

QSPR Study for Predicting Critical Micelle Concentrations of Gemini Surfactants

Minji Lee¹, Eun Sil Jin¹ and Sung Kwang Lee^{1,*}

^{1,*}Department of Chemistry, Hannam University, Daejeon, Korea, leesk@hnu.kr

Abstract: CMC (critical micelle concentration) is lowest concentration when micelle was created in aqueous solution of surfactant. In this study, we collected CMC data to develop the QSPR model of gemini surfactants from literatures. Gemini surfactants are composed of two monomeric surfactant molecules linked by a spacer chain. Gemini surfactants can self-assemble at concentrations almost a hundred-fold lower than that of corresponding conventional surfactants. Also, unlike conventional surfactants, Gemini surfactants have unique properties because of the high surface tension lowering ability, good solubility for water.

This study aimed to predict critical micelle concentration (CMC) of Gemini surfactant by using a quantitative structure-property relationship (QSPR) method. The CMC dataset were collected from a series of 130 Gemini surfactants. We tried to calculate each molecular descriptor using PreADMET program from Gemini surfactants substructure which subdivided in various ways. In this study, we split the structures of Gemini surfactants into two types of groups. First type, method 1 is that the structure was divided into two surfactant and spacer; connecting link between two surfactant. Second type, method 2 is that the structure was divided into ion1-spacer-ion2 part and two carbon chain. So, we performed three models which are composed of Method1, Method2 and undivided data model.

The forward selection and bootstrap sampling method were applied to determine the optimum descriptor of the multiple linear regression (MLR). The consensus model were tried for improving the predictive power of each model. It was possible to know the applicable range of the prediction model by applicability domain (AD) of the results of each model. Y-scrambling was performed to confirm chance correlation of the model. Lastly, we compared result of the three model for finding best model. The best model was a consensus model using multiple linear regression ($Data_R^2_{external\ test} : 0.955$, $Method1_R^2_{external\ test} : 0.960$, $Method2_R^2_{external\ test} : 0.962$)

Keywords: CMC, Gemini surfactants, QSPR