the major pituitary transcription factors that control the expression of growth hormone and prolactin genes in vertebrates (Bradford *et al.*, 2000). The presence of the sequence with high similarity to the Pit-1 binding site therefore suggests the
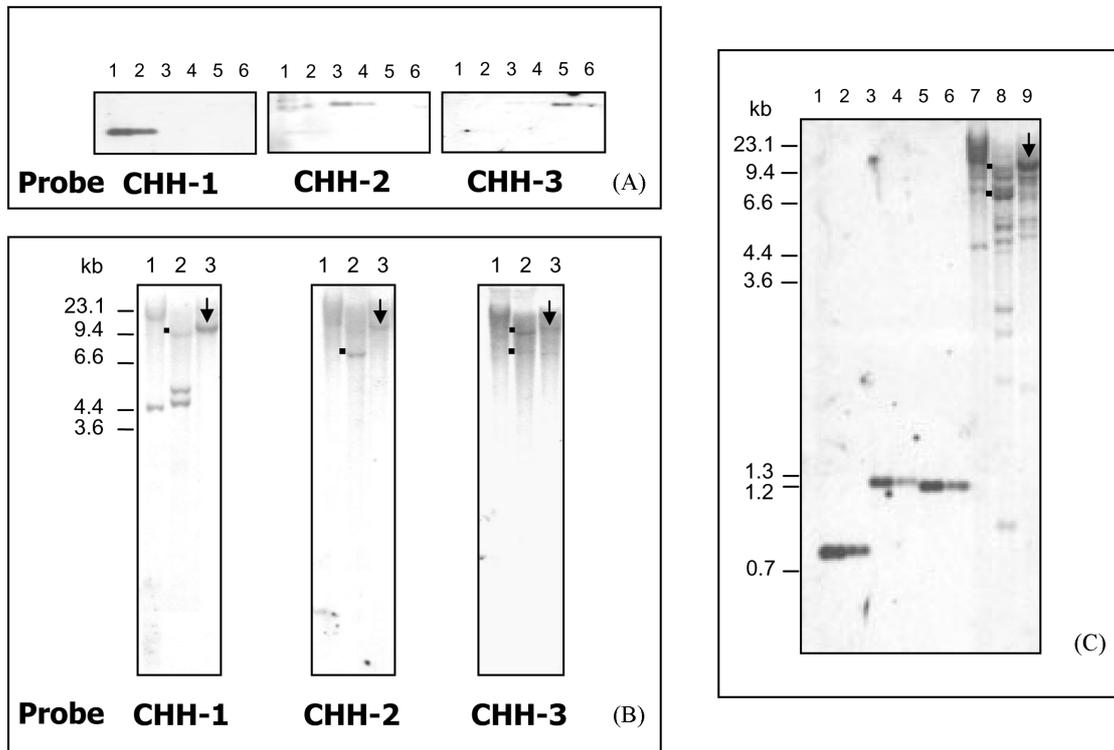
Fig. 5. Southern blot analysis of *Pem-CHH* genes of *P. monodon*. The same membrane was used in the hybridization with all DNA probes. (A) The DNA probes specific to each the *Pem-CHH* genes were hybridized to 30 pg and 10 pg of the *Pem-CHH1* genomic fragment (lanes 1 and 2) that are equivalent to three and one copies of *Pem-CHH1*, respectively. Lanes 3 and 4 represent three and one copy-equivalent of *Pem-CHH2;* lanes 5 and 6 represent three and one copy-equivalent of the *Pem-CHH3* gene, respectively. (B) About 10 μg of the *P. monodon*s genomic DNA that was digested with *Bgl* I, *Eco*R I, and *Hind* III were loaded in lanes 1 to 3, respectively. The filled arrowheads indicate the bands of the same size that were detected by CHH1- and CHH3-specific probes. Opened arrowheads indicate the bands of the same size that were detected by CHH2- and CHH3-specific probes. The bands that were detected by all of the CHH1-, CHH2-, and CHH3-specific probes are marked by arrows. (C) The membrane was re-hybridized with the mature CHH1 probe. Lanes 1 to 6 represent the three and one copy-equivalent of *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3*, respectively. Lanes 7 to 9 are genomic DNA of the *P. monodon* that was digested with *Bgl* I, *Eco*R I, and *Hind* III, respectively. The bands that were also detected by the CHH1-, CHH2-, and CHH3-specific probes are indicated as described in (B).

involvement of a tissue specific transcription factor in the regulation of the expression of the CHH-like genes. Another significant match is the homologue of the binding site for the activator protein 1 transcription factor (AP1). AP-1 is a heterodimer complex of two basic-region leucine-zipper (B-ZIP) factors, c-Fos and c-Jun (Steinmuller *et al.*, 2001). Ap-1 is involved in the control cell growth, differentiation, and development in *Drosophila* (Ciapponi and Bohmann, 2002). The presence of the binding site for AP-1 in the upstream regulatory region of CHH genes thus implies that the CHH expression may be involved in the growth and development of crustaceans via the regulation of the AP1-like factor. Binding sites for similar transcription factors were also found in other characterized CHH, MIH, and MOIH genes (Gu and Chang, 1998; Gu *et al.*, 2000; Lu *et al.*, 2000). This demonstrates that the genes in the CHH/MIH/GIH family might be under the control of similar mechanisms. Further investigation of these putative transcription factors in *P. monodon* and other crustaceans is thus promising for a better understanding of the

transcriptional control of the hormones in this family.

A Southern blot analysis showed that the genome of *P. monodon* contains at least three, one, and two copies of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes, respectively. In the shrimp *Metapenaeus ensis*, at least eighteen genes encoding two major types of CHH peptides, sixteen encoding MeCHH-A, and two encoding MeCHH-B were described. MeCHH-A and MeCHH-B share more than 98% identity in their amino acid sequences (Gu *et al.*, 2000). These genes were arranged into two different clusters, each cluster contained at least six MeCHH-A genes and one copy of the MeCHH-B gene (Gu *et al.*, 2000). In addition to CHH, other hormones in this family, MIH and MO-IH, of *Cancer pagurus* were also encoded from genes that are clustered within a 6.5 kb region of the genome (Lu *et al.*, 2000). Our results also give preliminary information on the clustering of Pem-CHH genes in the genome of *P. monodon*. A single genomic band above 10 kb that was obtained from the *Hind* III digestion hybridized to all of the Pem-CHH probes. This suggests that

at least one copy of each of the *Pem-CHH1*, *Pem-CHH2*, and *Pem-CHH3* genes are clustered within this Hind III-digested fragment. The DNA probe that was generated from the mature peptide-coding region of Pem-CHH1 can detect several extra bands in addition to those detected by each Pem-CHH-specific probe. This probe is 60-80% homologous in its nucleotide sequence to those of the other five CHH-like genes (Pm-sgp-I to -V) of *P. monodon* (Davey *et al*., 2000). It is therefore possible that the additional hybridizing bands correspond to the presence of Pm-sgp-I to -V and probably other genes in the CHH family.

In summary, this study showed that the *Pem-CHH2* and *Pem-CHH3* genes of *P. monodon* have similar structures to each other as well as to other genes in the same family that have been characterized, but is different from the previously reported structure of the *Pem-CHH1* gene. The putative binding sites for regulatory factors that were identified within the 5 upstream sequence of all the *Pem-CHH* genes suggested that these genes are regulated by similar mechanisms. Moreover, clustering of the three *Pem-CHH* genes within a distance of at least 10 kb was also demonstrated by a Southern blot analysis. Further studies on the genomic organization of CHH-like genes in *P. monodon* would provide an implication for their evolution as well as the control of their expression.

References

Bradford, A. P., Brodsky, K. S., Diamond, S. E., Kuhn, L. C., Liu, Y. and Gutierrez-Hartmann, A. (2000) The Pit-1 homeodomain and beta-domain interact with Ets-1 and modulate synergistic activation of the rat prolactin promoter. *J. Biol. Chem.* **275**, 3100-3106.

Chang, E. S. (1997) Chemistry of crustacean hormones that regulate growth and reproduction; in *Recent advances in marine biotechnology, vol.1, Endocrinology and reproduction*, Fingerman, M., Nagabhushanam, R. and Thompson, M-F. (eds.), pp.163-178, Science Publishers Inc., New Hampshire, USA.

Charmantier, G., Charmantier-Daures, M. and Van Herp, F. (1997) Hormonal regulation of growth and reproduction in crustaceans. in *Recent advances in marine biotechnology, vol.1, Endocrinology and reproduction*, Fingerman, M., Nagabhushanam, R. and Thompson, M-F. (eds.), pp. 109-161, Science Publishers Inc., New Hampshire, USA.

Ciapponi, L. and Bohmann, D. (2002) An essential function of AP-1 heterodimers in *Drosophila* development. *Mech. Dev.* **115**, 35-40.

Davey, M. L., Hall, M. R., Willis, R. H., Oliver, R. W., Thurn, M. J. and Wilson, K. J. (2000) Five crustacean hyperglycemic family hormones of *Penaeus monodon*: complementary DNA

sequence and identification in single sinus glands by electrospray ionization-fourier transform mass spectroscopy. *Mar. Biotechnol.* **2**, 80-91.

Escamilla-Chimal, E. G., Hiriart, M., Sanchez-Soto, M. C. and Fanjul-Moles, M. L. (2002) Serotonin modulation of CHH secretion by isolated cells of crayfish retina and optic lobe. *Gen. Comp. Endocrinol.* **125**, 283-290.

Gu, P. -L. and Chang, S. -M. (1998) The shrimp hyperglycemic hormone-like neuropeptide is encoded by multiple copies of genes arranged in a cluster. *FEBS Lett.* **441**, 397-403.

Gu, P. -L., Yu, K. L. and Chan, S. -M. (2000) Molecular characterization of an additional shrimp hyperglycemic hormone: cDNA cloning, gene organization, expression and biological assay of recombinant proteins. *FEBS Lett.* **472**, 122-128.

Huberman, A. and Aguilar, M. B. (1988) Single-step purification of hyperglycemic neurohormones from the sinus gland of *Procambarus bouvieri*: Comparative peptide mapping by means of high-performance liquid chromatography. *J. Chromatogr.* **443**, 337-342.

Huberman, A., Aguilar, M. B., Brew, K., Shabanowitz, J. and Hunt, D. F. (1993) Primary structure of the major isomorph of the crustacean hyperglycemic hormone (CHH-I) from the sinus gland of the Mexican crayfish *Procambarus bouvieri* (*Ortman*): interspecies comparison. *Peptides* **14**, 7-16.

Kegel, G., Reichwein, B., Weese, S., Gaus, G., Peter-Katalinic, J. and Keller, R. (1989) Amino acid sequence of the crustacean hyperglycemic hormone (CHH) from the shore crab *Carcinus maenas. FEBS Lett.* **255**, 10-14.

Kegel, G., Reichwein, B., Tensen, C. and Keller, R. (1991) Amino acid sequence of crustacean hyperglycemic hormone (CHH) from the crayfish, *Orconectes limosus*: emergence of a novel neuropeptide family. *Peptides* **12**, 909-913.

Keller, R. (1992) Crustacean neuropeptides: structures, functions and comparative aspects. *Experientia* **48**, 439-448.

Lu, W., Wainwright, G., Webster, S. G., Rees, H. H. and Turner, P. C. (2000) Clustering of mandibular organ-inhibiting hormone and moult-inhibiting hormone genes in the crab, *Cancer pagurus*, and implications for regulation of expression. *Gene* **253**, 197-207.

Martin, G., Sorokine, O. and Van Dorsselaer, A. (1993) Isolation and molecular characterization of a hyperglycemic neuropeptide from the sinus gland of the terrestrial isopod, *Armadillidium vulgare* (crustacea). *Eur. J. Biochem.* **211**, 601-607.

Silva, A. J., Kogan, J. H. and Frankland, P. W. (1998) CREB and memory. *Annu. Rev. Neurosci.* **21**, 127-148.

Soyez, D., Van Herp, F., Rossier, J., Le Caer, J. P., Tensen, C. P. and Lafone, R. (1994) Evidence for a conformation polymorphism of invertebrate neurohormones. *J. Biol. Chem.* **269**, 18295-18298.

Steinmuller, L., Cibelli, G., Moll, J. R., Vinson, C. and Thiel, G. (2001) Regulation and composition of activator protein 1 (AP-1) transcription factors controlling collagenase and c-Jun promoter activities. *Biochem. J.* **360**, 599-607.

Tensen, C. P., Janssen, K. C. P. and Van Herp, F. (1989) Isolation, characterization and physiological specificity of the crustacean hyperglycemic factor from the sinus gland of the lobster, *Homarus americanus* (Milne-Edwards). *Invert. Reprod. Dev.* **16**, 155-164.

Tensen, C. P., De Kleijn, D. P. V. and Van Herp, F. (1991a)

Cloning and sequence analysis of cDNA encoding two crustacean hyperglycemic hormone from the lobster, *Homarus americanus. Eur. J. Biochem.* **200**, 103-106.

Tensen, C. P., Verhoeven, A. H., Gaus, G., Janssen, K. P., Keller, R. and Van Herp, F. (1991b) Isolation and amino acid sequence of crustacean hyperglycemic hormone precursor-related peptides. *Peptides* **12**, 637-681.

Tierney, A. J. (2001) Structure and function of invertebrate 5-HT receptors: a review. *Comp. Biochem. Physiol. Part A*. **128**, 791-804.

Udomkit, A., Chooluck, S., Sonthayanon, B. and Panyim, S. (2000) Molecular cloning of a cDNA encoding a member of CHH/MIH/GIH family from *Penaeus monodon* and analysis of its gene structure. *J. Exp. Mar. Biol. Ecol.* **244**, 145-156.

Yang, W. J., Aida, K. and Nagasawa, H. (1995) Amino acid sequences of a hyper glycemic hormone and its related peptides from the kuruma prawn, *Penaeus japonicus. Aquaculture* **135**, 205-212.

Yang, W. J., Aida, K. and Nagasawa, H. (1997) Amino aids sequences and activities of multiple hyperglycemic hormones from the kuruma prawn, *Penaeus japonicus. Peptides* **18**, 479-485.

Yasuda, A., Yasuda, Y., Fujita, T. and Naya, Y. (1994) Characterization of crustacean hyperglycemic hormone from the crayfish (*Procambarus clarkii*): multiplicity of molecular forms by sterioinversion and diverse function. *Gen. Comp. Endocrinol.* **95**, 387-398.