ORIGINAL PAPER

QSAR study of amidino bis-benzimidazole derivatives as potent anti-malarial agents against *Plasmodium falciparum*

[a,b]Apilak Worachartcheewan, [a,b]Chanin Nantasenamat*,
[b]Chartchalerm Isarankura-Na-Ayudhya, [b]Virapong Prachayasittikul*

[a]*Center of Data Mining and Biomedical Informatics,* [b]*Department of Clinical Microbiology and Applied Technology,
Faculty of Medical Technology, Mahidol University, 10700 Bangkok, Thailand*

A data set of amidino bis-benzimidazoles, in particular $2'$-arylsubstituted-$1H,1'H$-[2,5′]bisbenzimidazolyl-5-carboximidine derivatives with anti-malarial activity against *Plasmodium falciparum* was employed in investigating the quantitative structure-activity relationship (QSAR). Quantum chemical and molecular descriptors were obtained from B3LYP/6-31g(d) calculations and Dragon software, respectively. Significant variables, which included total energy (E_T), highest occupied molecular orbital (HOMO), Moran autocorrelation-lag3/weighted by atomic masses (MATS3m), Geary autocorrelation-lag8/weighted by atomic masses (GATS8m), and 3D-MoRSE-signal 11/weighted by atomic Sanderson electronegativities (Mor11e), were used in the construction of QSAR models using multiple linear regression (MLR) and artificial neural network (ANN). The results indicated that the predictive models for both the MLR and ANN approaches using leave-one-out cross-validation afforded a good performance in modelling the anti-malarial activity against *P. falciparum* as observed by correlation coefficients of leave-one-out cross-validation ($R_{\text{LOO-CV}}$) of 0.9760 and 0.9821, respectively, root mean squared error of leave-one-out cross-validation (RMSE$_{\text{LOO-CV}}$) of 0.1301 and 0.1102, respectively, and predictivity of leave-one-out cross-validation ($Q^2_{\text{LOO-CV}}$) of 0.9526 and 0.9645, respectively. Model validation was performed using an external testing set and the results suggested that the model provided good predictivity for both MLR and ANN models with correlation coefficient of the external set (R_{Ext}) values of 0.9978 and 0.9844, respectively, root mean squared error of the external set (RMSE$_{\text{Ext}}$) of 0.0764 and 0.1302 respectively, and predictivity of the external set (Q^2_{Ext}) of 0.9956 and 0.9690, respectively. Furthermore, the robustness of the QSAR models is corroborated by a number of statistical parameters, comprising adjusted correlation coefficient (R^2_{Adj}), standard deviation (s), predicted residual sum of squares (PRESS), standard error of prediction (SDEP), total sum of squares deviation (SSY), and quality factor (Q). The QSAR models so constructed provide pertinent insights for the future design of anti-malarial agents.
© 2013 Institute of Chemistry, Slovak Academy of Sciences

Keywords: quantitative structure-activity relationship, malaria, *Plasmodium falciparum*, multiple linear regression, artificial neural network, anti-malarial agent, data mining

Introduction

Malarial infection caused by the *Plasmodium* genus is a public health concern as it gives rise to morbidity and mortality in tropical and subtropical areas. The disease is transmitted by the parasitic female anopheline mosquito (Farooq & Mahajan, 2004; Greenwood et al., 2005; White, 2004). Two of the five species commonly found in malarial infection are *Plasmodium falciparum* and *Plasmodium vivax* (Greenwood et al., 2005; White, 2004). In particular, *P. falciparum* has been documented as causing severe cerebral malaria

---

*Corresponding author, e-mail: chanin.nan@mahidol.ac.th, virapong.pra@mahidol.ac.th