Why the *in silico* Drug Design and Discovery

Daslav Hranueli, *Faculty of Food Technology and Biotechnology*

Abstract

Our approach to the *in silico* drug design and discovery links the DNA sequences encoding modular enzymes that synthesise natural products of clinical importance, polyketides and non-ribosomal peptides, to their *in silico* medicinal chemistry. This can be done because there is a direct correlation of the gene cluster DNA sequence, by virtue of amino acid sequences of modules and catalytically active domains of these multi-functional and multi-modular proteins, to the chemical structure of their biosynthetic product. Therefore, polyketide or peptide chemical structures can be predicted from gene-cluster DNA sequences.

It is well known that more than 50% of drugs - that are today in clinical use has - as their biologically active principls - polyketide or peptide compounds. To them we should add antiparasitics, coccidiostatics, animal growth promotants and natural insecticides used in food and agro-industries. Polyketides and peptides were, therefore, very important for the discovery of present drugs and will continue to be important for the novel drug discovery. Since 1940s the pharmaceutical companies and academic institutions have screened culturable microorganisms from different habitats for their natural products. They discovered a significant number of biologically active polyketide and peptide compounds with important pharmacological properties, some of which has found their way to the market. However, during the last 30 years or so, culturable microorganisms with novel biologically activities become exhausted and the discovery of novel drugs dropped down considerably. The major question now is: are there any alternative strategies to exploit the chemical potential of small molecule natural products?

The answer might lie in recent development of rapid and relatively in-expensive DNA sequencing technologies, which have delivered more than nine hundred sequenced bacterial genomes, with more than thirteen hundred genome sequencing projects in progress. Polyketide and peptide gene clusters were analysed within this genomes. The most prolific producers of these natural products are Actinomyces species. It appeared that genome of every *Actinobacteria* has more than one gene cluster for the synthesis of these secondary metabolites. For instance, the sequenced genome of *Streptomyces avermitilis* has, along the gene clusters for the polyketide antiparasitic avermectin and the polyketide antibiotic oligomycin, another nine polyketide and eight peptide gene clusters. Today, there are a large number of sequenced polyketide, peptide and hybrid gene clusters of unknown products in public databases and their number grows exponentially with every new bacterial genome sequenced. There is, therefore, significant chemical diversity potential in these sequenced genomes and gene clusters.

Even more important, these developments have also opened up new opportunities allowing DNA sequencing of unculturable microorganisms. Today they permit the DNA sequencing of metagenomes, in which a large number of unculturable microorganisms may be present, that can not be grown in the laboratory and whose biosynthetic potential can not be assessed by the traditional methods. Computational tools are now needed to rapidly mine such massive datasets to hunt for gene clusters encoding entirely novel compounds which could be the medicines of the future. This is where we hope our approach will become important. We are developing two generic computer program packages ***ClustScan*** and ***CompGen***. ***ClustScan*** and ***CompGen*** are acronyms of "***Clust***er ***Scan***ner and ***Comp***ound ***Gen***erator". The currents status of their development will be described at this meeting.