Comparative analysis of celesticetin and lincomycin biosynthetic gene clusters

Markéta Koběrská

The introductory part of the thesis presents isolation and sequencing of lincomycin gene cluster from type strain *Streptomyces lincolnensis* ATCC 25466. Two relatively extensive sequence changes and several hundred point mutations were identified if compared with the previously published sequence of the lincomycin industrial strain *Streptomyces lincolnensis* 78-11. Analysis of the cluster flanking regions revealed its localization within the genome of *S. lincolnensis* ATCC 25466. The cluster-bearing cosmid was integrated into the chromosome of lincomycin non-producing strains *Streptomyces coelicolor* CH 999 and *Streptomyces coelicolor* M 145. The modified strains heterologously produced lincomycin, but the level dropped to approximately 1-3% of the production in *S. lincolnensis* ATCC 25466.

The exact sequence of lincomycin gene cluster from the type strain allowed isolation and sequence analysis of the gene cluster of structurally related celesticetin. The analysis revealed 24 putative genes, 18 of them homologous with the genes participating in lincomycin biosynthesis. Four celesticetin specific genes are encoding enzymes involved in the salicylate biosynthesis and attachment, one is coding for celesticetin ornamenting methyltransferase and one is a putative resistance gene. The genetic information for biosynthesis of salicylate and amino sugar building units of celesticetin is prefabricated, *i.e.* organized in compact gene blocks. Analysis of biochemically related antibiotic gene clusters shows that lincomycin biosynthetic cluster originated from a fusion of two ancestral biosynthetic gene clusters, one coding for proline-incorporating lincosamide, the other for pyrrolobenzodiazepine compound with incorporated 4-propyl-L-proline derivative. Both related lincosamide biosynthetic clusters share compactly transferred four-gene module *ccbC/lmbC-ccbF/lmbF*. Results of gene inactivation and screening of protein-protein interaction proved that their protein products form the core of the crucial multimeric N-demethyllincosamide synthetase (NDLS) complex. The sequence coding for the remaining part of NDLS, the peptidyl carrier protein (PCP), forms 5´-terminal part of the gene *lmbN* in the lincomycin cluster but 3´-terminal part of the *ccbZ*, the adjoining gene of the *lmbN* counterpart in the celesticetin cluster. Consequently, PCP-coding sequence jumped between two adjoining genes during the evolution. Both NDLSs involved in celesticetin and lincomycin biosynthesis consist of an identical set of mutually homologous subunits. The evolutionary adaptation of the enzyme substrate specificity to the newly emerging unusual amino acid 4-propyl-L-proline in lincomycin biosynthesis thus resulted only from point mutations of the NDLS subunits.