J. Microbiol. Biotechnol. (2009), **19**(2), 140–146
doi: 10.4014/jmb.0809.554
First published online 30 December 2008

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Development of an Analysis Program of Type I Polyketide Synthase Gene Clusters Using Homology Search and Profile Hidden Markov Model

Tae, Hongseok[1], Jae Kyung Sohng[2], and Kiejung Park[1]*

[1]*Information Technology Institute, SmallSoft Co., Ltd., Daejeon 305-343, Korea*
[2]*Department of Pharmaceutical Engineering, Sunmoon University, Asan 336-708, Korea*

Received: September 30, 2008 / Accepted: November 25, 2008

MAPSI (Management and Analysis for Polyketide Synthase Type I) has been developed to offer computational analysis methods to detect type I PKS (polyketide synthase) gene clusters in genome sequences. MAPSI provides a genome analysis component, which detects PKS gene clusters by identifying domains in proteins of a genome. MAPSI also contains databases on polyketides and genome annotation data, as well as analytic components such as new PKS assembly and domain analysis. The polyketide data and analysis component are accessible through Web interfaces and are displayed with diverse information. MAPSI, which was developed to aid researchers studying type I polyketides, provides diverse components to access and analyze polyketide information and should become a very powerful computational tool for polyketide research. The system can be extended through further studies of factors related to the biological activities of polyketides.

Keyword: ASMPKS, MAPSI, modular, PKS, polyketide

Polyketides, which are of considerable medicinal importance, with antibacterial, immunosuppressive, antitumor, and other pharmacological activities, are a large family of secondary metabolites produced by microorganisms. Polyketide synthesis has been found to resemble fatty acid biosynthesis [16, 26], in that both involve the stepwise condensation of short carbon precursors [4]. The carbon frameworks of polyketides such as avermectin [8, 17], erythromycin [2], and rifamycin [6] are synthesized by a class of enzymes termed polyketide synthases (PKSs).

Type I PKSs are multifunctional enzymes composed of so-called modules [19, 20], which process serial steps in polyketide chain elongation. Each module contains a catalytic domain that is involved in one step of polyketide chain extension from acyl-coenzyme A (CoA) substrates. Depending on the iteration of their activities, type I PKSs can be classified as modular [7, 25] or type I iterative [22] PKSs. Modular PKSs, which are mainly encoded by bacterial genomes, noniteratively process serial steps of polyketide chain elongation. Type I iterative PKSs, which are mainly found in fungi, are usually composed of a single module that is iteratively used in various condensation and chain elongation steps.

Each module of type I PKSs usually contains an acyltransferase (AT) domain, which selects and transfers extender units; an acyl carrier protein (ACP) domain, which tethers the growing polyketide chain to the PKS for condensation; and a ketoacyl synthase (KS) domain. These essential domains are complemented by a variable set of additional domains, including ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains. A thioesterase (TE) domain catalyzes the release of a polyketide product from the last PKS participating in chain elongation. Diverse carbon structures of polyketides are dictated by the order of the modules, each of which consists of sets of different domains.

PKS genes that synthesize a polyketide usually form a cluster on a genome. Thus, the chemical structure and activity of a polyketide can be easily modified through gene manipulations [13, 14]. Attributes of a polyketide, such as structure or activity, can be modified by changing the genes related to starting units [11, 15] and extender units [27], as well as genes involved in the reduction of carbonyl carbons [18], determining the length of the carbon chain [3], and post-processing. In addition, genetic engineering, whether by the fusion of different polyketides [10] or their biosynthesis in heterologous hosts [24], was recently found to improve polyketide activities.

Because of their pharmaceutical importance, numerous polyketides have been identified over the past 30 years. As a result of intense interest in identifying new kinds of polyketides, the volume of data on polyketides and PKSs

*Corresponding author
Phone: +82-42-864-2524; Fax: +82-42-385-9240;
E-mail: kjpark@smallsoft.co.kr

is rapidly increasing. Since our development of ASMPKS (Analysis System for Modular Polyketide Synthases) [23], which analyzes modular PKSs only, we have improved the components of this package and we have developed an advanced polyketide analysis system termed MAPSI (Management and Analysis for Polyketide Synthase Type I). The aim of the MAPSI system is to offer advanced computational methods to detect PKS gene clusters of type I polyketides. To detect PKS gene clusters, MAPSI uses homology search and profile HMM (Hidden Markov Model) whereas ASMPKS uses homology search only. The genome analysis component of MAPSI detects known PKS clusters by homology analysis of proteins, and predicts putative PKS clusters by identifying domains of proteins in microbial genome sequences. MAPSI also provides improved polyketide visualization and other analytic components, such as PKS assembly and domain analysis. The polyketide data and analysis components are accessible through Web interfaces and are displayed with diverse information.

METHODS

Domain Identification

Because each carbon precursor is modified according to the domain composition of a module in each step of chain extension during polyketide biosynthesis, identifying the domain composition from putative PKS proteins is critical in the prediction of polyketide structures. Although PKS domains are highly conserved, several, especially DH domains, are not detected by most domain identification programs or databases, including InterProScan [29], CDD (Conserved Domain Database) [12], and PKSDB (Polyketide Synthesis Database) [28].

MAPSI predicts domain composition from protein sequences, and HMMER [5] is used to identify each domain. The main algorithm of HMMER is a profile Hidden Markov Model (HMM), one of the most powerful methods used to detect conserved regions of proteins. Identification of domains from protein sequences using HMMER requires a matrix file containing profile information from HMM for aligned domains. Whereas most domains of each type, including KS, AT, DH, ER, and KR, are highly similar, some domains cannot be detected when a single profile matrix is used for each type. A hierarchical clustering algorithm [9] was therefore applied to construct multiple profile matrices of each type.

Hierarchical Clustering for Multiple Profile Matrices of Profile Hidden Markov Model

An average link method was used to measure distances of pairs and homologies of domain sequences, which were calculated by BLAST [1]. The first step of the clustering was to assign each domain sequence to an individual cluster. In the second step, distances between all possible pairs of clusters were compared to find the minimum distance. To compute the distance between two clusters, the average distance calculated from all cross-pairs of members in the clusters was taken. If the minimum distance was less than the cut-off value, the two clusters were merged. This second step was repeated until no distance less than the cut-off value remained. Clusters containing fewer than five members were also merged. Following this process, the HMMER profile matrices for the clusters were constructed.

Prediction of Polyketide Synthase Gene Clusters from Genome Sequences

In many cases, one genome produces several different polyketides, and one polyketide can be produced by several different genomes. Because the relationships between polyketides and genomes are important in studying polyketides, genome-based polyketide analysis is required. PKS genes producing one kind of polyketide typically form a locational cluster in a genome. MAPSI identifies PKSs in genome sequences and predicts gene clusters for the PKSs (Fig. 1). MAPSI accepts genome sequences in FASTA or GenBank format files. When only genome sequences are submitted, Glimmer [21] is used to predict genes, and their sequences are converted to proteins. The PKS gene clusters predicted by MAPSI are classified into two types. The first type produces known polyketides that have been included in the database. PKSs of these clusters are detected by measuring the homology between protein sequences of an annotated genome and all the PKS sequences in the database, using BLAST. If the identity match between two sequences is over 50% and their alignment length is more than 90% of both sequences, the protein and the PKS are regarded as synthesizing the same product. When a complete set of proteins corresponding to all PKSs for a polyketide in the database is detected in a cluster, they are regarded as producing a known polyketide in that genome.

The second type of PKS gene clusters consists of as yet unidentified clusters that produce putative polyketide candidates, which can be detected by domain identification. HMMER is used to identify domains from protein sequences of genomes. When the e-value of a protein fragment is less than e-10, it is regarded as a domain. If the protein has valid domain compositions, at least "KS-ACP" or "TE," it is annotated as a PKS. If the distance between neighboring proteins is less than 15 kilobases, they are included in the same cluster and the cluster is regarded as producing a novel polyketide.

Proteins that have PKS domains but do not belong to any cluster are annotated as ambiguous candidates of modular type PKSs in a bacterial genome or putative iterative type PKSs in a fungal genome.

RESULTS

Evaluation of Domain Identification

To assess the performance of the domain identification component developed in this study, two parameters, namely, *SN* (sensitivity) and *PPV* (positive predictive value), were used. *SN* is the ratio between the number of correctly predicted actual domains and the total number of actual domains in the test data set, and *PPV* is the ratio between the number of correctly predicted actual domains and the total number of predicted domains. Equations of the two parameters are as follows:

$$SN=\frac{TP}{TP+FN}$$

$$PPV=\frac{TP}{TP+FP}$$

where *TP* (true positives) is the number of correctly predicted domains. *FP* (false positives) is the number of non-domain regions that are predicted as domains, and *FN* (false negatives) is the number of domains that are not predicted as domains.

The performance of MAPSI was compared to ASMPKS. Table 1 shows the *TP*, *FP*, *FN*, *SN*, and *PPV* values comparing the two programs for 212 PKSs containing 2006 domains.

Table 1. Results of the performance comparison.

	TP	*FP*	*FN*	*SN* (%)	*PPV* (%)
ASMPKS	1985	7	19	99.6	98.9
MAPSI	2002	2	4	99.9	99.8

Database for Polyketides and Polyketide Synthase Gene Clusters

MapsiDB, which contains information on polyketides and related genome data, is a subsystem of MAPSI. MapsiDB consists of a polyketide database and a set of genomic data. The polyketide database contains polyketide data, including information on PKS genes, modules, domains, and assembly. To date, there are 66 data entries on type I polyketides, including 45 modular type and 21 iterative type polyketides. These data were collected by surveying journals and other online resources. The genomic data were gathered using the auto-annotation component, which analyzed about 800 chromosomes and plasmids, detecting 91 gene clusters producing 79 putative polyketides and 12 known polyketides from 52 genomes.

Overview of Web Interfaces

The MAPSI system provides various Web interfaces to access and analyze polyketide information. The structure of the Web interfaces consists of five parts; MapsiDB, genome analysis, MAPSI tools, domain analysis, and management (Fig. 2). The MapsiDB portion displays detailed information on modular and type I iterative polyketides. This segment shows a data summary, PKS composition, and other related information on each polyketide. The "genome analysis" part provides tools, including genome-based polyketide analysis, and displays gene clusters producing polyketides from genomes analyzed by the auto-annotation component. The MAPSI tools are composed of "PKS prediction," which identifies domains from submitted proteins, and "PKS assembly," which displays an expected polyketide chain produced by an artificial PKS that is assembled by users. The domain analysis part provides a "homology search" and "multi-alignment" of domains. The management portion is the component that manages the polyketide database and the auto-execution of genome analysis for extending and editing the database.

Access to Polyketide Information

MAPSI provides Web interfaces that display detailed information on both types of polyketides. The left menu of

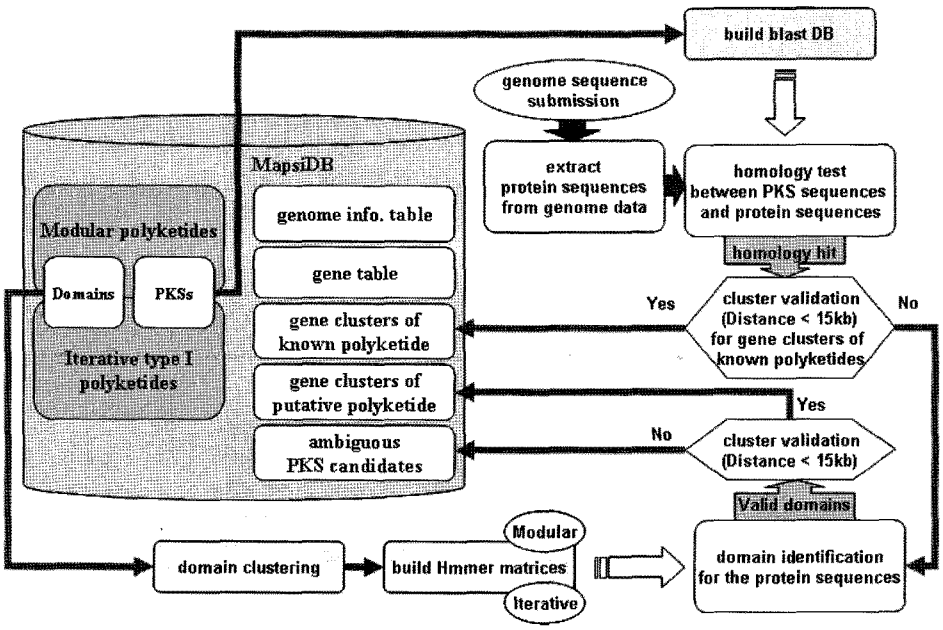

Fig. 1. Procedure used to search genome sequences for PKS gene clusters.
MAPSI searches genome sequences and predicts gene clusters for PKSs. MAPSI detects two types of PKS gene clusters. The first type, which produces known polyketides, can be detected by measuring the homology between protein sequences of an annotated genome and all PKS sequences. The second type, which produces putative polyketide candidates, can be predicted by identifying domains. If a protein has domains and does not belong to any cluster, it is regarded as an ambiguous candidate for the class of modular type PKSs.

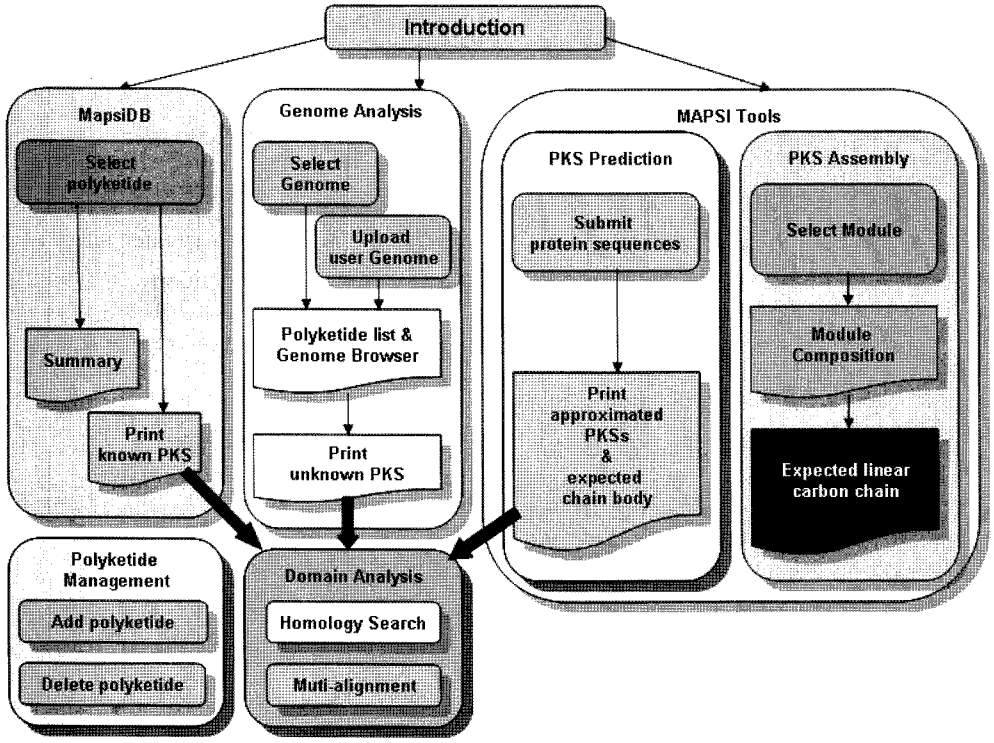

Fig. 2. Structure of Web interfaces.
The MapsiDB component displays detailed information on modular and type I iterative polyketides. The "genome analysis" component displays gene clusters producing polyketides from genomes, as analyzed by the auto-annotation component, and serves as an analysis component for genomes submitted by anonymous users. The "PKS prediction" component identifies domains from submitted proteins, and the "PKS assembly" component displays an expected chain body produced by an artificial PKS assembled by users. The domain analysis component provides "homology search" and "multi-alignment" of domains.

the "MapsiDB" page lists type I polyketides as modular or type I iterative (Fig. 3). These polyketides can be sorted according to chain lengths and starter units. The main page shows a linear carbon body, a description, PKS composition, and other polyketides related to the selected polyketide. Because one polyketide can be produced by several different genomes, the "PKSs" tab lists the names of the genomes that produce the polyketide. PKSs involved in biosynthesis of the polyketide are displayed in a table containing additional information, such as protein name, length, and GenBank accession number. Each PKS consists of modules, and each module contains two or more functional domains and processes one chain condensation reaction to synthesize a polyketide. The modules in the table are color-coded, with each short carbon unit of the linear chain body depicted on the left side of the table colored according to its corresponding module. The linear chain body is predicted by motif analysis of AT domains, which have a critical role in the selection of units.

Similarity between domains can be determined through homology analysis and multiple sequence alignment. If users click a domain figure, a menu with hyperlinks to homology analysis and multiple sequence alignment appears. BLAST is used for homology analysis between the domains of the same type and ClustalW is employed for multiple sequence alignments.

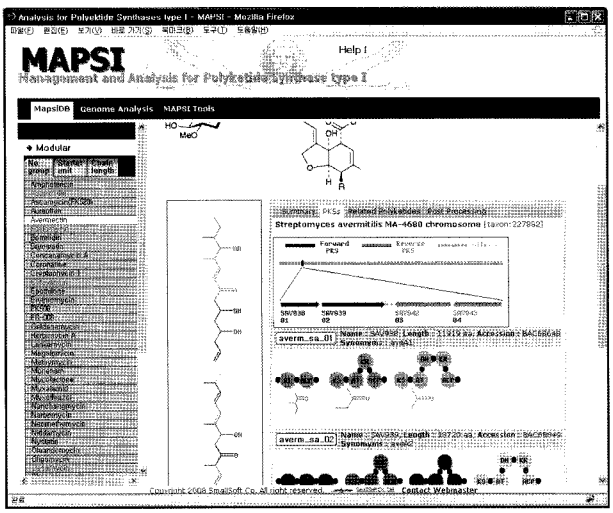

Fig. 3. Web interface for MapsiDB.
The left menu lists type I polyketides as modular or type I iterative, which can be sorted according to chain length and starter unit. The main page shows linear carbon body, data summary, PKS composition, and other polyketides related to the selected polyketide.

Visualization of Genome Analysis

The PKS gene clusters on a genome are displayed on a simple genome browser within MAPSI. If a gene cluster produces a known polyketide that is in MapsiDB, the genome browser is displayed on a "MapsiDB" page when the user selects a polyketide from the list on the left menu and clicks a genome name on the "PKS" tab in the polyketide information page.

The "genome analysis" page provides the results of auto-annotation executed by a manager that searches microbial genome sequences for PKS clusters. This page also has a genome browser, which shows the position and composition of the PKS cluster for each polyketide in a genome. The page lists genome names that are stored in MapsiDB and displays the gene clusters of known or putative polyketides, including ambiguous PKS candidates that have PKS domains but do not belong to any gene cluster. MAPSI also provides a tool to analyze new genomes submitted by users. If a user uploads a genome sequence file in FASTA or GenBank format, the MAPSI tool can extract protein sequences and perform genome analysis. The latter takes a few hours, and MAPSI shows progress in real time. The results page also shows gene clusters of known or putative polyketides, and ambiguous PKS candidates.

The order of substrates channeling in modular PKS biosynthetic clusters is not necessarily colinear with the order of ORFs on a genome. Substrates channeling is governed by protein-protein interactions, which may not be obvious. Hence, MAPSI provides an interface that allows users to decide the order of PKSs interacting in the biosynthesis of a putative polyketide, with visualization of the intermediates.

Web Interface for Domain Identification

The "PKS prediction" component takes multiple protein sequences from uploaded files or simply pasted strings on appropriate boxes (Fig. 4A). When multiple sequences are submitted, MAPSI identifies domains by executing HMMER. MAPSI then constructs modules of valid domain composition, each of which has responsibility for one elongation step, and predicts expected extender units by analyzing motifs of AT domains in the modules. Using predicted modules and extender units, a chain body of a polyketide is constructed, which can be synthesized by the submitted proteins. Because the domain identification results page also provides similarity analysis between domains, the user receives detailed information on the submitted PKSs and their products.

Assembly of an Artificial Polyketide Synthase

Modification of short carbon units, such as malonyl-CoA or methylmalonyl-CoA, is decided by the domain composition of modules. The "PKS assembly" component displays an expected chain body produced by an artificial PKS that is assembled by the interface (Fig. 4B). This can aid researchers in simulating the structures of newly found PKSs and to easily understand their properties. To assemble the artificial PKS, a starter unit should be selected first, followed by extender units. This order can lead to a new combination of modules and a new linear chain body. MAPSI provides 18 starters including 2-methylbutyryl-CoA, 3-methylbutyryl-CoA, 3,4-DHCHC-CoA, 3,5-AHBA-CoA, 3-amino-2-methylpropionate, acetoacetyl-CoA, acetyl-CoA, benzoyl-CoA, butyryl-CoA, cyclohexanecarboxylate, glycine, glycolate, *p*-aminobenzoate, *p*-coumaroyl-CoA, *p*-nitrobenzoate, phenylacetyl-CoA, propionyl-CoA, and trans-

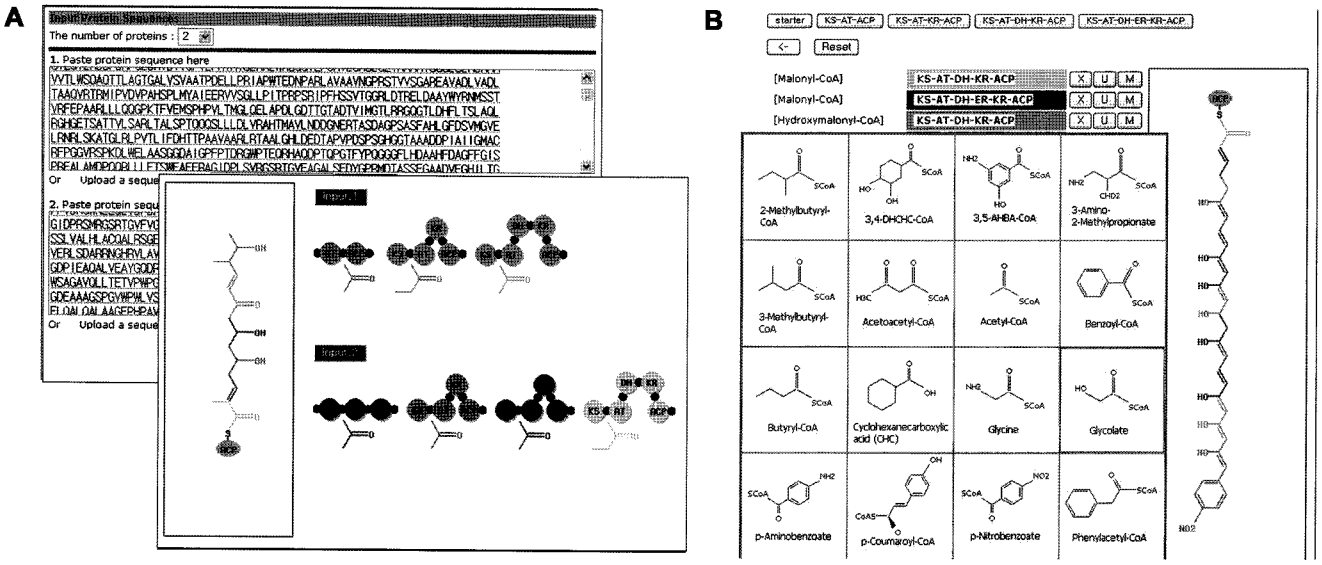

Fig. 4. Web interface for MAPSI tools.
When users submit multiple protein sequences, MAPSI identifies domains from the sequences and displays domain composition with an expected linear chain body. **B.** Users can assemble an artificial PKS by selecting modules and units, and the expected polyketide chain produced by the assembled PKS is then shown.

1,2-CPDA. MAPSI also provides 4 extenders: malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, and hydroxymalonyl-CoA. Moreover, MAPSI provides 4 types of domain composition including KS-AT-ACP, KS-AT-KR-ACP, KS-AT-DH-KR-ACP, and KS-AT-DH-ER-KR-ACP. By combining these elements, diverse chain bodies of polyketides can be constructed.

As polyketides are of medicinal importance, exhibiting antibacterial, immunosuppressive, antitumor, or other pharmacological activities, numerous polyketides have been identified over the past several years. MAPSI, which provides diverse components to access and analyze polyketide information, has been developed for computational analysis of type I PKSs against genome sequences. MAPSI shows detailed information on polyketides such as PKS gene clusters, chain length, and carbon precursors. In addition, it has user-friendly analysis interfaces, including genome analysis, new PKS assembly, and domain analysis, which may become very powerful computational tools for polyketide research. Finally, it provides management components for MapsiDB, which contains polyketide and genome information.

Because many factors are involved in polyketide biosynthesis, MAPSI could be improved through further development. Although some polyketides have similar chain bodies, they have different structures and activities because of the use of different starter units. Post-synthetic processing steps, such as *O*-methylation, alkylation, cyclization, and glycosylation, also give polyketides diverse attributes. These factors are critically related to the biological activities of many polyketides, making information on these factors necessary for comprehensive analysis of polyketides. MAPSI will later include components enabling analysis of these factors. Although type I polyketides are the most widely known, type II and type III polyketides and nonribosomal peptides are also of considerable medicinal importance. In addition, many polyketides are produced by PKS-NRPS hybrid enzymes. Useful analysis and database systems will be developed for each polyketide.

Availability

The MAPSI software is a Web application.
Project home page: http://gate.smallsoft.co.kr:8080/pks

Acknowledgments

This research was financially supported by the Ministry of Education, Science, and Technology and Korea Industrial Technology Foundation through the Human Resource Training Project, and by the Ministry of Knowledge Economy grants from the Intelligence Bioinformatics and Application Center at the Korea Research Institute of Bioscience and Biotechnology.

References

1. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.
2. Aparicio, J. F., P. Caffrey, A. F. A. Marsden, J. Staunton, and P. F. Leadlay. 1994. Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *J. Biol. Chem.* **269:** 8524–8528.
3. Broadhurst, R. W., D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, and K. J. Weissman. 2003. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10:** 723–731.
4. Cheng, Y. Q., G. L. Tang, and B. Shen. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **100:** 3149–3154.
5. Eddy, S. R. 1999. Profile hidden Markov models. *Bioinformatics* **14:** 755–763.
6. Floss, H. G. and T. W. Yu. 1999. Lessons from the rifamycin biosynthetic gene cluster. *Curr. Opin. Chem. Biol.* **3:** 592–597.
7. Gaitatzis, N., B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle, and R. Müller. 2002. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **277:** 13082–13090.
8. Ikeda, H., T. Nonomiya, M. Usami, T. Ohta, and S. Omura. 1999. Organization of the biosynthetic gene cluster for the polyketide antihelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. U.S.A.* **96:** 9509–9514.
9. Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys* **31:** 264–323.
10. Kim, B. S., D. H. Sherman, and K. A. Reynolds. 2004. An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. *Protein Eng. Des. Sel.* **17:** 277–284.
11. Long, P. F., C. J. Wilkinson, C. P. Bisang, J. Cortes, N. Dunster, M. Oliynyk, *et al.* 2002. Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.* **43:** 1215–1225.
12. Marchler-Bauer, A., A. R. Panchenko, B. A. Shoemaker, P. A. Thiessen, L. Y. Geer, and S. H. Bryant. 2002. CDD: A database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30:** 281–283.
13. McDaniel, R., A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach, and G. Ashley. 1999. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U.S.A.* **96:** 1846–1851.
14. McDaniel, R., C. M. Kao, S. J. Hwang, and C. Khosla. 1997. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4:** 667–674.
15. Moore, B. S. and C. Hertweck. 2002. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **19:** 70–99.
16. O'Hagan, D. 1993. Biosynthesis of fatty acid and polyketide metabolites. *Nat. Prod. Rep.* **10:** 593–624.
17. Omura, S., H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, *et al.* 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of

producing secondary metabolites. *Proc. Nat. Acad. Sci. U.S.A.* **98:** 12215–12220.
18. Pfeifer, B. A. and C. Khosla. 2001. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **65:** 106–118.
19. Rawlings, B. J. 2001. Type I polyketide biosynthesis in bacteria (Part A – erythromycin biosynthesis). *Nat. Prod. Rep.* **18:** 190–227.
20. Rawlings, B. J. 2001. Type I polyketide biosynthesis in bacteria (Part B). *Nat. Prod. Rep.* **18:** 231–281.
21. Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26:** 544–548.
22. Shen, B. and C. R. Hutchinson. 1996. Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A.* **93:** 6600–6604.
23. Tae, H., E. B. Kong, and K. Park. 2007. ASMPKS: An analysis system for modular polyketide synthases. *BMC Bioinformatics* **8:** 327.
24. Watanabe, K., C. C. Wang, C. N. Boddy, D. E. Cane, and C. Khosla. 2003. Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. *J. Biol. Chem.* **278:** 42020–42026.
25. Wiesmann, K. E., J. Cortes, M. J. Brown, A. L. Cutter, J. Staunton, and P. F. Leadlay. 1995. Polyketide synthesis *in vitro* on a modular polyketide synthase. *Chem. Biol.* **2:** 583–589.
26. Witkowski, A., V. S. Rangan, Z. I. Randhawa, C. M. Amy, and S. Smith. 1991. Structural organization of the multifunctional animal fatty-acid synthase. *Eur. J. Biochem.* **198:** 571–579.
27. Wu, K., L. Chung, W. P. Revill, L. Katz, and C. D. Reeves. 2000. The FK520 gene cluster of *Streptomycetes hygroscopicus var. ascomyceticus* (*ATCC 14891*) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251:** 81–90.
28. Yadav, G, R. S. Gokhale, and D. Mohanty. 2003. SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* **31:** 3654–3658.
29. Zdobnov, E. M. and R. Apweiler. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847–848.