

Genomics and Proteomics:

**A crayfish insulin-like-binding protein:
Another piece in the androgenic gland
insulin-like hormone puzzle is revealed**

Ohad Rosen, Simy Weil, Rivka Manor, Ziv
Roth, Isam Khalaila and Amir Sagi
J. Biol. Chem. published online June 17, 2013

GENOMICS AND
PROTEOMICS



Access the most updated version of this article at doi: [10.1074/jbc.M113.484279](https://doi.org/10.1074/jbc.M113.484279)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2013/06/17/M113.484279.DC1.html>

This article cites 0 references, 0 of which can be accessed free at

<http://www.jbc.org/content/early/2013/06/17/jbc.M113.484279.full.html#ref-list-1>

A crayfish insulin-like-binding protein: Another piece in the androgenic gland insulin-like hormone puzzle is revealed

Ohad Rosen^{1,2}, Simy Weil^{1,2}, Rivka Manor^{1,2}, Ziv Roth^{1,3}, Isam Khalaila³ and Amir Sagi^{1,2*}

¹Department of Life Sciences, ²National Institute for Biotechnology in the Negev and ³Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

* Corresponding author: P.O. Box 653, Beer-Sheva 84105, Israel. Tel: +972(8)6461364; Fax: +972(8)6479062; email: sagia@bgu.ac.il

Running title: IGFBP binds an insulin-like androgenic hormone in decapods

Key words: Insulin-like growth factor (IGF); Ligand binding protein; Physiology; Recombinant protein expression; Reproduction

Background: In *Crustacea*, a male-specific insulin-like androgenic hormone governs sexual differentiation.

Results: A crayfish insulin-like growth factor binding protein (Cq-IGFBP) was expressed and found to bind the androgenic hormone.

Conclusion: Cq-IGFBP is part of the scheme modulating the crustacean androgenic hormone.

Significance: Cq-IGFBP is the first IGFBP found that interacts with an insulin-like gender-specific ligand linking early evolution of insulin pathways with gender.

ABSTRACT

Across the animal kingdom, the involvement of insulin-like peptide (ILP) signaling in sex-related differentiation processes is attracting increasing attention. Recently, a gender-specific ILP was identified as the androgenic sex hormone in *Crustacea*. However, moieties modulating the actions of this androgenic insulin-like growth factor were yet to be revealed. Through molecular screening of an androgenic gland (AG) cDNA library prepared from the crayfish *Cherax quadricarinatus*, we have identified a novel insulin-like growth factor-binding protein (IGFBP) termed *Cq-IGFBP*. Based on bioinformatic analyses, the deduced Cq-IGFBP was shown to share high sequence homology with IGFBP-family members from both invertebrates and vertebrates. The protein also includes a sequence determinant proven crucial for ligand binding, which, according to 3D modeling, is assigned to the exposed outer surface of the protein. Recombinant Cq-IGFBP (rCq-IGFBP) protein was produced and, using a 'pull-down' methodology, was shown to specifically interact with the insulin-like AG hormone of the crayfish (Cq-IAG). Particularly, using both mass spectral analysis and an immunological tool, rCq-IGFBP was shown to bind the Cq-IAG prohormone. Furthermore, a peptide corresponding to residues 23-38 of the Cq-IAG A-chain was found sufficient for *in vitro* recognition by rCq-IGFBP. Cq-IGFBP is the first IGFBP family member shown to specifically interact with a gender-specific ILP. Unlike their ILP ligands, IGFBPs are highly conserved across

evolution, from ancient arthropods, like crustaceans, to humans. Such conservation places ILP signaling at the center of sex-related phenomena in early animal development.

INTRODUCTION

The insulin-like peptide (ILP) superfamily, found in both invertebrates and vertebrates, comprises some of the best studied peptides to date, such as insulin, relaxin, and bombyxin (1). While each of these hormones may be involved in its own well-defined pathway and thus in a distinctive physiological response, they all share fundamental structural traits that allow their classification as insulin-like peptides. Indeed, all ILP family members are thought to have evolved from an ancestral insulin-like gene (1). Generally, an ILP is encoded as a prohormone corresponding to a single-chained polypeptide containing a signal peptide followed by a B-chain, a connecting peptide (C-peptide) and an A-chain. Insulin-like growth factors (IGFs) are known to contain additional D- and E-domains at their C-terminus (1). Usually, a proteolytic cleavage event in which the C-peptide is removed occurs prior to secretion (2), except in the cases of IGFs which remain in a functional intact chain form. In all cases, the tertiary structure of the mature and functional hormone is acquired via the formation of a pair of disulfide bonds linking the A- and B-chains.

As proteinaceous hormones that exert their biological effects via dedicated membrane-spanning receptors, ILPs are released either to the local environment (in an autocrine or paracrine manner) or to distal target cells and tissues (i.e. at the endocrine level) via secretion into the circulatory system. While insulin and ILPs are generally associated with the systemic regulation of metabolism, IGFs are specifically involved in potentiating growth and are considered the safe-keepers of proper cell proliferation and survival processes, such as during cell differentiation, proliferation and migration of cells and in the case of apoptosis modulation (3). Imbalanced IGF signaling can result in developmental abnormalities, such as exaggerated or reduced growth and also cancer

(4-6). Thus, tight regulation of these highly potent hormones so as to render them freely available for receptor binding is essential. In contrast to insulin, which is known to freely circulate the blood system, IGFs move in the blood stream as part of complexes with specific high affinity IGF-binding proteins (IGFBP) (7). In humans, for example, the vast majority of IGF-I is transported in association with IGFBPs. These binding proteins subsequently regulate IGF-I signaling by acting as competitors to the designated receptors (8). Although all IGFBPs were found to be inhibitors of growth signaling via a sequestering of IGFs, some were also found to serve as positive regulators of IGF actions (9). IGFBPs bind IGFs with affinities that are equal to or greater than those of the various receptors (8) to dictate the availability of these hormones and accordingly, to inhibit their abilities to potentiate growth. Mechanistically, IGF-based signaling is enabled due to a reduction of this affinity, leading to IGF release at the target site either a result of IGFBP binding to the extracellular matrix or to an IGFBP-receptor, or due to IGFBP-specific proteolysis (7).

In addition to serving as regulators of metabolism and growth, insulin-like peptides have, in recent years, been assigned possible involvement in various aspects of sexual differentiation in a diverse set of animal species. Briefly, ILP signaling was shown to correlate with various sexual-related manifestations ranging from primary gonad development to secondary phenotypes (*e.g.*, dimorphic growth, weaponry, attractiveness, etc.) (10-13). With respect to ILP sex-dependency, a gender-specific insulin-like hormone has been recently recorded. In the red claw crayfish, *Cherax quadricarinatus*, a male-unique gland termed the androgenic gland (AG) was found to specifically express a male insulin-like hormone (Cq-IAG) (14). Since this pioneering study, numerous IAG-encoding transcripts have been identified in other decapod crustaceans (*e.g.*, crab, crayfish and shrimp) (15). IAG was shown to be the male sex hormone, inducing both masculine differentiation and maintenance (16) and was found to be crucial for the development and sustainability of the male gonad (17) as well as for secondary sex characteristics, such as male-

superior growth, aggression, courtship and others (18-20). With respect to sexual dimorphism, implantation of the AG into *C. quadricarinatus* females resulted in higher growth rates (similar to those of males), as compared with control females (20). This AG-related effect was later specifically linked to the crustacean IAG via an RNAi-based study, with such treatment resulting in a significant reduction of growth rates and whole body weights (21).

At present, the signaling schemes employed by IAG as an insulin-like growth-promoting factor remain unknown. Indeed, only a limited number of the conserved components from ILP growth pathways have been reported in crustaceans. These include two IGFBP homologs, although neither are linked to sexuality or growth regulation of the IGF-like AG hormone, instead being found to be associated with immunological responses (22,23). In current study, we report the identification of the first IGFBP homolog to be fully sequenced in the crayfish *C. quadricarinatus*, namely the organism in which the first decapod IAG transcript was discovered (14). The newly identified *C. quadricarinatus* IGFBP (Cq-IGFBP) is predicted to contain three key domains, namely an insulin-like binding domain, a kazal-type serine protease inhibitor domain and an immunoglobulin c-2-type domain, and based on the organization of these domains, was found to be one of many homologs found in invertebrate and vertebrate species. A recombinant version of the Cq-IGFBP protein was generated and purified and shown to specifically bind the crayfish insulin-like AG hormone.

EXPERIMENTAL PROCEDURES

Tissue dissection and extraction—Male crayfish were anesthetized in ice-cold water for about 10 min. Subsequently, the AG, green gland, hepatopancreas, muscle, sperm duct and testes were sampled. Following dissection, RNA was extracted using the EZ-RNA Total RNA Isolation Kit (Biological Industries, Beit Ha'emek, Israel). AG homogenates were also produced. Approximately 1,000 glands were dissected from eyestalk-ablated crayfish (24,25). The glands were manually homogenized in Tris-buffered saline (TBS) (150 mM NaCl, 10 mM

Tris-HCl, pH 7.5) supplemented with 2 mM AEBSF-hydrochloride protease inhibitor (AppliChem), yielding a total protein concentration of 10 µg/gland.

RT-PCR and rapid amplification of cDNA ends—cDNA was generated by a reverse transcriptase-catalyzed reaction containing 1 µg of total RNA using Expand reverse transcriptase (Roche), according to the manufacturer's instructions. The cDNA was then amplified by PCR. Amplification of *Cq-IGFBP* (accession no. KC952011) was performed using the forward primer 5'-AGAATGTGATAGGTCGAAATGTG-3' and the reverse primer 5'-TGGTCTTATCTTATCTTCTGGCTTG-3'. PCR products were separated by electrophoresis through a 1.3% agarose gel. Fragments were excised, purified (HiYield Gel/PCR DNA Fragments Extraction Kit, RBC Biosciences, Taiwan), and cloned (pGEM-Teasy vector, Promega). Selected clones were isolated, and plasmid DNA was purified (HiYield Plasmid Mini Kit, RBC Biosciences) and sequenced. The sequences of the 5' and 3' ends of *Cq-IGFBP* were obtained by 5' and 3' rapid amplification of cDNA ends (RACE) using the Clontech SMARTer RACE kit (Clontech) and following the manufacturer's protocol. PCR amplification of the 5' region was achieved using the gene-specific reverse primer *Cq-IGFBP* RACE-r (5'-GCATAGACCACAGTCCTCGCCATGA-3') and the Universal Primers Mix (UPM) provided with the kit. PCR amplification of the 3' region was performed with the UPM as a reverse primer and the gene-specific forward primer *Cq-IGFBP* RACE-f (5'-TAGTGCTGGATACTTGCGGATGCTGC-3'). The PCR products were cloned and sequenced as described above.

Expression and purification of recombinant Cq-IGFBP—*Escherichia coli* BL-21 strain cells were transformed with the pET-28a expression vector encoding an N-terminally polyhistidine-tagged version of *Cq-IGFBP* lacking its signal peptide (amino acids 23–248). The cells were grown overnight in liquid LB medium supplemented with 50 µg/ml kanamycin at 37°C under constant shaking (200 rpm). The cell culture was then diluted 1:100 into fresh LB with kanamycin and grown under the same

conditions until reaching an OD₆₀₀ of 0.6. An overnight incubation with 100 µM isopropyl-β-D-thiogalactopyranoside (IPTG) at 30°C ensued. The cells were pelleted by centrifugation at 12,000 g for 5 min, and resuspended in 10 ml of lysis buffer (50 mM Tris-HCl, pH 8, 2 mM EDTA, 100 µg/ml lysozyme). The suspension was frozen (−80°C) and thawed (37°C) three times, followed by DNase treatment (20 µg/ml) in the presence of 20 mM MgCl₂ for 30 min at 37°C. Inclusion bodies were separated from the cytosolic protein pool by centrifugation at 20,000 g for 30 min, and the inclusion bodies-containing pellet was solubilized with 4 ml of solubilization buffer (6 M Urea, 200 mM NaCl, 100 mM Tris-HCl, pH 8.3), supplemented with 10 µM β-mercaptoethanol and incubated overnight at 4°C under constant rotation. Thereafter, the sample was spun at 25,000 g for 30 min to remove any insoluble material and the His-tagged protein was purified using HisPur™ Ni-NTA Resin (Thermo Fisher Scientific) according to the manufacturer's protocol.

Binding assay—Biotinylation of recombinant *Cq-IGFBP* (r*Cq-IGFBP*) was performed using an EZ-Link HPDP-Biotin kit (Thermo Fisher Scientific), according to the manufacturer's protocol. Briefly, 2 mg of r*Cq-IGFBP* were dissolved in phosphate-buffered saline (PBS) (150 mM NaCl, 10 mM Na₂HPO₄, pH 7.4) and incubated with 400 µM HPDP-biotin for 2 h at room temperature (RT). Unbound biotin was removed using Vivaspin concentrating devices (Sartorius) with PBS. To examine whether *Cq-IGFBP* binds insulin-like peptides, three different avidin-agarose precipitations (AP) were performed. The biotinylated r*Cq-IGFBP* (25 µg) was incubated with a total AG protein homogenate (125 µg), human recombinant insulin (10 µg, Biological Industries) or human recombinant IGF-I (10 µg, PEPROTECH) for 2 h at RT in TBS (supplemented with 2 mM AEBSF) under constant agitation. Thirty microliters of agarose-bound avidin (high capacity neutravidin agarose resin, Thermo Fisher Scientific) were added to each of the above binding assays for an additional hour of incubation at RT. The reactions were then centrifuged at 8,000 g for 2 min, unbound supernatants were collected, and the pellets were washed three times with TBS.

All fractions (unbound and avidin-bound proteins) were separated on SDS-PAGE.

SDS-PAGE and immuno-blotting—Proteins were separated on 4-12% NuPAGE Bis-Tris Mini Gels (Invitrogen) using MES running buffer according to the manufacturer's instructions. Bands were visualized by either Coomassie brilliant blue or silver staining (26). For immuno-blot analyses, separated protein bands were transferred to a nitrocellulose membrane. Following blocking with 3% skim milk in TBS containing 0.1% Tween-20 (TBS-T), the membrane was incubated with the relevant primary antibodies, namely either His-probe (1: 500, Santa Cruz Biotechnology) or α -rec-Cq-IAG (1:1,000) antibodies for the blotting of rCq-IGFBP or Cq-IAG, respectively. After washing with TBS-T, the membrane was incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (1:15,000, Santa Cruz Biotechnology). Antibody binding was visualized using an EZ-ECL chemiluminescence detection kit (Biological Industries). Biotinylated rCq-IGFBP was directly blotted with HRP-conjugated streptavidin (1:1,000, Jackson ImmunoResearch Laboratories), with binding detected as described above.

Ligand dot blot—The membrane was initially dot-blotted with 2.5 μ g of rCq-IGFBP, 3 μ g of recombinant gastrolith protein 35 (rGAP35, unpublished data) and 30 ng of recombinant peptide A, corresponding to residues 23-38 of the Cq-IAG A-chain (synthesized by Adar Biotech). Upon blocking with 3% skim milk in TBS containing 0.1% Tween-20 (TBS-T), the membrane was incubated with ligand solution (100 μ g of peptide A in TBS-T) overnight at 4°C. The membrane was washed with TBS-T and incubated with a rabbit anti peptide A IgG (1:500, synthesized by Adar Biotech) primary antibody. Antibody binding was visualized using an EZ-ECL chemiluminescence detection kit. For staining of the spotted proteins, rCq-IGFBP was blotted with HRP-conjugated anti-His antibodies (1:6,000, Abcam), while rGAP35 was visualized using Ponceau stain (Sigma).

Mass spectrometry—Protein bands were excised, reduced, alkylated and trypsinized according to Roth *et al.* (27). The tryptic

peptides were purified on a C-18 column and dissolved in 0.1% formic acid. Nano-liquid chromatography and mass-spectrometry (MS) analysis were performed as described previously (28) using a 75- μ m internal diameter fused silica column, packed with C-18 resin (New Objective) connected to an Eksigent nano-LC system (Eksigent). Mass spectra were acquired using an LTQ-ORBITRAP XL (Thermo Fisher Scientific). Full MS and MS/MS fragmentation were performed in the data-dependent mode. After full MS acquisition, the multiply-charged six most abundant masses were chosen for CID MS/MS fragmentation in the LTQ performed at 35% collision energy and 30 ms activation time. Protein identification and validation were performed using Sequest and Mascot algorithms operated under Proteome Discoverer 2.0 software (Thermo Fisher Scientific) using an unpublished database containing the Cq-IAG and Cq-IGFBP sequences, allowing for 0.5 Da and 10 ppm tolerance.

Bioinformatics—The full length Cq-IGFBP cDNA sequence was computationally translated using the Expasy Proteomics Server (<http://web.expasy.org/translate/>). The deduced amino acid sequence was subsequently analyzed for the presence of conserved domains using the SMART program (<http://smart.embl-heidelberg.de>). Evolutionary analysis was conducted using MEGA, version 5.0 (29). Briefly, evolutionary history was inferred using the Neighbor-Joining method (30). The bootstrap consensus tree inferred from 5,000 replicates (31) was taken as representing the evolutionary history of the selected mature (i.e., signal peptide-lacking) IGFBPs among the taxa analyzed (31) (Table 1). Branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates were collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test are shown (31). Evolutionary distances were computed using the Poisson correction method. All positions containing gaps and missing data were eliminated, resulting in 139 positions being included in the final dataset.

3D modeling of Cq-IGFBP was performed using the ESyPred3D web server (<http://www.unamur.be/sciences/biologie/urbm/>

bioinfo/esypred/) (32) and processing by Swiss-PdbViewer software (33). The reference structures, which were chosen based on SWISS-MODEL (<http://swissmodel.expasy.org/>), were human HtrA1 (UniProt ID: Q92743, PDB 3TJQ) and the 5th immunoglobulin-like domain from human Roundabout homolog 2 (UniProt ID: Q9HCK4, PDB 2EDJ). Structures were superimposed based on their backbones, improved fit was applied and root-mean-square deviations (RMSD) were obtained in Angströms (Å) using the Swiss-Pdb Viewer. The final structural figures were generated using the PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).

RESULTS

In screening a previously constructed AG suppression subtractive hybridization (SSH) cDNA library prepared from male *C. quadricarinatus* crayfish (14), a contig encompassing a putative insulin-like-binding domain was identified. In this study, the complete sequence of this putative transcript was obtained and found to be just under 1,200 bp in length (Fig. 1A). The cDNA sequence includes 5' and 3' UTRs of 33 bases and ~400 bases, respectively, with the 3' UTR end terminating with a polyA tail (Fig. 1A). The open reading frame is 747 bases long and encodes a deduced protein of 248 amino acids. According to the SMART bioinformatics tool, the newly identified transcript encodes a deduced protein comprising an insulin-like-binding domain (IB), a kazal-type serine protease inhibitor domain (KI) and an immunoglobulin c-2 domain (IGc2), in that order (Fig. 1B). Based on the presence of these domains, the gene product is classified as an insulin-like growth factor-binding protein (IGFBP) homolog of the crayfish and was termed *C. quadricarinatus* IGFBP (Cq-IGFBP).

Assigned to the IGFBP family of peptides, Cq-IGFBP was considered in a multiple alignment with other IGFBPs sharing a similar domain organization (Fig. 1B) using the Clustal W algorithm (Supplementary Fig. S1). A phylogram, calculated from the obtained alignment (Fig. 2), emphasized the fairly limited number of invertebrate IGFBPs that have been fully sequenced to date and the relatively low sequence conservation level that they share. It

should be stressed that although these sequences had seemed to divert quite substantially one from another, they are nonetheless similarly organized, sharing the same domain-based architecture (Fig. 1B). By contrast, vertebrate IGFBP sequences share greater protein sequence similarity. Within vertebrates, fish and aves sequences were grouped together, while amphibian and reptile sequences were grouped together and found to be more similar to the mammalian group than to others. Cq-IGFBP showed the greatest sequence homology to that sequence from yet another crustacean, the copepod *Caligus rogercresseyi*. The sequence from a fellow arthropod, the lone star tick *Amblyomma americanum*, was found in the closest branch to that of the crustaceans, although the two branches are quite remote from one another.

As the linear organization of the three Cq-IGFBP signature domains was shared with selected IGFBPs (Fig. 2 and Supplemental Fig. S1), conservation at the level of 3D structure was next considered. Accordingly, two comparative homology models, one of the IB and KI domains (residues 28–133) (Fig. 3A) and the second of the IG domain (residues 143–248) (Fig. 3B), were built using SWISS-MODEL for identifying suitable templates. The template used for building the comparative model of the Cq-IGFBP IB and KI domains was the N-terminal domain of human HtrA1 (Fig. 3C) (sequence identity 31%; coverage approx. 95%, Supplemental Fig. S2), a protein previously shown to comprise an IGF binding domain which was structured in detail by X-ray crystallography (34). To generate a homology of the IG domain, the 5th IG-like domain from human Roundabout homolog 2 (Fig. 3D) (sequence identity 27%; coverage approx. 95%, Supplemental Fig. S2) was employed as template. The latter has been shown to be conserved in other invertebrate IGFBP-homologs (35,36), although its function with this respect is yet to be revealed. The final models of the IB and KI domains and of the IG domain were both found to resemble their reference structures with high accuracy and minimal structural variations, as indicated by 0.95 Å and 0.64 Å RMSD values, respectively. Other indicators of the accuracy of these models exist.

Previously, the Arg/Lys-X-Leu-X-X-Leu-Leu determinant was found to be crucial for insulin-like binding by human IGFBP-3 and IGFBP-5 (37,38). A highly similar determinant was identified within the IB domain of Cq-IGFBP (Arg₅₇-X₅₈-Leu₅₉-X₆₀-X₆₁-Leu₆₂). These residues are predicted to form a part of a β -sheet and a small loop, which are located in an exposed position (Fig. 3A, gold ribbon).

The tissue expression pattern of *Cq-IGFBP* was demonstrated by RT-PCR (Fig. 4). Although the primers used were devised for cDNA originally isolated from an AG SSH cDNA library, the transcript was shown to be amplified in all of the tissues considered, namely the AG, hepatopancreas, testes, muscle, sperm duct and green glands in mature male crayfish.

As a prerequisite for assessing Cq-IGFBP's ability to bind the Cq-IAG hormone, a recombinant version of the early (rCq-IGFBP) was produced using the pET28 expression system in *E. coli* cells. The 6xHis-tagged rCq-IGFBP protein had to be retrieved from inclusion bodies via solubilization prior to its purification via immobilized metal affinity chromatography. A Coomassie blue-stained SDS-PAGE gel showed that most of the proteins present in the solubilized inclusion bodies (Fig. 5A, BC) flowed through the nickel column, with minimal amounts of protein being removed from the resin by repetitive washing (Fig. 5A, FT-W₃). A single protein band of approx. 30 kDa was greatly enriched in the column eluate (Fig. 5A, E lanes). In a subsequent immuno-blot performed with anti-His antibodies, the purified protein was specifically labeled (Fig. 5B, arrowhead). Bovine serum albumin, which served as a negative control for the immuno-blot, was shown to be present in a replica gel stained with Coomassie blue (Fig. 5B, CB). The purified 30 kDa protein band (Fig. 5A, E₃) was excised from the gel and subjected to peptide analysis via tandem MS. A Sequest search performed against a database of the rCq-IGFBP protein that includes a His-tag and an adjacent thrombin cleavage site confirmed that the isolated band indeed corresponds to rCq-IGFBP. In these experiments, the generated tryptic fragments were shown to cover ~59% of the mature Cq-IGFBP sequence (i.e., lacking the signal peptide) (Fig. 5C).

Co-precipitation experiments were performed to test whether Cq-IGFBP specifically binds the insulin-like AG hormone (Cq-IAG). Biotinylated rCq-IGFBP (brCq-IGFBP) was incubated with AG homogenate, co-precipitated with streptavidin-conjugated agarose beads and resolved by SDS-PAGE. In the streptavidin-precipitated fraction, a band visualized using Coomassie stain (Fig. 6A, arrowhead) was excised from the gel and analyzed by MS. Such analysis of the tryptic digest of this band revealed two peptides corresponding to residues 48-59 of the Cq-IAG B-chain and residues 96-102 of the C-peptide (Fig. 6B₁). The scores calculated for these two peptides when doubly charged were 1.04 and 2.26, respectively. For the second peptide (FTVPDAR, corresponding to residues 96-102 of the C-peptide), the possible 'b' and 'y' ions are denoted (Fig. 6B₂) and an intensity map is provided (Fig. 6B₃).

The streptavidin-precipitated proteins were also resolved by SDS-PAGE and transferred onto nitrocellulose membranes for immuno-blot analysis. For this purpose, a specific antibody against recombinant proCq-IAG was previously raised (16). In the pull-down fraction of the immuno-blot, the antibodies specifically reacted with two peptide bands (Fig. 6C, arrowheads). The same peptides were also recognized in the AG homogenate, serving as a control (Fig. 6C, rightmost lane). The presence of the brCq-IGFBP in the pull-down fraction was verified by blotting using HRP-conjugated avidin (Fig. 6C, bottom). Using a ligand dot blot approach, the resolution of the IAG-IGFBP binding association was partially enhanced. It was shown that a portion of the Cq-IAG A-chain (residues 23-38) is recognized by recombinant IGFBP *in vitro* (Fig. 6D). The ability of brCq-IGFBP to bind other members of the insulin-like superfamily (e.g. human recombinant insulin and IGF-I) was also examined using streptavidin-based precipitation and SDS-PAGE separation, as above. The specificity of Cq-IGFBP for the IAG hormone was further established, as it completely did not bind human insulin, as it was present in the 'not bound' fraction and absent from the 'pulled-down' (PD) fraction, even though the gel was visualized by silver stain (Fig. 6E). A fairly low

amount of IGF-I might have in fact been bound to Cq-IGFBP, as indicated by a faint band in the PD lane (Fig. 6E).

DISCUSSION

The findings reported here solve another piece in the puzzle of what seems to be a complicated scheme involving a unique insulin-like peptide with clear androgenic activities. Indeed, these results contribute to an emerging body of evidence revealing the involvement of ILP signaling in primary and/or secondary sexual differentiation phenomena in the animal kingdom from the relatively early appearance of invertebrates to the later evolution of vertebrates (10,11,13,15,39,40).

In recent years, thanks in part to the giant leaps made in sequencing technology, many new invertebrate ILP homologs have been identified (41-45), including the crustacean insulin-like AG hormone (15). In the past, growth-related insulin-like peptides (namely IGFs) were thought to be exclusive to vertebrates. However, it is become ever clearer that this insulin-like growth-related family of peptides originated far earlier in the evolution of animals. Based on our knowledge and comparative studies showing that IGFs are secreted into the circulatory system where they interact with highly specific IGFBPs, it is less surprising that such binding proteins have been thus far identified in several invertebrate species, including crustaceans. Thus, our findings present exciting evidence adding new perspective to studies on the early evolution of the IGFBP family, which currently focus mainly on vertebrates (46). It should be emphasized that previous studies in decapods have linked these binding proteins to immune responses (22,23), rather than to reproduction and growth. With the present findings, Cq-IGFBP joins the list of other arthropod IGFBP homologs, namely Imp-L2 and *Sf*-IBP, previously reported in the fruit fly *Drosophila melanogaster* and the moth *Spodoptera frugiperda*, respectively (35,47). The insight gained here thus further supports the common incidence of insulin-like signaling, specifically in invertebrates and generally, in animalia. Nevertheless, in contrast to Cq-IGFBP, Imp-L2 was shown to bind various *Drosophila* ILPs with no gender specificity (35,48,49), while *Sf*-IBP was shown to bind various exogenous

human ILPs (47). Thus, crustacean IAG-signaling present a novel twist on these interactions, in that an IGFBP is suggested to serve as a pivotal component able to discriminatively bind primarily an endogenous male ILP sex-hormone. Cq-IAG has been suggested to have pleiotropic effects (16,20,21,50) and, therefore, must induce signaling in cells of various origin (e.g., muscle, testis and epithelium). It is noteworthy that the tissue expression pattern of *Cq-IGFBP* differs from that of *Cq-IAG* as previously described (14). The spatial profiling of *Cq-IGFBP*, revealing the wide expression of this transcript, supports the hypothesis that IAG-signaling is modulated by the expression/production of Cq-IGFBP in multiple tissues.

To date, a native proteinaceous IAG hormone was isolated in isopod but not in decapod crustaceans (51,52). In the latter, where IAG mRNA sequences were obtained, bioinformatics predict sites for C-peptide cleavage, although this has never been demonstrated. In addition, the deduced IAG peptide does not fall under the classic definition of an IGF, given that it lacks D and E domains and is suggested to undergo C-peptide removal, and is instead considered an ILP. As previously discussed, IGFs are secreted as a single-chained peptide (1) and transported in complex with IGFBP family members in the blood. We have shown that regardless of whether proteolysis occurs, the endogenous proCq-IAG hormone, which has been identified in this study for the first time, is specifically recognized by Cq-IGFBP, based on MS and immunological studies. These results thus join the observation made in *S. frugiperda* (47), and show that arthropod-IGFBPs are capable of binding proinsulin-like hormones. Moreover, our findings support those of Yamanaka *et al.* (53) arguing the common dogma suggesting that only IGFs interact with high affinity soluble binding proteins (8). As such, the findings reported here shed new light on the early development and co-evolution of the IGFBP superfamily and its diverse ligands.

As for the modeling of the IGF-IGFBP interaction, it was previously reported that such binding is primarily mediated via the B and A chains, with conflicting evidence regarding the

involvement of the C-peptide having been shown (54). Our usage of a recombinant Cq-IAG-based peptide has provided new details on the specific sites used for hormone recognition by crayfish IGFBP. It is evident that residues 23-38 of the IAG A-chain are sufficient for binding by Cq-IGFBP *in vitro*. As for the sites of interaction in IGFBP, numerous studies have shown the involvement of the N-terminal region in ligand binding, as recently reviewed (55). Specifically, site-direct mutagenesis efforts have shown the importance of the Arg/Lys-X-Leu-X-X-Leu-Leu motif for such binding (37,38). Cq-IGFBP provides another example of a family member bearing this sequence (excluding the final leucine). Based on homology modeling, the N-terminal domain of Cq-IGFBP encompassing the IB domain shows the highest structural homology to human HtrAI. Although the structure of HtrAI was solved in detail and found to present an IB domain, it was shown not to bind either human IGF-I or IGF-II (34). Thus, although both proteins possess structurally similar IB domains, only Cq-IGFBP is functional, i.e., capable of binding ILP. This can be rationalized by the presence of the conserved determinant Arg/Lys-X-Leu-X-X-Leu-Leu (37,38) that is absent in HtrAI. Furthermore, previous studies have clearly shown the necessity of the IGFBP C-terminal domain for ligand binding (55-57), which may also hold true for Cq-IGFBP.

In conclusion, ILP signaling is found at the core of sexual differentiation in various organisms. Cq-IGFBP, however, is the first family member involved in this process shown to specifically interact with a gender-specific androgenic ILP. It is striking that *Crustacea*, reaching back to the Cambrian (~500 million years ago) (58), present IGFbps that share high sequence similarity and physiological roles with homologous proteins across evolution (despite the low sequence conservation of ILP ligands). This suggests that parallel situations yet to be identified exist in other fundamental life forms.

Acknowledgments—We thank Mr. Omri Lapidot for his technical assistance with crayfish maintenance. We would also like to thank Dr. Raz Zarivach and Ms. Natalie Zeytuni for their assistance with protein modeling. This research was supported by Research Grant No. QB-9308-06 from BARD of the United States – Israel Binational Agricultural Research and Development Fund.

References

1. Chan, S. J., and Steiner, D. F. (2000) *Am. Zool.* **40**, 213-222
2. Steiner, D. F. (1978) *Diabetes* **27 Suppl 1**, 145-148
3. Pavelic, J., Matijevic, T., and Knezevic, J. (2007) *Indian J. Med. Res.* **125**, 511-522
4. Capoluongo, E. (2011) *Am. J. Pathol.* **178**, 26-31
5. Gallagher, E. J., and LeRoith, D. (2011) *Endocrinology* **152**, 2546-2551
6. Laron, Z. (2001) *Mol. Pathol.* **54**, 311-316
7. Hwa, V., Oh, Y., and Rosenfeld, R. G. (1999) *Endocr. Rev.* **20**, 761-787
8. Kelley, K. M., Schmidt, K. E., Berg, L., Sak, K., Galima, M. M., Gillespie, C., Balogh, L., Hawayek, A., Reyes, J. A., and Jamison, M. (2002) *J. Endocrinol.* **175**, 3-18
9. Baxter, R. C. (2000) *Am. J. Physiol. Endocrinol. Metab.* **278**, 967-976
10. Emlen, D. J., Warren, I. A., Johns, A., Dworkin, I., and Lavine, L. C. (2012) *Science* **337**, 860-864
11. Fedina, T. Y., Kuo, T. H., Dreisewerd, K., Dierick, H. A., Yew, J. Y., and Pletcher, S. D. (2012) *PLoS ONE* **7**, e49799
12. Kawamura, K., Kumagai, J., Sudo, S., Chun, S. Y., Pisarska, M., Morita, H., Toppari, J., Fu, P., Wade, J. D., Bathgate, R. A., and Hsueh, A. J. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7323-7328
13. Nef, S., Verma-Kurvari, S., Merenmies, J., Vassalli, J. D., Efstratiadis, A., Accili, D., and Parada, L. F. (2003) *Nature* **426**, 291-295
14. Manor, R., Weil, S., Oren, S., Glazer, L., Aflalo, E. D., Ventura, T., Chalifa-Caspi, V., Lapidot, M., and Sagi, A. (2007) *Gen. Comp. Endocrinol.* **150**, 326-336
15. Ventura, T., Rosen, O., and Sagi, A. (2011) *Gen. Comp. Endocrinol.* **173**, 381-388
16. Rosen, O., Manor, R., Weil, S., Gafni, O., Linial, A., Aflalo, E. D., Ventura, T., and Sagi, A. (2010) *PLoS ONE* **5**, e15281
17. Ventura, T., Manor, R., Aflalo, E. D., Weil, S., Rosen, O., and Sagi, A. (2012) *Biol. Reprod.* **86**, 90, 01-06
18. Barki, A., Karplus, I., Khalaila, I., Manor, R., and Sagi, A. (2003) *J. Exp. Biol.* **206**, 1791-1797
19. Barki, A., Karplus, I., Manor, R., and Sagi, A. (2006) *Horm. Behav.* **50**, 322-331
20. Manor, R., Aflalo, E. D., Segall, C., Weil, S., Azulay, D., Ventura, T., and Sagi, A. (2004) *Invertebr. Reprod. Dev.* **45**, 151-159
21. Ventura, T., Manor, R., Aflalo, E. D., Weil, S., Raviv, S., Glazer, L., and Sagi, A. (2009) *Endocrinology* **150**, 1278-1286
22. Castellanos, M., Jimenez-Vega, F., and Vargas-Albores, F. (2008) *Comp Biochem Phys D* **3**, 270-274
23. Gai, Y., Wang, L., Song, L., Zhao, J., Qiu, L., and Xing, K. (2010) *Fish Shellfish Immunol.* **28**, 10-17
24. Khalaila, I., Katz, T., Abdu, U., Yehezkel, G., and Sagi, A. (2001) *Gen. Comp. Endocrinol.* **121**, 242-249
25. Khalaila, I., Manor, R., Weil, S., Granot, Y., Keller, R., and Sagi, A. (2002) *Gen. Comp. Endocrinol.* **127**, 147-156
26. Arnott, D., O'Connell, K. L., King, K. L., and Stults, J. T. (1998) *Anal. Biochem.* **258**, 1-18
27. Roth, Z., Parnes, S., Wiel, S., Sagi, A., Zmora, N., Chung, J. S., and Khalaila, I. (2010) *Glycoconj. J.* **27**, 159-169

28. Glazer, L., Shechter, A., Tom, M., Yudkovski, Y., Weil, S., Aflalo, E. D., Pamuru, R. R., Khalaila, I., Bentov, S., Berman, A., and Sagi, A. (2010) *J. Biol. Chem.* **285**, 12831-12839
29. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) *Mol. Biol. Evol.* **28**, 2731-2739
30. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406-425
31. Felsenstein, J. (1985) *Evolution* **39**, 783-791
32. Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) *Bioinformatics* **18**, 1250-1256
33. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714-2723
34. Eigenbrot, C., Ultsch, M., Lipari, M. T., Moran, P., Lin, S. J., Ganesan, R., Quan, C., Tom, J., Sandoval, W., van Lookeren Campagne, M., and Kirchhofer, D. (2012) *Structure* **20**, 1040-1050
35. Honegger, B., Galic, M., Kohler, K., Wittwer, F., Brogiolo, W., Hafen, E., and Stocker, H. (2008) *J. Biol.* **7**, 10
36. Li, N., Zhang, Z., Zhang, L., Wang, S., Zou, Z., Wang, G., and Wang, Y. (2012) *Fish Shellfish Immunol.* **33**, 229-242
37. Hong, J., Zhang, G., Dong, F., and Rechler, M. M. (2002) *J. Biol. Chem.* **277**, 10489-10497
38. Imai, Y., Moralez, A., Andag, U., Clarke, J. B., Busby, W. H., Jr., and Clemmons, D. R. (2000) *J. Biol. Chem.* **275**, 18188-18194
39. Cheng, L. Y., Bailey, A. P., Leever, S. J., Ragan, T. J., Driscoll, P. C., and Gould, A. P. (2011) *Cell* **146**, 435-447
40. Holzenberger, M., Jarvis, E. D., Chong, C., Grossman, M., Nottebohm, F., and Scharff, C. (1997) *J. Neurosci.* **17**, 6974-6987
41. Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001) *Curr. Biol.* **11**, 213-221
42. Gulia-Nuss, M., Robertson, A. E., Brown, M. R., and Strand, M. R. (2011) *PLoS ONE* **6**, e20401
43. Okamoto, N., Yamanaka, N., Satake, H., Saegusa, H., Kataoka, H., and Mizoguchi, A. (2009) *FEBS J.* **276**, 1221-1232
44. Pierce, S. B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S. A., Buchman, A. R., Ferguson, K. C., Heller, J., Platt, D. M., Pasquinelli, A. A., Liu, L. X., Doberstein, S. K., and Ruvkun, G. (2001) *Genes Dev.* **15**, 672-686
45. Yan, X. C., Chen, Z. F., Sun, J., Matsumura, K., Wu, R. S., and Qian, P. Y. (2012) *PLoS ONE* **7**, e46513
46. Daza, D. O., Sundstrom, G., Bergqvist, C. A., Duan, C., and Larhammar, D. (2011) *Endocrinology* **152**, 2278-2289
47. Sloth Andersen, A., Hertz Hansen, P., Schaffer, L., and Kristensen, C. (2000) *J. Biol. Chem.* **275**, 16948-16953
48. Alic, N., Hoddinott, M. P., Vinti, G., and Partridge, L. (2011) *Aging Cell* **10**, 137-147
49. Okamoto, N., Nakamori, R., Murai, T., Yamauchi, Y., Masuda, A., and Nishimura, T. (2013) *Genes Dev.* **27**, 87-97
50. Sagi, A., Manor, R., Segall, C., Davis, C., and Khalaila, I. (2002) *Invertebr. Reprod. Dev.* **41**, 27-33
51. Martin, G., Sorokine, O., Moniatte, M., Bulet, P., Hetru, C., and Van Dorsselaer, A. (1999) *Eur. J. Biochem.* **262**, 727-736
52. Okuno, A., Hasegawa, Y., Nishiyama, M., Ohira, T., Ko, R., Kurihara, M., Matsumoto, S., and Nagasawa, H. (2002) *Peptides* **23**, 567-572
53. Yamanaka, Y., Wilson, E. M., Rosenfeld, R. G., and Oh, Y. (1997) *J. Biol. Chem.* **272**, 30729-30734

54. Denley, A., Cosgrove, L. J., Booker, G. W., Wallace, J. C., and Forbes, B. E. (2005) *Cytokine Growth Factor Rev.* **16**, 421-439
55. Forbes, B. E., McCarthy, P., and Norton, R. S. (2012) *Front. Endocrinol.* **3 (38)**, 1-13
56. Bach, L. A. (2004) *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists* **25**, 155-164
57. Firth, S. M., and Baxter, R. C. (2002) *Endocr. Rev.* **23**, 824-854
58. Walossek, D. (1998) On the Cambrian diversity of Crustacea. in: *Crustaceans and the biodiversity crisis* (Schram, F. R., and von Vaupel Klein, J. C. eds.), BRILL

Figure Legends

Figure 1. Complete cDNA and protein sequences of Cq-IGFBP. (A) The full-length cDNA (lower case) sequence comprises 1182 nucleotides (nt), including a 5'-UTR (nt 1–33) and a 3'-UTR (nt 780–1182). The ORF encodes a deduced protein of 248 amino acids (aa) (capital letters). The putative signal peptide (bold), insulin-like-binding domain (black open box, IB), kazal-type serine protease inhibitor domain (underlined, KI) and immunoglobulin c-2 type domain (gray-colored box, IGc2) are indicated. The stop codon is indicated by an asterisk. (B) Domain architecture organization of Cq-IGFBP (from the N-terminus to the C-terminus) as predicted by SMART (<http://smart.embl-heidelberg.de/>), showing the signal peptide (black box) and the IB, KI and IGc2 domains. Scale bar represents number of amino acids.

Figure 2. Evolutionary relationship of proteins across taxa having the same order of domains as Cq-IGFBP. Based on a CLUSTAL W algorithm, insulin-like-binding proteins having the same domains organization as Cq-IGFBP, from a wide array of organisms, are mapped on a phylogenetic tree (constructed using MEGA, version 5.0 (29)). This evolutionary history was inferred using the Neighbor-Joining method and a bootstrap consensus tree of 5000 replicates. The branch representing Cq-IGFBP is indicated (arrow) and the bar represents the number of amino acid substitutions per site.

Figure 3. 3D modeling of the three domains of Cq-IGFBP. Models of the Cq-IGFBP insulin-like-binding (IB), kazal-type serine protease inhibitor (KI) and immunoglobulin c-2 type (IG) domains were designed based on their relative similarities to human references, and are presented in a ribbon diagram. (A) The IB (blue, residues 28-102) and KI (purple, residues 103-133) domains. The R₅₇-X₅₈-L₅₉-X₆₀-X₆₁-L₆₂ (gold) sequence is the putative equivalent of the conserved determinant [R/K-X-L-X-X-L-L] found to be crucial for insulin-like peptide-binding in human IGFBPs. (B) The IG domain (residues 143-248). (C) The solved X-ray structure of human HtrA1 (34). (D) The solved solution NMR structure of the 5th IG-like domain from human Roundabout homolog 2.

Figure 4. Spatial expression of Cq-IGFBP. RT-PCR was performed on various tissues from male crayfish. *C. quadricarinatus 18S rRNA* (*Cq-18S*, Accession #AF235966.1) served as a positive control and doubly distilled water served as a negative control (NC). Androgenic gland: AG, hepatopancreas: Hp, testis: Ts, muscle: Ms, sperm duct: SD, green gland: GG.

Figure 5. Purification of recombinant Cq-IGFBP. (A) Mature recombinant Cq-IGFBP (rIBP) was retrieved from solubilized inclusion bodies subjected to a nickel-based chromatography. Fractions were separated on SDS-PAGE and stained with Coomassie blue (CB). (B) Purified rIBP and BSA separated by SDS-PAGE were either stained with CB or blotted with a rabbit anti-His IgG (IB). (C) The identity of rIBP was validated using mass spectrometry. On the full sequence of rIBP (positions are indicated), validated residues are highlighted in dark gray (high identification score) and light gray (medium score). The His-tag and thrombin sequences derived from the expression vector are underlined. MW: molecular weight standard, BC: Before column, FT: flow through, W: wash (1- 3 M Urea, 2- 1.5 M Urea, 3- No Urea), E: elution.

Figure 6. Cq-IGFBP specifically binds the Cq-IAG hormone. (A) Biotinylated recombinant Cq-IGFBP (brCq-IGFBP) was incubated with AG homogenate followed by protein precipitation. Fractions were separated on SDS-PAGE and stained with Coomassie blue. A band corresponding to a suspected Cq-IGFBP ligand is indicated (arrowhead). Streptavidin precipitation: AP, un-treated homogenate: Ho. (B) Identification of Cq-IAG by mass spectrometry. The suspected band (arrowhead in Fig. 6A) was shown to include two peptides (underlined) shown to be within the Cq-IAG prohormone sequence (B₁). One of the two peptides is shown as an example, with its relevant 'b' and 'y' ions indicated (B₂), as is an intensity map (B₃). (C) The interaction of brCq-IGFBP and Cq-IAG was examined using streptavidin

precipitation (P: Streptavidin) followed by immuno-blot (IB) using the anti-recombinant Cq-IAG IgG which specifically recognized the two protein bands (arrowheads). AG homogenate and brCq-IGFBP before precipitation (Bf), not bound (NB) and pulled down (PD). Hemocyanin and recombinant proCq-IAG served as negative (NC) and positive (PC) controls for the immuno-blot, respectively. The AG homogenate was also assayed (Ho). The presence of the ~30 kDa brCq-IGFBP is indicated in an avidin blot (AB). (D) Recombinant Cq-IGFBP (rIBP) was spotted onto a membrane, incubated with the peptide A ligand (residues 23-38 of Cq-IAG A-chain), and blotted with anti-peptide A antibodies (pA-LB). Recombinant GAP35 and peptide A served as negative (NC) and positive (PC) controls, respectively. Positive signals (PS) obtained using either anti-His antibodies or Ponceau stain for rIBP and the NC, respectively, validated the presence of spotted proteins. (E) The ability of rCq-IGFBP to bind other insulin-like peptides was also examined. Ten μ g of either human recombinant insulin or IGF-I were incubated with brCq-IGFBP, streptavidin-precipitated and resolved on SDS-PAGE. Before precipitation (Bf), not bound (NB) and pulled down (PD). The gel was silver-stained.

Table 1: Sequences used for phylogenetic analysis

| Organism | Common name | Accession no. |
|--------------------------------------|---------------------------|----------------------|
| <i>Amblyomma americanum</i> | Lone star tick | ADE06668.1 |
| <i>Xenopus laevis</i> | African clawed frog | ABF71729.1 |
| <i>Branchiostoma floridae</i> | Florida lancelet | XP_002608864.1 |
| <i>Rattus norvegicus</i> | Common rat | NP_001102442.1 |
| <i>Xenopus (Silurana) tropicalis</i> | Western clawed frog | XP_002940304.1 |
| <i>Takifugu rubripes</i> | Puffer fish | XP_003979427 |
| <i>Homo sapiens</i> | Human | NP_001007564.1 |
| <i>Anolis carolinensis</i> | Green anole | XP_003223380.1 |
| <i>Salmo salar</i> | Atlantic salmon | ACN11329.1 |
| <i>Sus scrofa</i> | Wild boar | XP_001924387 |
| <i>Bos taurus</i> | Bovine | NP_001092631.1 |
| <i>Rana catesbeiana</i> | American bullfrog | ACO51872.1 |
| <i>Branchiostoma belcheri</i> | Lancelet | BAB97382.1 |
| <i>Mus musculus</i> | House mouse | NP_061211.1 |
| <i>Tetraodon nigroviridis</i> | Spotted green puffer fish | CAF93099.1 |
| <i>Oncorhynchus mykiss</i> | Rainbow trout | NP_001118120.1 |
| <i>Caligus rogercresseyi</i> | Sea louse | ACO10479.1 |
| <i>Cherax quadricarinatus</i> | Red claw crayfish | KC952011 |
| <i>Gorilla gorilla gorilla</i> | Western lowland gorilla | XP_004048102.1 |
| <i>Papio anubis</i> | Olive baboon | XP_003911525.1 |
| <i>Meleagris gallopavo</i> | Wild turkey | XP_003205660.1 |
| <i>Gallus gallus</i> | Chicken | XP_420577.3 |
| <i>Biomphalaria glabrata</i> | Freshwater snail | ABL74453.1 |

Figure 1

A

```

nt 1  acgcggggcagtggtggaacacagtaaacatagtcacggcgaggacgtggtctaagctgtt
      M A R T V V Y A V aa 9
nt 61  tcggccgcctactagtagctgtggctgcctcagagatcgtcacatgtggagaatgt
      S A A L L V A V A A S Q S I V T C G E C aa 29
nt 121  gataggtcgaaatgtgcagttacagcctcgtgccgggtgggtttagtgctggaacttgc
      D R S K C A V T A S C P G G L V L D T C aa 49
nt 181  ggatgctgcagaggttggctcggggcctgggtcagctctgtgagaaaccaccactgag
      G C C E V C A R G L G Q L C E N T T T E aa 69
nt 241  gacactcccacaaccacacattatgggtgtctgtggccaagtacctgtgtctgcacccgc
      D T P T T T H Y G V C G Q Y L V C R T R aa 89
nt 301  actgatactgggtgttacaggtgagccacatgcagatctgaggactctggggcagtagt
      T D T G G T G E A T C E C E D S G A V C aa 109
nt 361  ggcagtgatggagtcacttaaccacccctgtgtcacttgcataatggaactgcagagagg
      G S D G V T Y T T L C H L L M E T A E R aa 129
nt 421  cctgatctcagcgttgcctgtcagaggcccttgcaaacagccctgtgatataccaag
      P D L S V A V R G P C K T A P V I K S K aa 149
nt 481  ccagaagataagataaacaccacttggagcatctcgtgttggatgtgagcctgtcgtt
      P E D K I R P L G S I L V L D C E A V G aa 169
nt 541  taacctgtgcctgacatcacatgggaactgaaacgcccagatggatcctcgttcaaattg
      Y P V P D I T W E L N R P D G S S F K L aa 189
nt 601  cctgggtgagactctcatttgcctgttcaagttcgtgggttcctgaaaccttcatgatt
      P G D D S S F A V Q V R G G P E P Y M I aa 209
nt 661  actgggtgggttcagatcatgcctgttacaaaagaaatctcgttatcta caattgtgtg
      T G W V Q I M R V T N E N L G I Y N C V aa 229
nt 721  gctaccaattctgagggagagccctcgtgcttctgtttctctcgtcagagaaacttga
      A T N S E G S A S C F C F S R Y R E T * aa 248
nt 781  gacagggcacaatgtcagctgtaaacactgccatagccatataatttttttttttccc
nt 841  tacaatttaagaaaataattttataataaaatctcaaatcttctctctatgaaatgta
nt 901  ttcaaatatgtaattcctcctaaagcaaccatgccataatagcaacaaaaaaatgtgta
nt 961  tgtctctacatttgataattactataagatgcccttctttagaataatttagtgtgtgta
nt 1021 atgcaaaactggtcaagtcactaggtcaagatttactattacaatggtatcattat
nt 1081 tagtaaaaacgttaggttataatttgtttaaatgtcaataaagctcaagaattggtaaag
nt 1141 tctaagaattggaaaaaaataaaataaaataaaataaaataaaataaaataaaataaa

```

B

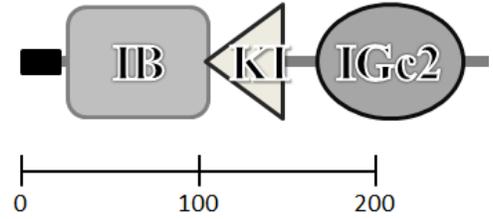


Figure 2

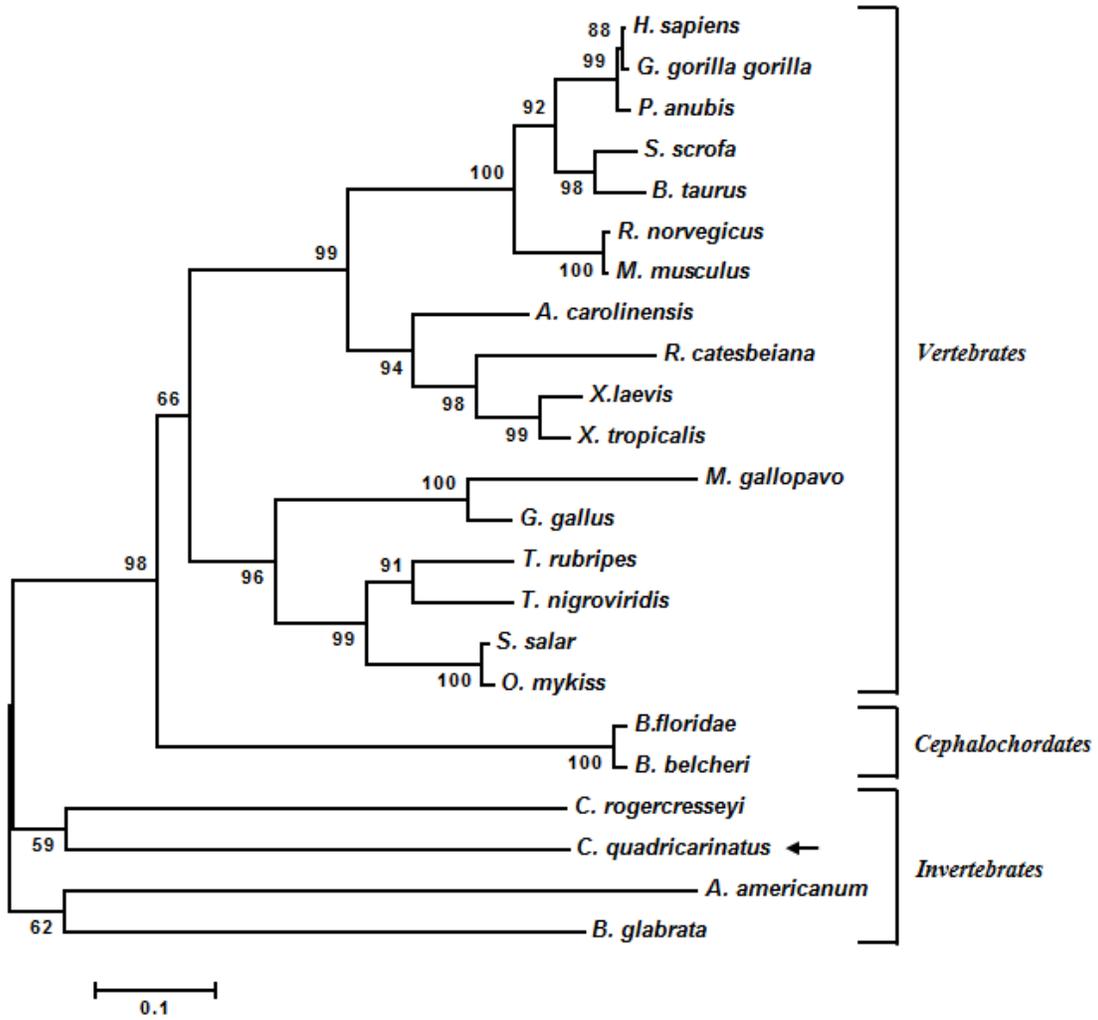
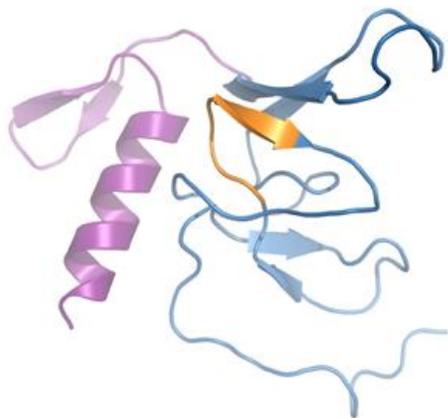
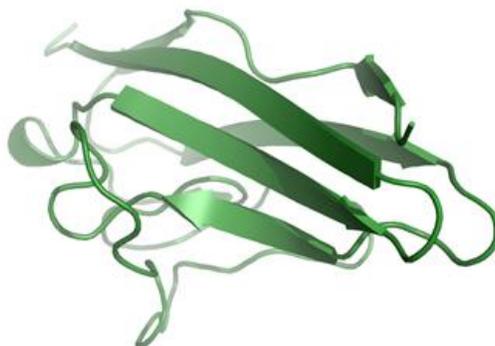


Figure 3

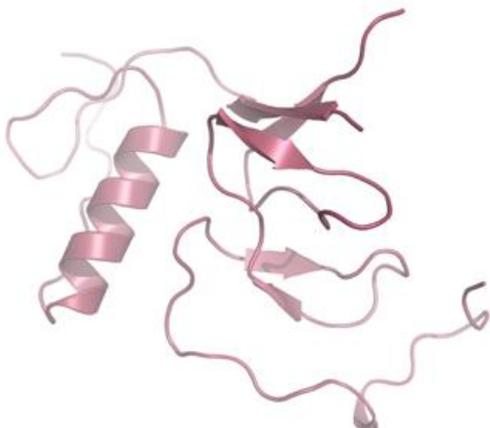
A



B



C



D

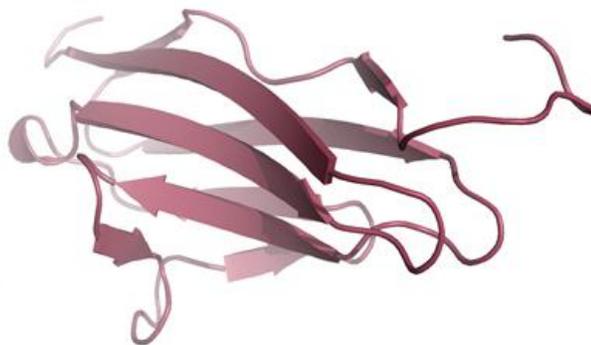


Figure 4

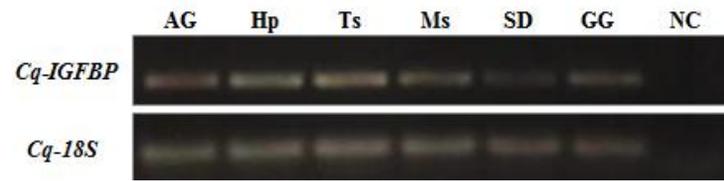


Figure 5

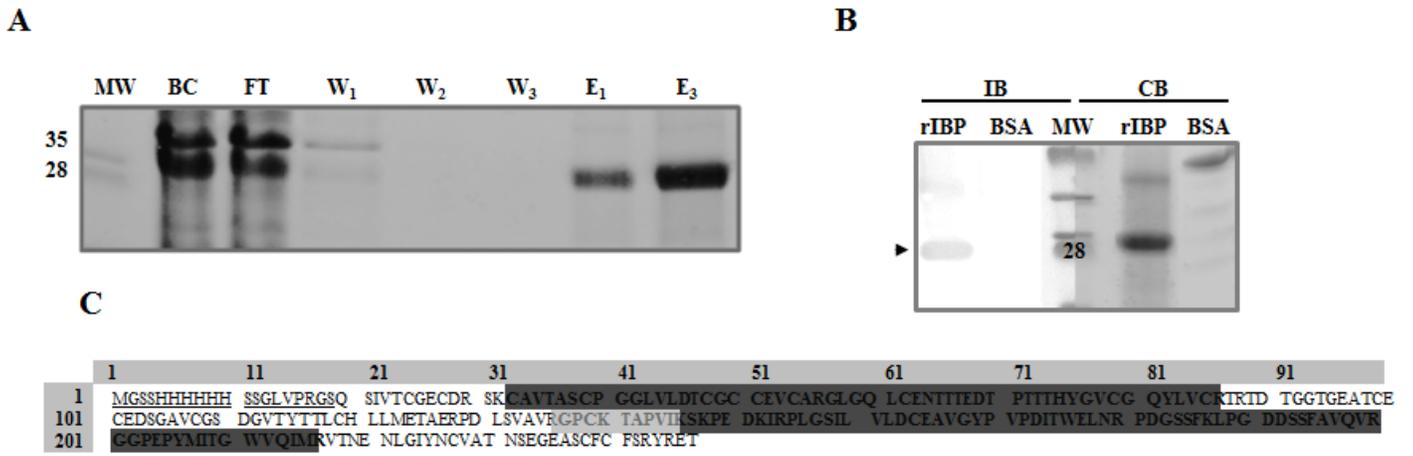


Figure 6

