ModPKSFinder

towards in silico discovery of novel polyketides

Natàlia Padilla Sirera and Stefan Günther, Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, University of Freiburg, Freiburg i. Br., Germany

BACKGROUND

Polyketides are a diverse family of pharmacological important natural products produced by an enzyme called polyketide synthase (PKS) present mostly in bacteria.

Beyond the pharmacological compounds found by traditional approaches of natural product discovery, the post-genomics era have revealed an unexpected wealth of polyketides by genome mining of well know, previously neglected or inaccessible bacteria.

Thanks to the remarkable conservation of both sequence and function of catalytic domains composing a PKS, the structure of a polyketide can be predicted based on the domains identified in the genome and the inference of their biosynthetic activities.

The polyketide structure is determined by the number and type of small organic acids selected by the AT domains and the modifications carried out in them by the KR, DH and ER domains.

OBJECTIVES

1. Polyketide structure prediction

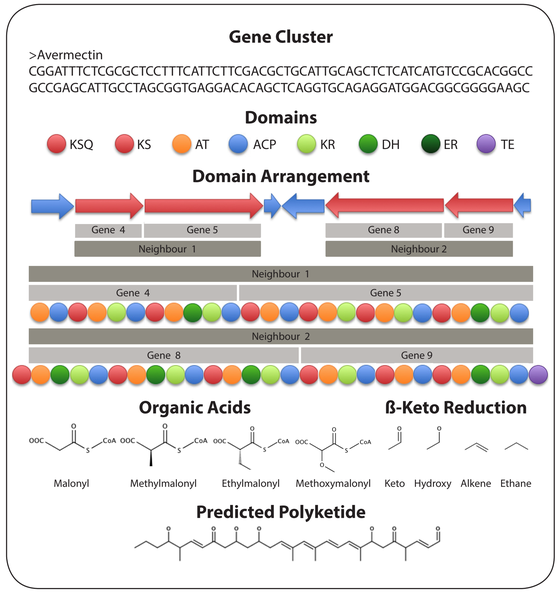
Based on the localisation of the domains in the genome, their ordering in the PKS protein and the inference of their catalytic activities, is accurately predict:

- Chain length, which depends on the number of organic acids incorporated in the polyketide chain. Each AT domain selects one single organic acid each time.
- Pattern of branching, which depends on the type of organic acid selected. The most common are malonyl and methylmalonyl, this last forms a methyl side branch.
- Reduction of the ß-keto group of the organic acids to hydroxyl, alkene or ethane, by the KR, DH or ER domains.

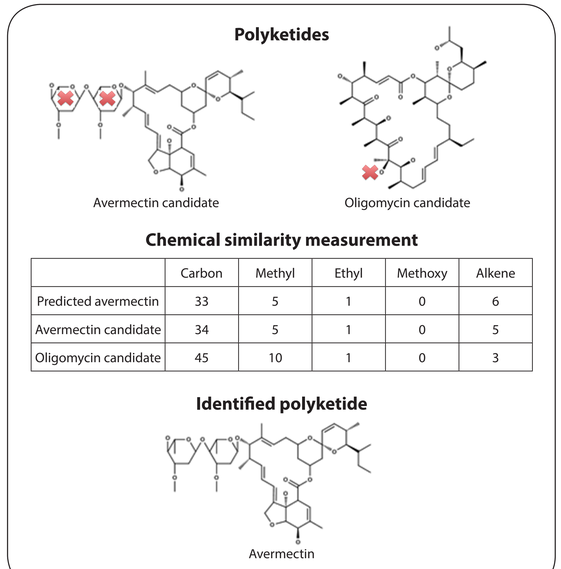
2. Polyketide identification

Based on the genome sequence and the natural products synthesised by a microorganism of interest, identify the polyketide produced by a query PKS gene cluster measuring the chemical similarity between the polyketide candidates and the estimated structure. On the other hand, identify the gene cluster responsible for the synthesis of a query polyketide within the genome of a microorganism.

DESIGN



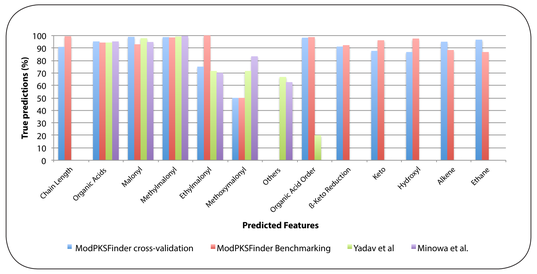
Workflow of polyketide structure prediction
Identification of the PKS gene cluster in the genome sequence of the microorganism of interest, parsing the active catalytic domains, prediction of the arrangement of the catalytic domains in the protein, deciphering the organic acids according to the substrate specificity of the AT domains and the reduction states of their ß keto group.



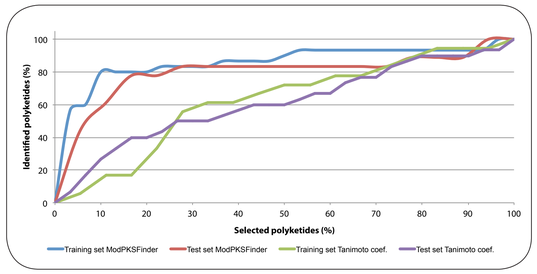
	Carbon	Methyl	Ethyl	Methoxy	Alkene
Predicted avermectin	33	5	1	0	6
Avermectin candidate	34	5	1	0	5
Oligomycin candidate	45	10	1	0	3

Workflow of polyketide identification
On the one hand, the structure of the polyketide is predicted from the query gene cluster, whereas on the other hand, the post-PKS modifications of the polyketide candidates are removed. Subsequently, the chemical similarity between the estimated structure and the polyketide backbounds is measured and the polyketide is identified.

RESULTS



Analysis of polyketide structure prediction
The number, type and order of organic acids composing the polyketide chain, and the reduction state of their ß-keto groups were analysed on a training set as leave-one-out cross-validation and on a new set as benchmarking validation. The results are in percentages. Moreover, the type and order of organic acids predicted by the algorithm of Yadav et al.[10] and Minowa et al.[8] are also provided.



Analysis of polyketide identification
The polyketide identification was analysed on a training set as leave-one-out cross-validation and on a new set as benchmarking validation. The results are shown as an cumulative frequency plot of identified polyketides over selected compounds. Moreover, the polyketide identification using the Tanimoto coefficient on the training and test set are also provided.

CONCLUSIONS

The recent surge in whole-genome sequencing projects has brought into the spotlight the opportunity of natural product discovery by genome mining of secondary metabolite gene clusters and structure prediction of their biosynthetic products.

ModPKSFinder is a powerful tool for structure prediction of polyketides biosynthesised by modular PKS. It accurately predicts the chain length, pattern of branching and ß-keto reduction of polyketide structures. New methodologies have been developed to estimate the substrate specificity of AT domains and to decipher the linear arrangement of protein subunits.

Further, ModPKSFinder allows to identify the gene cluster responsible for the synthesis of a polyketide of interest and to identify the polyketide synthesised by a query gene cluster among a set of natural products.

A future challenge is the prediction of post-PKS modifications, the last step to obtain the final bioactive structure.

REFERENCES

1. Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, et al. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic acids research 38: W487-W496
2. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Nucleic Acids Research 29: 2607-2618
3. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, et al. (2013) antiSMASH 2.0 —a versatile platform for genome mining of secondary metabolite producers. Nucleic acids research 41: W204-W212
4. Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic acids research 41: D402-D407
5. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic acids research 39: W29-W37.
6. Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, et al. (2013) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic acids research 41: D408-D414.
7. Lucas X, Senger C, Erxleben A, Grüning BA, Döring K, et al. (2013) StreptomeDB: a resource for natural compounds isolated from Streptomyces species. Nucleic acids research 41: D1130-D1136.
8. Minowa Y, Araki M, Kanehisa M (2007) Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. Journal of molecular biology 368: 1500-1517.
9. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, et al. (2011) Open Babel: An open chemical toolbox. Journal of cheminformatics 3: 1-14.
10. Yadav G, Gokhale RS, Mohanty D (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. Journal of molecular biology 328: 335-363

ACKNOWLEDGEMENTS

We would like to express our great gratitude to Xavier Lucas for his valuable comments and suggestions and we would also like to offer our special thanks to Kiran T. Telukunta for his valuable technical support.