Novel system for the design and discovery of potential drugs

Antonio Starčević, *Faculty of Food Technology and Biotechnology*

Abstract

Traditional screening has been immensely successful in identifying valuable secondary metabolites. Polyketides are a very important class of natural products and modular polyketide synthases (PKS) have a building block structure that makes them an inviting target for genetic manipulation. *In vitro* genetics has been successful in achieving desired compounds, but the yield is often extremely low. The natural evolution of modular systems must involve homologous recombination. We have been modeling recombination *in silico* in the expectation that homologous recombination will produce more compatible junctions than *in vitro* manipulations and, thus, better yields. Molecular genetic methods are available in *Streptomyces* and related genera, which would allow the laboratory realization of the *in silico* constructs. The program suite ***CompGen*** generates potential recombinants from PKS clusters. The clusters are held in a special database (***r***ecombinant ***C****lust****S****can* ***D***ata***B***ase; ***rCSDB***), which links DNA sequence to the biosynthetic order and, thus, allows prediction of the chemical structure of novel products. ***CompGen*** shows that the number of good sites for homologous recombination between clusters is usually low and, thus, produces a strongly selected set of product molecules, which are likely to mimic those that are available in nature. 16 well annotated clusters were used to test the program and 216 recombination sites were predicted in the total of 120 possible cluster pairs. A second program suite ***ClustScan*** was developed to analyze DNA sequences and annotate modular PKS clusters. The specificity of the modules is predicted from published knowledge about PKS proteins and new analyses are done to identify amino acid residues involved in specificity. This predicts the structure of the products including the stereochemistry. As the prediction is based on identification of critical amino acid residues rather than similarity with known PKS clusters, it is particularly useful for novel clusters that are not closely related to known clusters. ***ClustScan*** can output the predicted chemical structure as an isomeric SMILES (Simplified Molecular Input Line Entry System) allowing analysis by further chemistry-based programs. ***ClustScan*** was initially developed to build the database for ***CompGen***, but is useful in its own right for annotating the ever growing flood of PKS sequences in the DNA databases (***C****lust****S****can* ***D***ata***B***ase; ***CSDB***). The modular biosynthetic clusters in an actinomycete genome sequence can be annotated by a single person in about 3 hours. For those interested the ***ClustScan*** and ***CompGen*** program packages will be demonstrated online after the talk. To download ***ClustScan*** and access services, one should go to the Web site http://bioserv.pbf.hr/cms/.

Key words: ***ClustScan***, ***CompGen***, ***CSDB*** ("*C*lust*S*can *D*ata*B*ase"), ***r-CSDB*** ("*r*ecombinant *C*lust*S*can *D*ata*B*ase"), Web Portal: ***TMSS*** ("*T*hiotemplate *M*odular *S*ystems *S*tudies")