Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
JEZS 2017; 5(4): 1166-1178

Received: 15-05-2017
Accepted: 16-06-2017

Bhagath Kumar Palaka
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Tuleshwori Devi Sapam
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Anbumani Velmurugan Ilavarasi
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Shamik Chowdhury
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Sampath Kumar Rajendiran
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Mohd. Babu Khan
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Venkata Satyanarayana Nallala
Department of Biochemistry & Molecular Biology, School of Life Sciences, Pondicherry University, Puducherry, India

Dinakara Rao Ampasala
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Correspondence
Dinakara Rao Ampasala
Associate Professor,
Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, India

Molecular cloning and characterization of phosphoacetylglucosamine mutase from *Bombyx mori*

Bhagath Kumar Palaka, Tuleshwori Devi Sapam, Anbumani Velmurugan Ilavarasi, Shamik Chowdhury, Sampath Kumar Rajendiran, Mohd. Babu Khan, Venkata Satyanarayana Nallala and Dinakara Rao Ampasala

Abstract
The present study conducted at Centre for Bioinformatics, Pondicherry University for a period of six months during the year 2016 explored the molecular and structural features of phosphoacetylglucosamine mutase from *Bombyx mori*, a crucial enzyme of chitin metabolic pathway. A 1647 bp long full-length cDNA was amplified, cloned and gene-expression profile was studied which inferred the crucial role of *Bm*PAGM during larval molting. The ORF encoded a 548 amino acid long polypeptide with theoretical molecular mass of 59.859 kDa, pI of 5.65 and sequence similarity of >80% to other lepidopteran PAGMs. Sequence alignment and phylogenetic analysis of various PAGMs added an insight on their evolutionary relationship. The 3D structure of *Bm*PAGM was predicted and refined by molecular dynamics simulations to study various structural features. The structure composed of four domains with a centrally located active site and magnesium ion coordinated to three aspartic acids of the metal binding loop. This study will be useful to scientific community working on chitin metabolizing enzymes.

Keywords: PAGM, *Bombyx mori*, Cloning, Phylogenetic analysis, Modeling, Molecular Dynamics Simulation

1. Introduction
Chitin metabolic pathway is essential for survival and evolutionary success of various organisms like algae, fungi and certain invertebrates like nematodes, crustaceans, insects, cephalopods and mollusks [1, 2]. Chitin is an important nitrogenous polysaccharide and a major constituent of the fungal cell wall, nematode eggshell, crustacean and insect exoskeleton, cephalopod beaks and radulae of mollusks [2]. As chitin is a major component of the cuticle, trachea and peritrophic membrane of insects, chitin metabolic pathway is considered very essential for their growth and metamorphosis. Hence, understanding the molecular, structural and functional features of chitin metabolizing enzymes is a requisite [3-6]. The insect chitin biosynthetic pathway was presumed to be similar to the Leloir pathway of fungi and involves various enzymes, which convert initial substrate trehalose to chitin [7].

The chitin metabolic pathway includes various steps catalyzed by specific enzymes; the trehalose distributed in the extracellular matrix of insect tissues, is converted to glucose by the enzyme trehalase [8]. The glucose which enters the cytosol is converted in two steps to fructose-6-phosphate by two glycolytic enzymes namely hexokinase and glucose-6-phosphate isomerase. The amination of fructose-6-phosphate to glucosamine-6-phosphate was catalyzed by glutamine-fructose-6-phosphate amidotransferase followed by acetylation of the obtained glucosamine-6-phosphate to N-acetylglucosamine-6-phosphate (GlcNAc-6-P) by N-acetylglucosamine acetyltransferase. The next step involves the isomerization of GlcNAc-6-P to N-acetylglucosamine-1-phosphate (GlcNAc-1-P) by phosphoacetylglucosamine mutase (PAGM) [1]. The conversion of GlcNAc-1-P to a nucleotide sugar derivative, uridine diphosphate (UDP)-N-acetylglucosamine (UDP-GlcNAc) will be carried out by UDP-N-acetylglucosamine pyrophosphorylase (UAP) which supplies these activated UDP-GlcNAc molecules as substrate to chitin synthase for the synthesis of chitin, a linear biopolymer of these *N*-acetylglucosamine (GlcNAc) molecules connected by β-1,4-linkages [9, 10]. The enzymes of insect chitin degradation pathway include β-N-acetylglucosaminidases and chitinases that catalyze the hydrolysis of chitin polymers to monomers of UDP-GlcNAc.

Chitinolytic enzymes are also very crucial for proper growth and morphogenesis in other arthropods, for cell division and sporulation of fungal species [11-14].

Despite the high significance of chitin metabolic pathway in growth and development of insects, knowledge on structural and functional characteristics of these enzymes is inadequate. Though few reports exist on molecular and structural features of chitin synthases and chitinases, other enzymes of chitin metabolism need to be studied for better understanding of their role and mechanism of action [15, 16]. Few reports exist concerning the molecular and functional characterization of insect UAPs and their role in insect molting and metamorphosis. However, the structural and functional characteristics of insect PAGMs and their role in insect molting have not been reported. Hence in the present study, cloning, expression and *in silico* analysis of *Bombyx mori* (a typical lepidopteron insect) PAGM (*Bm*PAGM) was performed which helps in understanding the role of this enzyme in insect molting and metamorphosis and its evolutionary relatedness to other PAGMs.

The enzyme PAGM (EC 5.4.2.3) belongs to the family of isomerases that specifically transfer phosphate groups within a molecule (phosphotransferases or phosphomutases). This enzyme participates in metabolism of aminosugars and other glycosylation pathways and catalyzes the reversible conversion of GlcNAc-6-P to GlcNAc-1-P. The mechanism of UDP-GlcNAc biosynthesis is different in prokaryotes and eukaryotes. In prokaryotes intramolecularphosphoryl transfer step occurs before acetylation step whereas in eukaryotes it occurs after acetylation [17, 18]. Thus, PAGM is necessary and present only in eukaryotes but not in prokaryotes.

PAGM belongs to a superfamily of phosphohexomutases (Accession ID: cd03084) which includes many other related enzymes such as phosphoglucomutases viz., PGM1 (cd03085) and PGM2 (cd05799), phosphoglucosamine mutases (PNGM, cd03088), the bacterial phosphoglucosamine mutase (GlmM, cd05802), bacterial phosphomannomutases (ManB or PMM, cd03088) and bifunctional phosphomannomutase/phosphoglucomutases (PMM/PGM, cd03089). All these enzymes play diverse role in carbohydrate metabolism in various forms of life. PAGMs were found to have similar active site architecture compared to other enzymes of phosphohexomutase superfamily and share similar substrate recognition strategies and catalytic mechanisms [19].

Structural studies on few representative members of the superfamily were available viz., PGM from *Salmonella typhimurium* [20], *Oryctolagus cuniculus* [21], PMM/PGM from *Pseudomonas aeruginosa* [22] and PAGM from *Candida albicans* (*Ca*PAGM) in both apo and ligand bound forms which provided new insights into enzyme mechanism and substrate recognition. An earlier study on *Ca*PAGM (PDB ID: 2DKA) reported four highly conserved loop regions viz., active serine loop, metal binding loop, sugar binding loop and phosphate binding loop that are observed to be highly essential for interaction with substrate and metal ion and ultimately the activity of the enzyme [18]. The metal ion was observed to be coordinating to three aspartic acid residues of the metal binding loop that are crucial for the activity of the enzyme. This is the only crystal structure of PAGM available to date and scientific reports available are limited. As the structure of insect PAGMs has not been reported yet, it is necessary to determine the 3D structure of this enzyme to explore various structural features and to elucidate the catalytic mechanism. Determining the structure would help in exploring the possible role of PAGM as drug target and in identifying novel inhibitors by virtual screening approach. Hence, the present study was aimed at exploring the molecular, structural and functional features of phosphoacetylglucosamine mutase from *Bombyx mori.*

2. Materials and Methods

2.1 Insect Culture

The Pure Mysore x CSR2 hybrid variety of Silkworm eggs were procured from Central Sericulture Research and Training Institute (CSRTI), Mysore, Karnataka, India and reared as per the provided guidelines. Eggs were maintained at standard environmental conditions of 28 ºC and 85% relative humidity in dark. After hatching, the worms were reared on four feeds of V1 variety of mulberry leaves per day under normal 12 h light and 12 h dark conditions at 8 AM, 1 PM, 06 PM and 10 PM. Silkworm samples were collected every day at early hours for all the experimental works. The complete study discussed here was conducted at Centre for Bioinformatics, Pondicherry University, Puducherry, for six months from March 2016 until August 2016.

2.2 RNA Isolation, cDNA Synthesis and PCR

Second instar molting larvae were picked and immediately frozen in liquid nitrogen to isolate total RNA using RNA Iso plus reagent (Takara, Japan) following guanidium thiocyanate method [23]. Two micrograms of the isolated total RNA was used for the synthesis of first-strand cDNA using Prime script 1st strand cDNA synthesis kit (Takara, Japan), following manufacturer instructions. This cDNA was used for the amplification of *Bm*PAGM gene using gene specific primers (Forward primer: 5' ATGCCGCCGAGTAGCTTA 3' and Reverse primer: 5' TTAAGCTGGCAACTCTGGC 3') designed based on a UAP-like sequence predicted from a whole genome shot gun sequence of *B. mori* using GENSCAN web server (http://genes.mit.edu/GENSCAN.html). PCR reactions were carried out using PCR master mix containing Taq DNA polymerase following the recommendations of manufacturer (Takara, Japan) in an Eppendorf Gradient Master-thermo cycler (Eppendorf, Germany). The PCR protocol optimized for amplification is initial denaturation at 94 °C for 3 min, tracked by 35 cycles of amplification (94 °C for 60 s, 55 °C for 90 s, and 72 °C for 1 min) and a final extension at 72 °C for 10 min.

2.3 Developmental and Tissue Specific Expression Studies

Expression profile of *Bm*PAGM mRNA during the growth and molting of silkworm larvae were examined. Samples were collected for 11 days at every 24 h interval starting from a day before molting of first instar larvae until a day after molting of fourth instar. Fifth instar, third day larvae were used for tissue specific experimental studies and various tissues viz., cuticle, epidermis, fat bodies, head capsule, foregut, midgut, hindgut, and malpighian tubules were collected and frozen at -80 ºC until further use. Total RNA extraction was performed using TRIZOL method and 2 µg of the extracted RNA samples were used for cDNA library construction using first strand cDNA synthesis kit (Takara, Japan). The obtained cDNA was used for amplification of *Bm*PAGM gene using same gene specific primers and PCR conditions described in earlier section.

2.4 Multiple Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignment (MSA) and phylogenetic analysis were performed to explore the evolutionary relationship of PAGMs with other members of phosphohexosemutase super family. Protein sequences of PAGMs, PGMs, PGNMs and PMMs from various domains of life viz., bacteria, fungi, plants and animals were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). MSA was performed using ClustalW program (http://www.ebi.ac.uk/clustalw/) and phylogenetic trees were constructed using MEGA 6.0 software [24]. The maximum likelihood method was used based on the Jonathan Taylor Thomas model with a discrete gamma distribution. A bootstrap analysis of 1000 replications was carried out on the trees inferred from the maximum likelihood method and bootstrap values were shown at each branch of the tree. The accession numbers of all the proteins used are listed in Supplementary Table 1.

2.5 Predicting the Three Dimensional Structure of *Bm*PAGM

The protein sequence of *Bm*PAGM (UniProtKB-A0A089PRC4) was retrieved from UniProt database (http://www.uniprot.org) and submitted for BLASTp search against the Protein Data Bank (http://www.rcsb.org/pdb/). Based on sequence similarity, crystal structure of *Ca*PAGM (PDB ID: 2DKA), was selected as a template protein to predict the 3D model of *Bm*PAGM using MODELLER 9v12 [25] and the best model was selected based on the discrete optimized protein energy (DOPE) scoring function and molPDF score. Validation of the model quality was carried-out using SAVES server (http://nihserver.mbi.ucla.edu/SAVS/), which evaluates the quality of the protein model using PROCHECK [26], ERRAT [27], and Verify3D [28]. Molecular visualization and analysis was performed using PyMol and Chimera [29, 30].

2.6 Molecular Dynamics Simulations of *Bm*PAGM

Molecular dynamics (MD) simulations were performed for *Bm*PAGM protein model using GROMACS 5.0.4 Package [31]. Protein topology was defined using GROMOS96 43a1 force field and solvated using SPC216 water model. Protein is enclosed in a cubic box and neutralized by adding appropriate number of counter ions followed by energy minimization using steepest descent algorithm for 50,000 steps. Linear Constraint Solver (LINCS) algorithm [32] was used to constrain all types of bond angles and PME29 electrostatics were used to treat long range electrostatics. The system was equilibrated for a time period of 1000 ps at constant volume (NVT) by applying V-rescale temperature coupling method to attain 300 K temperature and for 1000 ps at constant pressure (NPT) of 1.0 bar by applying Parrinello–Rahman pressure coupling method [33]. Finally, MD run was performed for a period of 70 ns and corresponding trajectories were further analyzed to study the root mean square deviations (RMSD), root mean square fluctuations (RMSF) and radius of gyration (Rg) of the protein. Further, principal component analysis (PCA) was carried out and free energy landscape (FEL) was plotted using two principle components with low cosine content value to identify the minimum energy basin and to retrieve the most stable protein conformations.

2.7. Statistical Analysis

The developmental and tissue specific expression studies were performed in triplicate and the band intensities of the amplified gene were visualized under UV light. Quantity one software (Bio-Rad, USA) was used for Quantitative analysis of gel images. Actin3 gene concentration was used as a control for estimating reaction efficiency. The RMSD, RMSF and Rg of the protein structure during the MD simulations were calculated at every 10 ps interval and plotted using xmgrace tool.

3. Results and Discussion

3.1 Cloning and *in silico* Sequence Analysis of *Bm*PAGM

Scouting the NCBI database for *B. mori* PAGM resulted in identification of a gene sequence (Gene ID: LOC101743243) of about 15050 bp long, located on a whole genome shotgun sequence (NCBI Reference Sequence: NW_004582032.1) of *B. mori* strain p50T. The obtained gene sequence was submitted to GENSCAN server which predicted the exons and introns present in a given gene sequence. The LOC101743243 was found to have 12 exon regions and spans a length of 2164 bp with a 130 bp 5' UTR, 1647 bp long protein coding region and a 387 bp long 3' UTR. Gene specific primers were designed with an aim to amplify, clone and confirm the full-length protein-coding region of the predicted mRNA sequence of *Bm*PAGM. Total RNA was isolated from the third instar molting larvae and used as a template for amplification of *Bm*PAGM. A full-length coding region of 1647 bp with a stop codon at 3' end was amplified and cloned using pTZ57R/T vector into DH5α strain of *Escherichia coli* and confirmed by restriction digestion (Supplementary Fig. 1). The plasmid was isolated and sequenced using M13 forward and reverse primers to confirm the insert sequence. The obtained ORF encodes a polypeptide of 548 amino acids with a theoretical molecular mass of 59.859 kDa and an estimated pI of 5.65. BLAST search of the deduced protein sequence showed a sequence similarity of 86%, 85% and 83% to PAGM of lepidopteron insects viz., *Cnaphalocrocis medinalis* (GenBank: AJG44540.1), *Papilio xuthus* (GenBank: KPJ03163.1) and *Danaus plexippus* (GenBank: EHJ68938.1), respectively. Thus, it was confirmed that the sequence is a full-length cDNA of *B. mori* PAGM (*Bm*PAGM, GenBank: KX130001.1) and was submitted to GenBank database using sequin software. No signal peptide was identified when the protein sequence of *Bm*PAGM (GenBank ID: AON76440.1) was submitted to SignalP 4.1 server [34] and no transmembrane regions were identified by TMpred server [35]. The sub cellular localization of the protein was predicted using Psort [36], Protcomp9 [37] and ESLpred [38] servers which predicted the location of the protein as 'cytoplasmic' with high probability. SOPMA server [39] predicted the percentage of various secondary structure elements to be 42.52% of helices, 22.08% of strands, 24.82% of coils and 10.58% of β-turns. Submitting the protein sequence to NCBI conserved domain database (https://www.ncbi.nlm.nih.gov/cdd) revealed that the protein has a conserved phosphoglucomutase 3 domain (PGM3 domain, CDD Accession ID: cd03086) also known as phosphoacetylglucosamine mutase domain (CDD Accession ID: PLN02895) between residues 13-544. The enzymes with this domain catalyze the conversion of GlcNAc-6-P to GlcNAc-1-P leading to the biosynthesis of UDP-GlcNAc, an essential metabolic precursor of glycoproteins and mucopolysaccharides. The enzymes were reported to have four domains and one loop from each domain contributes for the formation of a centrally located active site.

3.2 Developmental and Tissue Specific Expression Studies: Silkworm larval samples from a day before molting of first instar to a day after molting of fourth instar were collected to study the expression profile of *Bm*PAGM. The Actin3 gene (BmA3, NCBI Accession Number: NM_001126254) of *B. mori* was used as a reference gene, to estimate the relative expression levels of *Bm*PAGM. On observation, it was revealed that the expression levels were periodically high during premolting and molting stages and decreased after ecdysis during intermolt stage, which indicates that the expression of *Bm*PAGM is essential for larval molting (Fig. 1A). To date no reports exist on the developmental and tissue-specific expression profile of insect PAGMs. However, earlier reports on the relative transcript levels of other chitin metabolic pathway enzymes in *Locusta migratoria* showed a regular pattern of increase during molting stage and decrease during intermolt stage in expression of UAP1 and CHS1 [1]. Similar results were also reported in the developmental expression profile of *T. castaneum* UAP and of *B. mori* CHS [4, 40]. A study on trehalase gene (TRE) of *Nilaparvata lugens* demonstrated that TRE silencing decreased the expression of other chitin metabolizing enzymes significantly and affected the regulation of chitin metabolism causing deformities during molting [41]. Thus, the relative high expression of PAGM observed during the molting of *B. mori* larvae indicates that the enzyme is essential for insect molting. Among various tissues analyzed for tissue specific expression profile, mRNA levels were observed to be higher in integument and least in head capsule (Fig. 1B). Based on our literature survey no tissue-specific expression studies were reported for insect PAGMs, but very few reports exist for other chitin metabolizing enzymes such as UAP and CHS. A high level of CHSA gene expression was reported in the integument and head tissues of fifth instar day 3, *B. mori* larvae, with relatively low expression levels in the gonad and fat body [4]. A study on *T. castaneum* UAP reported that the expression was higher in cuticle forming tissues of adult elytra [40]. Similarly, the expression profile of Esterase-A4 (Ea4) of *B. mori* determined by RT-PCR and gene microarray analysis demonstrated that Ea4 displayed stage-specific and tissue-specific expression during postembryonic stages with high Ea4 mRNA in spinning and eclosion stages in integument and head, but low in gonads of fifth-instar day-3 larvae [42]. These studies support that the chitin metabolizing enzymes associated with molting behavior in insects show a relatively high expression level in integument tissue.

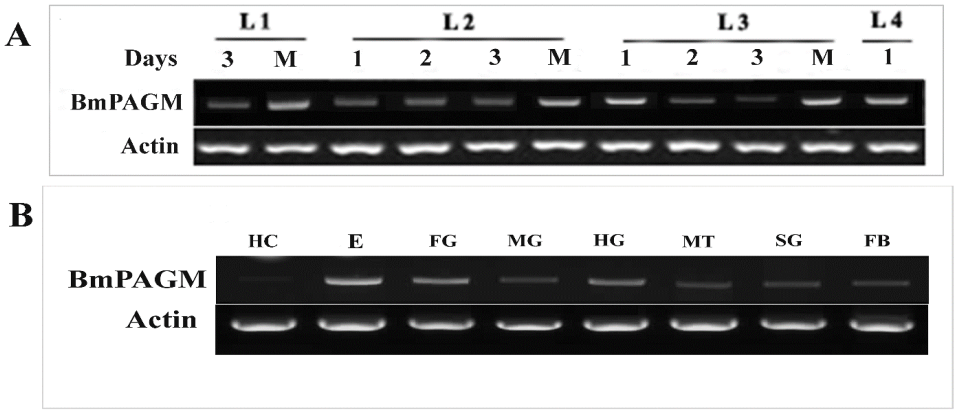


Fig 1: A) Developmental expression profile of *Bm*PAGM and Actin3 mRNA observed from a day before molting of first instar to a day after molting of fourth instar larvae. Numbers on the lanes indicate the days during the development of silkworm on which the RNA was isolated. L1- First instar larva, L2- Second instar larva, L3-Third instar larva, L4- Fourth instar larva. B) Tissue specific expression studies performed in head capsule, epidermis, foregut, midgut, hindgut, malpighian tubules, salivary glands and fat bodies. HC—head capsule; E—integument; FG—foregut; MG—mid-gut; HG—hindgut; MT— malpighian tubules; SG—salivary glands; FB—fat bodies;

3.3 Multiple Sequence Alignment

Sequence alignment of all available insect and other eukaryotic PAGMs revealed the conserved regions present in the protein across various domains of life. Amino acid residues that are extended along the full-length of *Bm*PAGM viz., Y27-33, R53, G63, M65, T67-N71, D75-G77, K79, D82, G85, M87-L88, W92-E93, N101, D133, R135, T164-P165, H168, N177, Y187, D215, N218-G219, G221, G255, D257, D281-D283, D285-R286, D302-D304, L309, Q332-Y335, Y337, T341, T357, G358, K360, L362, A366, G372, Y374-E376, N378-G381, G423-A425, S427, W446, Y450, P454, P480, Q484, I487, R498, F500, R502-D508 and R511-E515, were found to be completely conserved in various insect and other PAGMs considered for the study. The high conservation of the above mentioned residues indicate that they are very essential for the enzymes structure and function (Supplementary Fig. 2). The residues in the four-conserved loop regions present around the active site of *Ca*PAGM [18], were also found to be well conserved in *Bm*PAGM and all other PAGMs considered for the study with few conservative substitutions.

Sequence alignment of PAGMs with other members of the phosphohexomutase family viz., PGM, PNGM, and PMM was performed to identify the regions conserved among these protein families across various domains of life. To date, PAGMs were identified only in eukaryotes but not in prokaryotes, PGMs and PMMs were observed across all domains of life whereas PGNMs were observed in bacteria, fungi and few plants species but not in animal kingdom [43]. The four protein families vary in length within a range of 390 to 630 residues, except for the eukaryotic PMMs, which have a length of approximately 200-250 residues. Analyzing the sequence alignment revealed that one arginine and the three aspartic acid residues in the metal binding loop were highly conserved in all the four protein families across various forms of life, except in eukaryotic PMMs. The residues present in the active serine and phosphate binding loops were found to be conserved in all members across all domains of life with class specific variations, whereas the residues in the sugar binding loop were found to be varying in various protein families. Eukaryotic PMMs were shorter in length compared to other protein families and did not share much sequence

similarity.

The class specific variants of active serine motif present across various phosphohexomutase protein families include; "TASHN" in PAGMs, eukaryotic PGMs and prokaryotic PMMs and "SASHN" in prokaryotic PGNMs and PGMs, while the motif is not observed in eukaryotic PMMs. Though this motif is a highly conserved, the location of it in PAGMs is different from other protein families, which was reported to be either due to circular permutation or translocation in PAGMs [18, 19]. The "DXDXDR" motif forms a metal binding loop that is conserved in almost all the family members except eukaryotic PMMs. The class specific variant of this motif in PAGMs and prokaryotic PMMs is "DGDADR" while in PGNM and PGM2 proteins, it is "DGDGDR" and in PGM1 proteins, it is "DPDADR". The sugar binding loop is very diverse among various protein families which can be attributed to their wide substrate specificity [19]. The phosphate-binding motif (RXSXTX) is highly conserved among various protein families except in eukaryotic PMMs.

3.4 Phylogenetic Analysis

A phylogenetic tree of insect PAGM protein sequences was constructed to explore the evolutionary relationship among various insect orders. The analysis revealed that *Bm*PAGM has grouped with other lepidopteron PAGMs. PAGMs of various other insect orders has segregated into distinct clades each supported by high bootstrap values (Fig. 2). This indicates that, though there is a significant similarity among various insect PAGMs they have distinct variations among different orders. Representative PAGMs from fungal, plant, amphibian and mammalian species included in the data have evolved as a distinct out-group. A phylogenetic tree generated to study the evolutionary distance between insect and various other PAGM protein sequences revealed that the PAGMs has segregated into two distinct clades each supported by high bootstrap values (Supplementary Fig. 3). All fungal PAGMs has grouped into clade-I whereas other PAGMs into clade-II. The second clade further showed the evolution of insect and other vertebrate PAGMs into definite groups, which further clustered into distinct subgroups in accordance to their orders. The above analysis strongly suggests that dipteran, coleopteran and lepidopteron PAGMs are more closely related compared to hymenoptera PAGMs and as a whole insect PAGMs are more closely related to vertebrate PAGMs compared to fungal PAGMs.

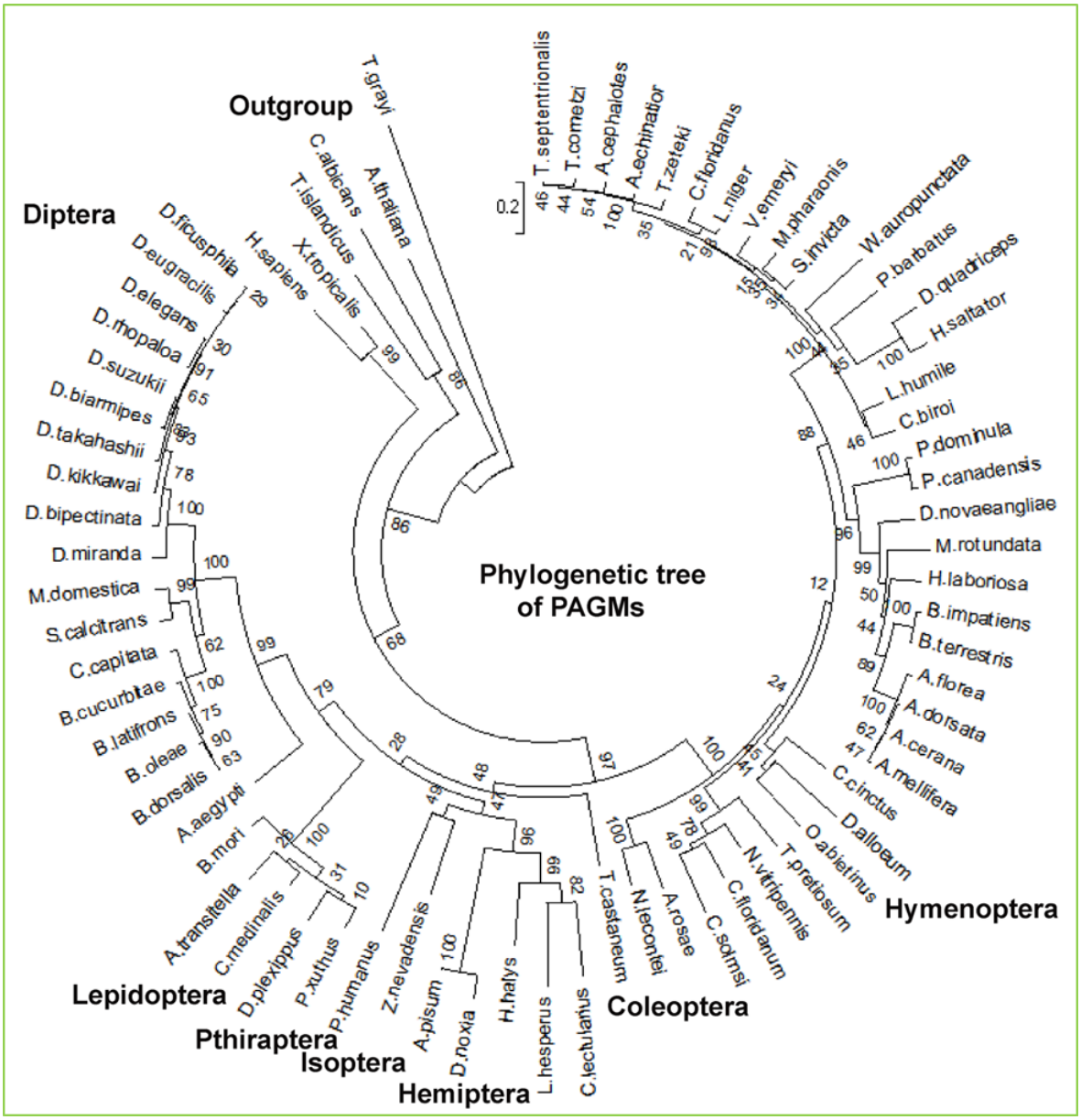


Fig 2: Phylogenetic tree of various insect PAGMs constructed by using Maximum Likelihood method. Bootstrap percentages of 1000 resampled datasets are shown at nodes.

The PAGMs were also compared with other related enzymes of the phosphohexomutase superfamily that catalyze the reversible intramolecular transfer of phosphoryl group on sugar substrates. PGM and PMM proteins are present in both prokaryotes and eukaryotes, whereas PNGM sequences are exclusively present in prokaryotes and PAGMs in eukaryotes [43]. An earlier study on *Trypanosoma brucei*, a eukaryote, revealed the complete absence of PGM protein and its activity was replaced by PMM and PAGM proteins [44]. Analyzing the phylogenetic tree revealed that the proteins segregated into two clades in which PGMs, PGNMs, PAGMs and prokaryotic PMMs have clustered into one clade and eukaryotic PMMs being short in length and most divergent among all the proteins has clustered into different clade (Supplementary Fig. 4). All PAGMs corresponded well into a separate group with high bootstrap value of 98%. Interestingly, prokaryotic PMMs are closer to PGNMs and PGMs than to eukaryotic PMMs. PGMs (PGM1 and PGM2) are divergent among themselves and separated into two different groups with insect PGMs being distributed in both the groups. As a whole, it was observed that insect proteins of phosphohexomutase family were closer to chordates compared to bacterial, fungal and plant proteins.

3.5 Structure and Dynamics of *Bm*PAGM

As PAGMs play a very essential role in molting and metamorphosis of insects, it is indispensable to understand the structural details of insect PAGMs. Hence, 3D structure of *Bm*PAGM was determined through comparative modeling method using 2DKA as template. As PAGMs require magnesium ion to achieve maximum activity, the 3D model of *Bm*PAGM was modeled along with a magnesium ion in the active site in order to explore its microenvironment and understand its interaction profile. The PROCHECK results infer that 94.50% residues were present in the most favored regions and Verify3D results infer that 94.71% of the residues had an averaged 3D-1D score ≥ 0.2 which suggest that the model quality is good. ERRAT value of 86.05% suggests that the overall quality factor of the protein was good and reliable. Superimposition of the template and target structures revealed an average RMSD of 0.24 nm and the predicted protein shared all the essential structural features of the template and displayed high conservation in the active site architecture. The validated model of *Bm*PAGM was subjected to MD simulation studies to observe the stability and fluctuations during 70 ns simulation time. Examining the RMSD plot of *Bm*PAGM during the simulation revealed that the protein has high deviation initially which decreased after 40 ns and attained stability (Fig. 3A). Decrease in the Rg value (Fig. 3B) towards the end of simulation infer that the protein has attained compactness during the 70 ns simulation time scale and analyzing the RMSF plot revealed that none of the amino acids displayed high fluctuations (Fig. 3C). PCA was performed and FEL was generated to obtain the least energy confirmation of the protein. The final structure obtained after simulation was found to be compact and reliable compared to the initial structure (Fig. 4A and B). Superimposition of target and final 3D structure of *Bm*PAGM revealed that this protein potentially retained all basic structural features required for the function though minute variations exist with respect to strand order in the β sheets of various domains.

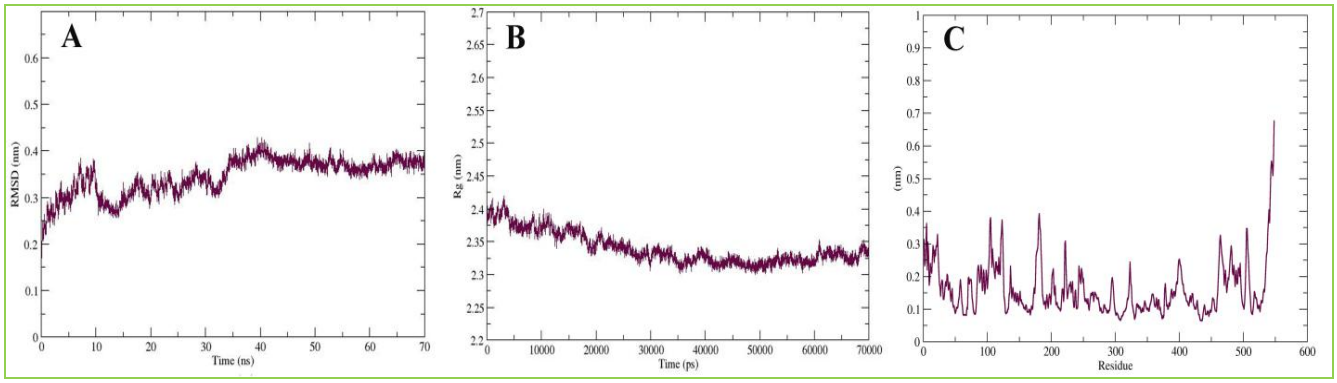


Fig 3: MD simulation results of *Bm*PAGM (A) The backbone RMSD plot (B) The Rg plot (C) The residue RMSF plot.

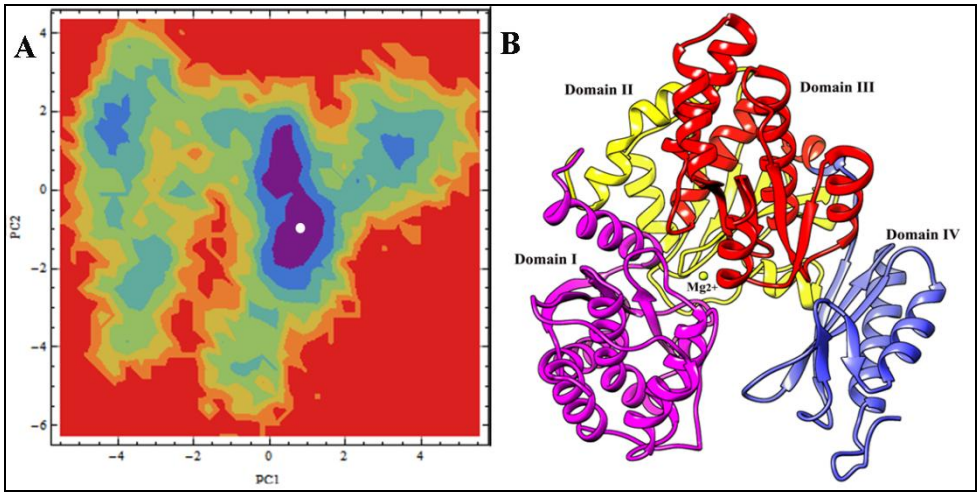


Fig 4: A) The free energy landscape of *Bm*PAGM protein as a function of principal components with least cosine content, the white dot denotes the minimum energy region from which the final structure has been extracted. B) Low energy conformation of *Bm*PAGM extracted from the most populated free energy minimum cluster.

Exploring the 3D structure of *Bm*PAGM revealed the presence of 19 helices, 19 strands, 3 β-α-β units, 6 β-hairpins, 2 β-bulges, 28 helix-helix interactions, 42 β-turns and 5 γ-turns. The protein contains four domain regions extended between residues 1-183 (Domain 1), 184-302 (Domain 2), 303-446 (Domain 3), and 447-548 (Domain 4) which are arranged in heart shaped manner. Similar to template structure, domain 4 of *Bm*PAGM is observed to be flexible and has fewer interactions with other domains. Each domain contains a central β-sheet, flanked on both sides by α-helices, and a β-hairpin from β-sheet of each domain combine to form an active site, which is present in a deep cavity at the center of the enzyme. A number of loops were identified surrounding the active site of *Bm*PAGM that may be responsible for proper positioning of the substrate and thus play a role in catalytic mechanism of the enzyme. The loops include, an active serine loop (T67-E71) that was reported to be a supplier of phosphate ion to the substrate, a metal binding loop (G281-R286) that interacts with magnesium ion, a sugar binding loop (E376-H380) that binds to hydroxyl groups of sugar moiety of the substrate and a phosphate binding loop (R501-A509) that binds to phosphate group of the substrate, which play a key role in opening and closing of the active site entrance (Fig. 5). Apart from the above mentioned conserved loop regions, there is another loop region (YGTXGFR) between residues 27-33 of *Bm*PAGM, which was observed to be well conserved among all the PAGMs considered for the study and reported to have a role in catalysis of the enzyme.

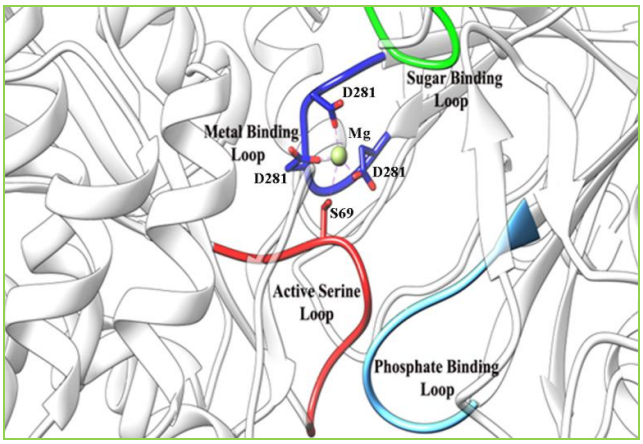


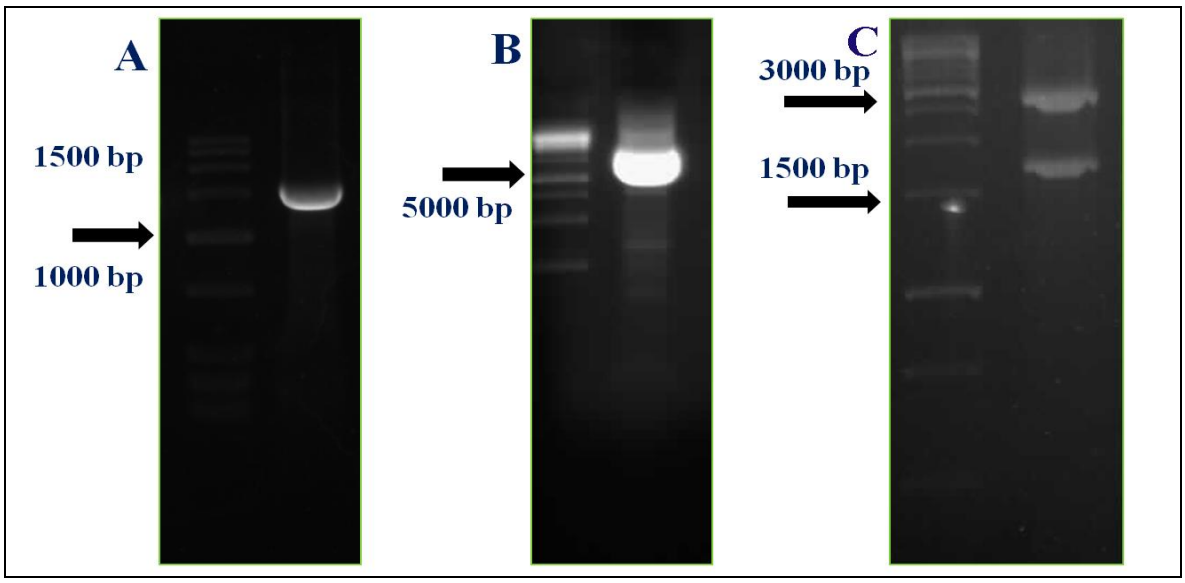
Fig 5: The active site of *Bm*PAGM shown in a cartoon representation specifying the functionally important motifs of protein in different colors: Active serine loop [red], Metal binding loop [blue], Sugar binding loop [green] and Phosphate binding loop [cyan]. Central metal ion in depicted as a green sphere and coordinate bonds as magenta sticks.

Supplementary Table 1: Details of various protein sequences used for phylogenetic analysis

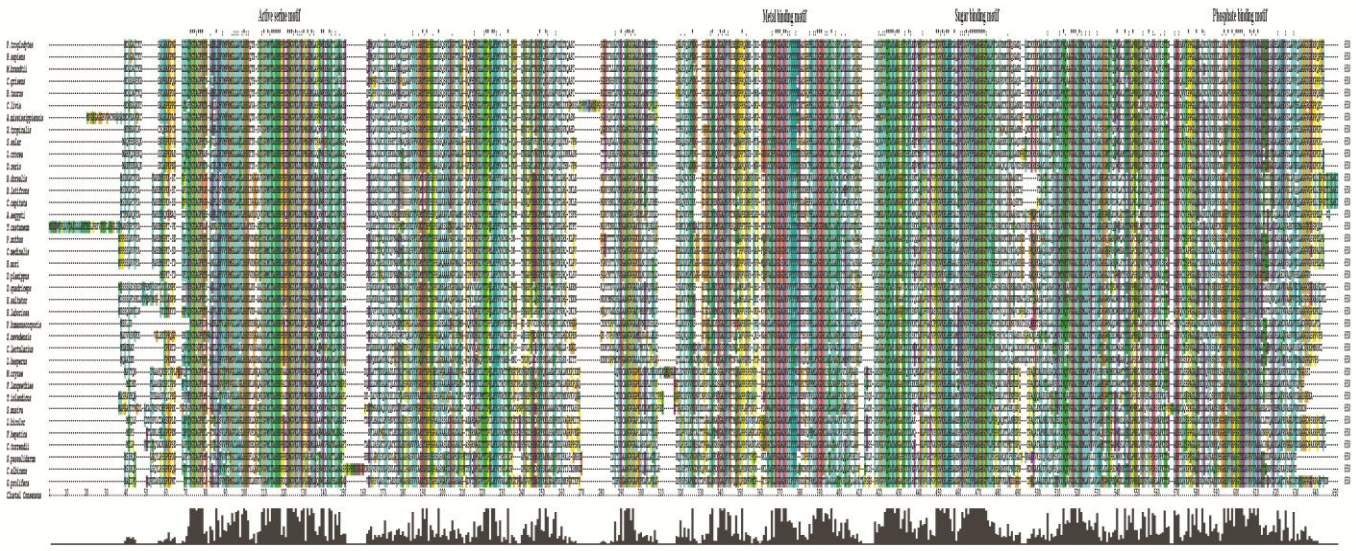
S. No.	Species	Accession Numbers	Sequence ID
1	Halyomorpha halys	XP_014287806.1	PAGM-H.halys
2	Bactrocera oleae	XP_014098249.1	PAGM-B.oleae
3	Diachasma alloeum	XP_015113928.1	PAGM-D.alloeum
4	Diuraphis noxia	XP_015375027.1	PAGM-D.noxia
5	Aedes aegypti	AAX47077.1	PAGM-A.aegypti
6	Tribolium castaneum	XP_973346.2	PAGM-T.castaneum
7	Polistes dominula	XP_015179219.1	PAGM-P.dominula
8	Polistes canadensis	XP_014603535.1	PAGM-P.canadenis
9	Dinoponera quadriceps	XP_014470489.1	PAGM-D.quadriceps
10	Cimex lectularius	XP_014242302.1	PAGM-C.lectularius
11	Trichogramma pretiosum	XP_014226477.1	PAGM-T.pretiosum
12	Copidosoma floridanum	XP_014210141.1	PAGM-C.floridanum
13	Habropoda laboriosa	KOC65411.1	PAGM-H. laboriosa
14	Amyelois transitella	XP_013185087.1	PAGM-A.transitella
15	Papilio xuthus	XP_013165770.1	PAGM-P. xuthus
16	Stomoxys calcitrans	XP_013116066.1	PAGM-S. calcitrans
17	Bombyx mori	XP_004933071.1	PAGM-B. mori
18	Monomorium pharaonis	XP_012525464.1	PAGM-M.pharaonis
19	Apis florea	XP_003692751.1	PAGM-A.florea
20	Orussus abietinus	XP_012272091.1	PAGM-O.abietinus
21	Athalia rosae	XP_012259264.1	PAGM-A.rosae
22	Linepithema humile	XP_012216594.1	PAGM-L.humile
23	Ceratitis capitata	XP_004525477.1	PAGM-C.capitata
24	Atta cephalotes	XP_012055177.1	PAGM-A.cephalotes
25	Wasmannia auropunctata	XP_011689802.1	PAGM-W.auropuncta
26	Pogonomyrmex barbatus	XP_011647842.1	PAGM-P.barbatus
27	Ceratosolen solmsi marchali	XP_011501168.1	PAGM-C.solmsi marchali
28	Cerapachys biroi	XP_011350617.1	PAGM-C.biroi
29	Musca domestica	XP_005181704.1	PAGM-M.domestica
30	Camponotus floridanus	XP_011257529.1	PAGM-C.floridanus
31	Solenopsis invicta	XP_011170522.1	PAGM-S.invicta
32	Harpegnathos saltator	XP_011151721.1	PAGM-H.saltator
33	Zootermopsis nevadensis	KDR17892.1	PAGM-Z.nevadensis
34	Drosophila rhopaloa	XP_016991351.1	PAGM-D.rhopaloa
35	Drosophila elegans	XP_017128684.1	PAGM-D.elegans
36	Drosophila bipectinata	XP_017100274.1	PAGM-D.bipectinata
37	Drosophila biarmipes	XP_016959204.1	PAGM-D.biarmipes
38	Drosophila eugracilis	XP_017070943.1	PAGM-D.eugracilis
39	Drosophila ficusphila	XP_017043445.1	PAGM-D.ficusphila
40	Drosophila miranda	XP_017135421.1	PAGM-D.miranda
41	Drosophila takahashii	XP_016994057.1	PAGM-D.takahashii

42	Drosophila kikkawai	XP_017016411.1	PAGM-D.kikkawai
43	Drosophila suzukii	XP_016932459.1	PAGM-D.suzukii
44	Apis cerana	XP_016919232.1	PAGM-A.cerana
45	Nasonia vitripennis	XP_016837507.1	PAGM-N.vitripennis
46	Apis mellifera	XP_016770126.1	PAGM-A.mellifera
47	Acyrthosiphon pisum	XP_016661082.1	PAGM-A.pisum
48	Trachymyrmex zeteki	KYQ46653.1	PAGM-T.zeteki
49	Trachymyrmex septentrionalis	KYN37087.1	PAGM-T.septentrionalis
50	Trachymyrmex cornetzi	KYN18387.1	PAGM-T.cornetzi
51	Lygus hesperus	JAQ06797.1	PAGM-L.hesperus
52	Neodiprion lecontei	XP_015523926.1	PAGM-N.lecontei
53	Dufourea novaeangliae	XP_015436643.1	PAGM-D.novaeangliae
54	Bactrocera latifrons	JAI38365.1	PAGM-B.latifrons
55	Bactrocera cucurbitae	JAD13128.1	PAGM-B.cucurbitae
56	Bactrocera dorsalis	JAC41694.1	PAGM-B.dorsalis
57	Bombus impatiens	XP_012236477.1	PAGM-B.impatiens
58	Bombus terrestris	XP_012173920.1	PAGM-B.terrestris
59	Megachile rotundata	XP_003702003.1	PAGM-M.rotundata
60	Vollenhovia emeryi	XP_011860744.1	PAGM-V.emeryi
61	Pediculus humanus corporis	XP_002431812.1	PAGM-P.humanuscorporis
62	Danaus plexippus	EHJ68938.1	PAGM-D.plexippus
63	Papilio xuthus	XP_013175542.1	PAGM-P.xuthus
64	Cephus cinctus	XP_015609819.1	PAGM-C.cinctus
65	Lasius niger	KMQ96669.1	PAGM-L.niger
66	Cnaphalocrocis medinalis	AJG44540.1	PAGM-C.medinalis
67	Apis dorsata	XP_006612374.1	PAGM-A.dorsata
68	Trypanosoma grayi	XP_009314520.1	PAGM-T.grayi
69	Arabidopsis thaliana	NP_568359.2	PAGM-A.thaliana
70	Homo sapiens	BAB00613.1	PAGM-H.sapiens
71	Talaromyces islandicus	CRG92507.1	PAGM-T.islandicus
72	Candida albicans	BAB00614.1	PAGM-C.albicans
73	Xenopus tropicalis	NP_001121416.1	PAGM-X.tropicalis
74	Fusobacterium nucleatum	WP_008796995.1	PGNM-F.nucleatum
75	Fusobacterium hwasookii	WP_005918224.1	PGNM-F.hwasookii
76	Streptococcus pneumoniae	CJP15950.1	PAGM-S.pneumoniae
77	Ktedonobacter racemifer DSM 44963	EFH85047.1	PAGM-K.racemifer
78	Laccaria bicolor S238N-H82	XP_001873786.1	PAGM-L.bicolor
79	Sphaerulina musiva SO2202	XP_016759679.1	PAGM-S.musiva
80	Spathaspora passalidarum NRRL Y-27907	XP_007375880.1	PAGM-S.passalidarum
81	Gonapodya prolifera JEL478	KXS16763.1	PAGM-G.prolifera
82	Arachis duranensis	XP_015963707.1	PAGM-A.duranensis
83	Cajanus cajan	KYP67026.1	PAGM-C.cajan
84	Lupinus angustifolius	XP_019424051.1	PAGM-L.angustifolius
85	Jatropha curcas	XP_012082839.1	PAGM-J.curcas
86	Vitis vinifera	XP_002281987.1	PAGM-V.vinifera
87	Salmo salar	NP_001133363.1	PAGM-S.salar
88	Danio rerio	NP_001007054.1	PAGM-D.rerio
89	Bos taurus	NP_001039387.1	PAGM-B.taurus
90	Alligator mississippiensis	KYO27998.1	PAGM-A.mississippiensis
91	Pan troglodytes	NP_001266726.1	PAGM-P.troglodytes
92	Rattus norvegicus	NP_001102242.1	PAGM-R.norvegicus
93	Bacteroides fragilis	WP_005795303.1	PGM-B.fragilis
94	Methylothermaceae bacteria B42	KXJ41420.1	PGM-M.bacteria
95	Cyanobacterium aponinum	WP_015220449.1	PGM-C.aponinum
96	Pseudanabaena biceps	WP_009628872.1	PGM-P.biceps
97	Stanieria cyanosphaera	WP_015191892.1	PGM-S.cyanosphaera
98	Rhizoctonia solani	CUA69810.1	PGM-R.solani
99	Talaromyces islandicus	CRG86074.1	PGM-T.islandicus
100	Sphaerulina musiva SO2202	XP_016759698.1	PGM-S.musiva
101	Gonapodya prolifera JEL478	KXS15004.1	PGM-G.prolifera
102	Coccidioides immitis RS	XP_001246071.1	PGM-C.immitis
103	Danaus plexippus	EHJ67776.1	PGM-D.plexippus
104	Spodoptera exigua	ACY69180.1	PGM-S.exigua
105	Drosophila melanogaster	NP_524675.1	PGM-D.melanogaster
106	Locusta migratoria	ACM78949.1	PGM-L.migratoria
107	Bombyx mori	XP_004923967.1	PGM-B.mori
108	Camponotus floridanus	XP_011254716.1	PGM-C.floridanus
109	Anopheles sinensis	KFB53136.1	PGM-A.sinensis
110	Culex quinquefasciatus	XP_001867120.1	PGM-C.quinquefasciatus
111	Trypanosoma grayi	XP_009309683.1	PGM-T.grayi
112	Arabidopsis thaliana	NP_199995.1	PGM-A.thaliana
113	Micromonas commoda	XP_002507519.1	PGM-M.commoda
114	Homo sapiens	NP_060760.2	PGM2-H.sapiens

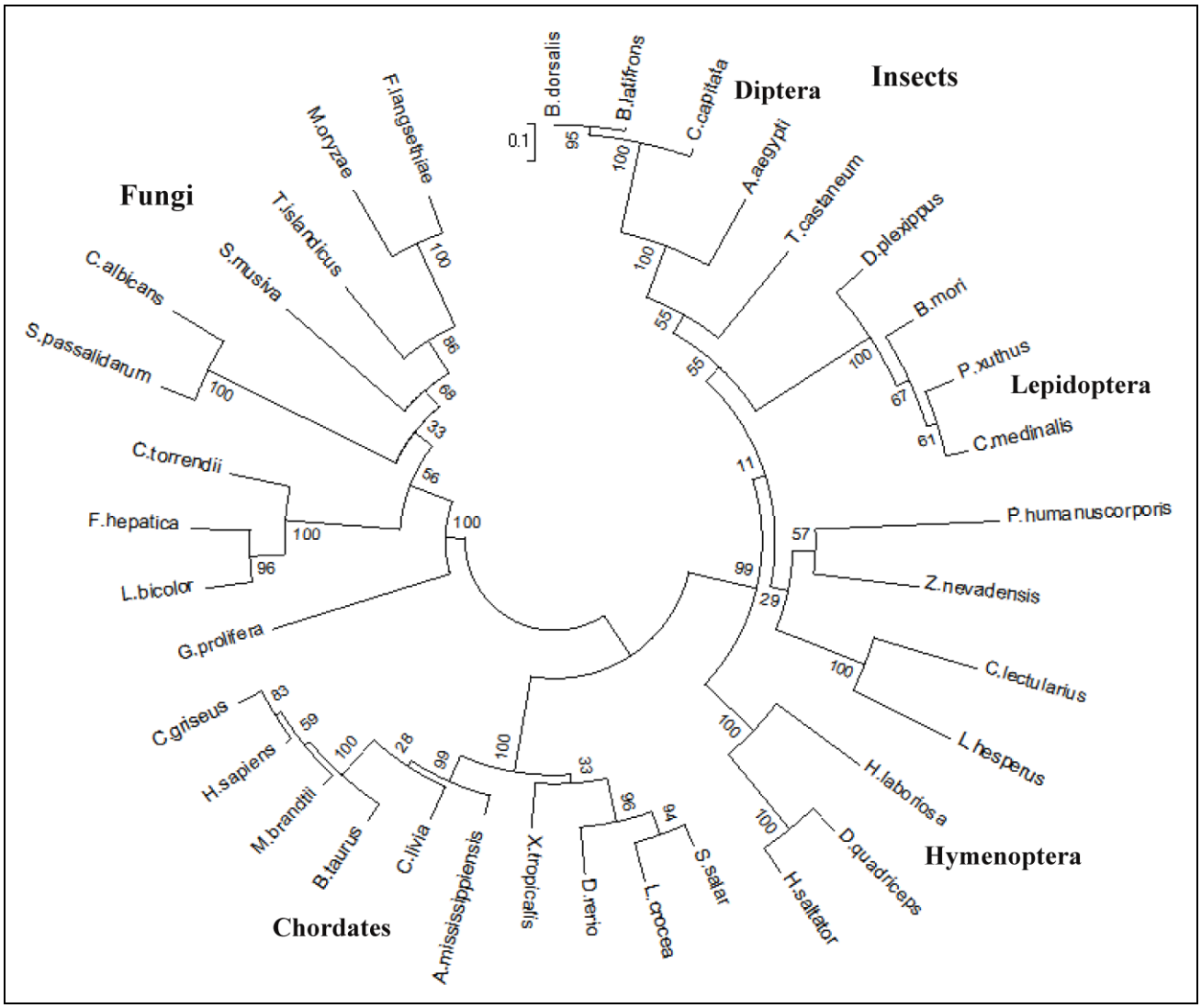
115	Mus musculus	NP_079976.1	PGM1-M.musculus
116	Pan troglodytes	XP_016806931.1	PGM2-P.troglodytes
117	Gallus gallus	AAQ10887.1	PGM1-G.gallus
118	Danio rerio	NP_998051.1	PGM2-D.rerio
119	Lasius niger	KMQ91532.1	PGNM-L.niger
120	Diplodia corticola	OJD35660.1	PGNM-D.corticola
121	Wickerhamomyces ciferrii	XP_011271771.1	PGNM-W-ciferrii
122	Madurella mycetomatis	KXX73599.1	PGNM-M.mycetomatis
123	Kitasatospora cheerisanensis KCTC 2395	KDN84753.1	PGNM-K.cheerisanensis
124	Methylothermaceae bacteria B42	KXJ41182.1	PGNM-M.bacteria
125	Bacillus amyloliquefaciens	WP_021495333.1	PGNM-B.amyloliquefaciens
126	Meiothermus silvanus	WP_013158636.1	PGNM-M.silvanus
127	Sulfobacillus acidophilus DSM 10332	AEW06229.1	PGNM-S.acidophilus
128	Deinococcus maricopensis	WP_013556190.1	PGNM-D.maricopensis
129	Desulfobacca acetoxidans	WP_013705309.1	PGNM-D.acetoxidans
130	Lasius niger	KMQ84614.1	PMM-L.niger
140	Acyrthosiphon pisum	NP_001280334.1	PMM-A.pisum
141	Anopheles darlingi	ETN61424.1	PMM-A.darlingi
142	Agrilus planipennis	XP_018325482.1	PMM-A.planipennis
143	Atta colombica	XP_018056053.1	PMM-A.colombica
144	Neodiprion lecontei	XP_015516832.1	PMM-N.lecontei
145	Harpegnathos saltator	XP_011147803.1	PMM-H.saltator
146	Sulfurimonas autotrophica DSM 16294	ADN08696.1	PMM-S.autotrophica
147	Ktedonobacter racemifer	WP_007909030.1	PMM-K.racemifer
148	Arcobacter nitrofigilis	WP_013133918.1	PMM-A.nitrofigilis
149	Desulfobulbus propionicus	WP_015723266.1	PMM-D.propionicus
150	Thermaerobacter marianensis DSM 12885	ADU51978.1	PMM-T.marianensis
151	Mortierella elongata AG-77	OAQ33705.1	PMM-M.elongata
152	Rhizoctonia solani	CUA78087.1	PMM-R.solani
153	Trypanosoma grayi	XP_009310361.1	PMM-T.grayi
154	Talaromyces islandicus	CRG87939.1	PMM-T.islandicus
155	Candida albicans SC5314	XP_721436.1	PMM-C.albicans
156	Spathaspora passalidarum NRRL Y-27907	XP_007372278.1	PMM-S.passalidarum
157	Malpighia glabra	ACN54046.1	PMM-M.glabra
158	Camelina sativa	XP_010506558.1	PMM-C.sativa
159	Raphanus sativus	XP_018466921.1	PMM-R.sativus
160	Brassica napus	XP_013734824.1	PMM-B.napus
161	Brassica rapa subsp. chinensis	AET14211.1	PMM-B.rapa.chinensis
162	Prunus avium	APO15251.1	PMM-P.avium
163	Homo sapiens	AAC51117.1	PMM-H.sapiens
164	Mus musculus	NP_058577.1	PMM2-M.musculus
165	Rattus norvegicus	NP_001100443.1	PMM2-R.norvegicus
166	Macaca fascicularis	NP_001272018.1	PMM1-M.fascicularis
167	Xenopus tropicalis	NP_001135677.1	PMM1-X.tropicalis
168	Ictalurus punctatus	NP_001187806.1	PMM2-I.punctatus
169	Danio rerio	NP_956378.1	PMM2-D.rerio
170	Macaca mulatta	NP_001244362.1	PMM2-M.mulatta
171	Bos taurus	NP_001069551.1	PMM1-B.taurus
172	Gallus gallus	NP_001026639.1	PMM2-G.gallus



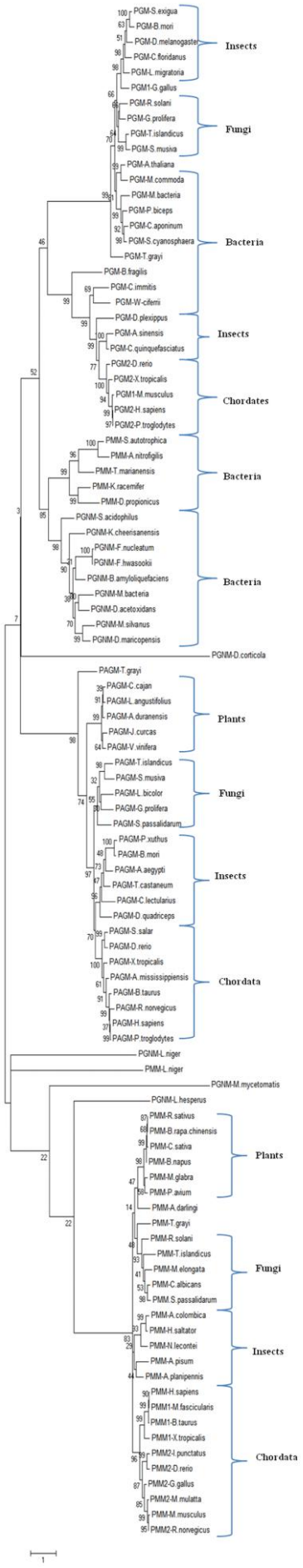
Supplementary Fig 1: A) Agarose gel electrophoresis image of amplified PCR product of *Bm*PAGM cDNA. B) Isolated plasmid sample after cloning into DH5α cells of *Bm*PAGM. C. Insert release after restriction digestion. Lane 1: molecular weight marker; Lane 2: Product.



Supplementary Fig 2: Multiple sequence alignment of various PAGMs.



Supplementary Fig 3: Phylogenetic tree of various vertebrate, invertebrate and fungal PAGMs constructed by Maximum Likelihood method. Bootstrap percentages of 1000 resampled datasets are shown at nodes.



Supplementary Fig 4: Phylogenetic tree of various phosphohexomutase family proteins collected from various domains of life viz., bacteria, fungi, plants and animals; constructed by Maximum Likelihood method.

4. Conclusions

In the present study, a full-length protein coding sequence of *Bm*PAGM was amplified, cloned and characterized. Semi-quantitative gene expression analysis revealed that the expression is abundant in molting larvae than in feeding larvae, and expression is higher in epidermis compared to other tissues, indicating that the enzyme has critical role in molting. *In silico* analysis suggested that the deduced protein sequence is "cytoplasmic" with no signal peptide. MSA of various phosphohexomutase protein families helped in identifying conserved and class specific motifs. Phylogenetic analysis revealed the evolutionary relatedness between counter parts of the phosphohexomutase family and their high homology in different species across various domains of life, thus suggesting that all enzymes might share similar protein folding and catalytic mechanisms. Since no crystal structure of insect PAGMs was available, the theoretical 3D model of *Bm*PAGM was built by homology modeling and validated by performing MD simulation studies that helped in understanding the structural architecture of *Bm*PAGM. Though the final 3D model of *Bm*PAGM aligns well with template structure, minute variations were observed in the secondary structural elements that may not affect the function majorly. Further, molecular docking, *in vitro* and *in vivo* studies will provide useful information regarding the catalytically important interactions and sheds light on the possibility of considering the protein as a drug target in controlling insect pests.

5. Disclosure

The authors declare that they have no conflict of interest.

6. Acknowledgments

Sincere thanks to the DST (SERB/F/8214/2015-16), UGC (F.No. 42-206/2013(SR)), New Delhi, for financial support. Thanks to Centre for Bioinformatics, Pondicherry University for providing computational and laboratory facilities.

7. References

1. Liu X, Li F, Li D, Ma E, Zhang W, Zhu KY *et al*. Molecular and functional analysis of UDP-N-acetylglucosamine pyrophosphorylases from the migratory locust, Locusta migratoria. PloS one, 2013. 8(8):e71970.
2. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. Journal of Experimental Biology. 2003; 206(24):4393-4412.
3. Muthukrishnan S, Merzendorfer H, Arakane Y, Kramer KJ. 7 Chitin Metabolism in Insects. Insect molecular biology and biochemistry, 2012.
4. Zhuo W, Fang Y, Kong L, Li X, Sima Y, Xu S. Chitin synthase A: a novel epidermal development regulation gene in the larvae of Bombyx mori. Molecular biology reports. 2014; 41(7):4177-4186.
5. Cohen E. Chitin synthesis and degradation as targets for pesticide action. Archives of insect biochemistry and physiology. 1993; 22(1-2):245-261.
6. Hirose T, Sunazuka T, Ōmura S. Recent development of two chitinase inhibitors, Argifin and Argadin, produced by soil microorganisms. Proceedings of the Japan Academy, Series B. 2010; 86(2):85-102.
7. Merzendorfer H. The cellular basis of chitin synthesis in fungi and insects: common principles and differences. European journal of cell biology. 2011; 90(9):759-769.

8. Becker A, Schlöder P, Steele J, Wegener G. The regulation of trehalose metabolism in insects. Experientia. 1996; 52(5):433-439.
9. Kramer K, Muthukrishnan S. Chitin metabolism in insects. Insect development: morphogenesis, molting and metamorphosis/edited by Lawrence I. Gilbert, 2009.
10. Muthukrishnan S, Merzendorfer H, Arakane Y, Yang Q. *Chitin Metabolic Pathways in Insects and Their Regulation*, in *Extracellular Composite Matrices in Arthropods*. Springer. 2016, 31-65.
11. Passonneau JV, Williams CM. The moulting fluid of the Cecropia silkworm. Journal of Experimental Biology. 1953; 30(4):545-560.
12. Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J *et al*. Chitinases: an update. Journal of Pharmacy and Bioallied Sciences. 2013; 5(1):21.
13. Correa JU, Elango N, Polacheck I, Cabib E. Endochitinase, a mannan-associated enzyme from Saccharomyces cerevisiae. Journal of Biological Chemistry. 1982; 257(3):1392-1397.
14. Kuranda MJ, Robbins PW. Chitinase is required for cell separation during growth of Saccharomyces cerevisiae. Journal of Biological Chemistry, 1991; 266(29):19758-19767.
15. Kramer KJ, Muthukrishnan S. Insect chitinases: molecular biology and potential use as biopesticides. Insect biochemistry and molecular biology, 1997; 27(11):887-900.
16. Herrera-Estrella A, Chet I. Chitinases in biological control. EXS-BASEL-, 1999; 87:171-184.
17. Mio T, Kokado M, Arisawa M, Yamada-Okabe H. Reduced virulence of Candida albicans mutants lacking the GNA1 gene encoding glucosamine-6-phosphate acetyltransferase. Microbiology. 2000; 146(7):1753-1758.
18. Nishitani Y, Maruyama D, Nonaka T, Kita A, Fukami TA, Mio T *et al*. Crystal structures of N-acetylglucosamine-phosphate mutase, a member of the α-D-phosphohexomutase superfamily, and its substrate and product complexes. Journal of Biological Chemistry. 2006; 281(28):19740-19747.
19. Shackelford GS, Regni CA, Beamer LJ. Evolutionary trace analysis of the α-D-phosphohexomutase superfamily. Protein Science. 2004; 13(8):2130-2138.
20. Mehra-Chaudhary R, Mick J, Tanner JJ, Henzl MT, Beamer LJ. Crystal structure of a bacterial phosphoglucomutase, an enzyme involved in the virulence of multiple human pathogens. Proteins: Structure, Function, and Bioinformatics. 2011; 79(4):1215-1229.
21. Liu Y, Ray W, Baranidharan S. Structure of rabbit muscle phosphoglucomutase refined at 2.4 Å resolution. Acta Crystallographica Section D: Biological Crystallography. 1997; 53(4):392-405.
22. Lee Y, Villar MT, Artigues A, Beamer LJ. Promotion of enzyme flexibility by dephosphorylation and coupling to the catalytic mechanism of a phosphohexomutase. Journal of Biological Chemistry, 2014; 289(8):4674-4682.
23. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical biochemistry. 1987; 162(1):156-159.
24. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. 2013; 30(12):2725-2729.
25. Webb B, Sali A. Protein structure modeling with MODELLER. Protein Structure Prediction, 2014, 1-15.
26. Laskowski R, MacArthur M, Thornton J. PROCHECK: validation of protein structure coordinates. International Tables of Crystallography, Vol. F. Crystallography of Biological Macromolecules. Kluwer Academic Publishers, The Netherlands, 2001, 722-725.
27. Colovos C, Yeates T. ERRAT: an empirical atom-based method for validating protein structures. Protein Science. 1993; 2:1511-1519.
28. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods in enzymology. 1997; 277:396-404.
29. DeLano WL. The PyMOL molecular graphics system, 2002.
30. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC *et al*. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004; 25(13):1605-1612.
31. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B *et al*. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 2015; 1:19-25.
32. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. Journal of computational chemistry, 1997; 18(12):1463-1472.
33. Berendsen HJ, Postma Jv, van Gunsteren WF, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. The Journal of chemical physics. 1984; 81(8):3684-3690.
34. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods. 2011; 8(10):785-786.
35. Hofmann K, Stoffel W. TMbase-A database of membrane spanning protein segments, 1993.
36. Nancy YY, Wagner JR, Laird MR, Melli G, Rey S, Lo R *et al*. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 2010; 26(13):1608-1615.
37. Klee EW, Ellis LB. Evaluating eukaryotic secreted protein prediction. BMC bioinformatics, 2005; 6(1):256.
38. Bhasin M, Raghava GP. ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. Nucleic Acids Research, 2004; 32(Web Server issue):W414-9.
39. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer applications in the biosciences: CABIOS. 1995; 11(6):681-684.
40. Arakane Y, Baguinon MC, Jasrapuria S, Chaudhari S, Doyungan A, Kramer KJ *et al*. Both UDP N-acetylglucosamine pyrophosphorylases of Tribolium castaneum are critical for molting, survival and fecundity. Insect biochemistry and molecular biology. 2011; 41(1): 42-50.
41. Zhao L, Yang M, Shen Q, Liu X, Shi Z, Wang S *et al*. Functional characterization of three trehalase genes regulating the chitin metabolism pathway in rice brown planthopper using RNA interference. Scientific reports,

2016; 6:27841.
42. Xu L, Liang H, Niu Y, Wang Y, Sima Y, Xu S. Differential Expression Of The Bombyx Mori Diapause-Termination Timer Gene Ea4 In Diapause-Inducing Temperature And Photoperiod. Archives of insect biochemistry and physiology, 2012; 79(3):182-194.
43. Mehra-Chaudhary R, Mick J, Beamer LJ. Crystal structure of Bacillus anthracis phosphoglucosamine mutase, an enzyme in the peptidoglycan biosynthetic pathway. Journal of bacteriology. 2011; 193(16):4081-4087.
44. Bandini G, Mariño K, Güther M, Wernimont AK, Kuettel S, Qiu W *et al*. Phosphoglucomutase is absent in Trypanosoma brucei and redundantly substituted by phosphomannomutase and phospho-N-acetylglucosamine mutase. Molecular microbiology, 2012; 85(3):513-534.