# Chapter 18
# Computational Methods for Identification of Novel Secondary Metabolite Biosynthetic Pathways by Genome Analysis

**Swadha Anand**
*National Institute of Immunology, India*

**Debasisa Mohanty**
*National Institute of Immunology, India*

## ABSTRACT

*Secondary metabolites belonging to polyketide and nonribosomal peptide families constitute a major class of natural products with diverse biological functions and a variety of pharmaceutically important properties. Experimental studies have shown that the biosynthetic machinery for polyketide and nonribosomal peptides involves multi-functional megasynthases like Polyketide Synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) which utilize a thiotemplate mechanism similar to that for fatty acid biosynthesis. Availability of complete genome sequences for an increasing number of microbial organisms has provided opportunities for using in silico genome mining to decipher the secondary metabolite natural product repertoire encoded by these organisms. Therefore, in recent years there have been major advances in development of computational methods which can analyze genome sequences to identify genes involved in secondary metabolite biosynthesis and help in deciphering the putative chemical structures of their biosynthetic products based on analysis of the sequence and structural features of the proteins encoded by these genes. These computational methods for deciphering the secondary metabolite biosynthetic code essentially involve identification of various catalytic domains present in*

*this PKS/NRPS family of enzymes; a prediction of various reactions in these enzymatic domains and their substrate specificities and also precise identification of the order in which these domains would catalyze various biosynthetic steps. Structural bioinformatics analysis of known secondary metabolite biosynthetic clusters has helped in formulation of predictive rules for deciphering domain organization, substrate specificity, and order of substrate channeling. In this chapter, the progress in development of various computational methods is discussed by different research groups, and specifically, the utility in identification of novel metabolites by genome mining and rational design of natural product analogs by biosynthetic engineering studies.*

## INTRODUCTION

Polyketides and non-ribosomal peptides constitute the largest family of small molecule natural products biosynthesized by microbes, fungi and plants as secondary metabolites (Linne *et al.*, 2003; Schwarzer *et al.*, 2003; Shen, 2003). These small molecule natural products not only show enormous diversity in their chemical structure, they also have a variety of biomedical and pharmaceutical applications in view of their therapeutic potentials. The elucidation of polyketide and nonribosomal peptide biosynthetic machinery by pioneering genetic and biochemical studies have revealed that these secondary metabolites are biosynthesized by multi functional megasynthases like Polyketide Synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), using an assembly line mechanism which resembles fatty acid biosynthesis. The availability of complete genome sequences in an increasing number of organisms has opened up the possibility of discovering novel secondary metabolite natural products by genome mining (Van Lanen & Shen, 2006). Major advances in biosynthetic engineering during the last decade have also demonstrated the feasibility of obtaining novel engineered natural products by rational manipulation of known secondary metabolite biosynthetic pathways using biosynthetic engineering approaches (Baltz, 2006; Zhang & Wilkinson, 2007). Hence, during the last decade, the research on PKS and NRPS biosynthetic pathways in various organisms has been pursued with two major goals, namely, iden-

tification and experimental characterization of new secondary metabolites in various microbial and fungal species and production of novel rationally designed natural products by manipulation of known PKS/NRPS biosynthetic machinery using a biosynthetic engineering approach.

The remarkable conservation of secondary metabolite gene clusters across organisms has offered abundant scope for obtaining novel insights into the secondary metabolite biosynthetic code by computational analysis. Hence, development of computational methods for relating the chemical structure of the complex secondary metabolites to the amino acid sequence of their corresponding biosynthetic proteins has been an area of active research. Such computational methods (Minowa *et al.*, 2007; Yadav *et al.*, 2003a) have played a major role in guiding various experimental approaches involving genetics, biochemistry, proteomics and metabolomics for discovery of new secondary metabolites by genome mining and reprogramming of known biosynthetic pathways for producing novel natural products by a rational design approach.

In this article, we attempt to give a brief overview of various computational methods which have facilitated easy correlation of the chemical structure of the secondary metabolites to the amino acid sequence of the various PKS and NRPS megasynthases present in the corresponding biosynthetic gene clusters. We first provide background information on various different polyketide and non-ribosomal peptide biosynthetic paradigms. This is followed by a description of different

# Related Content

Artificial Higher Order Neural Network Models
Ming  Zhang (2017). *Nature-Inspired Computing: Concepts, Methodologies, Tools, and Applications  (pp. 241-311).*
www.irma-international.org/chapter/artificial-higher-order-neural-network-models/161032

A Computational Model of Mitigating Disease Spread in Spatial Networks
Taehyong Kim, Kang Li, Aidong Zhang, Surajit Senand Murali Ramanathan (2011). *International Journal of Artificial Life Research (pp. 77-94).*
www.irma-international.org/article/computational-model-mitigating-disease-spread/54749

Sticker P System
A. S. Prasanna Venkatesan, V. Masilamaniand D. G. Thomas (2012). *International Journal of Natural Computing Research (pp. 28-43).*
www.irma-international.org/article/sticker-system/72870

Unsupervised Methods to Identify Cellular Signaling Networks from Perturbation Data
Madhusudan Natarajan (2011). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications  (pp. 337-354).*
www.irma-international.org/chapter/unsupervised-methods-identify-cellular-signaling/52323

Conversion of Tactile Sign Language into English for Deaf/Dumb Interaction
Urmila Shrawankarand Sayli Dixit (2017). *International Journal of Natural Computing Research (pp. 53-67).*
www.irma-international.org/article/conversion-of-tactile-sign-language-into-english-for-deafdumb-interaction/188782