A corazonin G protein-coupled receptor gene in the tick Ixodes scapularis yields two splice variants, each coding for a specific corazonin receptor

Hauser, Frank; Pallesen, Malte; Lehnhoff, Janna; Li, Shizhong; Lind, Anna; Grimmelikhuijzen, Cornelis J.P.

Published in:
Biochemical and Biophysical Research Communications

DOI:
10.1016/j.bbrc.2023.04.087

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
A corazonin G protein-coupled receptor gene in the tick *Ixodes scapularis* yields two splice variants, each coding for a specific corazonin receptor

Frank Hauser, Malte Pallesen, Janna Lehnhoff, Shizhong Li, Anna Lind, Cornelis J.P. Grimmelikhuijzen*

Section for Cell and Neurobiology, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100, Copenhagen, Denmark

**A R T I C L E  I N F O**

Article history:
Received 20 April 2023
Received in revised form 21 April 2023
Accepted 24 April 2023
Available online 2 May 2023

**Keywords:**
Corazonin
GnRH
AKH
ACP
GPCR
Lyme disease

**A B S T R A C T**

We have identified a corazonin G protein-coupled receptor (GPCR) gene in the tick *Ixodes scapularis*, which likely plays a central role in the physiology and behavior of this ectoparasite. This receptor gene is unusually large (1.133 Mb) and yields two corazonin (CRZ) receptor splice variants, where nearly half of the coding regions are exchanged: CRZ-Ra (containing exon 2, exon 3, and exon 4 of the gene) and CRZ-Rb (containing exon 1, exon 3, and exon 4 of the gene). CRZ-Ra codes for a GPCR with a canonical DRF sequence at the border of the third transmembrane helix and the second intracellular loop. The positively-charged R residue from the DRF sequence is important for coupling of G proteins after activation of a GPCR. CRZ-Rb, in contrast, codes for a GPCR with an unusual DQL sequence at this position, still retaining a negatively-charged D residue, but lacking a positively-charged R residue, suggesting different G protein coupling. Another difference between the two splice variants is that exon 2 from CRZ-Ra codes for an N-terminal signal sequence. Normally, GPCRs do not have N-terminal signal sequences, although a few mammalian GPCRs have. In the tick CRZ-Ra, the signal sequence probably assists with inserting the receptor correctly into the RER membrane. We stably transfected Chinese Hamster Ovary cells with each of the two splice variants and carried out bioluminescence bioassays that also included the use of the human promiscuous G protein G16. CRZ-Ra turned out to be selective for *I. scapularis* corazonin (EC50 = 10⁻⁸ M) and could not be activated by related neuropeptides like adipokinetic hormone (AKH) and AKH/corazonin-related peptide (ACP). Similarly, also CRZ-Rb could only be activated by corazonin, although about 4-fold higher concentrations were needed to activate it (EC50 = 4 x 10⁻⁸ M). The genomic organization of the tick corazonin GPCR gene is similar to that of the insect AKH and ACP receptor genes. This similar genomic organization can also be found in the human gonadotropin-releasing hormone (GnRH) receptor gene, confirming previous conclusions that the corazonin, AKH, and ACP receptor genes are the true arthropod orthologues of the human GnRH receptor gene.

© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**1. Introduction**

Ticks, belonging to the genus *Ixodes*, such as *I. scapularis* and *I. ricinus*, are part of the phylum Arthropoda and subphylum Chelicerata, to which also spiders, mites, and scorpions belong. These ticks are dangerous to humans, because they are ectoparasites and disease vectors, carrying various pathogens. One of these pathogens is the spirochete bacterium *Borrelia*, which ticks transfer to their human hosts during a blood meal. If not killed by antibiotics shortly after its transfection, *Borrelia* will cause borreliosis (Lyme disease), which is a debilitating, chronic, neurological disease that affects 300,000 people yearly in the USA and 85,000 in Europe. Besides bacteria, *Ixodes* can also transmit other pathogenic microorganisms, such as protozoans belonging to the genus *Babesia* that invade red blood cells (causing babesiosis), and a variety of viruses that cause encephalitis (TBE, tick-borne encephalitis).

Ticks have an unusual life cycle and understanding the biology...
and endocrinology of ticks might help researchers to find novel possibilities for reducing the populations of ticks and, thereby, reduce or even eliminate the incidence of tick-borne diseases.

Using bioinformatics, we have previously annotated a number of neuropeptide GPCRs in the newly sequenced genome from *I. scapularis* [1]. Comparing the number of neuropeptide GPCRs genes occurring in ticks to that of other arthropods, showed us that a few neuropeptide GPCR genes in ticks had specifically increased their copy numbers [1]. These findings suggested that these neuropeptide GPCRs would be important for ticks and that they might control tick-specific processes like liquid excretion (removing the watery remainder from large volumes of ingested blood), or starvation. Therefore, these GPCRs would be excellent drug targets to fight ticks [1].

We previously annotated a number of exons, separated by unusually long (>600 kbp) introns, apparently belonging to two incomplete genes for corazonin GPCRs in ticks, while most other arthropods only had one copy of this receptor [1–3]. In our current paper, we investigated these exons to establish their complete gene structures and to deorphanize their encoded GPCRs.

### 2. Materials and methods

RNA was purified from dead tick tissue preserved in RNA later (Qiagen, Hilden, Germany). No experiments with live animals have been carried out and comply, therefore, with the ARRIVE guidelines for animal experiments. RNA was purified using the RNeasy Mini kit (Qiagen, Hilden, Germany) from a mixture of 10 male and 10 female *I. scapularis* (Wickel strain). cDNA was made with Superscript III (Invitrogen, Carlsbad, USA) or Q5 High-Fidelity DNA Polymerase (New England Biolabs, Waltham, USA) or Hotstart DNA polymerase (Agilent Technologies, Inc., Wilmington, USA). All used PCR primers are listed in Supplementary information.

Using bioinformatics, we have previously annotated a number of exons, separated by unusually long (>600 kbp) introns, apparently belonging to two incomplete genes for corazonin GPCRs in ticks, while most other arthropods only had one copy of this receptor [1–3]. In our current paper, we investigated these exons to establish their complete gene structures and to deorphanize their encoded GPCRs.

### 3. Results

#### 3.1. Cloning of two tick corazonin receptor splice variants

We searched the sequenced genome from the tick *I. scapularis* with queries corresponding to several insect corazonin GPCRs [2,3]. These searches identified four exons: Exon 1, exon 2, exon 3, and exon 4 (Fig. 1). Exon 1 (highlighted in red in Fig. 1) is 537 bp long and contains the extracellular N-terminus and the first 3/4 transmembrane helices of a GPCR. Exon 2 (highlighted in blue in Fig. 1) is 579 bp long and contains the N-terminus and the first 3/4 transmembrane helices of a different GPCR. Exon 1 and exon 2 are separated by an unusually large intron 1 (623 bp). Exon 3 (highlighted in yellow in Fig. 1) is 184 bp long and contains the second half of the fourth transmembrane helix and a complete transmembrane helix #5. Also exon 3 is separated from exon 2 by a large intron 2 (432 bp). Finally, exon 4 is 611 bp long and contains the transmembrane helices #6 and #7 and the intracellular C-terminus of a GPCR. It is separated from exon 3 by an intron of 76 bp.

To establish which exons are part of a functional mRNA, we carried out nested PCR, using primers annealing to consecutive 3'-end regions.
which is 104 times larger than the average gene in mammals. The gene has four exons that are separated by three introns. Intron 1 (623 kbp), intron 2 (432 kbp), and intron 3 (76 kbp) are large and not drawn to scale. The exons, however, are drawn to scale. The gene produces two splice variants: Transcript CRZ-Ra contains exon 1, exon 2, exon 3, and exon 4; transcript CRZ-Rb contains exon 1, exon 3, and exon 4. The roman numbers above each exon refer to the numbering of the transmembrane helices. The exons are highlighted by their own color (red, blue, yellow and green for exon 1, exon 2, exon 3, exon 4, respectively). However, the region in exon 2 coding for the signal peptide (SP) is highlighted in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Regions of exon 2 and nested primers annealing to consecutive 3’-regions of exon 4 (Fig. 1). These experiments showed that exon 2, exon 3, and exon 4 contributed to one mRNA, coding for a corazonin receptor splice variant, which we named CRZ-Ra (represented as a cartoon in Fig. 1, upper panel).

In our previous I. scapularis genome paper [1], we assumed that exon 1 was part of a second corazonin receptor gene, due to the long distance between exon 1 and exon 2 (623 kbp). However, in our current investigations, we were unable to identify a second corazonin receptor gene, using the methods described above. Therefore, we used another approach, where we carried out 3’-Rapid Amplification of CDNA Ends (3’-RACE) PCR, using a set of nested primers annealing to consecutive 5’-regions of exon 1 and universal 3’-RACE primers to extend the 3’-end of exon 1. These 3’-RACE experiments were successful and showed that there existed a second mRNA coded for by exon 1, exon 3, and exon 4, which we named CRZ-Rb (Fig. 1, lower panel). Thus, one corazonin receptor gene gave rise to two splice variant, CRZ-Ra and CRZ-Rb (Fig. 1). Due to its very large introns, the entire GPCR gene spans 1.133 Mb, which is 104 times larger than the average gene in I. scapularis, which was mentioned to be 10.589 bp [1].

3.2. Protein sequences of the two corazonin receptors

Fig. 2 shows an alignment of the two corazonin receptor transcripts CRZ-Ra, and CRZ-Rb. This alignment shows that, although the N-terminal receptor parts corresponding to exon 1 and exon 2 are structurally related, there are also remarkable differences. Exon 2 has an unusual property for a GPCR-coding exon. According to the widely-used “Deep Learning Model for Transmembrane Topology Prediction” [8], exon 2 codes for a CRZ-Ra with an additional N-terminal signal sequence (see Supplementary file 3). Normally, GPCRs have seven transmembrane helices (see Supplementary file 4 for CRZ-Rb), which are sufficient to insert them correctly into the RER membrane and, subsequently, into the cell membrane. However, 5–10% of all mammalian GPCRs have an additional N-terminal signal sequence, which apparently helps the N-terminus to be properly inserted into the RER membrane [10]. Our finding of a signal sequence in the tick CRZ-Ra receptor shows, to our knowledge for the first time, that this GPCR insertion mechanism also is used in arthropods (Fig. 2).

The N-terminus of CRZ-Ra has two cysteine residues (indicated by yellow stars in Fig. 2) that likely form a cystine bridge resulting in a small extracellular loop that also includes six other amino acid residues (CX6C sequence). Within these X6 sequence, there is an NXT consensus sequence for N-glycosylation, showing the likely presence of a glycosylated Asn residue within this extracellular loop (upper line of Fig. 2).

Similar to CRZ-Ra, also CRZ-Rb has two cysteine residues in its N-terminus that likely form a cystine bridge (Fig. 2). This cystine bridge makes a loop including four other amino acid residues (CX6C sequence), of which three residues form the consensus sequence NXT for N-glycosylation. Thus, also this loop in CRZ-Rb is probably glycosylated.

CRZ-Ra has a canonical DRF motif at the border of the third transmembrane helix and the second intracellular loop (encircled in Fig. 2). In CRZ-Rb, this DRF motif has been replaced by an unusual DQL sequence. DRF or DRY motifs at the intracellular border of the third transmembrane helix play an essential role in G protein coupling of the GPCR after activation by its ligand [11,12]. Therefore, we can expect that the second messenger cascades might be different between CRZ-Ra and CRZ-Rb.

3.3. Phylogenetic tree analyses

Fig. 3 shows a phylogenetic tree analysis of the two tick corazonin receptor splice variants together with other cloned and deorphanized insect and molluscan corazonin receptors [2,3,13–16]; the human and mouse GnRH receptors [17,18]; a collection of cloned and characterized insect and molluscan AKH receptors [3,6,19–21]; and several insect ACP receptors [3,22–24]. This analysis shows that the four receptor types form clusters that are clearly evolutionarily related. Fig. 3 also shows that the mammalian GnRH receptors are about equally related to either the
corazonin, AKH, or ACP receptors.

3.4. Similar genomic organizations of the tick corazonin receptor and mammalian GnRH receptor genes

Fig. 4A and B give an alignment of the genomic organizations of the human and mouse GnRH receptor genes [17,18], the tick corazonin receptor gene from Fig. 1, the bivalve (Mollusca) Crassostrea gigas corazonin receptor gene, and a collection of insect corazonin receptor genes from Drosophila melanogaster, Anopheles gambiae, and Nasonia vitripennis [2,3] (for accession numbers, see: Supplementary file 2). The human, mouse, and bivalve genes have a first exon that codes for the extracellular N-terminus, transmembrane helix I, II, III, and the first half of transmembrane helix IV; a second exon, coding for the second half of transmembrane helix IV, and a complete transmembrane helix V; and a third exon, coding for transmembrane helices VI and VII, and the intracellular C-terminus. The tick gene has a similar organization (Fig. 4B), but the situation is a bit more complex, because the first exon has been duplicated, giving rise to two splice variants (Fig. 1). In addition to identical overall structures of their exons, the phasing of their introns is also the same in all seven GnRH/corazonin GPCR genes: Zero for the first intron (for ticks: Also for the second intron); and one for the second intron (for ticks: The third intron) (Fig. 4A and B).

Fig. 4C shows a collection of insect AKH receptor genes, aligned with the gene structures from the above-mentioned GnRH receptor and corazonin receptor genes. Also these AKH receptor genes have introns in common with each other and with the GnRH and corazonin receptor genes, including the same phasing. For example, the third intron in the Drosophila AKH receptor gene, the fourth intron in the Anopheles AKH receptor gene, and the third intron in the Nasonia AKH receptor gene lie at exactly the same positions as in the GnRH and corazonin receptor genes.

Fig. 4D shows that we can see the same pattern in the insect ACP receptor genes. Thus, not only are the amino acid sequences of the GnRH, corazonin, AKH, and corazonin GPCRs clearly evolutionary related (Fig. 3), but also are their gene structures (Fig. 4). These results confirm the close evolutionary relationships between the GnRH, corazonin, AKH, and ACP receptor genes.

The results from Fig. 4 also suggest that from the collective corazonin/AKH/ACP receptor group, the corazonin receptor genes are the most closely related to the mammalian GnRH receptor genes. This is best illustrated by the I. scapularis and C. gigas corazonin receptor genes (Fig. 4B).

3.5. The tick corazonin preprohormones

During searches of several I. scapularis transcriptome databases, we discovered that always two slightly different transcripts were present, coding for two slightly different corazonin preprohormones. The gene coding for these preprohormones consisted of four exons (Fig. 5A). The two transcripts were generated by alternative splicing of only the last exon, exon 4. Thus, the N-
termini of both preprohormones, which also contained the immature corazonin sequences, were identical, meaning that both variants would yield the same biologically active corazonin.

The immature corazonin sequence QTQYSRGWTNG (highlighted in red) is located just after the N-terminal signal sequence (highlighted in orange) in Fig. 5B. This N-terminal signal sequence is cleaved off during the transport of the preprohormone over the RER membrane. After removal of the signal sequence, the immature corazonin sequence is liberated from the prohormone by prohormone convertase (PC1/3), which cleaves C-terminally from the sequence KRR (Fig. 5B). After the KRR sequence has been activated by the prohormone convertase (PC1/3), which cleaves C-terminally from the immature corazonin sequence, the signal sequence is removed by a basic residue-specific carboxypeptidase, the immature corazonin sequence has obtained a C-terminal G residue that is known to be converted into a C-terminal amide group. Finally, the sequence KRR (Fig. 5B) is located just after the N-terminal signal sequence. The accession numbers for the receptors are given in Supplementary Fig. 2. The sequences for the receptors are shown in different colors: insect and arthropod AKH receptors (highlighted in green); and arthropod ACP receptors (highlighted in blue). The scale at the bottom of the figure shows the number of amino acid substitutions divided by the length of the receptor sequence. The accession numbers for the receptors are given in Supplementary Table 2. Only bootstrap values above 50 are given. Species abbreviations: Aaeg, Anopheles gambiae; Mus, Mus musculus; Rpro, Rhodius prolaxis; Isca, Tribolium castaneum. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.6. Stable expression of the two receptor transcripts in Chinese Hamster Ovary (CHO) cells

We cloned the two corazonin receptor splice variants into a pIRES expression vector that, in addition to either CRZ-Ra, or CRZ-Rb, cDNAs, also contained a coding sequence for Green Fluorescent Protein. After transfection of CHO cells with the pIRES vector, those cells that had the highest Green Fluorescent Protein expression were kept for further propagation and final cell clone selection. In this way, we obtained cell clones that stably expressed one of the two receptor transcripts.

3.7. Functional characterization of the two receptor transcripts

The CHO cells that we used (see above paragraph) also stably expressed G16, which is a promiscuous human G protein that interacts with virtually all GPCRs and forces their second messenger pathways into the IP3/Ca2+ cascade [2–6]. One day before testing the CHO cells that stably expressed transcript CRZ-Ra, these cell variants were transiently transfected with cDNA coding for the bioluminescence protein aequorin and 2 h before the bioassay, coelenterazine was added to the cell medium, so that at the actual time of testing, sufficient biologically active aequorin would be present inside the cells. At the start of the test, nanomolar concentrations of tick corazonin were added to the cell medium. This neuropeptide activated the receptor, which could be monitored as a prominent increase of bioluminescence (Fig. 6A). Cells that did not contain the CRZ-Ra transcript (the “empty CHO cells”) did not react to the addition of corazonin (Fig. 6A). Cells containing the CRZ-Ra transcript could only be activated effectively by corazonin and not in a significant way by related neuropeptides, such as AKH and ACP (Fig. 6C). The EC50 of the receptor activation by corazonin was 10−8 M (Fig. 6A).

CHO cells expressing the CRZ-Rb transcript were treated in the same way as described above for CRZ-Ra. Also these cells could only be activated effectively by tick corazonin and not by AKH or ACP (Fig. 6D). However, about four times higher concentrations of tick corazonin were needed to obtain the same receptor activation as for CRZ-Ra (EC50 was 4 x 10−8 M, Fig. 6B).

Fig. 6E shows the amino acid sequences of the peptides used in our bioassays of Fig. 6C and D. Only peptides that structurally resembled tick corazonin were effective agonists of the two receptors.

4. Discussion

In our present study, we found that one large (1.133 Mb) corazonin receptor gene from the tick I. scapularis generated two splice variants, the CRZ-Ra and CRZ-Rb transcripts (Fig. 1). The GPCRs encoded by these transcripts differ in their amino acid sequences, especially in the first halves of their seven transmembrane regions (Fig. 2). A remarkable difference between the two GPCRs was the presence of a canonical DRF sequence at the border of transmembrane helix #3 and the second intracellular loop of CRZ-Ra, and an unconventional DQL sequence at this position of CRZ-Rb (encircled in Fig. 2). This unusual DQL sequence likely has an influence on the G protein-coupling to the activated CRZ-Rb receptor. This would explain why 4-fold higher concentrations of corazonin were needed to activate CRZ-Rb (EC50 = 4 x 10−8 M) compared to CRZ-Ra (EC50 = 10−8 M) (Fig. 5).

The corazonin receptor gene and the AKH and ACP receptor genes form a receptor family that is orthologous to the human GnRH receptor [6,19,22,26–28]. This is also shown in Fig. 3, where
the protein sequences from CRZ-Ra and CRZ-Rb are included in a phylogenetic tree analysis together with the human and mouse GnRH receptors and some arthropod and molluscan AKH, corazonin, and ACP receptors. These receptor orthologies (Fig. 3), which are based on amino acid sequence alignments, are confirmed by a completely different and independent approach, namely by alignments of the intron/exon organizations of their receptor genes (Fig. 4). These alignments show that the corazonin receptor genes from the tick *I. scapularis* and bivalve *C. gigas* have intron/exon organizations that are either identical (*C. gigas*) or very similar. The differences are located in their C-termini and are indicated by blue and yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
(I. scapularis) to those from the human and mouse GnRH receptor genes (Fig. 4).

Although in many organisms, the corazonin receptor, AKH receptor, ACP receptor, and GnRH receptor genes have introns and exons at corresponding positions (Fig. 4), only in very few cases these exons have been duplicated, thereby opening the possibility for alternative splicing. In the bivalve C. gigas, the AKH receptor gene has duplicated its first exon and by this it has acquired an exon 1 and an exon 2, which are completely orthologous to exon 1 and exon 2 that we find in the tick corazonin receptor, including the existence of alternative splicing (Fig. 1)[19,27]. The phyla Mollusca and Arthropoda have a common ancestor that lived in the early Cambrian period, about 500 million years ago (MYA) [29]. This suggests that the molecular mechanisms responsible for the specific duplication of the first exon and the subsequent alternative splicing of the AKH/coronin receptor gene products have been conserved for more than 500 MYA. However, this conclusion does not imply that all arthropod or molluscan species have this alternative splicing of their AKH/coronin receptor gene products, because many of them have lost this property.

Corazonin excerts a number of different actions in various arthropods, but it is sometimes hard to recognize a common nominator for them. Originally, the peptide was isolated from cockroaches based on its cardio-stimulatory actions on the isolated cockroach heart [25]. However, this action on the cockroach heart was only observed in a few cockroach species and not in other insects [30]. Later, corazonin was proposed to be associated with nutritional stress [30], because it induced gregarization-associated darkening of the cuticle of free-living locusts, living in fields under starving conditions [31,32]. In addition to metabolic stress, the peptide was also found to regulate osmotic stress by inhibiting diuresis [33]. Furthermore, corazonin induces copulation and lengthens sperm transfer in Drosophila males [34]. Finally, ejaculation stimulated by corazonin and its receptor are an essential part of the mating reward system in male Drosophila [35]. Thus, corazonin and its receptor play an important role in fly reproductive behavior, just as GnRH and its receptor are central for mammalian reproduction.

It is unknown, how corazonin and its receptor are acting in the tick I. scapularis. This tick has a complex life cycle that includes short periods of excessive blood intake (up to 100 times its own weight) followed by very long periods of starvation. Thus, one possibility is that the corazonin/receptor couple in ticks might be involved in coping with starvation, either by finely-regulating metabolic stress [30], or counter-acting desiccation by blocking diuresis [33]. The two splice variants CRZ-Ra and CRZ-Rb could have different roles in these processes.

**Funding**

This work was supported by Carlsberg Foundation (grant number CF18-0101 to CJP). This funding source had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article.
for publication.

CRediT author statement

FH and CJPG: conceptualization, supervision, resources; FH, MP, JL, SL and AL: investigation, formal analysis, validation; FH: visualization; CJPG: writing - Original Draft, funding acquisition; CJPG and FH: writing - Review & Editing, project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Cornelis Grimmelikhuijzen reports financial support was provided by Carlsberg Foundation.

Acknowledgements

We thank Drs. Stephen K.K. Wickel, University of Texas Medical Branch, USA, and Girish Neelakanta, Old Dominion University, Norfolk, USA, for kindly supplying us with several stages of I. scapularis, and the biology students Camilla S. Hoffmann, and Malte S.L. Schlosser (University of Copenhagen) for their pilot experiments during early stages of our research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2023.04.087.

References