Computational Approach for the Analysis of Post-PKS Glycosylation Step

Ki-Bong Kim[1]* and Kiejung Park[2]

[1]Department of Bioinformatics Engineering, Sangmyung University, Cheonan 330-720, Korea, [2]SmallSoft Co., Deajeon 305-728, Korea

Abstract

We introduce a computational approach for analysis of glycosylation in Post-PKS tailoring steps. It is a computational method to predict the deoxysugar biosynthesis unit pathway and the substrate specificity of glycosyltransferases involved in the glycosylation of polyketides. In this work, a directed and weighted graph is introduced to represent and predict the deoxysugar biosynthesis unit pathway. In addition, a homology based gene clustering method is used to predict the substrate specificity of glycosyltransferases. It is useful for the rational design of polyketide natural products, which leads to *in silico* drug discovery.

Availability: The trial version of this system is available via WWW at http://sm.hacklib.com/.

Keywords: glycosylation, post-PKS tailoring step, glycosyltransferase, deoxysugar biosynthesis unit pathway

Introduction

The progress of microbial genetics in recent years has greatly enhanced our ability to understand and thereby rationally alter the biosynthesis of various biologically important natural products, especially polyketides (Yadav *et al.*, 2003). Polyketides are a large family of natural products with diverse chemical structures and a broad range of pharmacological effects, including antibacterial, antifungal, antiparasitic, anticancer and immunosuppressive activities (Staunton and Weissman, 2001). These secondary metabolites have been characterized in a wide range of organisms. Polyketide biosynthesis resembles fatty acid assembly and is catalyzed by a collection of enzyme activities called polyketide synthases (PKSs) (Hopwood and Sherman, 1990). The

structural variety of polyketides is given by the options in choice of starter/extending units, number of condensations, and degree of subsequent reductive cycles, all of which being conferred by polyketide synthase and closely related enzymes. Many of these compounds and their semi-synthetic derivatives are used today in the clinic.

The area of biosynthetic engineering of the enzymes (*i.e.* PKSs) involved in polyketide biosynthesis has advanced recently to obtain various compounds of polyketide origin (Tae *et al.*, 2007). The biosynthesis of a typical polyketide can be divided into three separate steps, including (*i*) the choice of starter unit, (*ii*) the choice of extender units and the degree of beta-keto group reduction, and (*iii*) the post-PKS tailoring of the basic polyketide backbone. The tailoring steps catalyzed mainly by oxidoreductases and group transferases are crucial for the addition of important functional groups to polyketide skeletons and are key to the structural diversity and biological activity of this class of natural products (Rix *et al.*, 2002).

Among group transfer reactions, glycosyltransfer ones are perhaps the most important biotransformation. Glycosyltransferases are responsible for the attachment of sugar moieties, often deoxysugars, which add important features to the shape and the stereo-electronic properties of a molecule and often play an essential role in the biological activity of many natural product drugs (Rix *et al.*, 2002). Thus, glycosyltransferases are and will become more and more important tools for combinatorial biosynthetic approaches. In this respect, our study focuses on developing the computational tools for the substrate prediction of a given glycosyltransferase (GT) and the prediction of deoxysuguar biosynthesis unit pathway. The deoxysugar is synthesized by diverse biosynthetic enzymes and functions as a substrate of GTs.

Features and Results

Our computational system developed in this work has 3-tier architecture which is composed of a client, an application server, and a back-end database server (Fig. 1). The application server consists of major three modules - pathway analysis module, GT analysis module and back-office module. The pathway analysis module is involved in SBPD (Sugar Biosynthesis Pathway Database) search and sugar biosynthesis unit pathway prediction and drawing. The main procedure of pathway
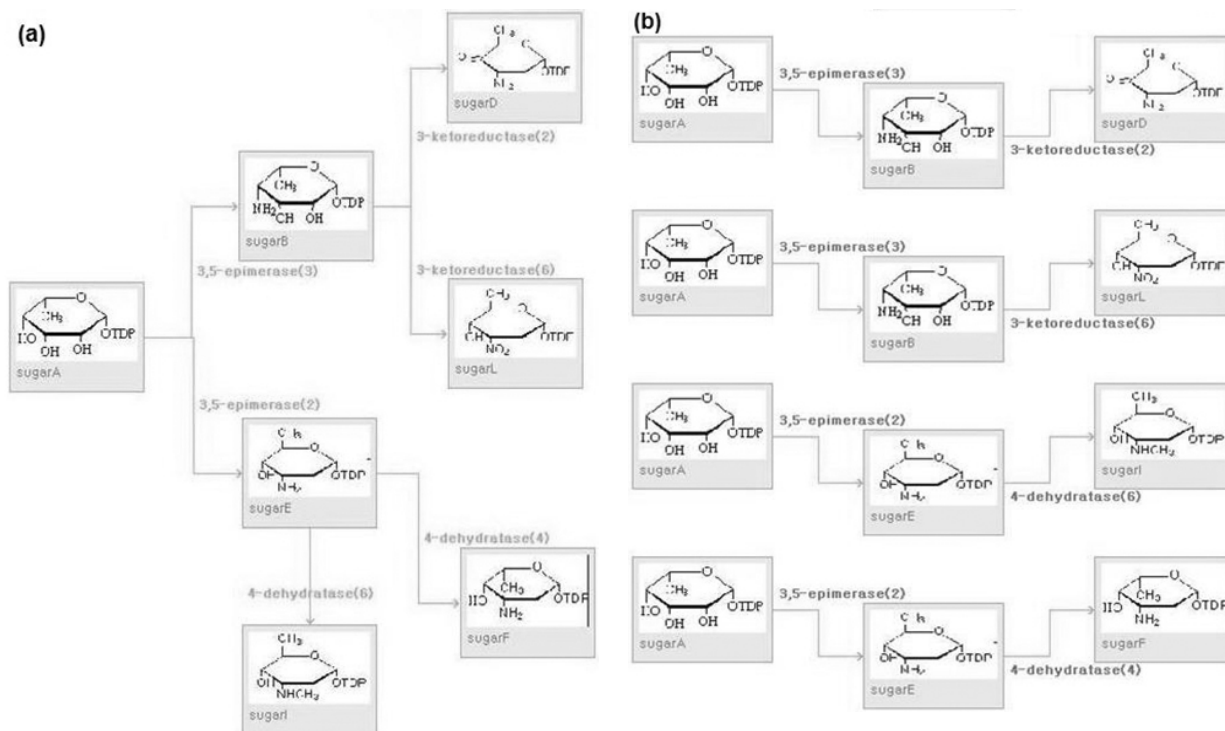
*Corresponding author: E-mail kbkim@smu.ac.kr
Tel +82-41-550-5377, Fax +82-41-550-5184
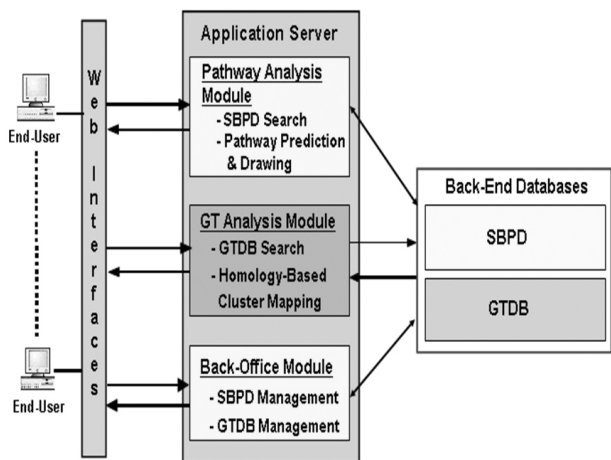
Fig. 1. The system frame with 3-tier client-server architecture. It has three essential components - a client, an application server, and a back-end database server.
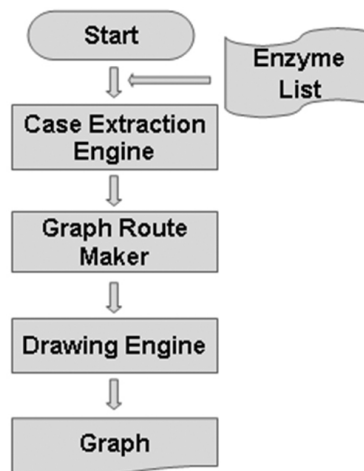

Fig. 2. The main procedure of the prediction and drawing of sugar biosynthesis unit pathways.

Fig. 3. An exemplary pathway maps for sugar biosynthesis unit. The drawing engine in pathway analysis module provides not only an integrated pathway map (a) but also all possible component paths (b).

prediction and drawing can be described as follows (Fig. 2): (A) Given an enzyme as a start node and a list of enzymes as input data, the case extraction engine calculates all possible combinatorial routes of input enzymes. (B) The graph route maker generates node (*i.e.* sugar) and edge (*i.e.* enzyme) data after applying the route information in the SPDB. (C) Finally, the drawing engine generates graph data objects from graph route data and draws the pathway maps (Fig. 3). A directed and weighted graph theory is employed to represent and predict all possible biosynthesis unit pathways with a given enzyme set.
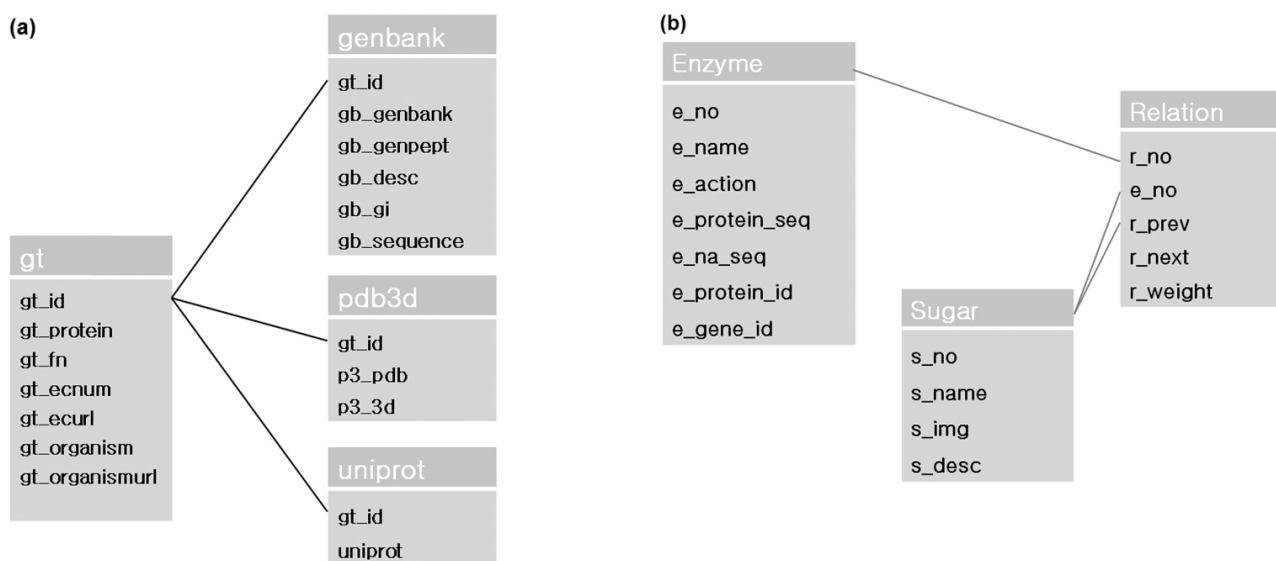
Fig. 4. The schemata of GTDB and SBPD (a). The GTDB consists of four tables, each of which contains the information extracted from CAZY, GenBank, PDB and UniProt respectively (b). The SBPD has three tables that are Enzyme, Relation and Sugar.

The GT analysis module undertakes GTDB (Glycosyl Transferase Database) search and GT-encoding gene clustering and mapping. We used BLAST program (Altschul *et al.*, 1990) to cluster and map GT-encoding genes on the input genomic sequence, where GTDB is used as a target database of BLAST program. The third part is back-office module that is in charge of user authentication, SBPD management and GTDB management. Using this module, system manager can access, update and manage the back-end databases with ease and security.

The back-end databases consist of SBPD and GTDB that are implemented as a relational database using MySQL database management system. The SBPD database is specifically designed for the sugar biosynthesis related to post-PKS tailoring step and has three main tables - Enzyme, Relation, and Sugar (Fig. 4). The information in SBPD is collected from two different sources: experimental data produced in our joint laboratory and reviewed publications. Using back-office module, system manager and authenticated users can also add new data to this database. The GTDB database is composed of four tables (*i.e.* gt, genbank, pdb3d, and uniprot) (Fig. 4) and the information in this database is mainly collected from CAZY system (http://www.cazy.org) by means of parsing and reengineering. The CAZY database describes the families of structurally-related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds. It features five classes of enzymes - glycoside hydrolases (GH), glycosyl trans-

ferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), and carbohydrate binding modules (CBM).

Discussion

Polyketides are the components of a vast variety of microbial and plant secondary metabolites, including biologically active compounds such as antibiotics. As mentioned above, the structural variety of polyketides is given by the options in choice of starter/extending units, number of condensations, and degree of subsequent reductive cycles, all of which being conferred by PKS and closely related enzymes. In addition, subsequent post-PKS tailoring steps provide the final product with further structural variations. In this context, we developed the computational system for the analysis of post-PKS glycosylation step.

Our system carries the prediction and drawing of deoxysugar biosynthesis unit pathway, SBPD and GTDB searches, GT-encoding gene clustering, and back-office management. The deduced sugar biosynthesis unit pathway can allow us to determine and design the appropriate sugar that can be a substrate for glycosyltransferases and be employed in the glycosylation of polyketides. The practical use of analysis result can provide researchers with the opportunity to conduct metabolic engineering of deoxysugar genes in collaboration with experimental work. The GT-encoding gene clustering and mapping may give clue to glycosyltransferase genes, each of which can be assigned to a specific GT

reaction in biosynthesis. They usually possess regional and mechanistic specificities: α-(1,3)-glycosylation for UrdGT1c; β-(1,4)-glycosylation for UrdGT1b. A series of their chimeric GT genes can be constructed, and expressed in a mutants lacking corresponding glycosyl-transfer activity to find the region determining the substrate specificities. Metabolic engineering to the region, including DNA shuffling, gives rise to the novel GT specificities leading to unnatural derivatives. Our system may be useful for experiment researchers to rationally design a novel polyketide and derivatives with biological activities.

Acknowledgements

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 213, 403-410.

Hopwood, D.A., and Sherman, D.H. (1990). Molecular genetics of polyketides and its comparison with Fatty acid biosynthesis. *Annu. Rev. Genet.* 24, 37-66.

Carbohydrate Active Enzymes database site. *http://www.cazy.org.*

Rix, U., Fischer, C., Remsing, L., and Rohr, J. (2002). Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* 19, 542-580.

Staunton, J., and Weissman, K. (2001). Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* 18, 380-416.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2003). Computational Approach for prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *J. Mol. Biol.* 328, 335-363.