Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4593–4597, May 1992
Neurobiology

Evolution of the vasopressin/oxytocin superfamily: Characterization of a cDNA encoding a vasopressin-related precursor, preproconopressin, from the mollusc *Lymnaea stagnalis*

(gastropod mollusc/neuropeptide/conopressin/neurophysin/prohormone organization)

RONALD E. VAN KESTEREN*†, AUGUST B. SMIT*, ROELAND W. DIRKS‡, NICO D. DE WITH*, WIJNAND P. M. GERAERTS*, AND JOOS JOOSSE*

*Faculty of Biology, Unit Molecular Neurobiology/Endocrinology, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands; and ‡Department of Cytochemistry and Cytometry, University of Leiden, Wassenaarseweg 72, 2333 AC Leiden, The Netherlands

Communicated by Berta Scharrer, December 27, 1991

ABSTRACT **Although the nonapeptide hormones vasopressin, oxytocin, and related peptides from vertebrates and some nonapeptides from invertebrates share similarities in amino acid sequence, their evolutionary relationships are not clear. To investigate this issue, we cloned a cDNA encoding a vasopressin-related peptide, Lys-conopressin, produced in the central nervous system of the gastropod mollusc *Lymnaea stagnalis*. The predicted preproconopressin has the overall architecture of vertebrate preprovasopressins, with a signal peptide, Lys-conopressin, that is flanked at the C terminus by an amidation signal and a pair of basic residues, followed by a neurophysin domain. The *Lymnaea* neurophysin and the vertebrate neurophysins share high sequence identity, which includes the conservation of all 14 cysteine residues. In addition, the *Lymnaea* neurophysin possesses unique structural characteristics. It contains a putative N-linked glycosylation site at a position in the vertebrate neurophysins where a strictly conserved tyrosine residue, which plays an essential role in binding of the nonapeptide hormones, is found. The C-terminal copeptin homologous extension of the *Lymnaea* neurophysin has low sequence identity with the vertebrate counterparts and is probably not cleaved from the prohormone, as are the mammalian copeptins. The conopressin gene is expressed in only a few neurons in both pedal ganglia of the central nervous system. The conopressin transcript is present in two sizes, due to alternative use of polyadenylylation signals. The data presented here demonstrate that the typical organization of the prohormones of the vasopressin/oxytocin superfamily must have been present in the common ancestors of vertebrates and invertebrates.**

The vasopressin/oxytocin hormone superfamily includes vasopressin, oxytocin, and related peptides in vertebrates and in invertebrates (ref. 1; Fig. 1). Invertebrate members include the vasopressin-like diuretic hormone from the locust *Locusta migratoria* (2), the conopressins from the gastropod molluscs *Conus geographus, Conus striatus* (3), and *Lymnaea stagnalis* (unpublished observations) and cephalotocin from the cephalopod mollusc *Octopus vulgaris* (4). All of these hormones are nonapeptides that share close similarities in primary and tertiary structure. Biological activity may differ, however, mainly due to the amino acid residue in position 8, which is a basic amino acid in vasopressin and vasopressin-related peptides and a neutral amino acid in oxytocin and oxytocin-related peptides.

In vertebrates, vasopressin, oxytocin, and related peptides are encoded by two types of structurally related genes and synthesized as part of larger precursor molecules (5). It is thought that these genes evolved from an ancestral gene about 400 million years ago (6, 7). The prohormones contain, in addition to the nominal peptide domain, a highly conserved cysteine-rich neurophysin domain, which probably is involved in binding of the hormones during axonal transport. The mammalian vasopressin precursor further includes a copeptin domain that is cleaved off and may play a role as a prolactin releasing factor (8). This copeptin domain is entirely absent in the oxytocin precursor and is represented as a C-terminal highly divergent extension of the neurophysin domain in the vasopressin-related precursors of lower vertebrates.

The high identity in amino acid sequence of the vertebrate and invertebrate peptides has led to the hypothesis that the vasopressin/oxytocin superfamily may have evolved from a common ancestral form, of which the nature of the precursor and gene is unknown. To further investigate the evolutionary origin of the vasopressin/oxytocin superfamily, we have now cloned and sequenced the cDNA encoding conopressin from *Lymnaea*.§ Our analysis of the deduced amino acid sequence for *Lymnaea* preproconopressin indicates that it is organized much like the vasopressin (related) precursors of the vertebrates, with a signal sequence followed by conopressin and a remarkably conserved neurophysin domain having a divergent copeptin-homologous C-terminal sequence. Thus, we could unequivocally demonstrate that the typical architecture of the precursors of the vasopressin/oxytocin hormone superfamily must have been present in the Archaemetazoa, a stem group from which both the vertebrates and invertebrates diverged about 600 million years ago.

MATERIALS AND METHODS

Animals. Adult *Lymnaea stagnalis* (shell height, 28–34 mm), bred in the laboratory under standard conditions (9), were used.

PCR. Two degenerate oligonucleotides, oligo 1 and oligo 2, were synthesized, based on the amino acid sequence of *Lymnaea* conopressin [oligo 1, 5′-CCAAGCTTTG(CT)TT-(CT)AT(ACT)(AC)G(GATC)AA(CT)TG(CT)CC-3′; oligo 2, 5′-CCAAGCTTAA(CT)TG(CT)CC(GATC)AA(GA)GG-(GATC)GG-3′]. Both oligonucleotides were provided with *Hin*dIII restriction site extensions on the 5′ ends and were used as primers in a PCR, with cDNA isolated from a λgt10 cDNA library of the central nervous system (CNS) (10) as template. As reverse primer, the λ oligo was used that is complementary to a sequence in the left arm of λgt10 (λ oligo, 5′-AGCAAGTTCAGCCTGGTTAAGTCC-3′). In the first

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CNS, central nervous system; nt, nucleotide(s).

†To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M86610).

vasopressin(-related) peptides											
Cys	Phe	Ile	Arg	Asn	Cys	Pro	Lys	Gly-NH_2	Lys-conopressin	*Lymnaea stagnalis, Conus geographus*
Cys	Ile	Ile	Arg	Asn	Cys	Pro	Arg	Gly-NH_2	Arg-conopressin	*Conus striatus*
Cys	Leu	Ile	Thr	Asn	Cys	Pro	Arg	Gly-NH_2	diuretic hormone	*Locusta migratoria*
Cys	Tyr	Ile	Gln	Asn	Cys	Pro	Arg	Gly-NH_2	vasotocin	non-mammalian vertebrates
Cys	Tyr	Phe	Gln	Asn	Cys	Pro	Arg	Gly-NH_2	Arg-vasopressin	mammals
Cys	Tyr	Phe	Gln	Asn	Cys	Pro	Lys	Gly-NH_2	Lys-vasopressin	mammals
Cys	Phe	Phe	Gln	Asn	Cys	Pro	Arg	Gly-NH_2	phenypressin	mammals
oxytocin(-related) peptides											
Cys	Tyr	Phe	Arg	Asn	Cys	Pro	Ile	Gly-NH_2	Cephalotocin	*Octopus vulgaris*
Cys	Tyr	Ile	Asn	Asn	Cys	Pro	Leu	Gly-NH_2	Aspargtocin	cartilaginous fishes (sharks)
Cys	Tyr	Ile	Gln	Asn	Cys	Pro	Val	Gly-NH_2	Valitocin	cartilaginous fishes (sharks)
Cys	Tyr	Ile	Ser	Asn	Cys	Pro	Gln	Gly-NH_2	Glumitocin	cartilaginous fishes (rays)
Cys	Tyr	Ile	Ser	Asn	Cys	Pro	Ile	Gly-NH_2	Isotocin	bony fishes
Cys	Tyr	Ile	Gln	Asn	Cys	Pro	Ile	Gly-NH_2	Mesotocin	mammals, birds, reptiles, amphibians, lungfishes
Cys	Tyr	Ile	Gln	Asn	Cys	Pro	Leu	Gly-NH_2	Oxytocin	mammals
1	2	3	4	5	6	7	8	9		

FIG. 1. Compilation of the primary sequences of various peptides of the vasopressin/oxytocin superfamily. Residues at positions 1, 5, 6, 7, and 9 that are shared by all members of the superfamily are indicated in boldface type. Each peptide is indicated by name and by species/group in which it has been identified.

PCR, cDNA was amplified between oligo 1 and the λ oligo by using 30 cycles: 94°C for 1 min, 45°C for 1 min, and 72°C for 2 min. After amplification, 5 μl of the PCR mixture were reamplified under the same conditions, but using oligo 2 and the λ oligo. Amplified cDNA was digested with *Eco*RI and *Hin*dIII, subcloned, and sequenced.

Labeling of the Conopressin cDNAs. Radiolabeled cDNA was synthesized by primer extension on a single-stranded M13 clone containing a conopressin cDNA derived either by PCR or from λgt10. An M13-specific primer was used, and [α-^{32}P]dATP was incorporated during the reaction. The specific activity of the probes was $>2 \times 10^8$ dpm/μg.

Screening of the cDNA Library. Approximately 2×10^5 recombinant λgt10 phages from a CNS-specific cDNA library (10) were plated at 7×10^4 plaques per 245 × 245 mm dish and absorbed to Hybond-N filters (Amersham). After prehybridization, the filters were hybridized with the radiolabeled conopressin cDNA (derived by the PCR) for 16 h, washed in 2× SSC (1× SSC = 150 mM NaCl/15 mM sodium citrate, pH 7.4)/0.1% SDS at 65°C for 30 min, and autoradiographed. Positive clones were rescreened at a lower plaque density.

Nucleotide Sequence Analysis. Fragments from *Eco*RI and *Eco*RI/*Bam*HI-digested λgt10 cDNA clones were subcloned into M13mp18 and sequenced (Fig. 2*A*).

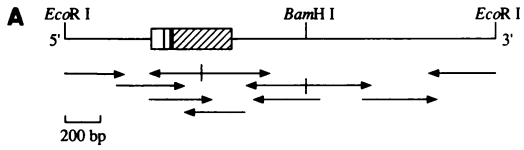
Size Determination of Conopressin mRNA. Poly(A)$^+$ RNA (7 μg) from the CNS was glyoxylated, fractionated on a 1.2% agarose gel, and transferred to a Hybond-N filter (Amersham). After prehybridization for 4 h, the filter was hybridized at 65°C for 16 h with a full-length radiolabeled conopressin cDNA, washed in 0.2× SSC/0.1% SDS at 65°C for 30 min, and autoradiographed.

***In Situ* Hybridization.** The procedure was performed as described (12). Cryostat sections (8 μm) of the CNS of *Lymnaea stagnalis* were mounted on poly(L-lysine)-coated glass slides, fixed in modified Carnoy's solution at 4°C for 30 min, and hybridized in 50% (vol/vol) formamide/3× SSC at 37°C for 16 h with the conopressin cDNA, labeled with a digoxigen-labeled dUTP, digoxigenin-11-dUTP, by standard nick-translation. Sections were rinsed three times in 50% formamide/2× SSC at 42°C and two times in 0.15 M NaCl/0.1 M Tris·HCl, pH 7.5. Hybrids were visualized by successive incubations with mouse anti-digoxigenin, sheep anti-mouse conjugated with digoxigenin, and sheep anti-digoxigenin conjugated with fluorescein isothiocyanate (all from Boehringer Mannheim).

RESULTS AND DISCUSSION

To clone the *Lymnaea* conopressin cDNA, we utilized the PCR to first amplify a cDNA fragment corresponding to the 3′ part of the mRNA encoding the conopressin region and the hypothetical neurophysin region of the preprohormone. A degenerate oligonucleotide primer that corresponds to the N-terminal part of conopressin and an oligonucleotide primer that is complementary to a sequence in the left arm of λgt10 were used to amplify cDNA from a λgt10 cDNA library of the CNS of *Lymnaea stagnalis*. Analysis of the PCR mixture by electrophoresis on an agarose gel revealed many products. To increase the specificity of the PCR, a degenerate oligonucleotide primer that corresponds to the C-terminal part of conopressin and the same λ oligonucleotide primer were used to amplify the cDNA mixture of the first round of amplification. Analysis of the second PCR mixture on agarose gel revealed a major product of 700 base pairs, which was cloned in M13mp18 and sequenced. It appeared to contain the PCR primer followed by an open reading frame encoding the endoproteolytic processing site Lys-Arg and a large cysteine-rich peptide. The first 9 amino acids of this peptide had only poor sequence identity with the nonconserved N termini of vertebrate neurophysins. The following part, however, contained all the characteristic amino acid residues previously found in the neurophysins of many vertebrate species. For instance, all the cysteine, glycine, and proline residues that are important in the formation of the tertiary structure of neurophysins (13, 14) appeared to be present at identical positions. To identify the full sequence of preproconopressin, we used this cDNA fragment to screen a *Lymnaea* CNS cDNA library. Five clones were isolated, and one, clone λVT4, contained a 2.5-kilobase (kb) insert, which was subcloned into M13mp18 and sequenced.

Nucleotide sequence analysis of several fragments (Fig. 2*A*) revealed that λVT4 contained an open reading frame encoding 155 amino acids flanked by a 498-base-pair 5′ untranslated leader sequence and a long 3′ untranslated trailer sequence of 1512 base pairs (Fig. 2*B*). The 3′ untranslated region is incomplete as indicated by the absence of a poly(A) tail; however, several polyadenylylation signals are present at nt 1226, 2294, and 2452. The deduced amino acid sequence of the *Lymnaea* conopressin precursor reveals that it is organized very similarly to the prohormones of the vasopressin/oxytocin superfamily with a hydrophobic signal peptide, a nominal peptide, a neurophysin domain, and a C-terminal copeptin homologous extension of the neurophysin domain. Translation may be initiated at one of the methionine residues at position 1, 2, or 8. As the signal peptide will be cleaved after the alanine residue at position 26, a sequence of 26, 25, or 19 residues is generated depending on the initiation site used. Conopressin contains 9 amino acids flanked by an amidation signal (glycine) and a pair of



B

```
                                                                                                    5'-CGACGACGCGACAAAACAAAGAC    23
AUUUAUAAAGUAGAGAUAGCAUCAGGAAGAGGCAGUCAUUUUCUGACUAAGAACAUUUACUCCGAAAUUUCAAGUCGAAAUUUUCUCACCUUAGU   118
UUAUUAAAUAUUUGAAUUUUUUUUUUCAAAAUUAUAAACUUUCUAACAACUUUGAUUGUUUUAAUUUUGAUAGAGUAGAUAUUUUUUUAAAAAAA   213
UAGAUUAUUUUUUGACGUAAGAGAUACGUUCCGAAAUUAAUUGAUUUCCCUUCCCAUGAUGCCACGCGGCUUGUGAGGACUAUAUCAGAGUCGUU   308
CAGGAUCUAUUCCGACUCCACGAGAUCGGACCAGACGUCUGCGGCAACGGAUUCAACAAGAGAUUCGUCGACUGUUUGCAUAGCCGCGACCCUAG   403
GGUGCAGUGGAUUCGUGUUGUCACCCCUGACACAAAAGCACGGGAGAUUUCAACAGGUCUUAUUCAAUUCCACGAUAGUACGACCACUCAGAAAA   498

1 ⟶ signal peptide                    10                                  20
AUG AUG UCC UCU CUG UGU GGU AUG CCA CUG ACC UAU CUG CUG ACC GCC GCU GUC CUG UCA CUG UCA CUG ACG   570
Met Met Ser Ser Leu Cys Gly Met Pro Leu Thr Tyr Leu Leu Thr Ala Ala Val Leu Ser Leu Ser Leu Thr

      ⟶ Lymnaea conopressin          35                  ⟶ neurophysin          45
GAC GCU UGU UUC AUC AGG AAC UGU CCA AAA GGU GGA AAG CGA UCG UUA GAC ACG GGC AUG GUG ACG UCA CGU   642
Asp Ala Cys Phe Ile Arg Asn Cys Pro Lys Gly Gly Lys Arg Ser Leu Asp Thr Gly Met Val Thr Ser Arg

    50                                  60                                  70
GAG UGC AUG AAG UGU GGG CCA GGU GGC ACC GGA CAG UGU GUC GGA CCG AGC AUC UGC UGU GGU CAG GAC UUU   714
Glu Cys Met Lys Cys Gly Pro Gly Gly Thr Gly Gln Cys Val Gly Pro Ser Ile Cys Cys Gly Gln Asp Phe

                        80                                  90
GGC UGU CAU GUC GGG ACA GCG GAG GCG GCG GUA UGC CAA CAG GAG AAC GAC AGC UCG ACC CCG UGC CUG GUC   786
Gly Cys His Val Gly Thr Ala Glu Ala Ala Val Cys Gln Gln Glu Asn Asp Ser Ser Thr Pro Cys Leu Val

          100                                 110                                 120
AAG GGG GAG GCG UGU GGG UCA AGG GAU GCG GGC AAC UGU GUG GCA GAC GGC AUA UGC UGU GAU UCA GAA UCA   858
Lys Gly Glu Ala Cys Gly Ser Arg Asp Ala Gly Asn Cys Val Ala Asp Gly Ile Cys Cys Asp Ser Glu Ser

                    130                                 140
UGU GCU GUG AAU GAC CGA UGC CGU GAU CUA GAU GGA AAC GCC CAG GCC AAC CGG GGU GAC CUC AUU CAG UUG   930
Cys Ala Val Asn Asp Arg Cys Arg Asp Leu Asp Gly Asn Ala Gln Ala Asn Arg Gly Asp Leu Ile Gln Leu

                150
AUA CAC AAA CUC CUG AAA GUG AGG GAU UAC GAC UAA CCGGAAGUCAACUAAGCCGCCAAGCUCUGCUUCUCAGCCAUGCGCGU  1013
Ile His Lys Leu Leu Lys Val Arg Asp Tyr Asp ***

CAGUGGUGGAAAACAUGGCGGGAAGUGGUCCUCAAGUCUCGCUUCUUUUGGCAUUUAGAUAACGUGGGGAAGUUGGUUGUCUCAAACAAAUGAUU  1108
GUCUCAAACAGUUGCUUGUCACAAACAGAUGAUUGUCUCAAACAGUUGAUUGCCACAAACAGCUGUUUGUCUCCAGCUGGUGUAAUGUACAUGAC  1203
AGUGAGAAUGUGGUGUAUAUGAAUAAAGAAUGGUCCAUGAAGGAGAAAUACACACCGUGAAUGAACGGACUGGGCAAUUCUUGAAAUCGUACCA  1298
CCCAUGAAACGUCACUCAAAGAUGGGGUUGCUGUGGGUUGUGCUGUGCUUGGCGGCUGUCCGUGCGGCGUCACAGGACGGGGUCGUGAACGAUCA  1393
CCGGGAUCCCUACGACAUCUACGAUUACAAAGAUCACCCAUGUCGAAGGCCCUGCGAGGCAGGUGCCCCGCCCAUGCUCUGUCGUUAUUAUUUCC  1488
UCUUGGAGCACUACACCACCAUGUCGUCAGCCUGCUACGACUGCCAACCAUUGGUGCCUACAACUGGACACAUUUUACCGUGGACGACUACCAUU  1583
UCCUUGUGGUGGCCAACGCUUUCACUGGCGAAAGCACUCUGGCCUUCUCGGUCCUCUACAUCUGGCAGGGGGACAAAUGGGUGGAGUUUCAGACA  1678
AUGGAGUUGAAGUUGUGAAAACAAAAUGGCAAAAUUUGAGGGACACUUUUCGUAAAGAACAUAUGAAAAUCUCAAACUAGAAAACUGGUAGUGCA  1773
GCUGUAUGUGAUACCCACAUGUCUUCAUGGAAGUUUUACUACCAACUUUUAUUCCUCGUGGACAUUUGUUCUUCAAGACAGACAAAGACCAAUCU  1868
ACCUGAAACAAGUCAGUGUUCCCUGUUAUCAGACUUAAAUUCACACCAAGAAGAAAGUAUGAUGGCAAGUAAUCUCAAUAUAAAUAAUGAUAUUG  1963
AUUCCACCUCCUUUAGUGGUGACACACAGGAUGCUGAAAUGUUUGGAGACACUGCUCAAAAAGACCAUAUCCCUAUUAAAAAUCCCGUAACCUUU  2058
ACAAACACAAAUAUGCAAAUGAAAAGAAAAAAAGCUGUUGGUCGGUCUGAUGCUAUCAAUCAGCUUCUAGAAAUGGAACACAAGAAGAUUGAAUG  2153
UAUGGUAUAAAAUUAACAGCAGAAGAGAUAGAGAUGAAGAAGAAAAAAAAGAUGAUGACUAUCAUUUCCUUAUGAGGUUGCUUCCACAUUUUCGU  2248
GAAAUUCCAAAGAGGUUGAAACGUGAAACAAAAAUUGAAAUACAAAUAAAUUUUGCUAAAUGCGGUUAGAGCUCAAGAAGUGGAAAGCUCAAAU  2343
UUACCAGAUUGAUCAUUGAAAUGUAACUGUUAUUAUUCAACUUUAUGACCUAACAUGACCCAAAUGUAUAUUAGCCUGGGCGUUCAGGACAUUUU  2438
CAGGACUUUGUUAAAUAAAUGAUAUUAAAAAUUUGAAAUAC-3'                                                      2479
```

FIG. 2. (*A*) Sequence strategy for the λVT4 cDNA clone. The entire cDNA (*Eco*RI–*Eco*RI) and the two *Eco*RI–*Bam*HI restriction fragments were subcloned in both orientations and sequenced by the dideoxynucleotide chain-termination method (11) as indicated by arrows, using an M13-specific sequence primer and several primers based on the cDNA sequence information. The 5′ and 3′ untranslated sequences are shown as lines and the open reading frame encoding preproconopressin are shown as a box containing a signal peptide (stippled box), conopressin (open box), a Lys-Arg processing site (solid box), and neurophysin (hatched box). (*B*) mRNA nucleotide sequence and derived amino acid sequence of the *Lymnaea* conopressin preprohormone, as deduced from the λVT4 cDNA clone. The cDNA sequence includes a 5′ untranslated leader sequence of 498 nt, an open reading frame encoding the conopressin preprohormone of 155 amino acids, and a 3′ untranslated trailer sequence of 1512 nucleotides (nt). Nucleotide positions are indicated at the right hand side of each line; amino acid positions are indicated between the lines. Arrows, start of the various peptide domains; underlining, conopressin domain; boxed, endoproteolytic processing site; boldface type, putative polyadenylylation sites. bp, Base pairs.

basic residues (Lys-Arg) at the C terminus. Protein chemistry has shown that endoproteolytic processing occurs at these paired basic residues generating an amidated conopressin (Fig. 1). The neurophysin/copeptin homologous domain consists of 117 amino acids and is probably not further processed to yield a neurophysin and a copeptin, as in mammals, since the only putative processing site (Arg-138) does not match the consensus for monobasic processing (15).

Northern blot hybridization of *Lymnaea* poly(A)$^+$ RNA from the CNS, using as a probe the conopressin cDNA, revealed a major 1.4-kb mRNA and a less-abundant transcript of 3.4 kb (Fig. 3). We estimate that the 1.4-kb mRNA is about 50 times more abundant. The 1.4-kb transcript is generated by the use of the polyadenylylation signal at position 1226. This could be demonstrated by the cloning of the corresponding cDNA (data not shown), which is identical to the 5′ part of the cDNA shown in Fig. 2*B* and contains a poly(A) tail at position 1244. The 2.5-kb cDNA (Fig. 2*B*) probably represents the 3.4-kb transcript and is lacking 900 nt at the 3′ end including the poly(A) tail. Therefore, it seems entirely possible that the presence of two conopressin transcripts in the *Lymnaea* CNS results from the use of two alternative polyadenylylation signals. We speculate that the use of the distinct polyadenylylation sites is regulated by physiological conditions and may affect the stability of the mRNA. Alternatively, the differential use of polyadenylylation signals may be neuron-specific. Also in rats, depending on physiological conditions, regulation of vasopressin transcript length has been demonstrated. However, in this case the variation involves only a variation in the length of the poly(A) tail (16, 17), not the use of alternative polyadenylylation signals.

Sequence alignment of the *Lymnaea* proconopressin and the vertebrate prohormones of the vasopressin/oxytocin superfamily (Fig. 4) shows that conopressin is identical to both vasopressin and oxytocin in 5 of 9 residues (56%). Also,

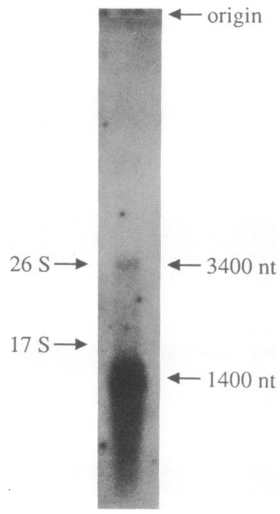


FIG. 3. Northern blot analysis of mRNA from the CNS of *Lymnaea stagnalis*. A Northern blot is shown, containing 7 μg of poly(A)+ RNA from the CNS of *Lymnaea stagnalis*, hybridized with a radiolabeled conopressin cDNA. Yeast 17S (1750 nt) and 26S (3390 nt) rRNAs were used as molecular size markers. Two conopressin transcripts are present in the *Lymnaea* CNS, with estimated lengths of 1400 and 3400 nt.

the comparison of the neurophysin domain of proconopressin with the homologous region of the vertebrate precursors shows a high degree of amino acid identity. For example, the neurophysin region of the *Lymnaea* precursor (amino acid residues 13–104) is identical to human vasopressin neurophysin (neurophysin II) in 45 of 91 residues (49%) and with human oxytocin neurophysin (neurophysin I) in 41 of 91 residues (45%). The *Lymnaea* copeptin homologous region (amino acid residues 114–139), however, is identical with human copeptin in only 5 of 26 residues (19%).

The conserved residues found in *Lymnaea* neurophysin include all residues that are necessary to form the tertiary conformation of neurophysin. Among those residues are the 14 cysteines that are involved in disulfide pairing (13, 14). The complete conservation of the cysteine residues suggests a tertiary conformation of *Lymnaea* neurophysin that is highly similar to that of vertebrate neurophysins. The identity of the cysteine positions in two similar domains indicates that neurophysins have been generated by an internal gene duplication. Interestingly, residue 79 in neurophysins of fish and residues 79 and 80 in neurophysins of higher vertebrates have been deleted after the primordial duplication. The presence of a duplicated neurophysin domain in *Lymnaea* indicates that the internal duplication probably was an early event that occurred in the common ancestors of vertebrates and invertebrates.

A highly conserved region found in vertebrate neurophysins involves the sequence Glu-Glu-Asn-Tyr-Leu-Pro-Ser (residues 61–67). This region, in particular Tyr-64, is important for binding the nonapeptide hormones (14, 24, 25). In the *Lymnaea* neurophysin, however, Tyr-64 and Leu-65 are replaced by Asp-Ser, introducing a putative N-linked glycosylation site (Asn-Asp-Ser) at position 63 (Asn-Xaa-Ser/Thr represents the consensus sequence for N-linked glycosylation). Since this site occurs in the most hydrophilic region of the molecule, at the end of a predicted α-helix, it is probably exposed to the surface and, therefore, probably used for N-linked glycosylation. If so, *Lymnaea* neurophysin might not be involved in binding of conopressin, in contrast to vertebrate neurophysins.

The most conserved region in the *Lymnaea* copeptin homologous domain is the hydrophobic core sequence, residues 128–133, in which Ile-129 and Lys-131 are conserved substitutions. It is only in mammals that the copeptin domain is cleaved off the neurophysin, whereas, in all lower vertebrates investigated so far, no processing occurs. Because of the absence of a proteolytic processing site, we predict that this is true in *Lymnaea* as well. The *Lymnaea* copeptin homologous sequence is remarkably shorter than those found in vertebrates and, in addition, it lacks the consensus sequence for N-linked glycosylation (Asn-Xaa-Ser/Thr), which is present in most vertebrate copeptin (homologous) domains, except for those of fish (20, 21). Since extensive variation exists with respect to length, amino acid sequence, and the presence of a glycosylation site, it is assumed that the copeptin homologous sequences do not have a biological function, except in mammals (8). In addition, in most oxytocin-related prohormones, except for those of the white

```
Vasopressin(-related)
                1         10          20         30         40         50         60
                *    *                      *  *       *      **       *        ___  *
Lymnaea       CFIRNCPKGG |KR| SL-DTGMVTSRECMKCGPGGTGQCVGPSICCGQDFGCHVGTAEAAVCQQENDSSTPC
Human         CYFQNCPRGG |KR| AMSDLEL---RQCLPCGPGGKGRCFGPSICCADELGCFVGTAEALRCQEENYLPSPC
Rat           CYFQNCPRGG |KR| ATSDMEL---RQCLPCGPGGKGRCFGPSICCADELGCFLGTAEALRCQEENYLPSPC
Ostrich      (CYIQNCPRGG |KR|)ALADAAL---RQCMPCGPGDRGNCFGPSICCGAELGCYVGTAETLRCAEENYLPSPC
Frog         (CYIQNCPRGG |KR|)SYPDTEV---RQCIPCGPGNRGNCFGPNICCGEDLGCYIGTPETLRCVEENYLPSPC
Chum salmon   CYIQNCPRGG |KR| ALQDTGI---RQCMTCGPGDQGHCFGPSICCGEGLGCWMGSPETARCFEENYLPTPC
White sucker  CYIQNCPRGG |KR| ALLEPVS---RQCLACGPGDKGRCLGPSICCGEEIGCLVGSPWMARCQEEEYLPSPC

Oxytocin(-related)
Human         CYIQNCPLGG |KR| AAPDLDV---RKCLPCGPGGKGRCFGPNICCAEELGCFVGTAEALRCQEENYLPSPC
White sucker  CYISNCPIGG |KR| AIQDSPS---RQCMSCGPGDRGRCFGPSICCGEGLGCLLGSPETQRCLEEDFLPSPC

Vasopressin(-related) continued
70        80        90         100        110          120        130        140        150
     *        *     **    *     *
LVKGEACGSRDAGNCVADGICCDSESCAVNDRCRD------  -  --LDGNAQANRGDLIQLIHKLLKVRDYD
QSGQKACGS--GGRCAAFGVCCNDESCVTEPECREGFHRRA |R| ASDRSNATQLDGPAGALLLRLVQLAGAPEPFEPAQPDAY
QSGQKPCGS--GGRCAAAGICCSDESCVAEPECREGFFRLT |R| AREQSNATQLDGPARELLLRLVQLAGTQESVDSAKPRVY
RAGGQPCGA--GGRCAAPGICCSDETCSLEPACLEEAGERG  G  EPAQKNLTGLDASAGDFLLKLMHLAANRQQQGGKGPLL
EAGGKPCGA--GGRCAAPGVCCNDQSCTMDSSCLDEDSERQ  R  VSPDQNMTQMNGSASDLLLRLMHMA-NRQQQQTKHY
QTGGRPCGS-DAGRCAAPGVCCDSESCVLDPDCLSE--SRY  H  SPADHSAGATSDSPGELLLRLLHFA-TRGQSEYKQ
QTAGKLCGS-DAGPCAAPGVCCGTEGCKLDPNCSED--SES  E  EPADQNT--LGASPGELLLRLLHPN-NRKHNQYQSGQK
                                                    ___
Oxytocin(-related) continued
QSGQKACGS--GGRC-ALGLCCSPDGCHADPACDAEATFSQ  R
EAGGKVCGY--EGRCAAPGVCCDSEGCSVDQSCVD------  -  GDGDATAVSQPASSQDLLLKLLHLSNPAHPYRLHQ
```

FIG. 4. Alignment of the prohormones of *Lymnaea* conopressin and vasopressin, oxytocin, and related peptides from several vertebrates. The *Lymnaea* conopressin prohormone is aligned with the prohormones of human and rat vasopressin (5), ostrich vasotocin (18), frog vasotocin (19), chum salmon vasotocin II (20), white sucker vasotocin II (21), human oxytocin (5), and white sucker isotocin I (22). Optimal alignment was achieved by computer analysis (23). The primary structures of ostrich and frog neurophysin are known from peptide sequence data only. The relationship to vasotocin is assumptive. Asterisks, cysteine residues; boxed, endoproteolytic processing sites; horizontal bar below the sequence, glycosylation sites in the copeptin domains of human, rat, ostrich, and frog; horizontal bar above the sequence, putative glycosylation site in *Lymnaea* neurophysin.

FIG. 5. Localization of preproconopressin mRNA in the pedal ganglia of *Lymnaea stagnalis* by *in situ* hybridization. An example of hybridization of preproconopressin mRNA in a single neuron of the left pedal ganglion is shown. (Bar = 20 μm.)

sucker isotocins (22), a copeptin homologous domain is entirely absent.

In situ hybridization showed that the conopressin gene is expressed in a few neurons in both pedal ganglia (Fig. 5). Immunocytochemistry has shown a wide distribution of the axons of these neurons throughout the CNS (26), strongly suggesting a central role for *Lymnaea* conopressin. This suggestion is supported by experiments that show effects of vertebrate neurohypophysial hormones on the electrical activity of central neurons of the marine snail *Aplysia californica* (27) and of the land snail *Otala lactea* (28). Further studies showed that extracts of snail CNS indeed contain appropriate peptide ligands for receptors on these neurons (29). Conopressin-containing axons also leave the CNS, suggesting that conopressin may also affect peripheral targets in *Lymnaea*. It has been demonstrated that the pleuropedal ganglia of several gastropod snails contain a vasopressin immunoreactive substance and, furthermore, that extracts of these ganglia stimulate water excretion by the skin (30, 31), suggesting that vasopressin homologues function in regulating the water balance in molluscs, as in vertebrates and insects (2).

The presence of vasopressin-related peptides in invertebrates was first proven by their structural identification in insects (2) and molluscs (3). Here, we have demonstrated that the *Lymnaea* precursor of vasopressin-related conopressin is organized very much like the precursors of the vertebrate peptides of the vasopressin/oxytocin superfamily. Thus the typical vasopressin/oxytocin preprohormone must have been present already in the Archaemetazoa, the common ancestors from which vertebrates and invertebrates diverged 600 million years ago. The presence of an oxytocin-related peptide in a cephalopod mollusc (*Octopus vulgaris*; ref. 4) suggests that also the evolutionary origin of oxytocin-related peptides may date back this far. This might imply that the gene duplication leading to a vasopressin- and an oxytocin-like lineage, which has been proposed to have occurred 400 million years ago (6, 7), in fact is an event that took place before vertebrates and invertebrates diverged.

This work was supported by the Foundation for Fundamental Biological Research (BION), which is subsidized by the Netherlands Organization for Scientific Research (NWO).

1. Acher, R. (1985) in *Neurosecretion and the Biology of Neuropeptides*, eds. Kobayashi, H., Bern, H. A. & Urano, A. (Japan Sci. Soc. Press, Tokyo/Springer, Berlin), pp. 11–25.
2. Proux, J. P., Miller, C. A., Li, J. P., Carney, P. L., Girardie, A., Delaage, M. & Schooley, D. A. (1987) *Biochem. Biophys. Res. Commun.* **149**, 180–186.
3. Cruz, L. J., de Santos, V., Zafaralla, G. C., Ramilo, C. A., Zeikus, R., Gray, W. R. & Olivera, B. M. (1987) *J. Biol. Chem.* **262**, 15821–15824.
4. Reich, G. (1992) *Neurosci. Lett.* **134**, 191–194.
5. Rehbein, M., Hillers, M., Mohr, E., Ivel, R., Morley, S., Schmale, H. & Richter, D. (1986) *Biol. Chem. Hoppe-Seyler* **367**, 695–704.
6. Acher, R. (1980) *Proc. R. Soc. London B* **210**, 21–43.
7. Hyodo, S., Kato, Y., Ono, M. & Urano, A. (1991) *J. Comp. Physiol. B* **160**, 601–608.
8. Nagy, G., Mulchahey, J. J., Smyth, D. G. & Neill, J. D. (1988) *Biochem. Biophys. Res. Commun.* **151**, 524–529.
9. van der Steen, W. J., van der Hoven, N. P. & Jager, J. C. (1969) *Neth. J. Zool.* **19**, 131–139.
10. Vreugdenhil, E., Jackson, J. F., Bouwmeester, T., Smit, A. B., van Minnen, J., van Heerikhuizen, H., Klootwijk, J. & Joosse, J. (1988) *J. Neurosci.* **8**, 4184–4191.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Dirks, R. W., Van Gijlswijk, R. P. M., Vooijs, M. A., Smit, A. B., Bogerd, J., van Minnen, J., Raap, A. K. & Van der Ploeg, M. (1991) *Exp. Cell Res.* **194**, 310–315.
13. Burman, S., Wellner, D., Chait, B., Chaudhary, T. & Breslow, E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 429–433.
14. Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W., Furey, W. F., Sax, M. & Wang, B. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4240–4244.
15. Devi, L. (1991) *FEBS Lett.* **280**, 189–194.
16. Mohr, E. & Richter, D. (1989) in *Recent Advances in Basic and Clinical Neuroendocrinology*, eds. Casanueva, F. F. & Dieguez, C. (Elsevier, Amsterdam), pp. 95–106.
17. Robinson, B. G., Frim, D. M., Schwartz, W. J. & Majzoub, J. A. (1988) *Science* **241**, 342–344.
18. Saayman, C., Naude, H. S., Oelofsen, R. J. & Chretien, M. (1987) *Int. J. Pept. Protein Res.* **30**, 634–645.
19. Michel, G., Chauvet, J., Chauvet, M. T. & Acher, R. (1987) *Biochem. Biophys. Res. Commun.* **149**, 538–544.
20. Heierhorst, J., Mahlmann, S., Morley, S. D., Coe, I. R., Sherwood, N. M. & Richter, D. (1990) *FEBS Lett.* **260**, 301–304.
21. Morley, S. D., Schönrock, C., Heierhorst, J., Figueroa, J., Lederis, K. & Richter, D. (1990) *Biochemistry* **29**, 2506–2511.
22. Figueroa, J., Morley, S. D., Heierhorst, J., Krentler, C., Lederis, K. & Richter, D. (1989) *EMBO J.* **8**, 2873–2877.
23. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
24. Abercrombie, D. M., McCormick, W. M. & Chaiken, I. M. (1982) *J. Biol. Chem.* **257**, 2274–2281.
25. Peyton, D., Sardana, V. & Breslow, E. (1987) *Biochemistry* **26**, 1518–1525.
26. van Kesteren, R. E., Smit, A. B., de With, N. D., van Minnen, J., Dirks, R. W., van der Schors, R. C. & Joosse, J. (1992) *Prog. Brain Res.* **92**, 47–57.
27. Barker, J. L., Ifshin, M. S. & Gainer, H. (1975) *Brain Res.* **84**, 501–513.
28. Barker, J. L. & Gainer, H. (1974) *Science* **184**, 1371–1372.
29. Ifshin, M. S., Gainer, H. & Barker, J. L. (1975) *Nature (London)* **254**, 72–74.
30. Deyrup-Olsen, I., Martin, A. W. & Sawyer, W. H. (1981) *Am. Zool.* **21**, 930 (abstr.).
31. Sawyer, W. H., Deyrup-Olsen, I. & Martin, A. W. (1984) *Gen. Comp. Endocrinol.* **54**, 97–108.