Classification of LARGE-like GlcNAc-transferases of *Dictyostelium discoideum* by Phylogenetic Analysis

Kuo-Yuan Hwa  
Center for Biomedical Industries, Department of Molecular Science and Engineering  
National Taipei University of Technology, Taipei, Taiwan, ROC  
Institute of Medical Technology  
Taipei Medical University, Taipei, Taiwan, ROC  
e-mail:kyhwa@ntut.edu.tw

Te-Ling Pang  
Institute of Biochemistry and Molecular Biology, National Yang-Ming University School of Life Sciences, Taipei, Taiwan, ROC

Mei-Yu Chen  
Institute of Biochemistry and Molecular Biology, National Yang-Ming University School of Life Sciences, Taipei, Taiwan, ROC

Previously, we have identified four genes, gnt12, gnt13, gnt14, and gnt15 from *D. discoideum*, as GlcNAc-transferase [4]. Our annotation is based on the sequences similarity to human and mouse LARGE proteins based on BLASTP [5]. LARGE proteins have been annotated as GlcNAc-transferase [6]. Traditionally, glycosyltransferases are classified based on the enzymatic activities according to the Enzyme Commission (http://www.chem.qmul.ac.uk/ubmb/ enzyme/). Many families of GlcNAc-transferase are defined by the EC number [EC 2.4.1.x.x.]. However, classification of proteins based on biochemical evidence is a difficult process, especially for membrane proteins, such as glycosyltransferases. Furthermore, to purify glycosyltransferase is not a regular conduct for most of laboratories. On the other hand, with the recent advancement of genome projects, DNA sequence of an organism is readily available. And, bioinformatics annotation tools are commonly used by life researchers. Hence, new approaches based on bioinformatics tools for classifying glycosyltransferase have been proposed. The best known database is the CAZy database [7]. In the CAZy database, glycosyltransferases are classified as families, clans, and folds based on their structural and sequence similarities, and on their mechanistic investigation. Although efforts have been made to correlate the two classifications, there are still many unresolved issues. In our case, the newly identified four genes, gnt12, gnt13, gnt14, and gnt15 from *D. discoideum* are not presented in any of the GlcNAc-transferase families identified by the CAZy. In order to examine if Gnt12-15 are members of GT49, we have conducted the phylogenetic analysis. And, to support our previous finding and to understand the molecular mechanism of Gnt15, we have presented our recent results from cellular biochemical studies.

1. Introduction

Glycosylation is a vital cellular process of a cell. It has been estimated that 1% of the ORFs of a genome is dedicated to this process. Many different types of enzymes are involved in glycosylation. Glycosyltransferases are enzymes transfer sugars moieties from activated donor molecules onto various macro-molecules such as DNA, proteins and lipids. It is known that glycosylation regulates the structure, location, stability and function of a protein [1, 2, 3]. Glycosyltransferase are classified based on the type of activated donor, namely glycosyltransferase (Glc-transferase), mannosyltransferase (Man-transferase) and *N*-acetylgalactosaminyltransferases (GlcNAc-transferase).

2. Methods and Materials

2.1 Bioinformatics tools

Several of the bioinformatics tools were used to analyze the sequences of gnt12-gnt15. The workflow of
the analysis is shown in Figure 1. MEME (Multiple Expectation-maximization for Motif Elicitation) program (version 3.5.4; San Diego Supercomputer Center, UCSD [http://meme.sdsc.edu/meme/]) [8, 9]. TMHMM program (version 2.0; Center for Biological Sequence Analysis, Technical University of Denmark [http://www.cbs.dtu.dk/services/TMHMM-2.0/]) [10]. CAZy (Carbohydrate-active Enzymes) database ([http://afmb.cnrs-mrs.fr/CAZY/]) [7]. Pfam HMM search (Sanger Institute [http://www.sanger.ac.uk/Software/Pfam/search.shtml]) [11]. GlcNAc O-glycosylations in Dictyostelium discoideum proteins ([http://www.cbs.dtu.dk/services/DictyOGlyc/]) [12].

2.2. Phylogenetic analysis of GlcNAc transferase family

ClustalW, a multiple alignment programs align two or more sequences with each other to determine any significant consensus [13]. This approach can be used for searching patterns. In GlcNAc-transferase analysis, once multiple alignment of all GlcNAc-transferase has been made. It can be used to construct a phylogenetic tree. The evolutionary history inferred from phylogenetic analysis is usually depicted as branching, treelike diagrams represent an estimated pedigree of the inherited relationships among molecules. However, evolutionary relationships can be seen via viewing Cladograms [14] or Phylograms [15].

Figure 1 Bioinformatics workflow for identification of glycosyltransferase.

2.3 Lectin precipitation and binding assay

Cells were harvested and washed twice in 20 mM sodium phosphate buffer, pH 6.5, resuspended to 1 × 10^6 cells/ml, and lysed by passing through double layer 5 µm Nucleopore filters (Whatman) in 1-ml syringes. Crude extracts were centrifuged at 20000g for 30 min at 4°C. The resulting pellets, which we refer to as “membrane fractions”, were solubilized in detergent extraction buffer; samples of equal protein amount were precipitated with biotin-conjugated WGA or WBA (final conc. 100 µg/ml). Beads were washed and proteins bound to beads were separated on 10% SDS-PAGE gels. One of two identical gels was stained with Coomassie blue. Proteins on the other gel were transferred onto a PVDF membrane, and incubated with 10 µg/ml of biotin-labeled winged bean agglutinin (WBA) or wheat germ agglutinin (WGA). Detection was done by ECL after incubating the blots with streptavidin-HRP.

Mass spectrometric analysis—Proteins in the gel piece were subjected to tryptic digestion as described previously [16]. Briefly, the gel was washed and dried in a Speed-Vac (Savant), proteins were reduced by β-mercaptoethanol and modified by 4-vinylpyridine. The gel was dried and then incubated with modified trypsin (Promega) overnight. The enzyme digest was removed from the gel, and the gel residue was extracted by 200 µl of 0.1% formic acid. These two fractions were combined, dried in a Speed-Vac, and kept at -20°C for storage. The sample was resuspended in 0.1% formic acid and analyzed in a LC-MS/MS system consisting of Agilent 1200 nanoflow HPLC and LTQ-Orbitrap hybrid tandem mass spectrometer (ThermoFisher, USA). The MS/MS spectra were acquired with relative collision energy of 35 and an isolation width of 2.5 Da. Interpretation of the resulting MS/MS spectra was facilitated by the algorithm TurboSEQUENT [17] against the Dictostelium protein database downloaded from National Center for Biotechnology Information (USA).

3. Results

3.1 GNT12-15 in Dictyostelium discoideum are putative glycosyltransferase.

Our previous work suggested that GNT12-15 are similar to human LARG proteins via BLASTP [4]. However, none of the GNT12-15 are classified as glycosyltransferase in the CAZy database an evolving hierarchical family classification for glycosyltransferases. Therefore, we designed a new bioinformatics workflow to reanalyze these four genes. We first searched for aspartate-any residue-aspartate (DXD) motif, commonly found in glycosyltransferase. All four proteins have DXD motif (Figure 2). In several of glycosyltransferase families, the DXD motif is essential for the enzymatic activity [18]. Our result suggested that the four proteins might be glycosyltransferases. The second step in the workflow was to predict the presence of transmembrane domains by the TMHMM. All four proteins contain a single predicted transmembrane domain at the N-terminal region (Figure 2). Type II transmembrane domain is a feature commonly found for Golgi-residing glycosyltransferases. Our result further suggested that these four genes might be Golgi glycosyltransferases. We then used MEME (Multiple Expectation-maximization for Motif Elicitation) program to search for homology between the four proteins. Four homologous regions were found, marked as I-IV (Figure 2). We also performed Pfam HMM search. A putative GT2 glycosyltransferase domain was found in Dictostelium Gnt13 and Gnt15