Abstract
Coronavirus is a type of pneumonia caused by the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). This virus causes severe acute respiratory syndrome and has caused 2 million cases worldwide. Although SARS-CoV-1 was brought under control in 2003, a new strain of the SARS-CoV-2 virus emerged that proved to be more virulent than its predecessor. Regarding the design of a new inhibitor for this strain, the main protease (Mpro) of SARS-CoV was used as the target inhibitor. In the in silico development, the Quantitative Structure Activity Relationship (QSAR) method is used which is quite commonly used in predicting the biological activity of compounds that have not been tested. This study aims to build a QSAR model, using aromatic disulfide compounds as the main protease attenuating compound (Mpro) in SARS-CoV. By using the Genetic Algorithm (GA) selection model and the Support Vector Machine (SVM) prediction method, a combination of dominant descriptors will be obtained for aromatic disulfide compounds, which can be utilized in the development of SARS-CoV virus antiviral agents. The dataset used is data that contains features of aromatic disulfide compounds, along with information on the toxicity activity of the compounds. The descriptor information used is a feature column that has a deviation above 0.5. The best GA selection results obtained by RBF kernel with its smallest MSE value of 0.0273, also the best SVM prediction results obtained through the implementation of the polynomial model, which obtained scores for each internal validation (R^2_{train}) and external validation (R^2_{test}) 95.2% and 67.6%, respectively.

Keywords: *main protease (Mpro), quantitative structure activity relationship (QSAR), genetic algorithm, support vector machine (SVM)*