ABSTRACT

Quantitative structure-activity relationships studies (QSAR), over several decades, has become a tool providing acceleration and costs reduction of development of new substances with the desirable properties. Prediction of the potential activity of chemical compounds before they are synthesized is used in the search for substances with a broad spectrum of activity, including antimicrobial. The growing incidence of bacterial resistance to the use of antibiotics, drives to prompt development of new substances with high activity.

In this study, QSAR analysis of 266 bis-imidazolium chlorides belonging to the group of quaternary ammonium salts was performed. The aim of the study was to discover quantitative relationships between structure properties of analyzed compounds described by molecular parameters and their antimicrobial activity, expressed as logarithm of the minimal inhibitory concentrations of microbial growth (logMIC). Furthermore, an attempt to demonstrate the advantage of multiple calculations of predictive models over generating single equations was made.

Independent variables were calculated for structures subjected to the molecular modeling and optimization process carried out on the Tryton supercomputer nodes (CI TASK, Gdańsk University of Technology), using Gaussian 09 software. Dependent variables were MIC values obtained for six strains of microorganisms: *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Enterococcus faecalis* and *Candida albicans*.

Using bootstrap technique, QSAR models were generated repeatedly. The equations were calculated using statistical methods: LASSO regression, stepwise regression and partial least squares (PLS) regression. Calculations were carried out in MATLAB R2018a software. Iterative calculations enabled the assessment of the variability of models validation parameters and generation of rankings of the most frequently occurring descriptors correlating with the logMIC value.

The obtained results indicate importance of performing multiple calculations in QSAR studies. This approach allows increasing the reliability of the analysis. It also facilitates the choice of the best equations owing to high variability of its validation parameters, determining models predictive ability.

Equations with high predictive performance were calculated for data sets describing the activity of the tested compounds against Gram-positive bacteria: *Staphylococcus aureus* and *Enterococcus faecalis* and the fungal strain, *Candida albicans*. Statistical methods used in the analysis led to diversified results. The equations obtained using LASSO regression consisted of a few variables, while maintaining acceptable values of predictive parameters. Stepwise regression and PLS generated models with high prediction and low error values, but consisting of a large number of variables.

Results indicate that the descriptors calculated on the basis of the two-dimensional structure of bis-imidazolium chlorides, including the distances between heterocyclic rings or hydrogen bond acceptors, exhibit a high correlation with the antimicrobial activity of these derivatives.

Keywords: QSAR, bootstrap, regression analysis, bis-imidazolium chlorides, antimicrobial activity

Karolak