



Characterization and Prokaryotic expression of Glucuronyltransferase-S Gene in Silkworm *Bombyx mori*

Yukun Chen

Institute of Life Sciences, Jiangsu University

301 Xuefu Road, Zhenjiang 212013, China

Tel: 86-511-88972522 E-mail: akuncxws@163.com

KePing Chen (Corresponding author), Fang Bao, Qin Yao

Institute of Life Sciences, Jiangsu University

301 Xuefu Road, Zhenjiang 212013, China

Tel: 86-511-88791923 E-mail: kpchen@ujs.edu.cn

This project financed is by the National Natural Science Foundation of China (NO.30871826) and the National "863" project of China (2008AA10Z145)

Abstract

As genome of *B. mori* is available in GenBank, identification of novel genes of *B. mori* can be carried out. In this study, we used the *in silico* cloning method to obtain the Glucuronyltransferase-S (GlcAT-S) gene of *B. mori* and analysed with bioinformatics tools. The result was confirmed by RT-PCR, prokaryotic expression and western blot. The GlcAT-S cDNA contains a 843bp ORF and has three exons. The deduced protein has 280 amino acid residues, with the predicted molecular weight of 31842.02 Da, isoelectric point of 9.16, and contains conserved GlcAT domains. The protein shows high degrees of identity with that of some homologous protein from other species.

Keywords: GlcAT-S, *Bombyx mori*, Bioinformatics, RT-PCR, Prokaryotic expression

1. Introduction

Glycosylation is one of the major post-translational protein modifications that play important roles in a variety of cellular functions, including recognition and adhesion (Hideki, 2005, pp. 23876–23883). Glucuronyltransferases form a gene family and belong to Glycosyltransferase superfamily (Takashi, 1999, pp. 182–187). This enzyme accepts a wide range of substrates, including phenols, alcohols, amines, and fatty acids.

The HNK-1 carbohydrate epitope is characteristically expressed on a series of cell adhesion molecules and also on some glycolipids in the nervous system over a wide range of species from insect to mammal and is postulated to be associated with neural crest cell migration, neuron-to-glia cell adhesion and preferential outgrowth of neurites from motor neurons (Koji, 1997, pp. 6093–6098). Researchers believe that Glucuronyltransferases are the key enzymes involved in the biosynthesis of this carbohydrate epitope. Glucuronyltransferase-P deficient mice exhibited reduced long-term potentiation (LTP) at the Schaffer collateral-CA1 synapses and defects in spatial memory formation (Shinako, 2004, pp. 111–119). The *Drosophila melanogaster* genome contains three glucuronyltransferases which are involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans (Byung, 2003, pp. 9116–9124). In variety of biological processes such as cell proliferation, cell-cell adhesion and tissue morphogenesis, proteoglycans play an essential role.

In the paper, we cloned the Glucuronyltransferase-S of *B. mori* using data-mining techniques and analysed with bioinformatics tools, including its genomic organization and the deduced amino acid sequence. The gene was confirmed by RT-PCR, prokaryotic expression and western blot. The deduced protein sequence was compared with that of some homologous protein from other species.

2. Materials and methods

2.1 Materials

The silkworm *Bombyx mori* and *E. coli* (strain TG1 and BL21) were inbred in our lab. Silkworm strain306 were used for this study. Restrictases, T4 DNA ligase, PCR reagents and pMD18-T were obtained from TaKaRa Company (Dalian); primer and other reagents were obtained from Shanghai Sangon Bio-technology Corporation.

2.2 Data extraction of cDNA sequence of *B. mori* GlcAT-S gene

The NCBI's (<http://www.ncbi.nlm.nih.gov/>) EST database is a popular starting point for identifying expressed sequence tags (ESTs) of different species, and more than 110, 000 *B. mori* EST sequences are currently available in GenBank. We also used another silkworm cDNA database BGI (<http://silkworm.genomics.org.cn/>) in this study. Data-mining techniques and bioinformatics tools were applied to search for cDNA sequence of the *B. mori* GlcAT-S gene by repeated cycles of assembling and extending EST sequence.

2.3 Genome analysis

In order to establish the genomic organization, we blasted the cDNA sequence to the contigs of *B. mori* genome in GenBank. SIM4 (<http://pbil.univ-lyon1.fr/sim4.php>) was used to align the cDNA sequence with the genomic sequences to search potential introns.

2.4 Protein prediction and analysis

We used the ExpASY Translate tool (<http://au.expasy.org/tools/dna.html>) to deduce the cDNA's amino acid sequence, and similarity analysis was performed using the BLAST tool in GenBank. We used PLOC (<http://www.genome.jp/SIT/ploddir/>) to predict the subcellular location of *B. mori* GlcAT-S. Using InterPro Scan (<http://www.ebi.ac.uk/InterProScan/>) to analyse the deduced amino acid sequence.

2.5 RT-PCR

A pair of specific primers was designed based on the sequence we obtained. The forward primer (5'-GGGGATCCATGATATATTACATCACGCCGACTT-3') contained a *Bam*HI restriction site (underlined), and the reverse primer (5'-CCCTCGAGTACCAAAGTGGGCAGAAGGCT-3') contained a *Xho*II restriction site (underlined). cDNA was prepared from midgut RNA with M-MLV reverse transcriptase and an oligodT primer. PCR reaction was carried out with Taq polymerase for 35 amplification cycles (94°C for 45 sec; 58°C for 45 sec; 72°C for 1min). PCR product was examined by electrophoresis in 1% agarose gel with the ethidium bromide staining.

2.6 Cloning and sequencing

The PCR product was ligated into pMD18-T vector using T4 DNA ligase and then transformed into *E. coli* (TG1 strain). Plasmid was purified with MiniBEST Plasmid Purification Kit (Takara). The sequencing was performed using an automatic sequence: CEQ8000 (Beckman company).

2.7 Construction of expression plasmid

The plasmid pMD18-T/GlcAT-S was digested with *Bam*HI and *Xho*II, and then purified. The purified fragment was ligated with the *Bam*HI-*Xho*II digested His-pET30a vector and transformed into *E. coli* (BL21 strain). The transformants harboring the recombinant plasmid were confirmed by restriction enzyme analysis.

2.8 Expression of fusion protein in *E. coli* and SDS-PAGE

For expression of recombinant protein, a positive clone was cultured in LB medium supplement with Kanamycin (50µg/ml) overnight at 37 °C with shaking. This culture was added into fresh LB medium and cultured at 37 °C with vigorous shaking to A_{600} about 0.6. The culture was then induced with isopropyl-thio-β-D-galactopyranosi (final concentration of 0.2 mmol/L) and further cultured for another 10 hours. 15% SDS polyacrylamide gel was used to analyze the recombinant protein. SDS-PAGE was performed in the Mini-Protein system (Bio-Rad, USA). After electrophoresis, the gel was stained with Coomassie Brilliant Blue R250 to visualize the protein bands.

2.9 Western blot analysis

Western blot analysis of proteins was performed following the standard protocol as described previously (Kar, 2004,pp.387–399).

3. Result

3.1 Nucleotide sequence analysis

A 996 bp cDNA was obtained through data-mining techniques. The cDNA sequence included an open reading frame (ORF), beginning with the initiation codon ATG at position 1 and ending with a termination codon TAA at position 843 (Figure 1). The cDNA also included a TATA-box (Figure 1). BLASTing the cDNA to contigs of *B. mori* genome in GenBank revealed that contig049159 and contig022658 having a high similarity. Using SIM4, three exons were found

in the relevant DNA sequence. The length of the exons were 153bp, 270bp and 420bp (Figure 2).

3.2 Cloning and identification of *GlcAT-S*

PCR amplification of the midgut cDNA was performed using the two specific primers. After electrophoresis in 1% agarose gel, we obtained a band about 840bp, which was consistent with the expected molecular mass. The PCR product was ligated into pMD18-T vector and confirmed by restriction endonuclease digestion and DNA sequencing. The *GlcAT-S* fragment could be isolated from the pMD18-T vector after the recombinant plasmid was digested with *Bam*HI and *Xho*II (Figure 3).

3.3 Analysis of deduced amino acid sequence

The open reading frame encoded a protein of 280 amino acids, as deduced from the nucleotide sequence (Figure 1). The molecular weight of the encoded protein was predicted to be 31842.02Da with an isoelectric point of 9.16.

The deduced amino acid sequence was analyzed using InterPro Scan software in the ExPASy website. The results indicated that the deduced amino acid sequence had a domain named Glycosyltransferase (IPR005027). The subcellular location of *B. mori* *GlcAT-S* was predicted in the cytoplasm. Hydropathy analysis showed that *GlcAT-S* is not a typical type II transmembrane protein. Using BLAST software of NCBI, we obtained a Conserved *GlcAT* domain (Figure 4). Using BLAST software of NCBI to search for homology in the GenBank database, the deduced amino acid sequence showed an identity of 53%, 59%, 47%, 42%, 44%, 44% to the corresponding genes of *Aedes aegypti* (EAT43602), *Anopheles gambiae* (EAA10323), *Drosophila melanogaster* (NP_723476), *Homo sapiens* (NP_001032391), *Oryzias latipes* (CAI68028), *Rattus norvegicus* (NP_072131) respectively (Figure 5).

3.4 Construction of expression plasmid

The plasmid pMD18-T/*GlcAT-S* was digested with *Bam*HI and *Xho*II and ligated with His- pET30a which was also digested with the same restriction enzymes to generate His-*GlcAT-S*- pET30a. The *GlcAT-S* fragment could be isolated from the His-*GlcAT-S*-pET30a vector after the recombinant plasmid was digested with *Bam*HI and *Xho*II (Figure 6). The recombinant plasmid His-*GlcAT-S*-PET30a was successfully constructed (Figure 6).

3.5 SDS-PAGE and Western blotting analysis

IPTG induced the *E. coli* BL21 transformed with the His-*GlcAT-S*-PET30a plasmid to express the His-*GlcAT-S* recombinant fusion protein. Recombinant His-*GlcAT-S*-pET30a include a 36kDa band that is absent from pET-30a (compared lanes 1 and 2). This corresponds to the size expected for the His-*GlcAT-S* fusion protein. Immunoblot analysis with antiserum specific for His confirmed that this band includes His and *GlcAT-S* (Figure 7).

4. Discussion

Cell surface carbohydrate modulate a variety of cellular functions, including recognition and adhesion (Koji *et al* 1997). Glucuronyltransferases can transfer glucuronic acid from UDP glucuronic acid to the specific substrate and are involved in the biosynthesis of cell surface carbohydrate.

In this study, we identified a novel *B. mori* *GlcAT-S* gene through bioinformatics approaches, then RT-PCR, sequencing, prokaryotic expression and western blot were used to confirm the result. The gene has a 843bp ORF encoding 280 amino acid with molecular weight of 31.8 kDa, isoelectric point of 9.16 and was predicted to locate in cytoplasm. Hydropathy analysis showed that *GlcAT-S* is not a typical type II transmembrane protein which is different from other glycosyltransferases. Analysis of the deduced amino acid indicated that *B. mori* *GlcAT-S* has a high homology with the *GlcAT-S* sequences of other species, suggesting that *B. mori* *GlcAT-S* might be a *GlcAT-S* ortholog in *B. mori*. A conserved catalytic domain *GlcAT* was also found.

GlcAT-S were reported in many species, but not in *B. mori*. This is the first time this gene was cloned and analyzed in *B. mori*. As an economic and model insect of Lepidoptera, research for *B. mori* and related genes has attracted more and more attention. Specially, the *GlcAT-S* gene may have so many functions. The cloning and analysis can help us to do further study about this gene.

References

- Byung-T.K., K.T., John Lincecum, Hiroshi Kitagawa, K. S. & Merton Bernfield. (2003). Identification and Characterization of Three *Drosophila melanogaster* Glucuronyltransferases Responsible for the Synthesis of the Conserved Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *Biological Chemistry* 11, 9116–9124.
- Hideki, T., Y. K., Tomoko Ikeda, Satsuki Itoh, Nana Kawasaki, M. L. O., Hidetake Kurihara, Akihiro Tojo & Tatsuo Sakai. (2005). A Non-sulfated Form of the HNK-1 Carbohydrate Is Expressed in Mouse Kidney. *Biological Chemistry* 25, 23876–23883.
- Kar, A. K., M., Ghosh, & P. Roy. (2004). Mapping the assembly of Bluetongue virus scaffolding protein VP3. *Virology* 324, 387–399.

Koji, T.Y., S. O., Takatal, Y.K. MIKI, A. N & A. T. K. (1997). Cloning and functional expression of a novel glucuronyltransferase involved in the biosynthesis of the carbohydrate epitope HNK-1. *Biochemistry* 94, 6093–6098.

Shinako Kakuda ,S. O. & T. K. (2004). Purification and characterization of two recombinant human glucuronyltransferases involved in the biosynthesis of HNK-1 carbohydrate in Escherichia coli. *Protein Expression and Purification* 35 ,111–119.

Takashi Seiki, S. O., Koji, Terayama, Kimiyuki Imiya. & Toshisuke Kawasaki. (1999). Molecular Cloning and Expression of a Second Glucuronyltransferase Involved in the Biosynthesis of the HNK-1 Carbohydrate Epitope. *Biochemical and Biophysical Research Communications* 255, 182–187.

```

-152                               GCCACGGCAAATAATCTGTGCAGCTTGAATGGA
AGGCCACTGCATGTGAAGAGCAGATTGTGTGCATGTAGACACCGCCCTGATATCAAAAATG
GTTAATTTAATTGACGGCAGAAATCATAACGAAGC TATAGCGAATAAGAGCGATTTGAAA
1 ATGATATATTACATCACGCCGACTTATCCAAGACCAGAACAGATTCCAGAGTAACTAGG
M I Y Y I T P T Y P R P E Q I P E L T R
61 CTTGGCCACACATTGATGCACGTGCCTCGGATACATTGGATCATTGCAGACGCCAGTCT
L G H T L M H V P R I H W I I A D D Q S
121 TTGTGTTCAACCAATGTTCTCAATTTACTTAGGCGGACGGGCTTGCCATTACACATATT
L C S T N V L N L L R R T G L P F T H I
181 TCCAGTCCAAAGCCATACGTGTACAAAGGTACAAACTTCCCGCGTGGCGTTTCGAACAGA
S S P K P Y V Y K G T N F P R G V S N R
241 CGTGCAGCCCTGGTTTGGTTGCGGGAGAACGTTTCGAGAAGGAGTCATGTACTTTGGCGAT
R A A L V W L R E N V R E G V M Y F G D
301 GACGATAATACAGTAGACTTACAGCTATTCGATGAAATAAGGCGTACCAAGAAAGGTATCC
D D N T V D L Q L F D E I R R T K K V S
361 ATGTTTCCCGTAGGGCTCATTGGTATTACGGTATTTCCGGCGCCCATCATCAAAGACGGA
M F P V G L I G D Y G I S A P I I K D G
421 AAGGTCGTCGCCTTTTTCGACTCATGGCCGGTCTAGAACATTCCCTGTCGATATGGCA
K V V A F F D S W P G S R T F P V D M A
481 GGATTGCTGTCAACATAGAGTTCTGACTCCGACGGCCACAATGCCTTACTCAGCTGGC
G F A V N I E F L T P T A T M P Y S A G
541 CACGAGGAAGATAAGTTTTAATGAGTTTAGGAATAAAATTAGACGACATTGAACCGTTA
H E E D K F L M S L G I K L D D I E P L
601 GCCGATAATTGTTCAAAGGTTTTAGTTTGGCACACGAAAAGTAAAGTTTAAAAAACC
A D N C S K V L V W H T K T V K F K K P
661 AACCTTAAAATCGATATCGAAAGAATTAGCAAGTTACCGAAGTACGAGGATTTGCTAAT
N L K I D I E R I S K L P K Y E D F A N
721 CTTTTAAAAGAACTTCTAGTTAGGTATGGCCGGAATATTCTCGAAAAATGGTAGTAAA
L L K E T S R L G M A G I F S K N G S K
781 ACTTTCATAATTAAGATAGGAAGACCTTTGAACCCTTAAGCCTTCTGCCACTTTGGTATAA
T F I I K D R K T F E P L S L L P T L V -
    
```

Figure 1. Nucleotide sequence and deduced amino acid sequence. The predicted amino acid is represented by the one letter code designation below the nucleotide sequence. The initiate codes and the stop codes are framed. The TATA-box is italic and underlined.



Figure 2. DNA sequence frame of the GlcAT-S gene. Exons are black framed, and introns are white framed.

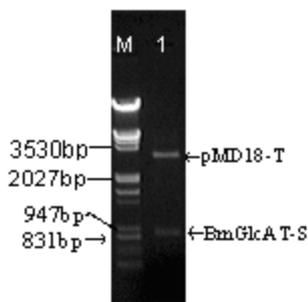


Figure 3. Identification of the recombinant plasmid pMD18-T/GlcAT-S. Lane 1, pMD18-T/ GlcAT-S digested with *Bam*HI and *Xho*II generated two fragments: pMD18-T (2.6 Kbp) and GlcAT-S (840 bp); M, DNA molecular mass maker.



Figure 4. The conserved superfamily domain in B.mori GlcAT

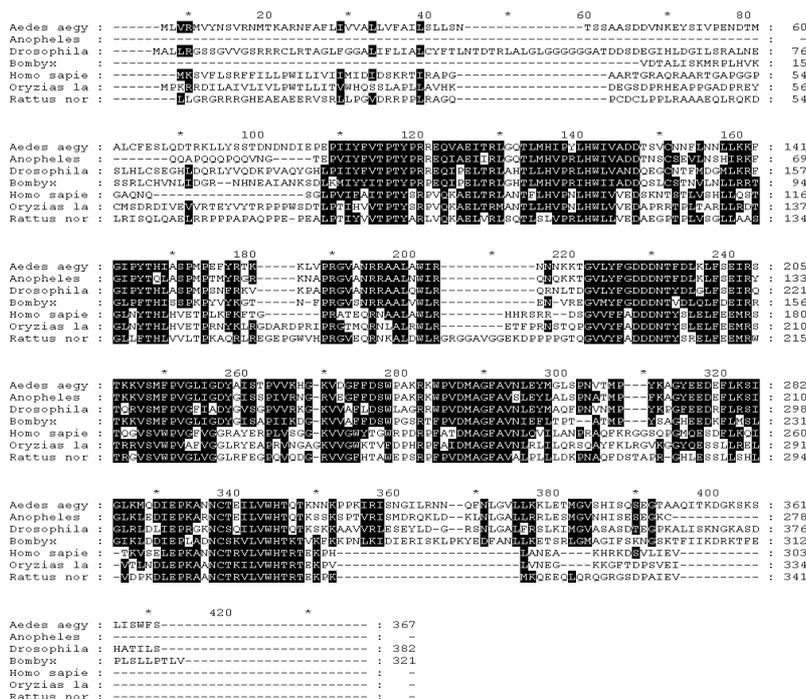


Figure 5. Multiple sequence alignment of B.mori GlcAT-S protein with the corresponding sequences from other six species, *Aedes aegypti* (EAT43602), *Anopheles gambiae* (EAA10323), *Drosophila melanogaster* (NP_723476), *Homo sapiens* (NP_001032391), *Oryzias latipes* (CAI68028), *Rattus norvegicus* (NP_072131) using DNASTar MegAlign program.

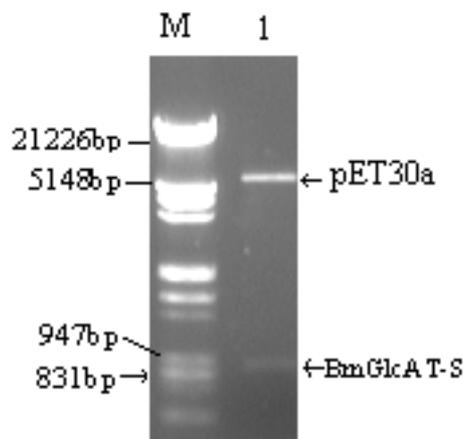


Figure 6. Identification of the expression vector His-GlcAT-S-pET30a. Lane 1, His-GlcAT-S-pET30a digested with *Bam*HI and *Xho*II generated two fragments: His-pET30a (5.4Kbp) and GlcAT-S (840 bp); M, DNA molecular maker.

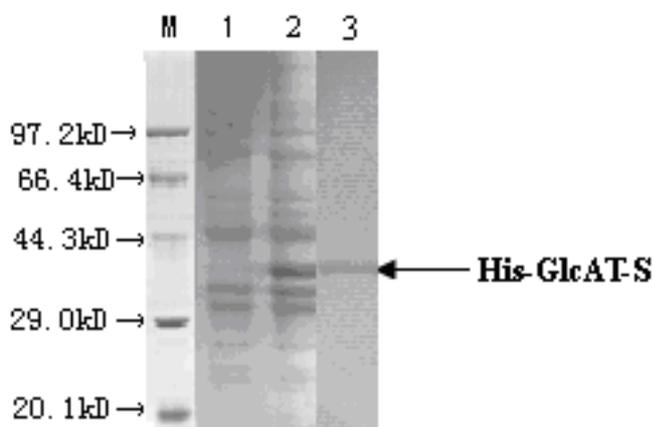


Figure 7. SDS-PAGE and Western blotting analysis. M, Protein marker; Lane 1, Protein of *E. coli* BL21 contained His-pET30a induced by IPTG; Lane 2, Protein of *E. coli* BL21 contained His-GlcAT-S-pET30a induced by IPTG; Lane 3, Western blot results of His-GlcAT-S fusion protein. The fusion protein bands were indicated by arrows.