BioOne COMPLETE

A NEW MEMBER OF THE GROWTH-PROMOTING GLYCOPROTEINS FROM DIAPREPES ROOT WEEVIL (COLEOPTERA: CURCULIONIDAE)

Authors: Huang, Zhipeng, Hunter, Wayne B., Cleland, Cathy A., Wolinsky, Murray, Lapointe, Stephen L., et al.

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

A NEW MEMBER OF THE GROWTH-PROMOTING GLYCOPROTEINS FROM DIAPREPES ROOT WEEVIL (COLEOPTERA: CURCULIONIDAE)

ZHIPENG HUANG[1], WAYNE B. HUNTER[2], CATHY A. CLELAND[3], MURRAY WOLINSKY[3],
STEPHEN L. LAPOINTE[2] AND CHARLES A. POWELL[4]

[1]Biotechnology Center, College of Life Science, Key Laboratory of Biopesticide and Chemical Biology
Fujian Agriculture & Forestry University, Fuzhou, Fujian 350002, P.R.China

[2]United States Department of Agriculture, Agricultural Research Service, U.S. Horticultural Research Laboratory
2001 South Rock Road, Fort Pierce, FL 34945, USA

[3]Los Alamos National Laboratory, Bioscience Division, P.O. Box 1663, MS M888, Los Alamos, NM 87545, USA

[4]University of Florida, Indian River Research and Education Center
2199 South Rock Road, Fort Pierce, FL 34945, USA

ABSTRACT

A new member belonging to the family of growth-promoting glycoproteins referred to as imaginal disc growth factors, IDGF, was identified from the root weevil *Diaprepes abbreviatus* (L.), (Coleoptera: Curculionidae). The imaginal disc growth factor full length cDNA transcript, designated as *idgf-DRW*, was cloned and identified from tissue of adult, teneral DRW females. Sequencing and subsequent homology comparisons of the nucleotide sequence (GenBank accession no. AY821658) indicated that the open reading frame (ORF) consisted of 1329 bases and encoded a putative protein of 442 amino acid residues with a calculated molecular weight of 49.5 kDa and a pI value of 6.68. BLASTX comparisons of the *idgf*-DRW cDNA sequence showed that the deduced amino acid sequence designated as IDGF-DRW (AAV68692.1) had 43% to 51% similarity with IDGF1-5 and DS47 from *Drosophila melanogaster*, *D. simulans*, and *D. yakuba*, 51% similarity with IDGF in *Pieris rapae*, and 51% similarity with an IDGF-like protein in *Bombyx mori*. Signal P analysis revealed that the predicted IDGF-DRW contained a signal peptide of 23 amino acid residues located at the N-terminus, similar to other known IDGF proteins. The structure of IDGF-DRW was predicted based on the characterized IDGF2 from *D. melanogaster* as the model. The deduced amino acid sequence for the IDGF-DRW protein had 48% similarity with *Drosophila melanogaster* IDGF2. The predicted IDGF-DRW displays the characteristics found in *Drosophila* IDGFs, the fold of family 18 glycosyl hydrolases, with an insertion (Gly304 to Phe392) in the beta barrel between strand β7 and helix alpha-7 that forms an additional α+β domain similar to that of *Serratia marcescens* chitinases A and B. An identified nucleotide change which results in an amino acid change within the active binding site in IDGF-DRW also was observed. The significant similarities of IDGF-DRW to other members within the family of IDGFs support its classification as a new member of the invertebrate growth factors and the first IDGF to be identified from a coleopteran.

Key Words: AAV68692, AY821658, cDNA, *Diaprepes abbreviatus*, gene expression, IDGF

RESUMEN

Identificamos a un nuevo miembro de la familia de glicoproteínas que promueven el crecimiento en invertebrados conocidas como factores de crecimiento de discos imaginales, IDGF (siglas en ingles), del gorgojo *Diaprepes abbreviatus* (L.) (Coleoptera: Curculionidae). El transcripto completo derivado del ADN del IDGF, identificado como idgf-DRW, fue clonado e identificado apartir del tejido de hembras adultas tenerales de *D. abbreviatus*. El secuenciamiento y subsecuente comparación de la homología a nivel de secuencia nucleotidíca (accesión AY821658 de GenBank) señalaron que el marco abierto de la lectura (ORF) consiste de 1329 bases y codifica para una proteína de 442 aminoácidos con un peso molecular calculado de 49.5 kDa y un valor pI de 6.68. Las comparaciones de BLASTP de la secuencia de cADN del idgf-DRW mostraron que la secuencia deducida de la proteina identificada como IDGF-DRW (AAV68692.1) tenía un 43% -51% de identidad con IDGF1-5 y DS47 de *Drosophila melanogaster, D. simulans* y *D. yakuba*, 51% de identidad con IDGF de *Pieris rapae*, y 51% de identidad con una proteina de *Bombyx mori* similar a IDGF. El análisis SignalP reveló que

Abbreviations: EST, expressed sequence tag; DS47, chitinase-like protein precursor; CHIT, chitinase; IDGF, imaginal disc growth factor protein; DRW, diaprepes root weevil; *idgf*, imaginal disc growth factor transcript.

el IDGF-DRW predicho contiene un peptido de señal de 23 aminoácidos situados en el N-término, similar a otras IDGF previamente caracterizadas. La estructura de IDGF-DRW fue predicha usando el IDGF2 de *D. melanogaster* como modelo. El IDGF-DRW predicho presenta el doblez característico de las glycosyl hydrolasas de la familia 18, con una inserción (Gly304 a Phe392) en el beta barrel entre la banda β7 y la hélice alpha-7 que forma un dominio adicional α+β similar al de las chitinases A y B de *Serratia marcescens*. También fue validado el cambio de un nucleótido que da lugar a un cambio de un aminoácido dentro del sitio activo de ligamiento de IDGF-DRW. Las significativas semejanzas de IDGF-DRW con otros miembros dentro de la familia IDGF apoyan su clasificación como un nuevo miembro de los factores de crecimiento de invertebrados. Este es el primer IDGF identificado en coleópteros.

Translation provided by the authors.

Information about the action of insect growth-regulating hormones is of interest for two reasons. First, the functions and cellular actions of invertebrate and vertebrate hormones have been shown to be remarkably conserved; thus, what we learn about insects may also enhance our understanding of hormonal regulatory processes in vertebrates. Second, by understanding how insect hormones function and interact in the regulation of insect development, we may be able to devise safe and specific agents to disrupt the insect life cycle, thus increasing the efficiency of efforts to manage agricultural pests and disease vectors. The pest system herein is the combination of *Diaprepes abbreviatus* (L.), *Diaprepes* root weevil, DRW, and *Phytophthora* which can cause severe tree decline and destroy groves within a few years (Graham et al. 1996, 2003). *Diaprepes* continues to be a major concern of citrus growers in Florida due to the difficulty of timely detection of larval infestations in the soil, and the paucity of effective management options (Lapointe 2000). To identify genes which regulate DRW development we constructed a gene expression cDNA library from DRW.

Little information on insect growth factors was available until 1996 when an insect growth factor was purified from the conditioned medium of an embryonic cell line, NIH-Sape-4, of *Sarcophaga peregrine* (flesh fly) (Homma et al. 1996). In 1999, a new developmental gene family of growth-promoting glycoproteins, referred to as the imaginal disc growth factors (*idgf*), was identified in *Drosophila melanogaster* (Kawamura et al. 1999). The IDGF proteins promote cell proliferation in the imaginal discs and were the first polypeptide growth factors to be reported from invertebrates. Throughout all developmental stages in variable patterns, IDGF proteins are expressed suggesting that *idgf* is an important gene having multiple functions during the entire life cycle (Kawamura et al. 1999). Though little is known about their precise mode of action, *idgf* genes appear to cooperate with insulin to stimulate the growth of imaginal disc cells. IDGF proteins are structurally related to chitinases from which they may have evolved, rather than to known growth factors, but have no known catalytic activity. Several nonenzymatic proteins with sequence homology to chitinases have been described in vertebrates indicating that the typical chitinase-like fold may be present in proteins with a wider range of biological functions other than chitin degradation (Hakala et al. 1993; Morrison & Leder 1994; Shackelton et al. 1995; Hu et al. 1996; Owhashi et al. 2000; Chang et al. 2001; Varela et al. 2002). The current *idgf* gene family in *Drosophila* is comprised of six members (*idgf1-5* and *ds47*) which encode proteins sharing 40-50% similarity to one another in amino acid sequence. Fortunately the IDGF2 member has had the crystal structure described from X-ray crystallography (Varela et al. 2002). Studies have shown that computational predictions of protein structures to elucidate functions and interactions have reached an acceptable level of accuracy, especially when there are related proteins that have been well characterized (Martelli et al. 2003; Aloy & Russell 2003; Bell & Ben-Tal 2003; Valencia 2003). So far, IDGF proteins have been reported from the lepidopterans *Bombyx mori* (Tsuzuki et al. 2001) and *Pieris rapae* (Asgari & Schmidt 2004), and the dipterans *Drosophila melanogaster*, *D. simulans*, and *D. yakuba* (Zurovcova & Ayala 2002). Rigorous evaluations of current hypotheses explaining growth promoting activation by IDGFs are still awaiting evidence from biochemical studies that will define the specificity of the ligand binding of these novel invertebrate growth factors; however, Varela et al. (2002) has postulated that the observed stimulation of imaginal disc cell proliferation by the cooperation of IDGFs and insulin may be a requirement to achieve optimal signaling of the insulin receptor. Invertebrate imaginal discs express an insulin receptor which is homologous to that of vertebrates (Garofalo & Rosen 1988; Fernandez et al. 1995; Ruan et al. 1995) and which is required for normal growth (Chen et al. 1996). An in-depth review is provided by Held (2002).

Herein we present the discovery of a coleopteran IDGF and its full length *idgf* transcript which was isolated and cloned from the *Diaprepes* root weevil, DRW. The genetic information from DRW will aid our understanding of the developmental and biological pathways important to the survival of DRW, and supports the development of

new management strategies against DRW by identifying genes critical in developmental pathways.

MATERIALS AND METHODS

Diaprepes Root Weevil Rearing and Collection

Larvae of DRW were obtained from a colony maintained at the U.S. Horticultural Research Laboratory (USHRL), Ft. Pierce, Florida. Individuals were reared in cups containing artificial diet at 26°C as described by Lapointe & Shapiro (1999). Callow teneral adults recently emerged from the pupal exuvium were selected and separated by gender for processing.

Library Construction

Seventeen whole teneral female DRW were used in the construction of an expression library. The insects were ground in liquid nitrogen and the total RNA extracted with guanidinium salt-phenol-chloroform procedure as previously described by Strommer et al. (1993). Poly(A)+RNA was purified with Micropoly(A) Pure™ according to the manufacturer's instructions (Ambion, Austin, TX, USA). A directional cDNA library was constructed in the Lambda Uni-ZAP® XR vector with Stratagene's ZAP-cDNA Synthesis Kit (Stratagene, CA, USA). The resulting DNA was packaged into Lambda particles with Gigapack® III Gold Packaging Extract (Stratagene, CA, USA). An amplified library was generated with a titer of 1.0×10^9 plaque-forming units per mL. Mass excision of the amplified library was accomplished by Ex-Assist® helper phage (Stratagene, CA, USA). An aliquot of the excised, amplified library was used for infecting XL1-Blue MRF' cells with subsequent plating on LB agar containing 100 µg/mL ampicillin. Bacterial clones containing excised pBluescript SK(+) phagemids were recovered by random colony selection.

Sequencing of Clones and Computer Analysis

pBluescript SK(+) phagemids were grown overnight at 37°C and 240 rpm in 96-well culture plates containing 1.7 mL of LB broth/well, supplemented with 100 µg/mL ampicillin. Archived stocks were prepared from the cell cultures with 75 µL of a LB-amp-glycerol mixture and 75 µL of cells. These archived stocks are held at the USHRL in an ultra low temperature freezer set at -80°C. Plasmid DNA was extracted by using the Qiagen 9600 liquid handling robot and the QIAprep 96 Turbo miniprep kit according to the recommended protocol (QIAGEN, Inc., Valencia, CA, USA). Sequencing reactions were performed with the ABI PRISM® BigDye™ Primer Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) along with a universal T3 primer. Reactions were prepared in 96-well format with the Biomek2000™ liquid handling robot (Beckman Coulter, Inc., USA). Sequencing reaction products were precipitated with 70% isopropanol, resuspended in 15 µL sterile water, and loaded onto an ABI 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Base confidence scores were calculated by TraceTuner® (Paracel, Pasadena, CA, USA). Low-quality bases (confidence score <20) were trimmed from both ends of sequences. Quality trimming, vector trimming, and sequence fragment alignments were executed by Sequencher® software (Gene Codes, Ann Arbor, MI, USA). cDNA sequences arising from rRNA and mitochondrial DNA were identified with BLASTN and were excluded from analysis along with sequences less than 200 nucleotides in length after both vector and quality trimming. Additional ESTs that corresponded to vector contaminants were removed from the dataset. To estimate the number of genes represented in the library and the redundancy of specific genes, ESTs were assembled into "contigs" by Sequencher®. Contig assembly parameters were set with a minimum overlap of 50 bases and 95% identity match.

Sequence Analysis and Structure Prediction

The *idgf* cDNA sequence was covered in eight overlapping clones and then bi-directionally sequenced five more times for sequence validation. The status of the gene sequence was determined based on TBLASTX homology searches by the National Center for Biotechnology Information BLAST server (http://www.ncbi.nlm.nih.gov) with the protein sequence comparisons made to protein databases (BLASTP). CLUSTALX was used for multiple alignments of amino acid sequences of the homologous proteins (Thompson et al. 1994), and an unrooted phylogenetic tree was constructed from this alignment with PAUP 4.0 and 1,000 bootstraps (Swofford 2002). The theoretical molecular weight and pI value of the predicted protein was calculated by ExPASy (http://au.expasy.org). SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP) was used to analyze the presence and location of potential signal peptide cleavage sites in the amino acid sequence. Structural modeling of the IDGF protein was performed with ROBETTA (http://www.robetta.com). ROBETTA is a full-chain protein structure prediction server which also performs domain parsing, 3-D modeling, fragment library generation, and protein-protein interaction studies by using an interface alanine scanning method. The programs utilize the Ginzu domain parsing and fold detection method developed by Dylan Chivian, David Kim, Lars Malmstrom, and David Baker, and the ROSETTA fragment insertion method (Simons et al. 1997). ROBETTA scans protein

chains to identify homologs with PDB-BLAST, FFAS03, 3D-Jury, and the Pfam-A protein family databases and assigns z-scores by MAMMOTH, a computer program that provides consistent protein model quality ranking by comparing modeled structures and their experimental counterparts (Ortiz et al. 2002). PSI-BLAST multiple sequence alignment then assigns regions of increased likelihood of including an adjoining domain with sequence clusters (Bowers et al. 2000; Rohl & Baker 2002). Loop regions are assembled from fragments and optimized to fit the aligned template structure. Protein domain predictions were identified by using the full prediction protocol (Kim et al. 2004). Graphic rendering of the predicted 3-D structures were done by the PDB file created by ROBETTA and Protein Explorer (http://www.molvis.sdsc.edu/protexpl/ frntdoor.htm).

A phylogenetic tree was generated by bootstrap neighbor joining, 1000×, PAUP 4.0, unrooted (Swofford 2002). Topology was similar when analyzed by maximum parsimony (branch-and-bound search method); data not shown. The phylogenetic tree of amino acid sequences was constructed with five members of *Drosophila* IDGF, and a bacterial chitinase.

RESULTS

Isolation of IDGF-DRW

An initial 10,000 clones were sequenced from the 5' end. These sequences were trimmed of vector and low-quality sequence and filtered for minimum length (200 contiguous bp, Phred 20 quality score) producing the final set of 8,480 high-quality ESTs. These ESTs were assembled by the program Sequencher® (Gene Codes, Ann Arbor, MI 48105) to produce 5,508 contiguous sequences (contigs) with 1,240 ESTs remaining as singlets. Of the assembled sequences analyzed, one contig (representing eight ESTs) was similar to the known IDGF protein sequences based on the TBLASTX results. The nucleotide sequence of this contig contained an intact ORF and was bidirectionally sequenced for sequence validation, providing an average 5× coverage for 90% of the transcript sequence, which was registered in GenBank (accession no. AY821658). The mRNA transcript was designated as *idgf-DRW*. A second cDNA library produced from DRW adults caught in the field also identified the *idgf-DRW* transcript (Hunter, data not shown).

Sequence Analysis of IDGF-DRW

Analysis of the nucleotide sequence revealed an ORF in *idgf-DRW*, consisting of 1329 bases and encoding a putative protein of 442 amino acid residues (Fig. 1) with a predicted molecular weight of 49.5 kDa and a pI value of 6.68. TBLASTX comparison and multiple sequence alignment of the *idgf-DRW* cDNA sequence (Fig. 2) indicated that the deduced amino acid sequence for the IDGF-DRW protein (accession no. AAV68692) had 43% similarity with *Drosophila melanogaster* IDGF1 (accession no. AAC99417), 48% with *Drosophila melanogaster* IDGF2 (accession no. AAC99418), 43% with *Drosophila melanogaster* IDGF3 (accession no. AAC99419), 51% with *Drosophila melanogaster* IDGF4 (accession no. AAC99420), 44% with *Drosophila melanogaster* IDGF5 (accession no. AAF57703), 51% with *Drosophila melanogaster* DS47 (accession no. AAC48306), 51% with *Pieris rapae* IDGF (accession no. AAT36640), and 51% with *Bombyx mori* IDGF-like protein (accession no. BAB16695). SignalP analysis of the predicted IDGF-DRW showed the most likely signal sequence cleavage site was between Ser_{23} and Ala_{24} (Fig. 1, ●), which will generate a signal peptide of 23 amino acid residues and a predicted mature protein with a calculated molecular weight of 47.1 kDa and a pI value of 6.76. Presence of a signal peptide was common to all IDGF proteins described to date although the length of the signal peptide varied (Fig. 2; refs for variation in signal sequence length: Chou 2002; Martoglio & Dobberstein 1998). Additionally, a single consensus motif for N-linked glycosylation (Asn_{227}), as previously reported for DS47 (Asn_{233}), also was identified (Fig. 1, ●).

Phylogenetic Relationship Between IDGF-DRW and Other Members of the IDGF Family

Analysis of the sequence and predicted protein structure of IDGF-DRW strongly supports its classification as a new member of the growth-promoting glycoproteins IDGF family. The IDGF proteins, encoded by members of the *idgf* gene family, were the first soluble polypeptide growth factors to be reported from invertebrates. These proteins are structurally related to chitinases rather than to known growth factors, and were shown to possess no catalytic activity (Kirkpatrick et al. 1995; Kawamura et al. 1999). The tree topology revealed that the *Drosophila* IDGF1, IDGF2, and IDGF3 members form a clade separate from the other insect IDGF4, IDGF5, and chitinase (Zurovcova & Ayala 2002) (Fig. 3). Herein, a phylogenetic tree of the multiple amino acid sequence alignment of IDGF-DRW, *Drosophila* IDGF, *P. rapae* IDGF, and *B. mori* IDGF-like proteins is shown (Fig. 3). In pairwise comparisons, these IDGF proteins showed 38% to 79% similarity. The phylogenetic tree showed that the three protein ortholog groups (IDGF1, IDGF2, and IDGF3) were more similar to each other than to the other six.

Structure was predicted by using ROBETTA, which is a full-chain protein structure prediction server (Fig. 4). The view of conserved residues (in green space-fill) predicted to be essential to main-

```
AATTCGGCACGAGGATTGAC ATG GAGTGTGTAAAAATTGTCCTCTTGGCAATTTTTGCCTTGGCGAGTTTCACCG    75
                      M   E  C  V  K  I  V  L  L  A  I  F  A  L  A  S  F  T  G     19

GCAAGACAGAATCAGCAACAGATAGCAAATTGGTGTGTTACTATGATAGCAGAGCATATAATAGACCAGGAAATG       150
 K  T  E  S ●  A  T  D  S  K  L  V  C  Y  Y  D  S  R  A  Y  N  R  P  G  N  G       44

GTAAATTCGACATTCCGTTCTTGGAACCCGCTTTGCAATTCTGCACTCACTTAATTTACGGATATGCAGGAATCA      225
 K  F  D  I  P  F  L  E  P  A  L  Q  F  C  T  H  L  I  Y  G  Y  A  G  I  R        69

GAGAGGACAATTTCAAAATATCACCATTGAACGAACCCCTGGACATCAACAAACAAAACTATAGACATATCACTG      300
 E  D  N  F  K  I  S  P  L  N  E  P  L  D  I  N  K  Q  N  Y  R  H  I  T  D        94

ATTTGAAGAGGAAATACCCTGGCTTGAGAGTTCTTCTTTCCGTTGGAGGAAATAATGATGTTACCGGAGAAGGCA      375
 L  K  R  K  Y  P  G  L  R  V  L  L  S  V  G  G  N  N  D  V  T  G  E  G  S       119

GTGAAAAGAATTTGAAATATAGAACTTTGCTAGAATCTGTAGAAAGCAGGTTAGCCTTCGTCAACTCTGCTCACG      450
 E  K  N  L  K  Y  R  T  L  L  E  S  V  E  S  R  L  A  F  V  N  S  A  H  D       144

ATTTAGTGAAAAACTATGGCTTTGATGGACTTGATCTCTCTTGGGAATTCCCAGAAAATAAACCAAAGAAAATCC      525
 L  V  K  N  Y  G  F  D  G  L  D  L  S  W  E  F  P  E  N  K  P  K  K  I  R       169

GTAACGCAGTTTCTTCATGGTTTTCCAAAATCAAACATAAAATTGTTGGTGAATCAGTAGTTGATGAAAAAGCCG      600
 N  A  V  S  S  W  F  S  K  I  K  H  K  I  V  G  E  S  V  V  D  E  K  A  E       194

AAGAACATAAAGAACAATTTACTGCTTTAGTTCGAGAACTTAAGAACGTATTCAGACATGATGGACTTCTATTAA      675
 E  H  K  E  Q  F  T  A  L  V  R  E  L  K  N  V  F  R  H  D  G  L  L  L  T       219

CAGTCTCTGTATTACCAAATGTTAACAGTTCAGTATACTTTGACCCCCGTCAGCTTGCTCCAAATATTGATTTTG      750
 V  S  V  L  P  N  V  N ●  S  S  V  Y  F  D  P  R  Q  L  A  P  N  I  D  F  A     244

CTACACTTGAAGCTTTTGACTACAGAACTCCTCAACGCAATCCTAAAGAATTAGATTATGTCGCTCCACTTTACG      825
 T  L  E  A  F  D  Y  R  T  P  Q  R  N  P  K  E  L  D  Y  V  A  P  L  Y  E       269

AACTTTTGGACAGAAAAGTTGATGAAAATGCTGATTACCAAGTTAGATATTGGTTGGGAGGTGGTTTACCAGCCA      900
 L  L  D  R  K  V  D  E  N  A  D  Y  Q  V  R  Y  W  L  G  G  G  L  P  A  N       294

ATAAACTTATCCTTGGCATTCCAACCTACGGTCGCGCCTGGAAACTTAATGATGACTCTGGACTGACTGGCGTTC      975
 K  L  I  L  G  I  P  T  Y  G  R  A  W  K  L  N  D  D  S  G  L  T  G  V  P       319

CACCGCTACTTACAGATGGGGCTGCAGATCCTGGCCCTTACAGCAATGAAGCTGGGCTATTAAGCTATCCAGAAA     1050
 P  L  L  T  D  G  A  A  D  P  G  P  Y  S  N  E  A  G  L  L  S  Y  P  E  I      344

TTTGCAGTAAAATTGCCACTCCAAAGGAAATTCAAGCCGGATATCTTGGAAAGTTAAGGAAAACTAATGATCCAA     1125
 C  S  K  I  A  T  P  K  E  I  Q  A  G  Y  L  G  K  L  R  K  T  N  D  P  T      369

CAAAAAGATATGGTTCATATGCATATCGTCTACCCGACAGTAATGGTGAAAATGGTATTTGGGTAGGATTTGAAG     1200
 K  R  Y  G  S  Y  A  Y  R  L  P  D  S  N  G  E  N  G  I  W  V  G  F  E  D      394

ATCCTGACACCGTTGGAAATAAAGCTGCATATTCCAAGGCTAAAGGACTTGGAGGTATTGCTATCGTTGACCTAA     1275
 P  D  T  V  G  N  K  A  A  Y  S  K  A  K  G  L  G  G  I  A  I  V  D  L  T      419

CACTAGATGATTTTAGAGGAACATGTTCACAAGATCATTTCCCATTGTTAAGAGCTGCGAAGTTTAGATTA TAA A   1350
 L  D  D  F  R  G  T  C  S  Q  D  H  F  P  L  L  R  A  A  K  F  R  L  *          442

TACCCCATTTCAATTTTTTTTATGTTAAATGTTAATAATTTCATATTATATCA                           1403
```

Fig. 1. Nucleotide and deduced amino acid sequences of cDNA encoding IDGF-DRW. The deduced amino acid sequence is shown below the nucleotide sequence and numbered accordingly. The protein sequence is numbered starting from the first **Met** residue. The signal peptide cleavage site is indicated by ● after Ser_{23}. The signal peptide is shown by underline. A single N-linked glycosylation site is indicated by ● at Asn_{227}. The start codon is bolded and encodes the residue Met. The termination codon is indicated in bold and asterisk in the nucleotide and protein sequence, respectively.

tain the barrel folding are shown. Residues are Gly_{109}, Gly_{110}, Asp_{152}, Gly_{153}, Leu_{218}, Asp_{242}, Lys_{295}, Gly_{412}, and Asp_{421}. The residue at position 159 is Glu (in brown space-fill) which is commonly re- placed by Gln in other known IDGF proteins. There are two disulfide bridges, Cys_{31}-Cys_{58} and Cys_{345}-Cys_{427}, in the predicted IDGF-DRW structure which are conserved in all known IDGF fam-

```
                          ←β1→           ←α1→        ←β2→   β2→  ←α2→   ←

DRW     IDGF      MECVKI---VLLAIFALASFTGKTESATDSKLVCYYDSRAYNRPGNGKFDIPPLEPALQFCTHLIYGYAGIREDNFKISPLNEPLDINK--QN  89
D.m.    IDGF2     MKAW------IWFTFVACLFAAS---TEAASNLVCYYDSSSYTRBGLGKLLNPDLEIALQFCSHLVYGYAGLRGENLQAYSMNENLDIYK--HQ  84
D.m.    IDGF1     MRFQ-----LFYLLGL-LSVTS--LTHAASNLICYYDSNSYLRQGLAKMHTNELDLALQFCTHLVYGYAGLKSGTLELFSLN--VDLDM--FY  82
D.s.    IDGF1     MRFQ-----LFYILGL-LSVTS--LTQAANNLCYYDSTSYLRQGLAKMHTNELDLALQFCTHLVYGYAGLKSGTLELFSLN--VDLDM--FY  82
D.y.    IDGF1     MRFQ-----LCYLLGL-LSVTS--LSHAASNLICYYDSTSYLRQGLAKMHTHELDLALQFCTHLVYGYAGLKAGTLELFSLN--VDLDM--FY  82
D.m.    IDGF3     MTGS-----LWLSLALSLAVLAQFKVSAAPNLVCFYDSQGSQRQGLAQFSMIDIELALQFCTHLVYGYAGVNADNYEMQSINKRLDLEQ--RH  87
D.s.    IDGF3     MSGS-----LWLSLALSLAVLAQFKVSAAPNLVCFYDSQGSQRQGLAQFSITDIELALQFCTHLVYGYAGVNADNFEMQSINKRLDLEQ--RH  87
D.y.    IDGF3     MTGS-----LWLSLALSLAVLAQFKVSAAPNLVCFYDSQGFQRQGLAQFSMTDMELALQFCTHLVYGYAGVNADNYEMQSINKRLDLEQ--RH  87
D.m.    IDGF4     MKLYAL---FSLLVGSLAIGQISAAGS--HHLLCYYDGNSFVREGLSKLILTDLEPALQYCTHLVYGYAGINPSSNKLVSNNEKPDLDLGSSL  88
D.m.    IDGF5     MMWIQKNPPLGLLLCSFLAFFQSTYAEVG-KLVCFYDAQSFVREGPAQMSLAELEPALQFCNFLVYGYAGIDAVTYKIKSLDPSLTNDR--QH  91
D.m.    DS47      MIIKALAI-VSLCLASIQASKVGAPQLPKKHLVCYYDSASFVKEGLGKLVIDELEPALQFCDYLVYGYAGIERDSHKAVSLNQQLDLDLGKGL  92
P.r.    IDGF      MK-------IILTLAGVLALAVSTPTPAQSKVVCYYDSKSYVRESQARMLPLDLDPALSFCTHLVYGYAGIQPDTYKMVPLNENLDVDRSHAN  86
B.m.    IDGF-like MK-------LFIALVGLLALAKALPAVTHSKVLCYYDSRSYVRESQARMLPLDLDPALSFCTHLLYGYAVIQPDTYKLVSLNENLDIDRTHDN  86
                           *         * **             ** *  * ****
```

```
                  α2'→   ←β3→               ←αa→   ←   α3        →  ←β4→

DRW     IDGF      YRHITDLKRKYPGLRVLLSVGGNNDVTGEGS--EKNLKYRTLLESVE-SRLAFVNSAHDLVKNYGFDGLDLSWEFPENKPKKIRNAVSSWFSK 178
D.m.    IDGF2     FSEVTSLKRKYPHLKVLLSVGGDHDIDP-----DHPNKYIDLLEGEKVRQIGFIRSAYELVKTYGFDGLDLAYQFPKNKPRKVHGDLGLAWKS 171
D.m.    IDGF1     YKDITALRQKFPQLKILLSVGDRDVDE-----AHPNKYVELLEANRTAQQNFIDSSMILLKRNGFDGLDLAFQLPRNKPRKVHGSLGSYWKS 169
D.s.    IDGF1     YKDITALRQKFPQLKILLSVGDRDVDE-----AHPNKYVELLEANRTAQQNFIDSSMILLKRNGFDGLDLAFQLPRNKPRKVHGSLGTYWKS 169
D.y.    IDGF1     YKEITALRQKFPQLKILLSVGDRDVDE-----AHPNKYVELLEANRTFQQNFIDSSMILVKRNGFDGLDLAFQLPRNKPRKVHGSLGTYWKS 169
D.m.    IDGF3     LAQITSMKERYPHIKFLLSVGGDADTN------EGNQYIKLLESGQQGHRRFIESARDLVRRYNFDGLDLALQLPRNKPRKVHGDVGSAWKS 172
D.s.    IDGF3     LAQYTSMKERYPHIKFLLSVGGDADTN------EGNQYIKLLESGQQGHRRFIESARDLVRRYNFDGLDLALQLPRNKPRKVHGDVGSAWKS 172
D.y.    IDGF3     LAQVSSLKERYPHIKFLLSVGGDADTN------EGNQYIKLLESGQQGHRRFIESARDLVRRYNPDGLDLALQLPRNKPRKVHGDVGSAWKS 172
D.m.    IDGF4     PRQVTGLKRKYPALKVLLSVGGDKDTVDPEN-----NKYLTLLESSNARIPPINSAH SLVKTYGFDGLDLGWQFPKNKPKKVHGSIGKFWKG 175
D.m.    IDGF5     YRHITALRKKYPHVRFLLSVGGDRDVNSEGV--ADSDKYLRLLEQSE-HRKSFQASVLAELNNNGFDGIDLAWQFPKNRPKLQQGVFKRVWGS 180
D.m.    DS47      YRTVTRLKRKYPNVKILLSVGGDKDIELDKDAKELPNKYLELLESPT-GRTRFVNTVYSLVKTYGFDGLDVAWQFPKNKPKKVHSGIGNLWKG 184
P.r.    IDGF      YRAITNFKSKYPGLKVLLSVGGDADK-------EEEQKYNLLLESPQ-ARTAFVNSGMLLAEQHGFDGIDLAWQFARIKPKKVRSTWGSIWHG 171
B.m.    IDGF-like YRAITSLKAKYPGLTVLLSVGGDADT-------EEPEKYNLLLESQQ-ARTAFINSGVLLAEQYGFDGIDLAWQFPRVKPKKIRSTWGSLWHG 171
                        *  ******      *     *   **        *   **  *    *
```

```
                  ←      σ4       →  ←β5→ αb→ α5→ β6→

DRW     IDGF      IKHKIVGESVVDEKAEEHKEQFTALVRELKNVFRHD-GLLTVSVLPNVNSSVYFDPRQLAPNIDFATLEAFDYRTPQRNPKELDYVAPLYEL 270
D.m.    IDGF2     IKKLFTGDFIVDPHAALHKEQFTALVRDVKDSLRAD-GFLLSLTVLPNVNSTWYFDIPALNGLVDFVNLATFDFLTPARNPEEADYSAPIYHP 263
D.m.    IDGF1     FKKLFTGDFVVDPQAEQHKSQFTDLVGNIKNAFRSA-NLMLSLTVLPNVNSTWYFDVPKLHPQFDYINLAAFDFLTPLRNPEEADFTAPIFFQ 261
D.s.    IDGF1     FKKLFTGDFVVDPLAEQHKSQFTDLVGNIKNAFRSA-NLMLSLTVLPNVNSTWYFDVPKLHPQFDYINLAAFDFLTPLRNPEEADFTAPIFFQ 261
D.y.    IDGF1     FKKLFTGDFVVDPLAEQHKSQFTDLVGNLKNAFRSA-NLMLSLTVLPNVNSTWYFDVPKLHPQFEYINLAAFDFLTPVRNPEEADFTAPIFFQ 261
D.m.    IDGF3     FKKFFTGDFIVDTESETHKGQVTALIKDLSAALKQN-DLLLSLTVLPNVNSSWYYDAPSIAPSLDFINLGTFDFLTPQRNPEEADFSAPTYEA 264
D.s.    IDGF3     FKKFFTGDYIVDTESETHKGQVTALIKDLSAALKQN-DLLLSLTVLPNVNSSWYYDAPSIAPSLDFINLGTFDFLTPQRNPEEADFSAPTYEA 264
D.y.    IDGF3     FKKFFTGDFIVDTDSETHKGQVTALIKDLSAALKQN-DLLLSLTVLPNVNSSWYYDAPSIAPSLDFINLGTFDFLTPQRNPEEADFSAPTYEA 264
D.m.    IDGF4     FKKIFSGDHLVDEKAEEHKEAFTALVRELKNAFRPD-GYILGLSVLPNVNSSLFFDVPAIINNLDYVNLHTYDFQTPERNNEVADFPAPIYEL 267
D.m.    IDGF5     LR-GWFSSSSVDEKAEEHREGFATLLEELQSDLRRG-GQLLTVSMLPHVSAELFIDVPKVLSNVDFVNLGTYDFQTPERDPKVADLPTPLYAM 271
D.m.    DS47      FKKVFSGDSIVDEKSEEHKEQFTALLRDVKNAFRPD-NLLLSTTVLPNVNSSLFYDIPAVVNYLDFVNLGTFDFFTPQRNPEIADYAAPIYEL 276
P.r.    IDGF      IK-KTFGTTLVDEKAEHREGYTALVRELKQAMTLK-QMQLSISVLPNVNSSIYYDVPAIINLVDIVNLHAYDYYTAARNPKEADYSAPTYKP 262
B.m.    IDGF-like IK-KTFGTTPVDEKESEHREGFTALVRELKQALIHKPKMQLGVTVLPNVNSTIYHDVPAIINLVDYVNVGAYDYYTPTRNNKEADYTAPIYTP 263
                    **  **    *                 *  *       **  *     *
```

```
                  ←  α6    →    ←β7→←βa→                          βb→←αc→

DRW     IDGF      LD---RKVDENADYQVRYWL--GGGLPANKLILG--IPTYGRAWKLNDDSGLTGVPPLLTD-GAADPGPYSN-EAGLLSYPEICSKIATP-KE 353
D.m.    IDGF2     DGSKDRLAHLNADVQVEYWL--SQGFPSNKINLG--VATYGNAWKLTKDSGLEGVPVVPETSGPAPEGFQSQ-KPGLLSYAEICGKLSNPQNQ 351
D.m.    IDGF1     DE-ENRLPHLNVEFQINYWL--QNHCPGQKLNLG--IASYGRAWKLSKGSGLSGAPIVHETCGVAPGGIQIQSAEGLLSWPEICSKLSQNASA 349
D.s.    IDGF1     DE-QNRLPHLNVEFQVNYWL--QNHCPGQKLNLG--IASYGRAWKLSKESGLSGAPIVHETCGAAPGGIQIQSAEGLLSWPEICSKLSQNASA 349
D.y.    IDGF1     DE-QNRLPHLNVEFQVNYWL--QNNCPGQKLNLG--IASYGRAWKLSKGSGLSGAPIVQETCGAAPGGIQIQSADGLLSWPEICSKLSQNASA 349
D.m.    IDGF3     VG-QNRLGHYNLNFQMEHWL--IATYGRSWKMSKDSGDSGMPVVPSTQGPAPAGPQSK-QEGLLNWAEICSLMPNPSNT 351
D.s.    IDGF3     VG-QNRLGHYNLNFQMEHWL--LQRVPANKINIG--IATYGRTWKMSKDSGDSGMPVVSSTQGPAPAGPQSK-QEGLLNWAEICSLMPNPSNS 351
D.y.    IDGF3     VG-QNRLGHYNLNFQTEHWL--LQRVPANKINIG--IATYGRTWKMTKDSGDSGMPVVPSTQGPAPAGPQSK-KEGLLNWAEICSLMPNPGNT 351
D.m.    IDGF4     NE---RNPEFNVNYQVKYWT--GNRAPAAKINVG--VATYGRAWKLTKDSGLTGLPPVAEADGVAPAGTQTQ-IPGLLSWPEVCAKLPANQ 352
D.m.    IDGF5     YD---RDPSHNVQYQVQYWMMNQTSEISVHKLHVG--VTSYGRAWNMTRNSGITGYPPIPAANGAAPPGRQTV-TPGLLSWPEICDLLQ---QQ 355
D.m.    DS47      SE---RNPEFNVAAQVKYWL--RNNCPASKINVG--VATYGRPWKLTDDSGDTGVPPVKDVKDEAPVGGNTQ-VPGIYSWPEVCALLPNQNNA 361
P.r.    IDGF      QN---RNELLNVDAAISYWL--TGGAPSSKIVVG--IATYARTWKLDSDSEISGVPPLHTD-GPGEAGPYTK-TEGLLSYPEVCAKLINP-NH 344
B.m.    IDGF-like QN---RNPLQNADAAVTYWL--TSGAPSQKIVLSXRLRSYLETGX---DSEIAGVPPIHTD-GPGEAGPYVK-TEGLLSYPEVCGKLINP-NQ 344
                        *     *       *    *           **          *       *         * **
```

```
                  ←βc→     ←βd→  ←βe→   ←    α7    →  ←β8→           ←   α8→

DRW     IDGF      IQAGYLGKLRKTNDPTKRYGSYAYRLPDSNGENGIWVGFEDPDTVGNKAAYSKAKGLGGIAIVDLTLDDFRGTCS-QDHFPLLRAAKFRL-- 442
D.m.    IDGF2     FLKGNESPLRRVSDPTKRFGGIAYRPVDGQITEGIWVSYDDPDSASNKAAYARVKNLGGVALFDLSYDDFRGQCS-GDKYPILRAIKYRL-- 440
D.m.    IDGF1     QYRGELAPLRKVTDLTQKYGNYALRPADDNGDFGVWLSFDDPDFAGLKAVYAKGKGLGGIALFDLSYDDFRGLCT-GQKYPILRSIKYFMG- 439
D.s.    IDGF1     QYRGELAPLRKVTDLTQKYGNYALRPADENGDFGVWLSFDDPDFAGIKAVYAKSKGLGGIALFDLSYDDFRGLCT-GQKYPIVRSIKYFMG- 439
D.y.    IDGF1     QYRGEMAPLRKVTDLTQKYGNYALRPADDNGDFGVWLSFDDPDFAGIKAAYAKGKGLGGVAIFDLSYDDFRGLCT-GQKFPIVRSIKYFMG- 439
D.m.    IDGF3     NARGPNAPVKRVVDPTKRYGSYAFRAADENGDHGLWISYDDPDSASSKAMYARARNLGGVALFDLTQDDFRGQCT-NDRFPMLRAIKYRLL- 441
D.s.    IDGF3     NARGPNAPVKRVVDPTKRYGSYAFRAADENGDHGLWISYDDPDSASSKAMYARARNLGGVALFDLTQDDFRGQCT-GDRFPMLRAIKYRLL- 441
D.y.    IDGF3     NARGPSAPVKRVLDPTKRYGSYAFRAADENGDHGLWISYDDPDSASSKAMFARVRNLGGVALFDLTQDDFRGQCT-NDRFPMLRAIKYRLL- 441
D.m.    IDGF4     HLKGADGPLRKVGDPTKRFGSYAYPSADDSGENGVWVGYEDPDTAAIKAEYVKREGLGGIAVVDLSFDDFRGGCT-GHDKFPILRQVKSKL- 442
D.m.    IDGF5     PQDREVPHLRKVGDPTKRFGIYAYRAADDQGENGLWVGYEDPLTAAIKAGFVHAQGLGGVAFHDLSMDDFRGQCA-GEKFPILRSIKFKL-- 444
D.m.    DS47      YLKGANAPLIKVQDPAKRFGSYAYRAADKKGDNGIWVSFEDPDTAADKAGYVRTENLGGVALFDLSYDDFRGLCT-NEKYPILRAIKYRLTN 452
P.r.    IDGF      -QKGMRPHLRKVTDPSKRFGTYAFRLPDDNGDPGLWVSYEDPDTAGQKASYVKSKNLGGVSIVDLSMDDFRGLCT-GDKYPILRAAKYRL-- 433
B.m.    IDGF-like -QKGMRPHLRKVTDPSKRFGTYAFRLPDDNGEGGIWVSYEDPDTAGQKAAYVKSKNLGGVAIVDLSLDDFRGLCT-GDKYPILRAAKYRL-- 433
                       *  *    *            **  **     **  ***   *  *  ***** *  * * *
```
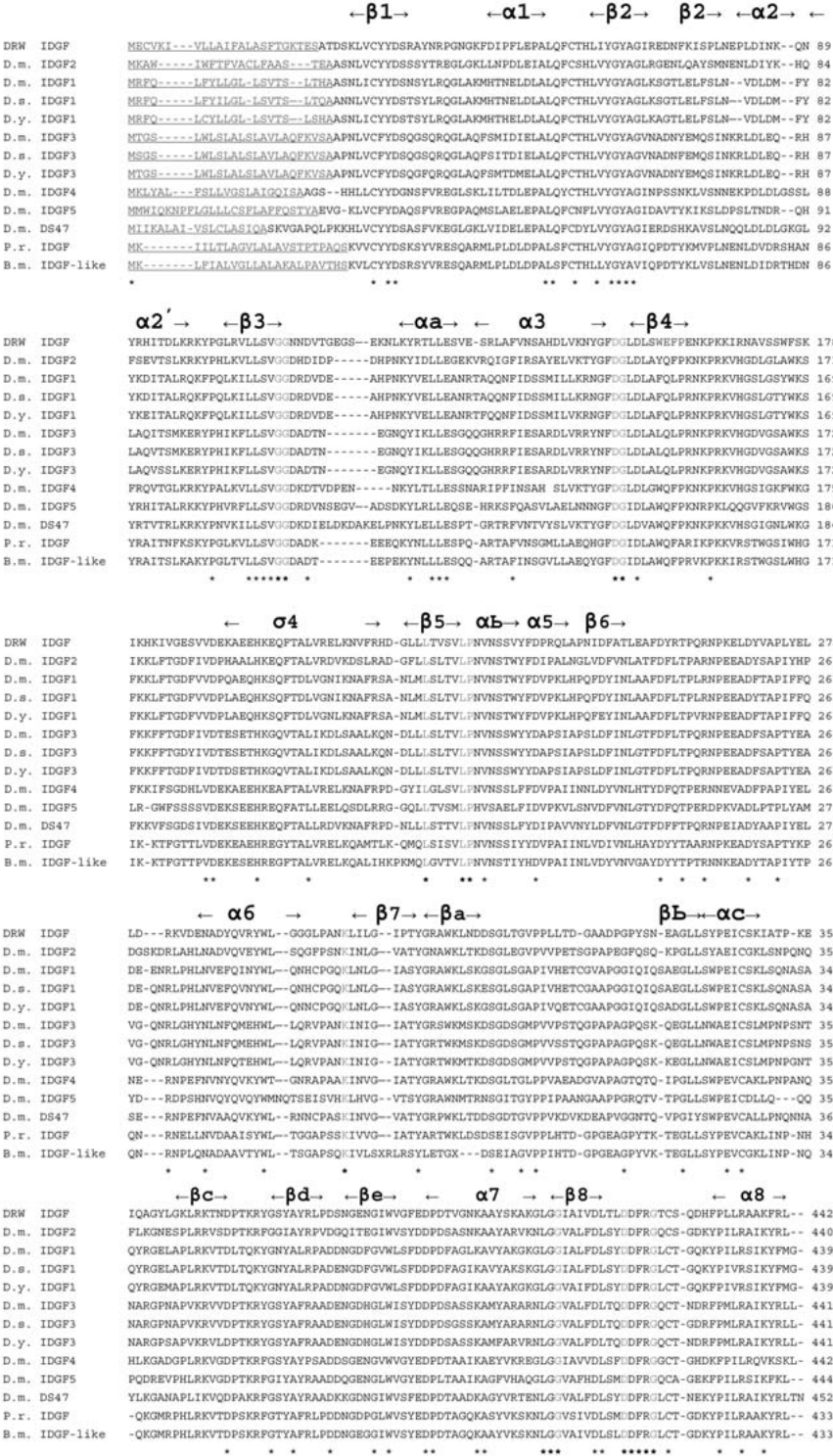
Fig. 2. Multiple sequence alignment of IDGF-DRW and all IDGF family members. *Diaprepes abbreviatus*. DRW, *Drosophila melanogaster*, D.m., *Drosophila simulans*, D.s., *Drosophila yakuba*, D.y., *Pieris rapae* P.r., and *Bombyx mori*, B.m. The signal peptide is underlined. The red characters show the conserved residues, * in bold are predicted to be necessary for barrel formation in IDGF's. The strictly conserved residues in all IDGF are shown by asterisk. IDGF-DRW secondary structure elements are depicted on top as assigned IDGF2 after Varela et al. (2002).
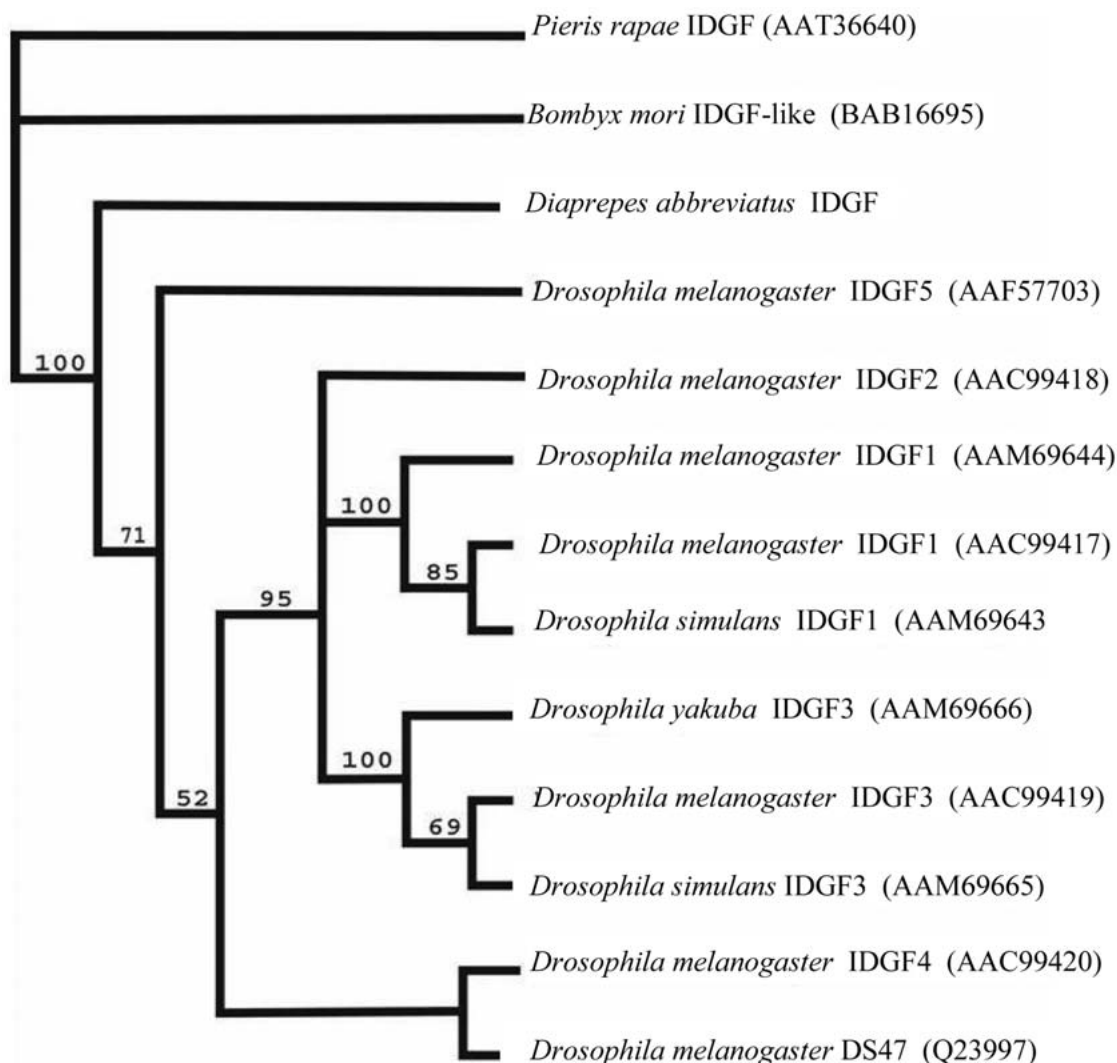
Fig. 3 Phylogenetic tree of the IDGF family. The accession number for sequences is shown inside parentheses. Phylogenetic tree was generated by bootstrap neighbor joining, 1000×, PAUP 4.0, unrooted (Swofford 2002). (*Diaprepes abbreviatus*, IDGF- DRW accession no. <u>AAV68692.1</u>).

ily members (Fig.2), in human chitotriosidase, and in mammalian chitinase-like proteins with no chitinase activity, but are not found in family 18 glycosyl hydrolases from plants or bacteria (Varela et al. 2002). The putative binding site of IDGF2 (Varela et al. 2002) is composed of the ten residues, Tyr_{65}, Asp_{111}, His_{112}, Gln_{159}, Phe_{160}, Lys_{162}, Asp_{250}, Tyr_{303}, Phe_{416}, and Tyr_{420}. These residues are not strictly conserved in all known IDGF family members, however, the predicted binding site of IDGF-DRW contains a subset of these residues (Tyr_{65}, Phe_{160}, Asp_{250}, and Tyr_{303}) and the remaining residues are replaced by the same or complementary category of amino acids.

DISCUSSION

Overall Structure of IDGF-DRW

The structure of IDGF-DRW was modeled against *Drosophila* IDGF2 (Varela et al. 2002). The formation of the clade of IDGF members from *Drosophila* species separate from the other insects, implies an evolutionary divergence of all the *Drosophila* sequences after the evolutionary divergence of the insects for which IDGF sequence data is presented. This would suggest a very recent expansion of the IDGF gene family. Further examination is needed before statements
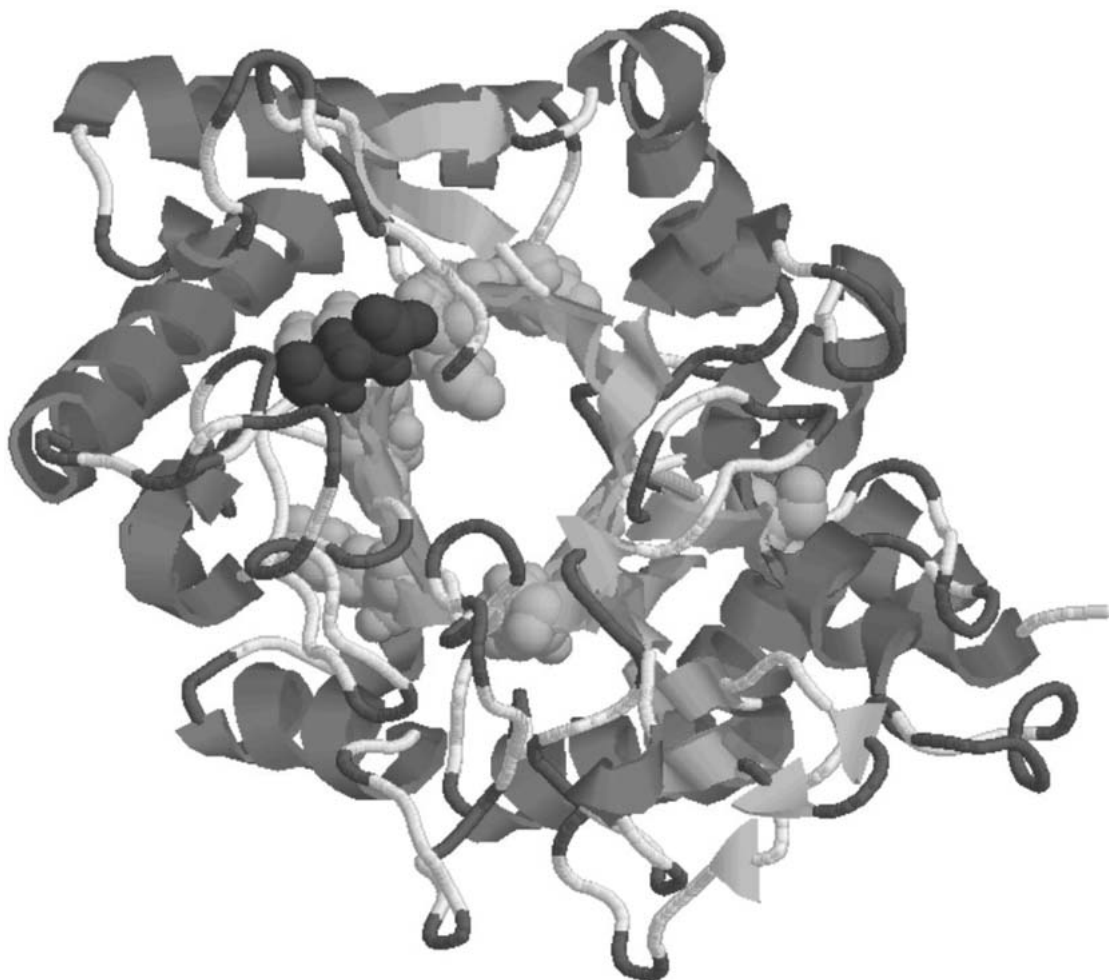
Fig. 4. Basic view of IDGF-DRW protein 3-D structure. The beta barrel motif shown in yellow in center. Structure was predicted with ROBETTA, a full-chain protein structure prediction server. View of conserved residues (in green space-fill) are apparently essential to maintain the barrel folding as predicted by interface alanine scanning (Van Scheltinga et al. 1996). Residues are Gly_{109}, Gly_{110}, Asp_{152}, Gly_{153}, Leu_{218}, Asp_{242}, Lys_{295}, Gly_{412}, and Asp_{421}. The residue at position 159 is Glu (in brown space-fill), which is replaced by Gln in other known IDGF proteins. Structure key: alpha helix (pink), beta strand (yellow), turn (blue) and other (grey).

of evolutionary divergence within the remaining insect species, DRW, *P. rapae*, and *B. mori* can be ascertained.

The structure of IDGF-DRW displays the characteristic fold of the family 18 glycosyl hydrolases. An insertion (Gly_{304} to Phe_{392}) in the beta barrel motif between strand β_7 and helix α_7 forms an additional $\alpha+\beta$ domain similar to that of *Serratia marcescens* chitinases A and B (Perrakis et al. 1994; Van Aalten et al. 2000). The feature is common to other IDGF proteins and to most chitinase-like proteins described to date, although the insertion length varies (Varela et al. 2002). A few conserved residues, present in family 18 glycosyl hydrolases and other IDGF proteins, also were found in IDGF-DRW and are apparently essential

to maintain the barrel folding residues shown in red (Fig. 2). Characteristically in IDGF2, there are three *cis* peptide bonds—Gly_{64}-Tyr_{65}, Pro_{319}-Val_{320}, and Phe_{416}-Asp_{417}. The first and third are conserved in all family 18 members and appear to be necessary for correct folding of the barrel. The second is located in the inserted $\alpha+\beta$ domain and is not conserved in *S. marcescens* chitinase A or B (Perrakis et al. 1994; Varela et al. 2002). The corresponding peptide bonds in IDGF-DRW are Gly_{64}-Tyr_{65}, Pro_{319}-$Pro_{320,}$ and Val_{416}-Asp_{417}. A conserved amino acid change in the third peptide bond, $Phe_{416} \rightarrow Val_{416}$, may preserve its function as valine is an aliphatic-hydrophobic amino acid and capable of binding substrate. Additionally, the aromatic residues (Tyr_{65} and Phe_{416}) involved in the

two conserved *cis* peptide bonds have been reported to be important for binding of substrates in all glycosyl hydrolases with triosephosphate isomerase barrel folds (Jabs et al. 1999). Thus the changes in amino acids were predicted to not affect the binding site structure.

The cooperation between IDGFs and insulin in promoting cell proliferation makes these interactions a possible genetic target for disruption, especially in a long-lived insect like DRW which has a subterranean larval stage. Disruption of critical developmental pathways, such as reduction of the signaling of the insulin receptor, may provide a means for the development of novel management methods to reduce DRW growth and/or survival.

Our work has identified a new member of growth-promoting glycoprotein IDGF protein family, IDGF-DRW. The *idgf-DRW* sequence was amplified from both cultured and field caught adult DRW. Further investigations are needed to examine the expression of these proteins during development to identify when they are maximally expressed, and their interactions with downstream effector signals in developmental pathways. Identification of the genes and proteins functioning in developmental pathways increases our understanding and aids the development of essential tools to conduct future experiments on the functions and interactions of growth-factor like receptors in the DRW.

REFERENCES CITED

ALOY, P., AND R. B. RUSSELL. 2003. Understanding and predicting protein assemblies with 3D structures. Comp. Functional Genomics 4: 410-415.

ASGARI, S., AND O. SCHMIDT. 2004. Isolation of an imaginal disc growth factor homologue from *Pieris rapae* and its expression following parasitization by *Cotesia rubecula*. J. Insect Physiol. 50: 687-694.

BELL, R. E., AND N. BEN-TAL. 2003. In silico identification of functional protein interfaces. Comp. Functional Genomics 4: 420-423.

BOWERS, P. M., C. E. STRAUSS, AND D. BAKER. 2000. *De novo* protein structure determination using sparse NMR data. Journal of Biomolecular NMR 18: 311-318.

CHANG, N. A., S. L. HUNG, K. Y. HWA, I. KATO, J. E. CHEN, C. H. LIU, AND A. C. CHANG. 2001. A macrophage protein, Ym1, transiently expressed during inflammation is a novel mammalian lectin. J. Biol. Chem. 276: 17497-17506.

CHEN, C., J. JACK, AND R. S. GAROFALO. 1996. The *Drosophila* insulin receptor is required for normal growth. Endocrinology 137: 846-856.

CHOU, K. C. 2002. Prediction of protein signal sequences. Current Protein Peptide Sci. 3(6): 615-22.

FERNÁNDEZ, R., D. TABARINI, N. AZPIAZU, M. FRASCH, AND J. SCHLESSINGER. 1995. The *Drosophila* insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. The EMBO Jour. 14: 3373-3384.

GAROFALO, R. S., AND O. M. ROSEN. 1988. Tissue localization of *Drosophila melanogaster* insulin receptor transcripts during development. Molec. Cell. Biol. 8: 1638-1647.

GRAHAM, J. H., C. W. MCCOY, AND J. S. ROGERS. 1996. Insect-plant pathogen interactions: preliminary studies of *Diaprepes* root weevil injury and *Phytophthora* infections. Proc. Florida State Hort. Soc. 109: 57-62.

GRAHAM, J. H., D. B. BRIGHT, AND C. W. MCCOY. 2003. *Phytophthora-Diaprepes* weevils complex: *Phytophthora* spp. relationship to citrus rootstocks. Plant Disease 87: 85-90.

HAKALA, B. E., C. WHITE, AND A. D. RECKLIES. 1993. Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family. J. Biol. Chem. 268: 25803-25810.

HELD, L. I, JR. 2002. Imaginal Discs. The Genetic and Cellular Logic of Pattern Formation. Developmental and Cell Biology Series, Cambridge University Press, Cambridge CB2 2RU, UK. 460 pp.

HOMMA, K., T. MATSUSHITA, AND S. NATORI. 1996. Purification, characterization, and cDNA cloning of a novel growth factor from the conditioned medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina* (flesh fly). J. Biol. Chem. 271: 13770-13775.

HU, B., K. TRINH, W. F. FIGUEIRA, AND P. A. PRICE. 1996. Isolation and sequence of a novel human chondrocyte protein related to mammalian members of the chitinase protein family. J. Biol. Chem. 271: 19415-19420.

JABS, A., M. S. WEISS, AND R. HILGENFELD. 1999. Non-proline cis peptide bonds in proteins. J. Molec. Biol. 286: 291-304.

KAWAMURA, K., T. SHIBATA, O. SAGET, D. PEEL, AND P. J. BRYANT. 1999. A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. Development 126: 211-219.

KIM, D. E., D. CHIVIAN, AND D. BAKER. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32 Suppl. 2: W526-31.

KIRKPATRICK, R. B., R. MATICO, D. E. MCNULTY, J. E. STRICKLER, AND M. ROSENBERG. 1995. An abundantly secreted glycoprotein from *Drosophila melanogaster* is related to mammalian secretory proteins produced in rheumatoid tissues and by activated macrophages. Gene 153: 147-154.

LAPOINTE, S. L., AND J. P. SHAPIRO. 1999. Effect of soil moisture on development of *Diaprepes abbreviatus* (Coleoptera: Curculionidae). Florida Entomol. 82: 291-299.

LAPOINTE, S. L. 2000. Particle film deters oviposition by *Diaprepes abbreviatus* (Coleoptera: Curculionidae). J. Econ. Entomol. 93: 1459-1463.

MARTELLI, P. L., P. FARISELLI, G. TASCO, AND R. CASA-DIO. 2003. The prediction of membrane protein structure and genome structural annotation. Comp. Functional Genomics 4: 406-409.

MARTOGLIO, B., AND B. DOBBERSTEIN. 1998. Signal sequences: more than just greasy peptides. Trends Cell Biol. 8(10): 410-15.

MORRISON, B. W., AND P. LEDER. 1994. Neu and ras initiate murine mammary-tumors that share genetic-markers generally absent in C-MYC and INT-2-initiated tumors. Oncogene 9: 3417-3426.

ORTIZ, A. R., C. E. M., STRAUSS, AND O. OLMEA. 2002. MAMMOTH (Matching molecular models obtained from theory): an automated method for model comparison. Protein Sci. 11: 2606-2621.

OWHASHI, M., H. ARITA, AND N. HAYAI. 2000. Identification of a novel eosinophil chemotactic cytokine (ECF-L) as a chitinase family protein. J. Biol. Chem. 275: 1279-1286.

PERRAKIS, A., I. TEWS, Z. DAUTER, A. B. OPPENHEIM, I. CHET, K. S. WILSON, AND C. E. VORGIAS. 1994. Crystal-structure of a bacterial chitinase at 2.3-angstrom resolution. Structure 2: 1169-1180.

ROHL, C. A., AND D. BAKER. 2002. *De novo* determination of protein backbone structure from residual dipolar couplings using Robetta. J. American Chem. Soc. 124: 2723-2729.

RUAN, Y., C. CHEN, Y. CAO, AND R. S. GAROFALO. 1995. The *Drosophila* insulin receptor contains a novel carboxyl-terminal extension likely to play an important role in signal transduction. J. Biol. Chem. 270: 4236-4243.

SHACKELTON, L. M., D. M. MANN, AND A. J. MILLIS. 1995. Identification of a 38-kDa heparin-binding glycoprotein (gp38k) in differentiating vascular smooth muscle cells as a member of a group of proteins associated with tissue remodeling. J. Biol. Chem. 270: 13076-13083.

SIMONS, K. T., C. KOOPERBERG, E. HUANG, AND D. BAKER. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulate annealing and Bayesian scoring functions. J. Molec. Biol. 268: 209-225.

STROMMER, J. N., R. GREGERSON, AND M. VAYDA. 1993. Isolation and characterization of plant mRNA, pp. 49-65 *In* B. R. Glick and J. E. Thompson [eds.], Methods in Plant Molecular Biology and Biotechniques. CRC Press, Boca Raton, FL.

SWOFFORD, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

TSUZUKI, S., M. IWAMI, AND S. SAKURAI. 2001. Ecdysteroid-inducible genes in the programmed cell death during insect metamorphosis. Insect Biochem. Molec. Biol. 31: 321-331.

VALENCIA, A. 2003. Multiple sequence alignments as tools for protein structure and function prediction. Comp. Functional Genomics 4: 424-427.

VAN AALTEN, D. M. F., B. SYNSTAD, M. B. BRURBERG, E. HOUGH, B. W. RIISE, V. G. H. EIJSINK, AND R. K. WIERENGA. 2000. Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9-angstrom resolution. Proc. Nat'l. Acad. Sci., USA 97: 5842-5847.

VARELA, P. F., A. S. LLERA, R. A. MARIUZZA, AND J. TORMO. 2002. Crystal structure of imaginal disc growth factor-2. J. Biol. Chem. 277: 13229-13236.

VAN SCHELTINGA, A. C. T., M. HENNIG, AND B. W. DIJKSTRA. 1996. The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. J. Molec. Biol. 262: 243-257.

ZUROVCOVA, M., AND F. J. AYALA. 2002. Polymorphism patterns in two tightly linked developmental genes, *idgf*1 and *idgf*3, of *Drosophila melanogaster*. Genetics 162: 177-188.