Using Linear Method and Genetic Algorithm for QSAR modeling of Bisphosphonates Based on DFT Method

M. A. Safarpour a, H. Mostofipour b

aDepartment of Physical Chemistry, Faculty of chemistry, Iran University of Science and Technology
bSchool of Science, Ardakan University of Payam noor
msafarpour@Iust.ac.ir

Abstract

In the present investigation, a set of bisphosphonate derivatives consisting of 29 molecules was used for QSAR study. The density functional theory (DFT) method at B3LYP/6-31G (d) level of theory was used to calculate molecular descriptors. Chemical and quantum chemical descriptors were used to derive a quantitative relationship between the antitumor activity of the drugs and structural properties. An equation containing five descriptors with good statistical quality was obtained by multiple linear regression (MLR) using stepwise method. Also, GA-MLR regression was used to model more accurately. The results showed that the results obtained by GA-MLR are similar to MLR.

Keywords: DFT, QSAR, Bisphosphonate, Genetic Algorithm

Introduction

Nitrogen-containing bisphosphonates have direct anticancer activity. They function by inhibiting the enzyme farnesyl pyrophosphate synthase (FPPS), the enzyme responsible for the synthesis of the FPP [1].

Quantitative structure-activity relationships (QSAR), mathematical equations relating chemical structure to their biological activity, are a major factor in drug design [2].

At present, many types of molecular descriptors have been proposed to describe the structural features of the molecules. Latest development, allows calculating quantum chemical descriptors at DFT methods, with higher accuracy including consideration of electron correlation effects. The classical QSAR methods rely principally on the mathematical technique of multiple linear regressions (MLR) [3]. Variable selection methods can be stepwise, genetic algorithm (GA) and etc.
Nowadays, GA is well known as an interesting and more widely used variable selection method. A GA is a stochastic method to solve optimization problems defined by a fitness criteria applying evaluation hypothesis of Darwin and different genetic functions, i.e crossover and mutation [4].

Materials and Methods

In this study, the biological data used are the antitumor activity in human tumor cell lines (IC\textsubscript{50}) [1]. All quantum chemical calculations in the present study were performed using the Gaussian 98 series of program. The molecular descriptors were calculated with DFT method at the hydride functional B3LYP and the medium-size basis set 6-31G(d) level. Quantum chemical descriptors were obtained as follow: \( E_{\text{HOMO}} \), \( E_{\text{LUMO}} \), difference between LUMO and HOMO orbital energies, dipole moment, molecular polarizability, electric field gradient (EFG), electrostatic potential, local charge (LC\textsubscript{i}) calculated according to mulliken population analysis (MPA) and natural population analysis (NPA) at each atom. Also, molecular volume, \( \log p \), molar refractivity and molecular surface area were calculated. Quantum chemical indices of hardness (\( \eta \)), softness (S), electrophilicity (\( \omega \)) and electronegativity (\( \chi \)) were calculated according to the method proposed by Thnikaivelan et al.

MLR analysis and correlation analysis were carried out by the statistics software SPSS 13.0 version. The correlation of each one of the descriptors with each other and with IC\textsubscript{50} was calculated. After correlation analysis, the remaining descriptors were used to construct the MLR model, in accordance with the stepwise method.

Also, the GA was used for the selection of the variables that resulted in the best-fitted models. Appropriate models with low standard errors and high correlation coefficient were obtained.

Finally, cross-validation was done by leave-one-out technique. The high cross-validated correlation coefficient (\( Q^2_{\text{cv}} \)) value obtained for the models.

Results and Discussion

MLR analysis

QSAR model was obtained by using all type of descriptors as following:
\[-\log \text{IC}_{50} = 7.644(\pm 2.385) - 5.273(\pm 1.119) \text{LC}_{C} + 9.464(\pm 2.560) \text{LC}_{C_4} - 17.602(\pm 8.445) \text{LC}_{C_7} - 10.129(\pm 2.203) \text{LC}_{P_{10}} + 1.208(\pm 1.031) \text{EFG}\]

Where \(N=29\), \(R=0.880\), \(R^2=0.774\), \(F=15.773\), \(Q^2_{cv}=0.636\)

The predicted activities by using this equation are plotted against the experimental values in figure 1.

This Equation contains five descriptors to describe the quantitative relationship between the structure and activity of the BP derivatives. Equation, which has high statistical quality, demonstrates that local charge is a major factor controlling the binding of the compounds to the receptor.

Variables used in equation can explain 77.4% of the variance in the biological activity of BP derivatives. Also, the equation suggests local charge at carbon atoms number 1 and 4 (\(\text{LC}_{C1}, \text{LC}_{C4}\)) are sites activation on the pyridine ring of the drugs. The negative coefficients of \(\text{LC}_{C1}, \text{LC}_{C7}\) and \(\text{LC}_{P_{10}}\) in equation, showed the lower values correspond to the effective dosage greater antitumor. And the positive coefficient of EFG reveals that the effective dosages of the drug decrease with increasing the electric field gradient.

GA-MLR

The GA was run many times with different parameters and initial population. The most relevant equations for different set of descriptors are similar to MLR analysis.

![Figure 1. Plot of the predicted activity against the corresponding experimental activity for the MLR and GA-MLR](image)

Conclusions

A QSAR study was employed to study the antitumor activity of twenty nine BP
derivatives. The DFT theory was used to calculate a diverse set of quantum chemical descriptors. In the MLR procedure, the descriptors concerning to the individual atoms in the molecule (local charge at atom number 1, 4, 7 and 10) were found to be important factors controlling the antitumor behavior of the molecules.

In the GA-MLR procedure, the obtained results were similar to the MLR procedure.

Reference
