

QSAR Study on the IC₅₀ of Thiosemicarbazone Derivatives as PC-3 Inhibitors Based on Mixed Kernel Function Support Vector Machine

Guoliang LI, Xianquan WANG, Anqi LI & Peijian ZHANG *

Qingdao University, No. 308, Ningxia road, QingDao,
Shandong, 266000, China

SUMMARY. Prostate cancer is the second common cancer for males, which has caused over 5 million deaths in last ten years. In the process of discovery of new anti-prostate cancer drugs, thiosemicarbazone derivatives (TDs) had been proved with great anti-prostate cancer activity. In order to screen and design new anti-prostate cancer drugs, four quantitative structure-activity relationship (QSAR) models were established based on a set of 41 TDs to predict their PC-3 (human prostate cell) IC₅₀ values. They were heuristic method model, random forest model, radial basis function kernel support vector machine model, and mixed kernel function support vector machine (MIX-SVM) model, respectively. The results indicate that double bonds, nucleophile reactivity, structural information, and partial charge are significant factors influencing the PC-3 IC₅₀ values of TDs. In addition, genetic algorithm (GA) was used to optimize the search process of optimum parameters of MIX-SVM model. Among four models, MIX-SVM model had the best prediction ability and model robustness: R² of training set was 0.924, R² of test set was 0.922, R²_{cv} of leave one out cross-validation was 0.922. What's more, obtained MIX-SVM model can precisely predict PC-3 IC₅₀ values of TDs and provide information for designing new PC-3 inhibitors that could be used as anti-prostate cancer drugs.

RESUMEN. El cáncer de próstata es el segundo cáncer común en los hombres, que ha causado más de 5 millones de muertes en los últimos diez años. En el proceso de descubrimiento de nuevos fármacos contra el cáncer de próstata, se ha demostrado que los derivados de tiosemicarbazona (TD) tienen una gran actividad contra el cáncer de próstata. Con el fin de examinar y diseñar nuevos fármacos contra el cáncer de próstata, se establecieron cuatro modelos de relación estructura-actividad cuantitativa (QSAR) basados en un conjunto de 41 TD para predecir sus valores IC₅₀ de PC-3 (células de próstata humana). Eran el modelo de método heurístico, el modelo de bosque aleatorio, el modelo de máquina de vector de soporte de función de núcleo radial y el modelo de máquina de vector de soporte de función de núcleo mixto (MIX-SVM), respectivamente. Los resultados indican que los dobles enlaces, la reactividad de los nucleófilos, la información estructural y la carga parcial son factores significativos que influyen en los valores PC-3 IC₅₀ de los TD. Además, se utilizó el algoritmo genético (GA) para optimizar el proceso de búsqueda de parámetros óptimos del modelo MIX-SVM. Entre cuatro modelos, el modelo MIX-SVM tuvo la mejor capacidad de predicción y robustez del modelo: R² del conjunto de entrenamiento fue 0,924, R² del conjunto de prueba fue 0,922, R²_{cv} de la validación cruzada de exclusión fue 0,922. Además, el modelo MIX-SVM obtenido puede predecir con precisión los valores PC-3 IC₅₀ de TD y proporcionar información para diseñar nuevos inhibidores de PC-3 que podrían usarse como medicamentos contra el cáncer de próstata.

KEY WORDS: genetic algorithm, prostate cancer, quantitative structure-activity relationship, random forest, support vector machine.

* Author to whom correspondence should be addressed. E-mail: peijianzh@126.com