

MLR, SVM, and ANN QSAR Studies of Leucine Ureido Derivatives as APN Inhibitors

Guogang TU *

School of Pharmacy, Jiangxi Medical College, Nanchang University, China

SUMMARY. Aminopeptidase N (APN), a promising antitumor target, is a zinc-dependent membrane bound exopeptidase. QSAR studies for a set of 81 leucine ureido derivatives having anti-APN activity were carried out by means of multivariate analysis method, which composed of multiple linear regression (MLR), support vector machine (SVM) and artificial neural network (ANN). The results showed that MLR provided reasonable statistical performance with squared correlation coefficient (R^2 , 0.709–0.726) and leave-one-out cross validation ($Q2\ LOO$, 0.624–0.657). More sophisticated ANN methods provided the highest accuracy with squared correlation coefficient (R^2 , 0.790–0.804) and 5-fold cross validation ($Q2\ L5O$, 0.635–0.693), while SVM gave reasonable performance with squared correlation coefficient (R^2 , 0.687–0.719) and 5-fold cross validation ($Q2\ L5O$, 0.509–0.619). Descriptor analysis according to the regression coefficients from the MLR models indicated that molecules should have high frequency of O and N atoms to form hydrogen bonds, and low molecule mass, and less polar substituent groups on R_1 , and more polar substituent groups on R_2 in order to provide potent biological activities for the leucine ureido scaffold derivatives as potent APN inhibitors.

RESUMEN. La aminopeptidasa N (APN), un prometedor objetivo antitumoral, es una exopeptidasa unida a la membrana dependiente de zinc. Se llevaron a cabo estudios QSAR para un conjunto de 81 derivados de ureido de leucina con actividad anti-APN mediante un método de análisis multivariado, que se componía de regresión lineal múltiple (MLR), máquina de vectores de soporte (SVM) y red neuronal artificial (ANN). Los resultados mostraron que la MLR proporcionó un rendimiento estadístico razonable con un coeficiente de correlación al cuadrado (R^2 , 0,709-0,726) y validación cruzada con exclusión de uno ($Q2\ LOO$, 0,624-0,657). Los métodos ANN más sofisticados proporcionaron la mayor precisión con coeficiente de correlación al cuadrado (R^2 , 0,790-0,804) y validación cruzada de 5 veces ($Q2\ L5O$, 0,635-0,693), mientras que SVM dio un rendimiento razonable con coeficiente de correlación al cuadrado (R^2 , 0,687-0,719) y validación cruzada de 5 veces ($Q2\ L5O$, 0,509-0,619). El análisis de descriptores según los coeficientes de regresión de los modelos MLR indicó que las moléculas deberían tener una alta frecuencia de átomos de O y N para formar enlaces de hidrógeno, y baja masa molecular, y grupos sustituyentes menos polares en R_1 , y grupos sustituyentes más polares en R_2 para proporcionar potentes actividades biológicas para los derivados del andamiaje de leucina ureido como potentes inhibidores de APN.

KEY WORDS: APN inhibitors, artificial neural network, multiple linear regression, QSAR, support vector machine.

* Author to whom correspondence should be addressed. E-mail: tugg110@163.com