Genomics & Informatics Vol. 6(4) 227-230, December 2008

Computational Approach for Biosynthetic Engineering of Post-PKS Tailoring Enzymes

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

Compounds of polyketide origin possess a wealth of pharmacological effects, including antibacterial, antifungal, antiparasitic, anticancer and immunosuppressive activities. Many of these compounds and their semisynthetic derivatives are used today in the clinic. Most of the gene clusters encoding commercially important drugs have also been cloned and sequenced and their biosynthetic mechanisms studied in great detail. The area of biosynthetic engineering of the enzymes involved in polyketide biosynthesis has recently advanced and been transferred into the industrial arena. In this work, we introduce a computational system to provide the user with a wealth of information that can be utilized for biosynthetic engineering of enzymes involved in post-PKS tailoring steps. Post-PKS tailoring steps are necessary to add functional groups essential for the biological activity and are therefore important in polyketide biosynthesis.

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

Keywords: pharmacological effects, gene clusters, biosynthetic engineering, post-PKS tailoring step

Introduction

Polyketides are important natural products exhibiting antibacterial (rifamycin), antifungal (erythromycin), antitumor (doxorubicin), immunosuppressant (FK506) and cholesterol-lowering activities (lovastin). They are produced mainly by *Streptomyces* species which belong to the large group of mycelially growing, filamentous bacteria from soil known as actinomycetes. These gram-positive, fungi-like bacteria are one of the best known producers of secondary metabolites used as naturally occurring antibiotics. Each core of the polyketide is synthesized biologically under the control of an exceptionally large, multifunctional enzyme called polyketide synthase (PKS), in a manner similar to that of fatty acid synthesis, where the carbon backbones of the molecules are assembled by the successive condensation of small acyl units.

There are presently three types of polyketides. Type I modular polyketides are built up by a PKS consisting of large multifunctional proteins with a different active site for each enzyme-catalyzed step in polyketide carbon chain assembly. Type II iterative polyketides also known as aromatic polyketide are normally synthesized by a single PKS built up by discrete polypeptides which carry active sites that are used more than once in the biosynthetic pathway. Type III iterative polyketides were characterized not long ago and the type III PKS is a member of the chalcone synthase (CHS) and stilbene synthase (STS) superfamily of PKS previously only found in plants.

The readily synthesized and cyclized polyketide is usually modified via hydroxylation, glycosylation, methylation and acylation. These tailoring, or post PKS modifications are believed to be crucial for addition of important functional groups to polyketide skeletons and to the structural diversity and biological activity of this class of natural products (Rix *et al.*, 2002; Kwan *et al.*, 2008). The most frequently found post-PKS modifications are catalyzed by oxidoreductases, a very broad group of enzymes consisting of oxygenases, oxidases, peroxidases, reductases (*e.g.*, ketoreductases), and dehydrogenases. In general, these enzymes introduce oxygen-containing functionalities, such as hydroxyl groups ('hydroxylases'), aldehyde or keto groups, and epoxides ('epoxidases') or modify such functionalities by addition or removal of hydrogen atoms, e.g. transforming a ketone into a secondary alcohol or an aldehyde into a carboxylic acid. Although oxidoreductases provide or modify relatively small functional groups, they can have a tremendous impact on the binding properties of a molecule with respect to a biological ligand molecule (receptor protein, enzyme, DNA etc.). The term group transferase refers to enzymes that possess transferase activity introducing novel functional groups and altered profiles on the product relative to the substrate. This enzyme group contains important enzymes such as amino transferases, alkyl (usually methyl) transferases, acyl (usually acetyl) transferases, glycosyltransferases (GTs)

*Corresponding author: E-mail kbkim@smu.ac.kr
Tel +82-41-550-5377, Fax +82-41-550-5184
Accepted 28 October 2008

and kinases.

Methyl- and glycosyltransferases are the most important post-PKS modifying enzymes. Glycosyltransferases perform perhaps the most important biotransformation on earth, at least in quantitative terms. GTs are responsible for attachment of sugar moieties, often deoxysugars, adding important features and properties to a molecule and they often play an essential role in the biological activity of many natural product-based drugs (Floss, 2001; Mendez *et al*., 2001; Trefzer *et al*., 1999; Xue *et al*., 2001). Some GTs have also shown to have unexpected inherent substrate flexibility either towards their acceptor substrate (usually alcohol or phenols) or (deoxy) sugar donor co-substrates, and sometimes even to both (Faust *et al*., 2000; Xue *et al*., 2001). Methyltransferases use S-adenosylmethionine (SAM) as cofactor and can methylate O, N or C-atoms. O and N-methylations increase the liphophilicity of a molecule and also remove hydrogen-bond donor sites. The methylation reactions can occur either at the polyketide derived aglycone moiety or on sugar residues, either prior to or after the glycosyl transfer (Rix *et al*., 2001).

Recent advances in combinatorial biosynthesis technology have facilitated biosynthetic engineering of the enzymes involved in polyketide biosynthesis. The area of biosynthetic engineering of the enzymes has emerged as an important field and been transferred into the industrial arena. In this respect, we have developed a computational system to provide the user with a wealth of information that can be utilized for biosynthetic engineering of enzymes, especially involved in post-PKS tailoring steps.

Features and Results

Our system developed in this work is organized in three main parts - graphical user interfaces, an application server and an integrated database. The system offers an attractive way to many of the specialized or single-task analyses over the Web. In particular, it has sophisticated graphical user interfaces (GUIs) with a wide range of well-tested analytical functions, database, and well-maintained user-support applications to make most aspects of post-PKS tailoring enzyme analysis simpler.

The core part of our system is PPTE (Post-PKS Tailoring Enzymes) database that we have constructed at local site through this work. The PPTE is implemented as a relational database using MySQL database management system. This is specifically designed for glycosyltransferases (GTs), methyltransferases (MTs) and hydroxylases (HLs) involved in post-PKS tailoring step and consists of ten main tables. The information in PPTE is collected from three different sources: InterPro, CAZY, and GenPept. InterPro is a database of protein families, domains, repeats and functional sites in which identifiable features found in known proteins can be applied to new protein sequences (http://www.ebi.ac.uk). 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 (http://www.cazy.org). GenPept is a genetics sequence databank which contains translations of coding sequences in the GenBank nucleotide sequence database (http://www.ncbil.nlm.nih.gov/).

The overall procedure for PPTE construction can be described in Fig. 1. PPTE follows the GenPept format and conventions as closely as possible. The production of PPTE starts with searching CAZY and GenPept with all possible queries that have been made up of combinations of polyketide names and tailoring enzyme names. At this stage all annotation you can find in a PPTE entry comes from the corresponding GenPept and CAZY entries. The first post-processing step is the removal of redundancy. The redundancy of the database was eliminated by merging separate entries corresponding to the same entity. The second post-processing step is to scan all retrieved protein sequences against InterPro databases with InterProScan and to get additional annotation related to protein families, domains, repeats and functional sites. InterProScan is a wrapper on top of a set of applications for scanning protein sequences against InterPro member databases and provides an efficient way to analyze protein sequences for known domains and functional sites by launching the

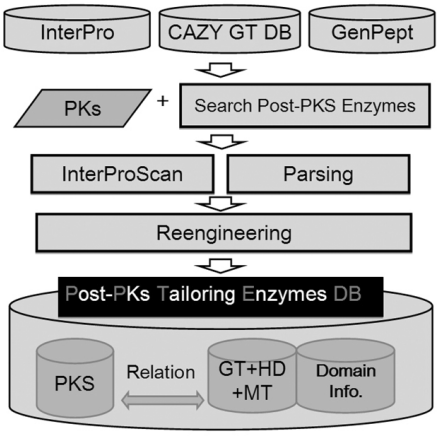

Fig. 1. The overall procedure for PPTE database construction. PPTE consists of ten main tables and its three different sources are InterPro, CAZY, and GenPept. Refer to detailed description in the text.

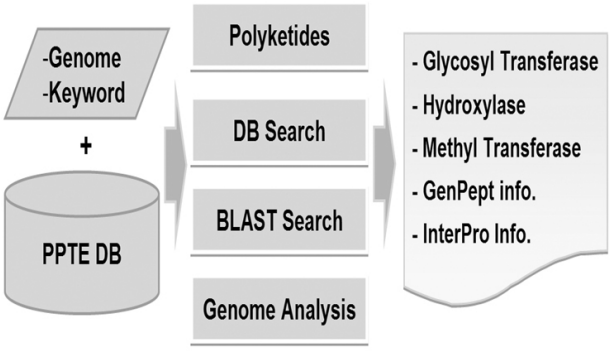

Fig. 2. The application server consists of four modules – Polyketides, DB Search, BLAST Search and Genome Analysis. The feature of each module is described in detail in the text. Through this application server, we can retrieve comprehensive information on tailoring enzymes, GenPept and InterPro from PPTE and also carry out various analyses such as prediction of functions of individual proteins, automatic annotation of newly sequenced proteins, searching for new protein families, yielding a hierarchical organization of gene clusters, etc.

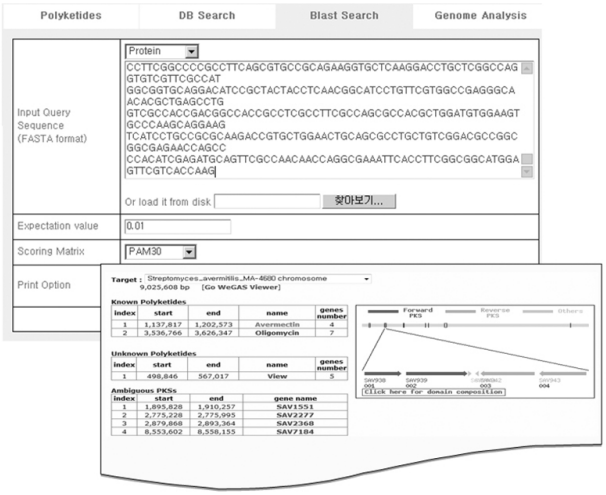

Fig. 3. Our system has sophisticated graphical user interfaces (GUIs) with four main menus – Polyketides, DB Search, Blast Search and Genome Analysis. GUIs offers an efficient way to many of the specialized or single-tank analyses over the Web.

applications in parallel. Finally, we have parsed the output of two previous steps and combined the results at the level of unified attributes into one representation. The database PPTE provides a comprehensive, non-redundant and up-to-date protein sequence database of tailoring enzymes with high information content.

The application server consists of four modules - Polyketides, DB Search, BLAST Search and Genome Analysis (Fig. 2). Our system has sophisticated graphical user interfaces with four main menus that correspond to four modules in application server (Fig. 3). Polyketide module is to search for polyketides and its related information in PPTE, including tailoring enzymes, GenPept information, and InterPro information. While working with full text searches one may use DB Search module to retrieve all information (*i.e.*, tailoring enzymes, GenPept information and InterPro information) from PPTE. BLAST Search module is developed for homology-based analysis of input genome sequence, which is based on analysis of all pairwise comparisons between input sequence and protein sequences of PPTE. The analysis can be carried out for different levels of protein similarity, yielding a hierarchical organization of clusters on the input sequence. This module can be used for prediction of functions of individual proteins, automatic annotation of newly sequenced proteins, searching for new protein families, yielding a hierarchical organization of gene clusters. Even though Genome Analysis module has the same features that BLAST Search module has, it has been developed to be optimized for whole genome analysis (Lee *et al.*, 2007), providing the framework for the presentation of genome-wide functional data.

Discussion

Combinatorial biosynthesis is defined by C. Khosla as to "involve genetically engineering biosynthetic pathways in such a way that they may be combinatorially reconstructed to produce libraries of novel small molecules that are appropriate for use in screening for new drugs" (Khosla and Zawada, 1996). Post-PKS tailoring steps are necessary to add functional groups essential for the biological activity and are therefore important in the biosynthesis. Targeted gene inactivation of post-PKS enzymes does not only present a very useful tool for the elucidation of biosynthetic pathways by analysis of accumulated intermediates or shunt products, but is also used in the context of combinatorial biosynthesis to generate new "unnatural" products with altered biological activities. Not only can a single modification be introduced into a polyketide structure, but even multiple alterations can be incorporated simultaneously through the rational approach of biosynthetic engineering. The technology of biosynthetic engineering of enzymes involved in polyketide synthesis has also disadvantages compared to the semi-synthetic chemistry approach. Although it may sound very lucrative to be able to replace various functional groups of a natural polyketide, it must be noted that not all of the modifications attempted have been successful.

As mentioned above, genetic manipulation of related genes provides useful information on the functions and characteristics of the encoded enzymes; the ultimate goal being rational design and generation of novel hybrid natural products. The progress toward this goal has been slow, as many experiments have been and still are conducted in a semi-random mix-and-match fashion, especially for post-PKS tailoring enzymes. In this respect, we have developed the well-established computational system for analyzing and retrieving post-PKS tailoring enzymes and related information. This system will readily lend itself to facilitating the use of genetic engineering to develop non-natural analogues of commercially valuable products. Furthermore, it will also pave the way to a more rational approach employing genes encoding post-PKS tailoring enzymes.

Acknowledgements

This work was supported by MOCIE grants from Intelligence Bioinformatics and Application Center, KRIBB.

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.

Faust, B., Hoffmeister, D., Weitnauer, G., Westrich, L., Haag, S., Schneider, P., Decker, H., Kunzel, E., Rohr, J., and Bechthold, A. (2000). Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in *Streptomyces fradiae* Tu2717. *Microbiology* 146, 147-154.

Floss, H.G. (2001). Antibiotic biosynthesis: from natural to unnatural compounds. *J. Ind. Microbiol. Biotechnol.* 27, 183-194.

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.

Kwan, D.H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-stark, J.C., Haydock, S.F., and Leadlay, P.F. (2008). Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chem. Biol.* 15, 1231-1240.

Khosla, C., and Zawada, R.J. (1996). Generation of polyketide libraries via combinatorial biosynthesis. *Trends. Biotechnol.* 14, 335-341.

Lee, D., Seo, H., Hahn, J., Kong, E., and Park, K. (2007). RISS: rice (Oryza sativa L. ssp. Japonica) genome information service system. *Genomics & Informatics* 5, 194-195.

Mendez, C., and Salas, J.A. (2001). Altering the glycosylation pattern of bioactive compounds. *Trends. Biotechnol.* 19, 449-56.

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.

Trefzer, A., Salas, J.A., and Bechthold, A. (1999). Genes and enzymes involved in deoxysugar biosynthesis in bacteria. *Nat. Prod. Rep.* 16, 283-299.

Xue, Y., and Sherman, D. H. (2001) Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in *Streptomyces venezuelas. Metab. Eng.* 3, 15-26.

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.