J Mol Evol (2002) 54:703–714
DOI: 10.1007/s00239-001-0036-7

JOURNAL OF MOLECULAR EVOLUTION

A Possible Molecular Ancestor for Mollusk APGWamide, Insect Adipokinetic Hormone, and Crustacean Red Pigment Concentrating Hormone

Francisco Martínez-Pérez,[1] Arturo Becerra,[1] Jesús Valdés,[2] Samuel Zinker,[3] Hugo Aréchiga[4]

[1] Posgrado en Ciencias Biológicas, Facultad de Ciencias, Universidad Nacional Autónoma de México, México, D.F., México
[2] Departamento de Bioquímica, Centro de Investigación y de Estudios Avanzados del IPN, México, D.F., México
[3] Departamento de Genética y Biología Molecular, Centro de Investigación y de Estudios Avanzados del IPN, México, D.F., México
[4] División de Estudios de Posgrado e Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, México, D.F., México

Received: 9 February 2001 / Accepted: 3 October 2001

Abstract. Precursor structures of various members of the neuropeptide family adipokinetic hormone/red pigment concentrating hormone (AKH/RPCH) of mandibular arthropods and the APGWamide family of mollusks were compared. Amino acid alignments showed a common overall architecture (signal peptide, active peptide, related peptide), with a similar α helix–random coil secondary structure. DNA sequence alignments revealed close similarities between the genes encoding for the peptides of the two families. The APGWamide genes are larger than the AKH/RPCH genes. The sequence environment occupied by introns is similar in AKH/RPCH and APGWamide genes. Such similarities suggest that these peptide families might have been originated by gene rearrangements from a common ancestor having either an AKH/RPCH/APGWamide-like structure or both an AKH/RPCH-like and an APGWamide-like structures. In the former model, DNA fragments could have been gained when the ancestor evolved to mollusks and it could have lost nucleotides when the progression to mandibular arthropods took place. In the second model, AKH/RPCH-like structures could have been fused during evolution toward mandibular arthropods, whereas in mollusks they could have been lost with the possible amplification of the APGWamide-like structure. Loss of domains in exon 1 may have originated the signal peptide and the first codon of the active RPCH. In exon 2, loss of domains possibly determined the junctions of codons 2 to 5 with the loss of a APGWamide copy; exon 3 underwent fewer variations. The similarity of the mollusk APGWamide precursors is closer to that of the RPCH family than the insect AKH family, indicating an earlier evolutionary departure.

Key words: Adipokinetic hormone — APGWamide — Crustaceans — Peptide evolution — Insects — Mollusks — Neuropeptides — Red pigment concentrating hormone

Introduction

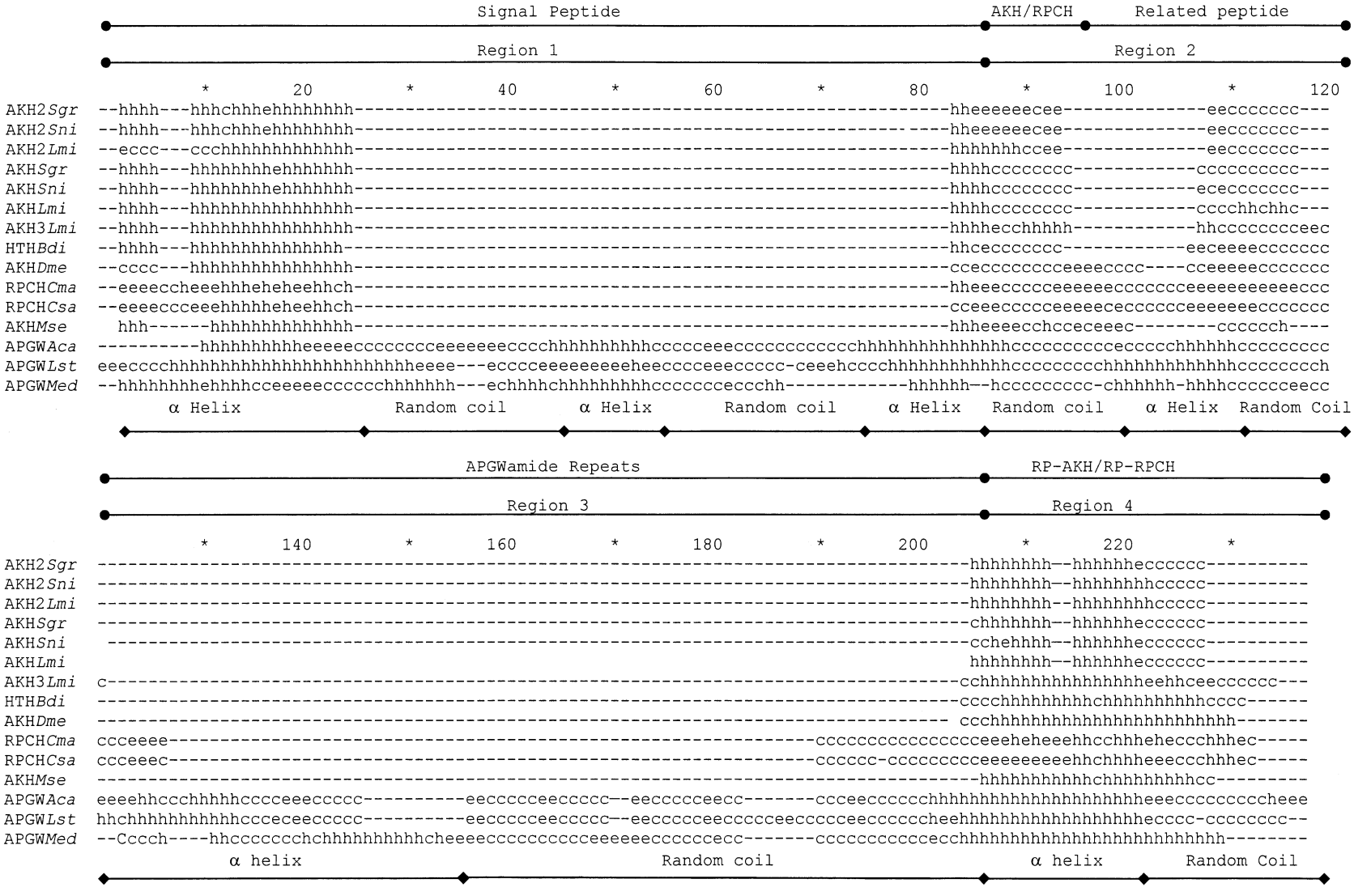
Elucidation of the chemical structure of bioactive peptides disclosed the existence of structurally related families, some members of which are present in a wide variety of zoological groups while other members appear to be unique to a given group. For instance, while peptides such as enkephalins have been identified in many species, from unicellulars to humans (Le Roith et al. 1982; O'Neill et al. 1988), the crustacean hyperglycemic hormone retains this denomination because it has been found only in crustaceans (Lacombe et al. 1999). Characterization of the peptide precursors, and more recently cloning of the genes encoding them, allows us to explore some molecular mechanisms underlying evolutionary trends (Cerff 1995).

Correspondence to: Hugo Aréchiga, Ciudad Universitaria, México, D.F. 04510, México; *email*: arechiga@servidor.unam.mx

A

Fig. 1. Alignments of the amino acid sequences of the precursors of the AKH/RPCH and APGWamide families. **A** The four regions are indicated by *lines with diamonds* (bottom). The different motifs of each precursor are indicated by *lines with circles* (top). All active peptides are indicated in *boldface*. Sequences corresponding to the signal peptide are residues 1–85 in AKH/RPCH and residues 1–54 in APGWamide. The basic amino acids that serve as proteolytic sites are *underlined*. The Cys participating in precursor dimerization is common to all species (*arrow* at residue) (226) and may form disulfide bonds with other cysteines in the related peptide or in the signal peptide (*italics, boldface*) (*underlined*). **B** All precursors show similar possible secondary structures, with an α helix (h) (region 1) followed by a random coil (c) (regions 2–3) up to the end of the related peptide, which shows an α helix followed by a random coil in the carboxyl terminus (region 4) (e, extended strand). *Sgr, Schistocerca gregaria; Sin, Schistocerca nitans; Lmi, Locusta migratoria; Bdi, Blaberus discoidalis; Dme, Drosophila melanogaster; Cma, Carcinus maenas; Csa, Callinectes sapidus; Mse, Manduca sexta; Aca, Aplysia californica; Lst, Lymnea stagnalis; Med, Mytilus edulis.*

B


Signal Peptide
AKH/RPCH
Related peptide
Region 1
Region 2
AKH2Sgr
AKH2Sni
AKH2Lmi
AKHSgr
AKHSni
AKHLmi
AKH3Lmi
HTHBdi
AKHDme
RPCHCma
RPCHCsa
AKHMse
APGWAca
APGWLst
APGWMed
α Helix
Random coil
α Helix
Random coil
α Helix
Random coil
α Helix
Random Coil
APGWamide Repeats
RP-AKH/RP-RPCH
Region 3
Region 4
α helix
Random coil
α helix
Random Coil

In the present work, we have taken as a model a neuropeptide family to date ascribed only to mandibular arthropods, particularly crustaceans and insects, with some differential features in each group. The crustacean member of the family is the red pigment concentrating hormone (RPCH), an octapeptide with the sequence pGlu–Leu–Asn–Phe–Ser–Pro–Gly–Trp–NH_2 (Fernlund and Josefsson 1972). This structure is common to all crustacean species in which it has been identified (Gauss et al. 1990). The initial physiologic effect it was known to exert, as indicated by its name, is the aggregation of pigment granules in the tegumentary erythrophores, thus regulating their color intensity. Other physiologic functions for RPCH have been described, such as the aggregation of pigments in retinal cells, thus participating in the control of photon flow to the retinal photoreceptors (Garfias et al. 1995), as well as in direct influences on the excitability of central neurons and the control of motor patterns (Swensen and Marder 2000).

In insects, a very similar peptide has been amply characterized, i.e., adipokinetic hormone (AKH). It has a primary structure varying from 8 to 10 amino acid residues, depending on the species. The most commonly identified physiologic effect of AKH is the control of lipid and carbohydrate metabolism, although it has some other functions (Gäde et al. 1997). The first described structure (AKH I) is pGlu–Leu–Asn–Phe–Thr–Pro–Asn–Trp–Gly–Thr–NH_2 (Stone et al. 1976), thus having the first four and the last amino acids in common with RPCH. Up to two other distinct sequences have been found in a single species, for example, AKH II (pGlu–Leu–Asn–Phe–Ser–Ala–Gly–Trp–NH_2) and AKH III (pGlu–Leu–Asn–Phe–Thr–Pro–Trp–Trp–NH_2) from *Locusta migratoria.* At present, the primary structures of 34 members of the family have been elucidated in 75 species. Thirty-three AKHs of the three varieties have been identified in insects, and only one RPCH structure in crustaceans (Gäde et al. 1997; Lee et al. 2000). Of particular interest is that biological cross-reactivity has been demonstrated between members of these two groups; thus, AKH may induce pigment aggregation in crustaceans and RPCH elicits adipokinetic effects in insects (Mordue and Stone 1977). These structural and functional similarities between RPCH and AKH led to the notion of an RPCH/AKH family (Gäde et al. 1997). Not all members appear to raise blood lipid levels in their native species; others, called hypertrehalosemic hormones (HTHs), stimulate the synthesis of trehalose, the main blood carbohydrate in many insect species. An example from the cockroach is a decapeptide with the following structure: pGlu–Val–Asp–Phe–Ser–Pro–Gly–Trp–Gly–Thr–NH_2 (Hayes et al. 1986). Other varieties have been identified (Gäde et al. 1997).

Study of the primary structure of peptides has proven useful for disclosing structural–functional relationships and phylogenetic analyses of peptide hormones. Among other spectroscopical techniques, circular dichroism (Paolillo et al. 1992; Brakch et al. 1993; Goldsworthy 1994; Goldsworthy et al. 1997) has been used to determine the amino acid secondary structure exposed to the surface of prohormones, such as β-turn promoting sequences (Rholam et al. 1986, 1990; Paolillo et al. 1992; Brakch et al. 1993) and Ω loops (Leszczynsky and Rose 1986; Bek and Berry 1990; Rayne and O'Shea 1993). Comparisons of peptide hormone secondary structure among distant zoological groups have been used in attempts to disclose similar functional activity (Goldsworthy 1994) or the ultimate localization of signal peptide amino acid sequences in *Escherichia coli* (Sjostrom et al. 1987). Mathematical models have been put forward to predict the relationships between hormone structures and their potencies (Lee et al. 2000). Physiological and biochemical assays include trehalose mobilization (Michalik et al. 1998), activation of glycogen phosphorylase in *Manduca* (Ziegler et al. 1998) lipid mobilization (Gäde 1993), and acetate uptake into fat bodies in locusts (Lee and Goldworthy 1995). Finally, comparison of peptide hormone primary and secondary structures among distant zoological groups has provided information for constructing phylogenetic trees (Gäde et al. 1994; Bogerd et al. 1995) and on evolutionary trends (Hoyle 1999).

A tetrapeptide has been identified in various molluscan species with the structure APGWamide, that is, with its last three residues in common with those in the C terminus of RPCH (Kuroki et al. 1990; Croll et al. 1991). It appears to play an important role in reproductive behavior. Its precursor has been sequenced (Smit et al. 1992; Favrel and Mathieu 1996; Fan et al. 1997) and bears interesting similarities to those of RPCH and AKH, which are discussed later.

In the search for a possible common ancestor of the AKH/RPCH and APGWamide peptide families, amino acid primary and secondary structures and the DNA sequences of the genes coding for these peptides were compared. Whether the sequence environments occupied by introns in the AKH genes were conserved in the RPCH and in the APGWamide genes was considered as an additional comparative criterion.

The structure of the precursors of these peptides has revealed a common overall architecture, with an initial signal peptide followed by the sequence of the active peptide and ending in the related peptide (see Fig. 1). However, a great diversity of structures was apparent, and between them and the RPCH precursors, thus suggesting a great genetic diversity (Linck et al. 1993). Some of the genes encoding for the precursors of the three AKHs have been cloned, and their architecture is quite similar (Bogerd et al. 1995).

The AKH/RPCH and APGWamide families are therefore a good model system to pose questions such as, Are these genes only present in arthropods or are there any similar genes in other zoological groups? In particular, it

is tempting to search for the possible existence of a common ancestor within earlier phylogenetic groups. This possibility has been suggested (Noyes and Schaffer 1993), as well as the likelihood that the two families have been derived from separate but convergent evolutionary lines (Smit et al. 1992).

In this paper, we present a comparative analysis of the structure of the precursors of the members of these three peptide families and their similarities and propose possible routes for their diversification.

Materials and Methods

For structural analysis comparisons were made for the cDNA sequences of AKH I, II, and III of *Locusta migratoria* (Bogerd et al. 1995), AKH I and II of *Schistocerca gregaria* (Schulz-Allen et al. 1989; Fisher-Lougheed et al. 1993), AKH I and II of *Schistocerca nitans* (Noyes and Schaffer 1990), and HTH from *Blaberus discoidalis* (Lewis et al. 1997). Comparisons were made to the only available cDNA sequence of RPCH precursors, *Carcinus maenas* (Linck et al. 1993) and that of *Callinectes sapidus* (Klein et al. 1995). Sequences containing APGWamide were from *Limnea stagnalis* (Smit et al. 1992), *Mytilus edulis* (Favrel and Mathieu 1996), and *Aplysia californica* (Fan et al. 1997). The AKH gene structures considered were those from *Manduca sexta* (Bradfield and Keeley 1989), *Schistocerca nitans* (Noyes and Schaffer 1993), and *Drosophila melanogaster* (Noyes et al. 1995).

The alignments of the precursors were carried out using the program CLUSTAL W (Thompson et al. 1994) with a 0.05 gap station and gap window p 9. Phylogenetic trees were made with the average distances tree, using PID (from CLUSTAL W). The secondary structure was determined following the methods devised by Gibrat et al. (1987) and Geourjon and Deléage (1994).

Results and Discussion

Alignment of the Amino Acid Structures of AKH/RPCH and APGWamide Precursors

As shown in Fig. 1A, the alignment shows the existence of the following four domains: the signal peptide (region 1, residues 1–85), the active peptide (region 2, residues 86–120), the APGWamide repeats (region 3, residues 121–205), and the carboxyl terminus (region 4).

In mollusks, the region of the signal peptide includes the first copy of the APGWamide; it is 43 to 57 residues longer than those of insects and crustaceans, which are more alike in this regard.

The second region contains the sequence coding for the active peptide of AKH and RPCH and the second APGWamide copy (residues 86–94). This is followed by the basic amino acids (residues 95–96 and 109–110), which are the cleavage sites for the processing of the prohormones. In addition, it contains 7–9 residues of the AKH-related peptide and the first 24 residues of the RPCH-related peptide. Among the AKH precursors, this region shows the greatest variations in size. While in *D. melanogaster* it is made up of 31 residues (including the active peptide), in other insects it contains only 18–23 residues. In this regard, crustaceans are more similar to mollusks, with 31–35 residues.

The third region shows the greatest differences among the three groups. While it contains most of the copies of the APGWamide stretch, only 20 amino acids correspond to the RPCH-related peptide in crustaceans and the whole region is absent in insects.

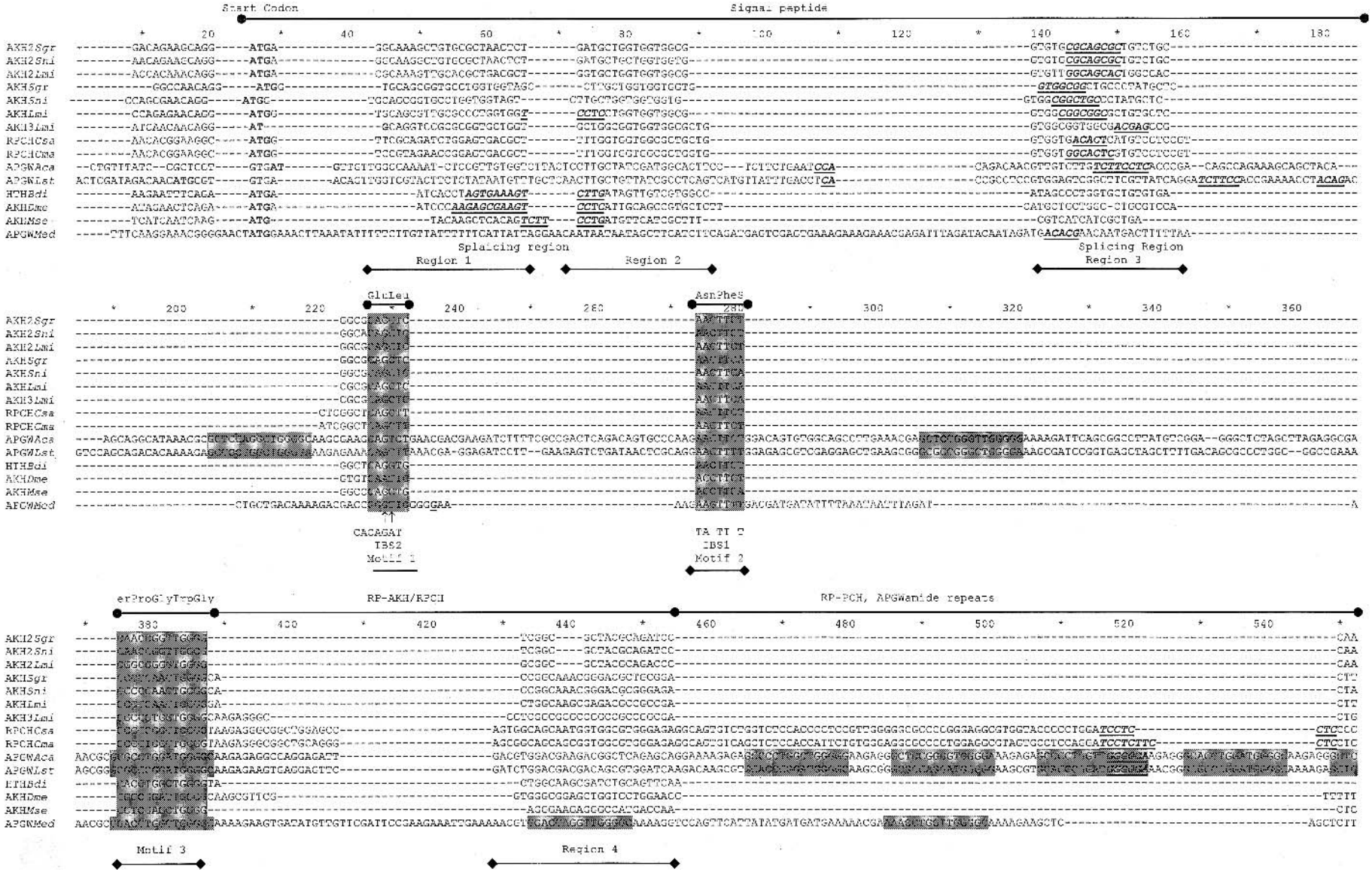
The fourth region contains the last amino acids of the AKH and RPCH-related peptides, and a similar domain is present in APGWamide. It contains the cysteine (residue 226), which has been shown to be necessary for AKH–AKH dimerization, a prerequisite for the processing of the pro-AKH in *L. migratoria* (Fisher-Lougheed et al. 1993). Additionally, Cys 226 may form a disulfide bond with another Cys in the third domain of crustaceans and mollusks (residue 201) and in the AKHs of *M. sexta, D. melanogaster*, and *B. discoidalis* and the AKH III of *L. migratoria* (residues 113 and 118, respectively).

To validate the aforementioned domains, an analysis was made of the possible secondary structure of the precursors of the AKH/RPCH family, following the methods devised by Geourjon and Deléage (1994) and Gibrat et al. (1987); both methods yielded the same results. As shown in Fig. 1B, the structure of all precursors is quite similar in all the species analyzed. Actually, the common structure is an α helix comprising region 1; however, the α helix in the mollusk is interrupted by a random coil structure (residues 20–44 and 54–74 in *A. californica*, residues 31–77 in *L. stagnalis*, and residues 15–28 and 44–65 in *M. edulis*). Region 2 forms a random coil in all species but in mollusks the structure is interrupted by an α helix formed by residues 99–111. The random coil structure is maintained in region 3 in crustaceans but at the beginning of the region in mollusks an α helix exists (residues 121–134 in *A. californica* and *L. stagnalis* and residues 141–154 in *M. edulis*). Finally, region 4 presents one α helix and ends with a random coil structure in all precursors.

As reported previously (Kuroki et al. 1990; Croll et al. 1991), both analyses showed good primary and secondary structural conservation of the signal peptide, active peptide, and C-terminus regions. In particular, the last three amino acids of the physiologically active AKH/RPCH peptide are well conserved with the second repeat of the APGWamide peptides in all species studied (Figs. 1A and B). For this reason, it was investigated whether this conservation is maintained at gene level.

DNA Sequence Alignments of the Precursors

Figure 2 shows that the codons for the signal peptide consist of three regions [nucleotides (nt), 1–223]: an immediate one, with those encoding for the last codon of the signal peptide, and the first two amino acids of AKH and RPCH (nt 224–232). This region is also present in


Start Codon
Signal peptide
Splicing region
Region 1
Region 2
Splicing Region
Region 3
GluLeu
AsnPheS
CACAGAT
IBS2
Motif 2
TA TT C
IBS1
Motif 2
erProGlyTrpGly
RP-AKH/RPCH
RP-PCH, APGWamide repeats
Motif 3
Region 4

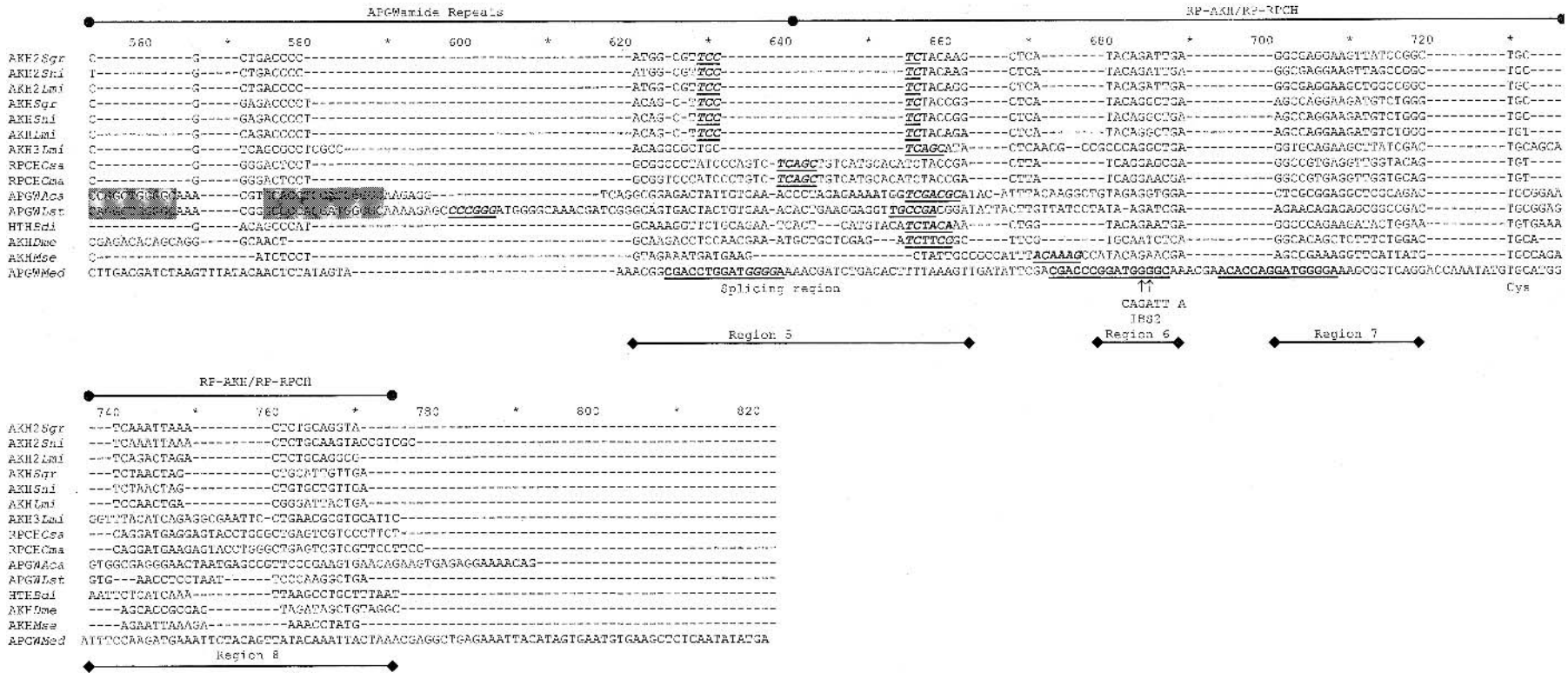


Fig. 2. Alignment of the nucleotide sequence precursors of the AKH/RPCH and APGWamide families. The homologous regions among hormones AKH, RPCH, and APGWamide are indicated below the alignments (*lines with filled diamonds*). Note that motif 1 corresponds to the first two codons of AKH and RPCH, while motif 2 contains codons 3 and 4 and the first nucleotide of codon 5. Finally, motif 3 contains the remaining two nucleotides of codon 5 and codons 6–8 of AKH, RPCH, and APGWamide (*boxed*). Motifs 1 to 3 are indicated by *open diamonds*. The *arrows* underneath indicate the nucleotides in which introns are localized in the AKH genes (*asterisks* indicate the reported sequences for precursor genes). Immediately upstream of these exon–exon junctions are located the possible SR protein-binding and splicing-enhancer motifs (*italics, boldface*, and *underlined*). Motifs 1 and 2 also correspond to two possible consensus sequences IBS2 and IBS1 for insertion or elimination of group II self-splicing introns. Codons for the active peptides are *boxed*. The Cys codon for precursor dimerization (nucleotides 730–732) is well conserved. Species abbreviations are as indicated in the legend to Fig. 1.

```
            Motif 1  Motif 2       Motif 3      Cut region
            ↓     ↓ ↓       ↓↓              ↓↓
AKH2Sgr     CAG CTC AAC TTC TCA ACC GGT TGG GGT CGG CGC
AKH2Sni     CAG CTC AAC TTC TCA ACC GGT TGG GGT CGG CGC
AKH2Lmi     CAG CTC AAC TTC TCG GCG GGG TGG GGG CGG CGC
AKHSgr      CAG CTC AAC TTC ACC CCC AAC TGG GGC ACC GGC AAA CGG
AKHSni      CAG CTC AAC TTC ACC CCC AAC TGG GGC ACC GGC AAA CGG
AKHLmi      CAG CTC AAC TTC ACC CCC AAC TGG GGG ACT GGC AAG CGA
AKH3Lmi     CAG CTC AAC TTC ACG CCG TGG TGG GGC AAG AGG
RPCHCsa     CAG CTT AAC TTC TCC CCC GGC TGG GGT AAG AGG
RPCHCma     CAG CTT AAC TTC TCC CCT GGC TGG GGT AAG AGG
APGWAca     CAG TCT AAC TTC TCG CCT GGA TGG GGC AAG AGA
APGWLst     CAG TTT AAC TTT TCG CCC GGA TGG GGC AAG AGA
HTHBdi      CAG GTG AAC TTC TCA CCT GGC TGG GGT ACT GGC AAG CGA
AKHDme      CAA TTG ACC TTC TCG CCG GAT TGG GGC AAG CGT
AKHMse      CAG CTG ACC TTC ACC TCG AGC TGG GGA GGG AAG AGG
APGWMed     GGG TTG AAG TTT TGA CCT GGC TGG GGC AAA AGA
             1   2   3   4   5   6   7   8   9  10  11  12  13

            Motif 1  Motif 2       Motif 3      Cut region
            ↓     ↓ ↓       ↓↓              ↓↓
AKH2Sgr     Gln-Leu-Asn-Phe-Ser-Thr-Gly-Trp-Gly-Arg-Arg
AKH2Sni     Gln-Leu-Asn-Phe-Ser-Thr-Gly-Trp-Gly-Arg-Arg
AKH2Lmi     Gln-Leu-Asn-Phe-Ser-Ala-Gly-Trp-Gly-Arg-Arg
AKHSgr      Gln-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-Thr-Gly-Lys-Arg
AKHSni      Gln-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-Thr-Gly-Lys-Arg
AKHLmi      Gln-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-Thr-Gly-Lys-Arg
AKH3Lmi     Gln-Leu-Asn-Phe-Thr-Pro-Trp-Trp-Gly-Lys-Arg
RPCHCsa     Gln-Leu-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Lys-Arg
RPCHCma     Gln-Leu-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Lys-Arg
APGWAca     Gln-Phe-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Lys-Arg
APGWLst     Gln-Val-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Lys-Arg
HTHBdi      Gln-Val-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Thr-Gly-Lys-Arg
AKHDme      Gln-Leu-Thr-Phe-Ser-Pro-Asp-Trp-Gly-Lys-Arg
AKHMse      Gln-Leu-Thr-Phe-Thr-Ser-Ser-Trp-Gly-Gly-Lys-Arg
APGWMed     Gly-Val-Lys-Phe-***-Pro-Gly-Trp-Gly-Lys-Arg
             1   2   3   4   5   6   7   8   9  10  11  12  13
```

Fig. 3. Sequence alignment of the regions encoding for AKH and RPCH with the precursors of the APGWamide. When linking the three motifs of nucleic acids conserved between the AKH/RPCH family and APGWamide (indicated by *arrows*), it can be seen that the three mollusk species (*boldface*) contain the sequenced encoding AKH and RPCH. The conceptual translation of the APGWamide of *A. californica* and *L. stagnalis* (in *boldface*) shows 91% identity with RPCH and AKH. The only difference is in codon 2 of *A. californica*, in which a cytosine substitutes for a thymine (see Fig. 2). In the APGWamide precursor of *M. edulis*, the union of the three motifs results in a stop codon TGA (codon 5, labeled ***). Species abbreviations are as indicated in the legend to Fig. 1.

the precursor of the three molluscan species (motif 1) and its consensus is CAGBTB (B = no adenine). The codons for amino acids 3 and 4 and the first nucleotide of codon 5 in AKH and RPCH (nt 274–280) are separated by 45 nt in the precursors of *A. californica* and *M. edulis.* In this region, the consensus in the precursors of mollusks, crustaceans, and insects is AACTTCW (W = A/T) (motif 2). Finally, the codons for amino acids 6–8 (nt 375–387) are separated by 94 nt in *A. californica* and *L. stagnalis.* This region contains the two nucleotides of amino acid 6 and the first two nucleotides coding for amino acid 9 of AKH and RPCH (motif 3). The whole motif has the consensus CVCCBDRNTGGGGN (R = A/T; N = any; B, D, and V = not A, C, and T, respectively) and includes the first copy of APGWamide of *M. edulis* and the third copy of APGWamide of *A. californica* and *L. stagnalis.* After motif 3, there are two homologous regions (nt 429–455 and 621–662) coding for the RP-AKH/RPCH amino acids and most of the repeats of APGWamide. The final portion of the nucleotide sequence of the precursors is a region encoding for the termination site, which show close similarities among all species. Regions 6, 7, and 8 (nt 680–689, 701–719, and 737–776, respectively) show the greatest homology at the 3′ end of the precursor, in particular, region 6, which contains the intron of AKH I and II of *S. nitans* and the codon for the Cys (nt 730) which participates in peptide dimerization.

The analysis we have presented thus far raises the issue of the origin of these genomic structures. At least two possibilities can be considered: (1) a common ancestor that could have had an AKH/RPCH/APGWamide-like structure, which possibly could have gained DNA fragments in the evolution to the molluscan lineage and could have lost nucleotides when the progression to insects and crustaceans occurred; and (2) a common ancestor that possibly had separate APGWamide-like and AKH/RPCH-like structures, which possibly could have fused during the evolutionary steps toward insects and crustaceans, whereas in mollusks the former might have been amplified, losing the majority of the AKH/RPCH-like fragments.

These possibilities are further supported when joining the three motifs obtained from the alignment of the cDNAs of the various AKH, RPCH, and APGWamides (nt 227–232, 274–280, and 374–387 in Fig. 2). Figure 3 shows the conceptual translation of this union, in which the APGWamide precursors in *A. californica* and *L. stagnalis* initiate the AKH/RPCH sequence. Two differences were observed: (1) codon 2 is occupied by Phe in *A. californica*, while in *L. stagnalis*, crustaceans, and insects a Val or Leu is substituted instead; and (2) in *A.*

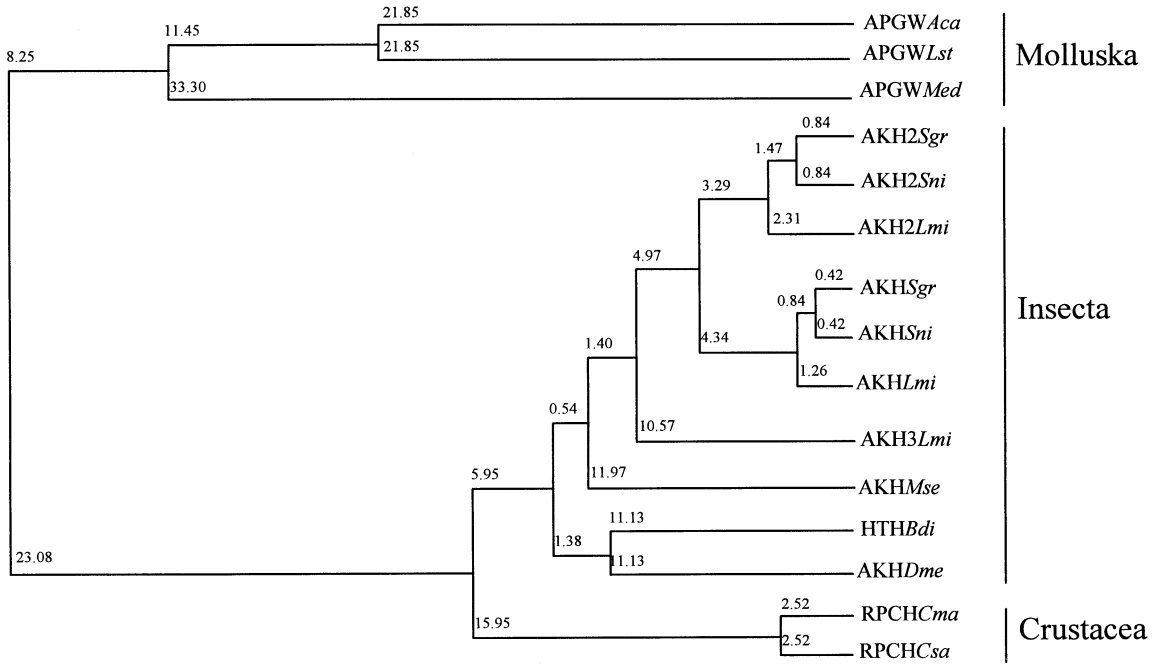


Fig. 4. Phylogenetic tree of the AKH/RPCH and APGWamide families. The amino acid phylogenetic tree shows that molluscan APGWamide precursors are closer to the possible ancestor than crustacean RPCH and insect AKH.

californica and *L. stagnalis*, codon 5 is occupied by the TCG codon for Ser. A similar result can be obtained by repeating the operation for *M. edulis.* The sequence for *L. stagnalis* is obtained, except for codon 1 (Gly) and the interruption of the open reading frame by a TGA stop codon in position 5.

The phylogenetic tree obtained from the amino acid primary sequences of the precursors of the APGWamide and RPCH/AKH families agrees with our results and previous findings (Gäde et al. 1994; Bogerd et al. 1995). The tree shows that the APGWamide precursor of mollusks, although similar, appeared earlier than those of crustacean RPCH and insect AKH precursors (Fig. 4).

A second feature that allows us to think about DNA rearrangements for the AKH/RPCH family and the APGWamide precursors of any possible ancestor is the position of the introns in the AKH gene. In *D. melanogaster*, the gene has an intron between the first and the second codons of the active peptide, while in the AKH genes of *S. nitans*, the intron is between codon 20 and codon 21 (AKH I) and between codon 52 and codon 53 (AKH II) of the AKH-related peptide (Noyes and Schaffer 1993); the AKH gene of *M. sexta* has no introns. Furthermore, AKH and APGWamide precursor alignments showed that the nucleotide environment occupied by introns is preserved in mollusks and crustaceans (Fig. 2).

Intron insertion might have occurred by reverse splicing of an excised intron (within a nonhomologous mRNA) followed by reverse transcription and homologous recombination; in addition, it could have occurred by invasion of self-splicing Group II introns (from organelles) into the nuclear genome followed by mutation, transforming intron II into a nuclear intron. A single mutation of the sequence flanking Group II introns (U/CA. . .GU) is required to produce the canonic sequence of nuclear introns (Rzhetsky et al. 1997).

A remarkable feature of the arrangement revealed by the nucleotide alignments is that the regions containing the AKH genes in *D. melanogaster* and *S. nitans* share an identity with the IBS2 motif (nt 225–231 and 681–689 in Fig. 2), which appears to be a region for insertion or elimination of self-splicing Group II introns (Mörl and Schmelzer 1990; Yang et al. 1996). By assuming such a function for these regions, and the fact that they have become nuclear introns, it is possible that even when the exon–exon junction site does not show 100% identity with the IBS2 motif, regions sharing similarities may be considered echoes of the insertion or elimination of both introns in the precursors of APGWamide, AKH, and RPCH.

According to this assumption, the exons of AKH genes lacking one or both introns might have contained sites for splicing enhancers and/or binding domains for splicing auxiliary factors, which may have participated in the processing of the pre-mRNA of the ancestral molecule. Although such motifs no longer participate in a splicing process due to the lack of an intron, their presence could reflect indirect evidence of their role in the processing of the eliminated intron. In this regard, all possible exons have the consensus proposed by Liu et al. (1998) and by Schaal and Maniatis (1999). The sequence AGAGC (nt 56–60) is present in the first exon (nt 9–229) of *D. melanogaster.* This is similar to the binding

motif for SRp40 proteins (sequence consensus ACDGS, where S = G/C and D = not T) and a class II motif F (nt 65–76), characteristic of pyrimidine-rich enhancers (TCCTC). This motif is also in the same position in the AKH of *L. migratoria*, whereas *M. sexta* has a class II motif E (TCTTC) in the same position, but the gene lacks exons (regions 1 and 2 in Fig. 2). Near the 3′ end of the first exon in the *D. melanogaster* AKH precursor (region 3), the *S. nitans* gene, which lacks an intron, has the motif SRSASGA (nt 142–148), similar to the binding site for SF2/AFS protein. This motif is also present in region 3 of the precursors of *S. gregaria, L. migratoria,* and *C. maenas.* In this region, precursors of *C. sapidus, L. stagnalis, M. edulis*, and AKH III of *L. migratoria* have motifs similar to those recognized by SRp40 proteins. *A. californica* and *M. edulis* precursors have sequences resembling those of splicing enhancers, that is, class II motif E (Fig. 2).

In a similar fashion, at the 3′ end of the second exon of AKH I and II of *S. nitans* (nt 683–684), what could be a class II motif F (nt 629–656) precedes the intron. This motif is also present in AKH I and II of *S. gregaria* and *L. migratoria.*

In region 5, motifs similar to the binding site of SRp40 and class II motif D (TCTCC) are present in the precursors of *M. sexta* and *D. melanogaster*, respectively. Neither species has introns in this position. Region 5 bears homology with the AKH precursors of *B. discoidalis*, AKH III of *L. migratoria*, the RPCH precursors of *C. sapidus* and *C. maenas*, and the APGWamide precursors of *A. californica* and *L. stagnalis*, in which a possible SRp40 motif is also present. Additionally, in *M. edulis* and in AKH I and II of *S. gregaria* and *L. migratoria*, possible class I (GGGGA) and class II A motifs are present, respectively (Fig. 2).

Although no motif mentioned thus far has been tested experimentally, these features might suggest that the ORF in the ancestral gene could possibly have been constituted by three exons separated by two introns. If so, the first exon could have contained part of the 5′UTR end, the start codon, the nucleotides encoding the signal peptide, and the first codon for RPCH. The second exon could have encoded the APGWamide copies. It might have been localized between the nucleotides coding for amino acid residues 2–4 of RPCH, preceding one of the APGWamide copies. Finally, the third exon might have contained the last codons of the related peptide and the 3′UTR.

From our analyses, we propose that the AKH/RPCH family could have originated from rearrangements of one ancestral gene (AKH/RPC/APGWamide-like) or by recombination of two ancestral genes (AKH/RPCH-like and APGWamide-like). In any event, loss of domains in the first exon might have formed what later became the signal peptide and the first codon for the Glu residue in RPCH. In the second exon, this loss of domains could result in the union of the nucleotides forming amino acid residues 2–5 with the subsequent copy of APGWamide, from which the amino acids forming RPCH are derived. The third exon possibly underwent fewer variations in all molecular species considered in this study. Selection pressure acted to maintain the peptidic domain conserving the required Cys for AKH dimerization and for the processing of the prohormone (Fischer-Lougheed et al. 1993).

Once the RPCH precursor was formed, the gene was inherited in crustaceans and preserved in insects, in which more variations have appeared; in some species, an intron was lost, as is the case for *D. melanogaster* and *S. nitans*, while in others, such as *M. sexta*, both introns were lost to form AKH. Future cloning of RPCH and APGWamide genes will allow a more precise understanding of intron movements in the AKH/RPCH family.

The search for the ancestral gene of AKH/RPCH and APGWamide may be conducted in other mollusks, as well as in other invertebrates. In the nematode *Panagrellus redvivus*, a peptide has been identified with physiological activity similar to that of the AKH/RPCH peptides (Davenport et al. 1991); however, its structure is still unknown. To interpret these physiologic similarities, one must bear in mind that, although orthologous structures or sequences may correspond to homogeneous molecules stemming from a common molecular ancestor, they do not necessarily retain the original function(s) (Fitch 1970; Goldsworthy 1994). The neuropeptides AKH, RPCH, and APGWamide are a good example of this functional diversification, because while APGWamide's main known function appears to be the regulation of reproductive behavior and muscle control (Favrel and Mathieu 1996; Smit et al. 1992; Fan et al. 1997), RPCH regulates pigment position and neuronal activity (Garfias et al. 1995; Swensen and Marder 2000) and AKH controls lipid and carbohydrate metabolism (Stone et al. 1976). This functional diversity suggests an independent evolution of the receptors to these peptides and has been documented (Goldsworthy 1994; Hoyle 1999): the same bioactive molecule carries out entirely different functions in various phyla. The evidence of biological cross-reactivity between these peptides suggests that some molecular similarities may be found among receptors (Mordue and Stone 1976, 1977; Dallman et al. 1981). Some structure–function correlations have been reported for members of the AKH/RPCH family (Lee et al. 2000), but to date sufficient information on receptor structure is still lacking.

Given the fast rate at which information on the genomic structure of various species is being accrued, our scope for future comparisons will widen and more detailed results will be produced.

Acknowledgments. We thank Ma. Teresa Pacheco Reyes for her expert secretarial assistance. This work was supported by Grants 30595-N (to J.V.) and 28089-N from the Consejo Nacional de Ciencia y Tecnología, México.

References

Bek E, Berry R (1990) Prohormonal cleavage sites are associated with Ω loops. Biochemistry 29:178–183

Bogerd J, Kooiman FP, Pijnenburg MAP, Hekking LHP, Oudejans RCHM, Van der Horst DJ (1995) Molecular cloning of the three distinct cDNAs, each encoding a different adipokinetic hormone precursor of the migratory locust, *Locusta migratoria.* J Biol Chem 270:23038–23043

Bradfield JY, Keeley LL (1989) Adipokinetic hormone gene sequence from *Manduca sexta.* J Biol Chem 264:12791–12793

Brakch N, Rholam M, Boussetta H, Cohen P (1993) Role of β-turn in proteolytic processing of peptide hormone precursors at dibasic sites. Biochemistry 32:4925–4930

Cerff R (1995) In: Gö M, Schimmel P (eds) Tracing biological evolution in protein and gene structures. Elsevier, New York, pp 205–227

Croll RP, Van Minnen J, Smit AB, Kits KS (1991) APGWamide: Molecular, histological and physiological examination of a novel neuropeptide. In: Kitts KS, Boer HH, Joose J (eds). Molluscan neurobiology. North-Holland, Amsterdam, pp 248–254

Dallmann SH, Herman WS, Carlsen J, Josefsson L (1981) Adipokinetic activity of shrimp and locust peptide hormones in butterflies. Gene Comp Endocrinol 43:256–258

Davenport TRB, Isaac RE, Lee DL (1991) The presence of peptides related to the adipokinetic hormone/red pigment-concentrating hormone family in the nematode, *Panagrellus redivivus.* Gen Comp Endocrinol 81:419–425

Fan X, Croll RP, Wu B, Fang L, Shen Q, Painter SD, Nagle GT (1997) Molecular cloning of a cDNA encoding the neuropeptides APGWamide and cerebral peptide 1: Localization of APGWamide-like immunoreactivity in the central nervous system and male reproductive organs of *Aplysia.* J Comp Neurol 387:53–62

Favrel P, Mathieu M (1996) Molecular cloning of a cDNA encoding the precursor of Ala-Pro-Gly-Trp amide-related neuropeptides from the bivalve mollusk *Mytilus edilus.* Neurosci Lett 205:210–214

Fernlund P, Josefsson L (1972) Crustacean color change hormone: Amino acid sequence and chemical synthesis. Science 177:173–175

Fischer-Lougheed J, O'Shea M, Cornish I, Losberger C, Roluet E, Schulz-Allen MF (1993) AKH biosynthesis: Transcriptional and translational control of two co-localized prohormones. J Exp Biol 177:223–241

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool 19:99–113

Gäde G (1993) Structure-activity-relationships for the lipid-mobilization action of further bioanalogs of the adipokinetic hormone red pigment-concentrating hormone family of peptides. J Insect Physiol 39:375–383

Gäde G, Reynolds SE, Beeching JR (1994) Molecular evolution of peptides of the AKH/RPCH family. In: Davey KG, Peter RE, Tobe SS (eds). Perspectives in comparative endocrinology. National Research of Canada, Ottawa, pp 486–492

Gäde G, Hoffmann KH, Spring JH (1997) Hormonal regulation in insects: Facts, gaps, and future directions. Physiol Rev 77:963–1032

Garfias A, Rodríguez-Sosa L, Aréchiga H (1995) Modulation of crayfish retinal function by red pigment concentrating hormone. J Exp Biol 198:1447–1454

Gauss G, Kleinholz LH, Kegel G, Keller R (1990) Isolation and characterization of red pigment-concentrating hormone (RPCH) from six crustacean species. J Comp Physiol 160B:373–379

Geourjon C, Deléage G (1994) SPOM: A self-optimized method for protein secondary structure prediction. Protein Eng 7:157–164

Gibrat JF, Garnier J, Robson B (1987) Further developments of protein secondary structure prediction using information theory. J Mol Biol 198:425–443

Goldsworthy GJ (1994) Adipokinetic hormones of insects: Are they insect glucagons? In: Davey KG, Peter RE, Tobe SS (eds). Perspectives in comparative endocrinology. National Research of Canada, Ottawa, pp 486–492

Goldsworthy GJ, Lee MJ, Luswata R, Drake AF, Hyde D (1997) Structures, assays and receptors for Locust adipokinetic hormones. Comp Biochem Physiol 117:483–496

Hayes TK, Keeley LL, Knight DW (1986) Insect hypertrehalosemic hormone: Isolation and primary structure from *Blaberus discoidalis* cockroaches. Biochem Biophys Res Commun 140:674–678

Hoyle CH (1999) Neuropeptide families and their receptors: Evolutionary perspectives. Brain Res 848:1–25

Klein JM, Mohrher CJ, Sleutels F, Janeke N, Riehm JP, Rao R (1995) A highly conserved red pigment concentrating hormone precursor in the blue crab *Callinectes sapidus.* Biochem Biophys Res Commun 212:151–158

Kuroki Y, Kanda T, Kubota I, Fujisawa Y, Ikeda T, Miura A, Minamitake Y, Muneoka YA (1990) Molluscan neuropeptide related to the crustacean hormone RPCH. Biochem Biophys Res Commun 167:273–279

Lacombe C, Greve P, Martin G (1999) Overview on the sub-grouping to the crustacean hyperglycemic hormone family. Neuropeptides 33:71–80

Lee MJ, Goldsworthy GJ (1995) Acetate uptake assay: The basis for a rapid method for determining potencies of adipokinetic peptides for structure-activity studies. J Insect Physiol 41:163–170

Lee MJ, de Jong S, Gäde G, Poulus C, Goldsworthy GJ (2000) Mathematical modeling of insect neuropeptides potencies: Are quantitatively predictive models possible? Insect Biochem Mol Biol 30: 899–907

Le Roith D, Shiloach J, Roth J (1982) Is there an earlier phylogenetic precursor that is common to both the nervous and endocrine systems? Peptides 3:211–215

Leszczynsky JF, Rose GD (1986) Loops in globular proteins: A novel category of secondary structure. Science 234:849–855

Lewis DK, Jezierski MK, Keeley LL, Bradfield JY (1997) Hypertrehalosemic hormone in a cockroach: Molecular cloning and expression. Mol Cell Endocrinol 130:101–108

Linck B, Kelin JM, Mangerich S, Keller R, Weidemann WM (1993) Molecular cloning of crustacean red pigment concentrating hormone precursor. Biochem Biophys Res Commun 195:807–813

Liu HX, Zhang M, Krainer A (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev 12:1998–2012

Michalik J, Szolajska E, Lombarska-Sliwinska D, Rosinski G, Konopinnska D (1998) Hypertrehalosemic insect peptide periplanetin CC-2 and its analogues: Synthesis and biological evaluation. Eur J Entomol 95:1–7

Mordue W, Stone JV (1976) Comparison of the biological activities of an insect and crustacean neurohormone that are structurally similar. Nature 264:287–289

Mordue W, Stone JV (1977) Relative potencies of locust adipokinetic hormone and prawn red pigment-concentrating hormone in insect and crustacean systems. Gen Comp Endocrinol 33:103–108

Mörl M, Schmelzer C (1990) Integration of group II intron bl1 into a foreign RNA by reversal of the self-splicing reaction in vitro. Cell 60:629–636

Noyes BE, Schaffer MH (1990) The structurally similar neuropeptides adipokinetic hormone I and II are derived from similar, very small mRNAs. J Biol Chem 265:483–489

Noyes BE, Schaffer MH (1993) The closely related neuropeptide genes encoding adipokinetic hormones I and II have very different 5′-flanking regions. DNA Cell Biol 12:509–516

Noyes BE, Katz FN, Schaffer MH (1995) Identification and expression of the *Drosophila* adipokinetic hormone gene. Mol Cell Endocrinol 109:133–141

O'Neill JB, Pert CB, Ruff MR, Smith CC, Higgins WJ, Zipser B (1988)

Identification and characterization of the opiate receptor in the ciliated protozoan, *Tetrahymena.* Brain Res 450:303–315

Paolillo L, Simonetti M, Brakch N, D'Auria G, Saviano M, Dettin M, Rholam M, Scatturin A, Di Bello C, Cohen P (1992) Evidence for the presence of a secondary structure at the dibasic processing site of prohormone: The pro-oxytocin model. EMBO J 11:2399–2405

Rayne RC, O'Shea M (1993) Structural requirements of pro-adipokinetic hormone I. FEBS Lett 217:905–911

Rholam M, Nicolas P, Cohen P (1986) Precursors for peptide hormones share common secondary structures forming features at the proteolytic processing sites. FEBS Lett 207:1–6

Rholam M, Cohen P, Brakch N, Paolillo L, Scatturin A, Di Bello C (1990) Evidence for β-turn structure in model peptides reproducing pro-oxytocin/neurophysin proteolytic processing site. Biochem Biophys Res Commun 168:1066–1073

Rzhetsky A, Ayala FJ, Hsu LC, Chang C, Yoshida A (1997) Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory. Proc Natl Acad Sci USA 94:6820–6825

Schaal TM, Maniatis T (1999) Selection and characterization of pre-mRNA splicing enhancers: Identification of novel SR protein-specific enhancers sequences. Mol Cell Biol 19:1705–1719

Schulz-Allen MF, Roulet E, Fischer-Lougheed J, O'Shea M (1989) Synthesis of a homodimer neurohormone precursor of locus adipokinetic hormone studied by in vitro translation and cDNA cloning. Neuron 2:1369–1373

Sjostrom M, Wold S, Wieslander A, Rilfors L (1987) Signal peptide amino acid sequences in *Escherichia coli* contain information related to final protein localization. A multivariate data analysis. EMBO J 6:823–831

Smit AB, Jimenez CR, Dirks RW, Croll RP, Geraerts PM (1992) Characterization of a cDNA clone encoding multiple copies of the neuropeptide APGWamide in the mollusk *Lymnaea stagnalis.* J Neurosci 12:1709–1715

Stone JV, Mordue W, Batley KE, Morris HR (1976) Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilization during flight. Nature 263:207–211

Swensen A, Marder E (2000) Multiple peptides converge to activate the same voltage-dependent current in a central pattern generating circuit. J Neurosci 20:6752–6759

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weighting matrix choice. Nucleic Acids Res 22:4673–4680

Yang J, Zimmerly S, Perlman PS, Lambowitz AM (1996) Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. Nature 381:332–335

Ziegler R, Cushing AS, Walpole P, Janensky RD, Morimoto H (1998) Analogs of *Manduca* adipokinetic hormone tested in bioassay and in a receptor-binding assay. Peptides 19:481–486