48. Computational approach for understanding substrate specificity of nonribosomal peptide synthetases

Mohd. Zeeshan Ansari, Rajesh S. Gokhale and Debasisa Mohanty
National Institute of Immunology,
New Delhi, INDIA

An automated computational protocol has been developed for identification of various catalytic domains in a putative nonribosomal peptide synthetase (NRPS) and prediction of their substrate preferences. Attempts have been made to model the 3D structure of a minimal NRPS module by computational docking of various NRPS domains.

Nonribosomal peptide synthetases (NRPS) are large multi-enzymatic, multi-domain megasynthases involved in biosynthesis of nonribosomal peptides, a diverse family of pharmaceutically important natural products. Understanding of the sequence to product relationship in NRPS system is essential for rational design of novel peptide antibiotics. Systematic analysis of domain organization has been carried out for a large number of experimentally characterized NRPS clusters for developing knowledge based methods for prediction of sequence to product relationship. Based on this analysis, an automated computational protocol has been developed for identification of NRPS domains in a query sequence. Structure based analysis of the active site residues of various adenylation domains with known specificity, has provided a knowledge base for prediction of substrate specificity of uncharacterized adenylation domains. Benchmarking on a test set of NRPS and PKS clusters has established that it can correctly identify domain boundaries and predict specificity for starter/extender precursor units with a very high accuracy. To the best of our knowledge, NRPS-PKS (1)is the only available tool which can correctly identify various NRPS and PKS domains and predict their specificities. This makes NRPS-PKS a valuable resource for providing leads to decipher the natural products biosynthesized by NRPS/PKS clusters found in newly sequenced microbial genomes.

Condensation domains (C) play a central role in NRPS by making peptide bond between two activated amino acid substrates. This domain along with adenylation and thiolation (T) domain constitutes a minimal module of NRPS, required to incorporate one amino acid in the growing peptide product. A separate class of specialized condensation domains called cyclization domain (Cy) are known to mediate formation of heterocyclic rings such as oxazolines and thiazolines along with making peptide bonds. These domains carry out cyclization along with the condensation, thus having two activities associated with the same domain. Epimerization domains (E), involved in changing the chirality of the activated amino acid substrates are also known to take a fold similar to C and Cy domains. Based on sequence and structural analysis of various types of C, Cy and E domains in experimentally characterized NRPS clusters, we have developed profile HMMs for identifying these domains in uncharacterized NRPS clusters. Our profiles are able to distinguish C, Cy, and E domains with high accuracy. These profiles can also distinguish C domains catalyzing bond between two L-amino acids, from those making peptide bond between D-amino acid and L-amino acid. These analyses can help in design of domain swapping experiment for producing novel nonribosomal peptide products. Crystal structure of PapA5 protein of Mycobacterium tuberculosis shows that it takes structural fold very similar to C domains. Using docking simulations, we have analyzed interaction of PapA5 and acyl carrier protein (ACP) domain, which is analogous to T domains. Our predictions of crucial residues involved in this inter-domain interaction have been experimentally validated by mutational studies(2).

Analysis of inter-domain linker sequences of various characterized NRPS clusters has indicated that they consist of very short stretches of amino acids ranging from 10 to 20 residues. This knowledge of short linkers can help in searching for the interacting surfaces of two domains. Docking simulations were performed between two individual NRPS domains and results were filtered using linker constraint along with other biological filters to produce meaningful complexes. Analysis of the structural models of NRPS modules would help in deciphering the role of linkers and inter-domain interactions in channeling of substrate during biosynthesis of nonribosomal peptides and will help for rational engineering of novel peptide antibiotics.

1. Ansari, M. Z., Yadav, G., Gokhale, R. S., and Mohanty, D. (2004) Nucleic Acids Res 32, W405-413
2. Trivedi, O. A., Arora, P., Vats, A., Ansari, M. Z., Tickoo, R., Sridharan, V., Mohanty, D., and Gokhale, R. S. (2005) Mol Cell 17, 631-643