Variable selection in peptide QSRR model development: Metaheuristic optimization algorithms approach

Petar Žuvela, 유 준†, Katarzyna Macur[1],
Tomasz Bączek[1]
부경대학교; [1]Medical University of Gdańsk
(jayliu@pknu.ac.kr†)

Quantitative Structure-Retention Relationships (QSRRs) present a useful model in prediction of retention time in Reversed Phase – Liquid Chromatography (RP-LC). In this work, five metaheuristic nature-inspired algorithms (Genetic Algorithm, Particle Swarm Optimization, Artificial Bee Colony, Firefly Algorithm, and Flower Pollination Algorithm) were applied in variable selection prior to development of theoretical QSRR models for 83 peptides originating from eight model proteins. They were coupled with three component – Partial Least Squares (PLS), and compared to classical interval PLS (iPLS). Selected subset size was varied, and algorithm parameter optimization was carried out. Fitness function used was Root Mean Square Error of Prediction (RMSEP). It was shown that for a population size of 450, and a mutation rate of 0.2 GA-PLS is superior. It yielded a low RMSEP of 5.534%, and outperformed the rest in terms of computational cost. The best model was further validated with a Y-randomization procedure and external validation on 106 peptides originating from Bacillus subtilis proteome.