Quantum Chemical Structure Activity Studies of Anticancer Activity of Seconucleoside Nitrosourea Analogs

Ibrahim Ali Noorbatcha[1], Nurhusna Samsudin[1], Hamzah Mohd Salleh[1] and Syed Zahir Idid[2]

1   Bioprocess and Molecular Engineering Research Unit (BPMERU), Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur, Malaysia

2   Department of Biomedical Science, Faculty of Science, International Islamic University Malaysia, Bandar Indera Mahota, Kuantan, Pahang, Malaysia

Corresponding author:

Ibrahim Ali Noorbatcha

✉   Ibrahiman@iium.edu.my

Bioprocess and Molecular Engineering Research Unit (BPMERU), Department of Biotechnology Engineering, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.

Abstract

Nucleosides can be considered as specific carriers for the nitrosourea moiety with altered pharmacological properties such as improved water solubility, lower myelotoxicity with enhanced cell membrane transport. Several seconucleoside nitrosourea compounds have been reported to have anticancer activity against solid tumors such as, murine adenocarcinomas of the colon 13 (MAC 13), MAC 15A, colon 38 and mammary carcinoma. In an effort to have a good understanding of the role of the structural features on the anticancer activity of these seconucleosides, we have constructed the quantitative structure activity relationship (QSAR) involving 26 seconucleoside nitrosourea compounds having anticancer activity against murine adenocarcinomas of the colon 15A (MAC 15). Traditional QSAR studies require knowledge of several physicochemical properties to obtain a good QSAR. To circumvent this problem, we have made use of quantum chemical methods to calculate several electronic and molecular properties of these compounds and used these properties to obtain the best QSAR using statistical procedures. Semi-empirical quantum chemical RM1 methods were used to optimize the molecular structures. From the optimized structure, hundreds of molecular and electronic properties were calculated and correlated with the anticancer activities of the compounds. The best correlation was obtained using heuristic and multi-linear regression methods. We were able to obtain a best QSAR with correlation coefficient, $R^2=0.8993$. We find that the average structural information contents of the second order, average bond order of oxygen atom and maximum columbic interaction for H-N bond play significant roles in influencing the anticancer activity of the selected compounds. The QSAR obtained in our study can be used to predict the anticancer activity of new seconucleoside nitrosourea analogs, prior to resorting to any experimental studies.

Keywords: QSAR; Anticancer activity; Nucleosides; Solid tumors

Introduction

Diverse names have been proclaimed in the design of efficient nitrosourea compounds capable of anticancer activity. Nucleoside can be considered as a specific carrier for the nitrosourea moiety with altered pharmacological properties such as increased water solubility, lower myelotoxicity and increased cell membrane transport. Besides, these nucleoside analogs can be one of the prodrugs of the nucleoside that permit the slow release of the active drugs in the cancer cells. They also can be considered as irreversible inhibitors for nucleotide-metabolizing enzyme.

Quantitative structure activity relationship (QSAR) which was first developed by Hansch [1] 40 years ago is one of the methods that is commonly used to correlate molecular structure to some kind of *in-vitro* or *in-vivo* biological properties. Such studies have become important for understanding the structure-activity relationships of bioactive compounds to guide discovery of new drugs and optimization of drug activity [2]. There are two goals of QSARs, one of which is to recognize the relation between structure and activity, and to identify which chemical properties will significantly influence the biological activities of a compound. The second goal is to forecast biological activity

for novel and sometimes yet unavailable compound. Due these reasons, QSAR has become an inevitable tool in pharmaceutical industry [3]. New drugs require strict approval to assure its efficacy and safety. Properties such as low toxicity and an appropriate absorption, distribution, metabolism, and excretion (ADME) profile need to be fulfilled. A lot of new drugs failed in the final phase of the developmental process because of poor ADME properties. The process to select the drugs that fulfill all the required characteristics is an expensive and time consuming procedure. QSAR provide an avenue to discover those properties even in the early phase of the drug development. On the basis of the developed QSAR model, drugs that are deficient in required chemical and physical properties can be removed from further

development. Besides, the most suitable compound satisfying all the required criteria can be carried forward for further evaluation [4]. In addition to saving time and cost in chemical syntheses and speeding up decision making process, QSAR model can drastically diminish the use of animals for experimental purpose and fulfill the requirement of international association to avoid animal testing [2].

Materials and Methods

A total of 26 chloroethyl nitrosourea analogs of seconucleosides with known anticancer activity against the mouse adenosarcoma of the colon 15A (MAC 15A) were selected for our analysis
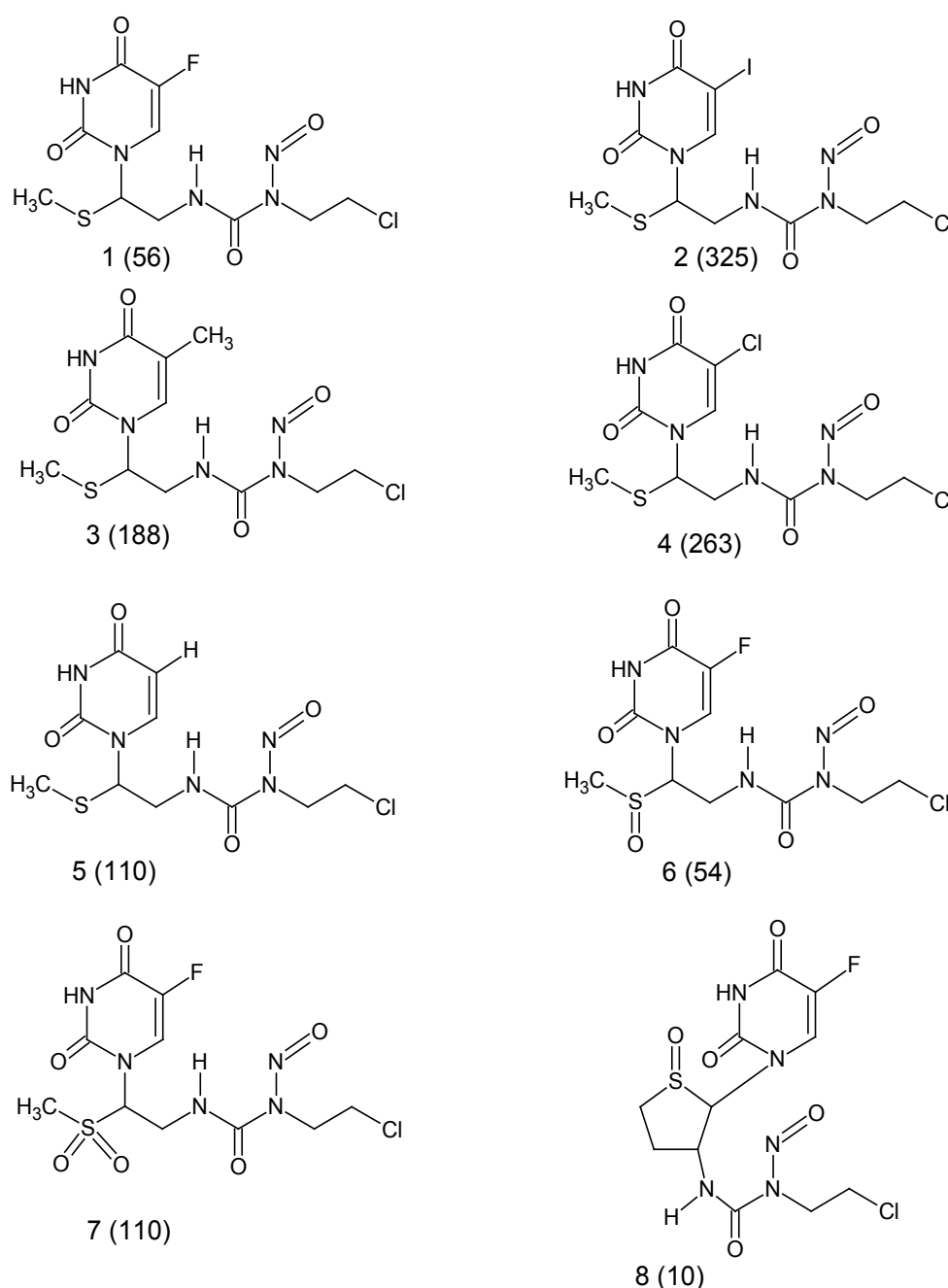

Figure 1a Structures of the compound numbers 1-8. Their respective %ILS values are shown in parenthesis.

(Figure 1). The anticancer activities of these compounds were collected from ref. [5-8].

The biological activities of CENU analogs of seconucleoside have been expressed in the percent increase in life span or %ILS which is calculated using:

$$\%ILS = \left[\frac{(T - C)}{C} \right] \times 100 \qquad (1)$$

Where, T is the median survival time of the treated animals and C is the median survival time of the control such as the untreated animals. The higher the number of T/C or %ILS, the better the drugs can inhibit the growth of cancer cells. A %ILS value of 125% indicates the minimum value necessary for a statistically significant antitumor activity [8].

A set of experimental data in numerical value and a set of corresponding structure were prepared in computer comprehensible format. A specialized molecular editor ACD/ChemSketch version 10.0 [9] was used for the conversion of drawing into the corresponding connectivity table.

3D geometry optimization

The most significant in forecasting and description of biological activities and molecular properties is the molecular shapes and
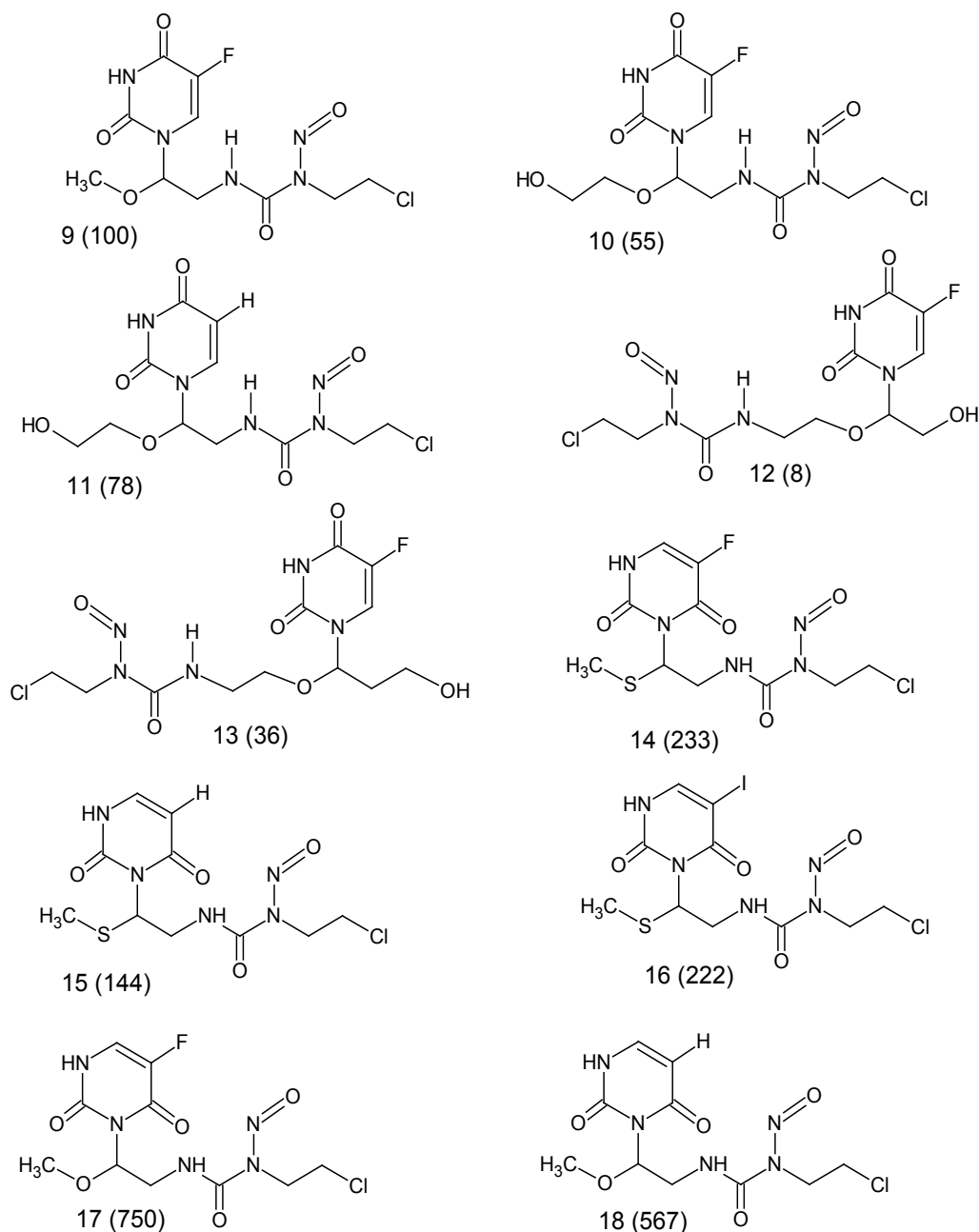

Figure 1b Structures of the compound numbers 9-18. Their respective %ILS values are shown in parenthesis.
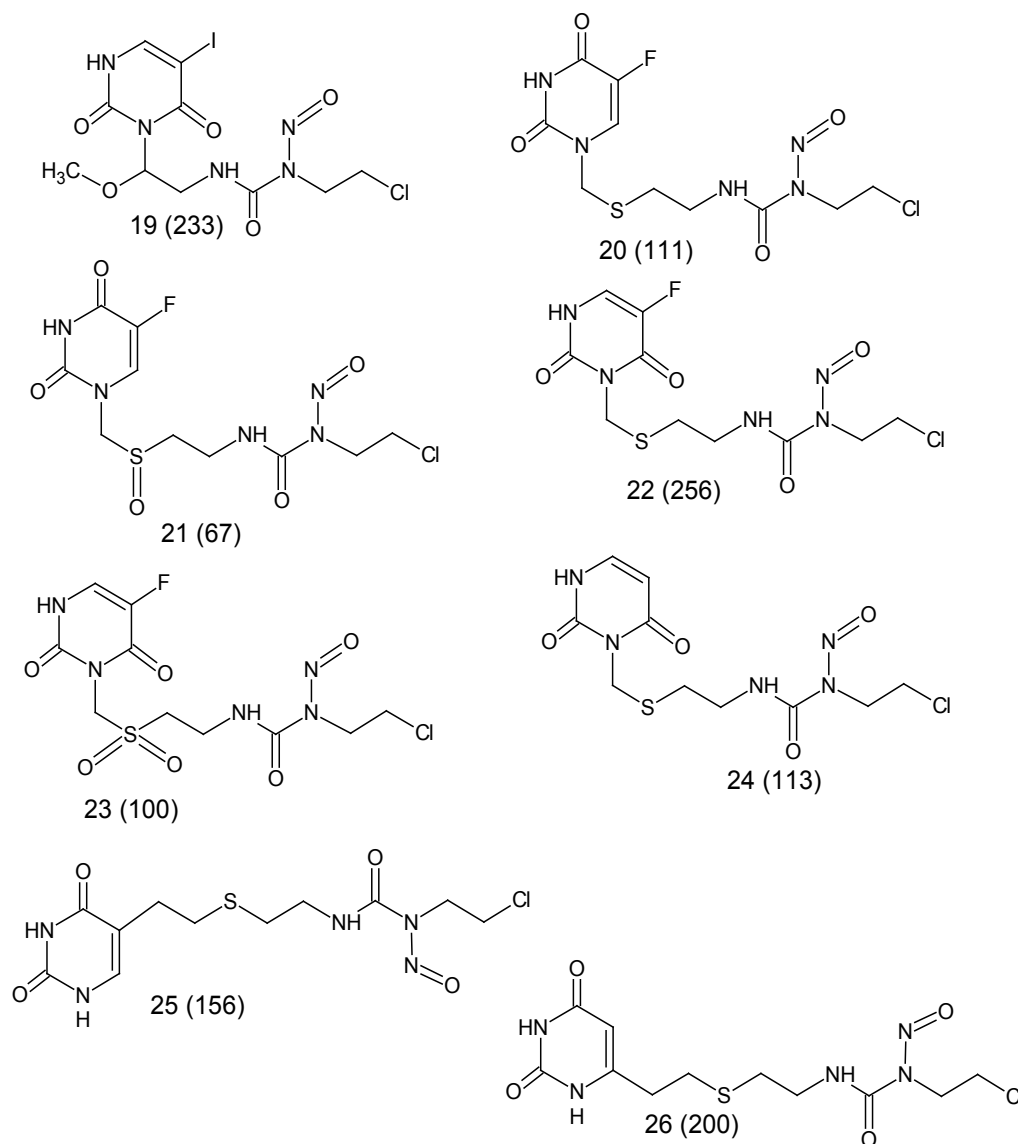
Figure 1c Structures of the compound numbers 19-26. Their respective %ILS values are shown in parenthesis.

conformations. CAChe WorkSystem Pro program [10] was used to convert the 2-D connectivity table into a 3-D image of the molecules with parallel geometry optimization. This program merges a molecular editor with a geometry optimization practice that is usually based on molecular mechanics. The optimized 3-D structures are then subjected to self-consistent field (SCF) calculation using MOPAC 6.0 [11] using RM1 semi empirical quantum chemical method [12]. The optimized geometries are verified with the available structural information.

Calculation of molecular descriptor

The Comprehensive Descriptor for Structural and Statistical Analysis (CODESSA) program [13-15] was used in the calculation of molecular descriptors. Four types of descriptors were calculated; the constitutional, topological, electrostatic, and quantum chemical molecular descriptors. The set of molecular descriptors were further treated with the experimental data by

best multilinear regression analysis. The square of the correlation coefficient R^2, the square of the leave one out cross validated correlation coefficient R^2_{cv} and the F-test were used to determine the best correlation model. The square of the correlation coefficient R^2 is calculated using Eq. 2.

$$R^2 = 1 - \frac{\sum_i^n \left(y_{ei} - y_{ci} \right)^2}{\sum_i^n \left(y_{ei} - y_a \right)^2}$$

(2)

In the above equation the symbols have the following meaning:

n=total number of samples in the data set

y_{ei}=experimental property of the sample "i"

y_{ci}=calculated property of the sample "i"

y_a=average experimental property of all the samples

Due to the limited number of samples available, the data set was not divided in to training set and test set. Instead the "leave one out" method was used to measure the prediction accuracy of the

QSAR. In this technique, each compound is left out once from the analysis and the model is then derived from the remaining objects. The model is then used to predict the activity values of left out compounds. The squared cross-validated correlation coefficient (R^2_{cv}), is calculated Eq. 3.

$$R^2_{cv} = 1 - \frac{\sum_i^n \left(y_{ei} - y_{pi}\right)^2}{\sum_i^n \left(y_{ei} - y_a\right)^2} \qquad (3)$$

where y_{pi} represents the predicted property of the sample "*i*", after this sample is left out of the model. Agreement between R^2 and R^2_{cv} is taken as the measure of the accuracy of predictive ability of the model.

Several strategies are available in the framework of *CODESSA* to develop the QSAR equations with the maximum predictive power. The search for the multi-parameter regression with the maximum predicting ability is performed using the following strategy.

1. Among all the possible descriptors only the orthogonal pairs of descriptors i and j (with $R^2_{ij} < R^2_{min} = 0.1$) are used for further analysis.

2. Next a two-parameter regression with the pairs of descriptors is carried out. All the pairs of descriptors with regression correlation coefficient greater than $R^2_{nc} = 0.65$ are chosen for performing the higher-order regression treatments.

3. For each of the above descriptor pair (I,j), a third descriptor, k (with $R^2_{ik} < R^2_{nc}$ and $R^2_{kj} < R^2_{nc}$), is added, and a new three parameter regression treatment is performed. If the Fisher criterion at a given probability level ($P < 0.05$), F, is higher than that for the two-parameter correlation, then the new three parameter regression is carried over for the next step. A new set of descriptor triples with highest regression correlation coefficients are chosen for the next step.

4. For each descriptor set from the previous step, an additional non-collinear descriptor is added by repeating the above step, till the maximum value of the Fisher criterion and with the highest cross-validated correlation coefficient is obtained.

In using this approach, it is crucial to decide when to stop further addition of descriptors to improve the QSAR equation. We have adopted the technique known as 'breaking point' technique [16] which helps to control the model expansion and thus in turn improves the statistical quality of the model. Addition of any independent variables in the QSAR model can lead to improvement in the R^2 value in the consequent regression. However, if the addition of the new descriptor does not significantly improve the R^2 value, then the added descriptor does not contribute any new information to the model. If the increase in R^2 value by the addition of the new descriptor is less than 0.02, then the QSAR model described without the addition of this new descriptor is considered as the best model.

Results and Discussion

The best QSAR for seconucleoside nitrosoureas and the fitted parameters with their statistical errors are given in **Table 1**. The

QSAR with these parameters can be written as:

$$\log\left(\frac{1}{\%\mathrm{ILS}}\right) = 2.6825(^2\mathrm{SIC}) - 2.9245(q_{\mathrm{NH}}) - 2.8587(b_{\mathrm{OX}}) + 1.1557(q_{\mathrm{N}}) + 5.1200$$

$$(4)$$

The average information content of order *k*, $^K IC$, (Bonchev, 1983) is defined as:

$$^K\mathrm{IC} = \sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n} \qquad (5)$$

where n_i is a number of atoms in the *i*-th class and *n* is a total number of atoms in the molecule. The average structural information content of second order (^2SIC) is defined as

$$^2\mathrm{SIC} = \frac{^2\mathrm{IC}}{\log_2 n} \qquad (6)$$

The empirical partial charge characteristics in the molecule are calculated using the method proposed by ref. [17-18].

The data set for anticancer activity contained structure belonging to nucleoside compounds. In **Figure 2** illustrates the result obtained from a QSAR calculation of 26 anticancer seconucleoside compounds which are the analogs of the 2-chloroethyl nitrosourea (CENU).

The regression presented in **Table 1** with 4 parameters has good squared correlation coefficient $R^2 = 0.8796$ and squared cross-validated correlation coefficient $R^2_{cv} = 0.8521$. The agreement between R^2 and R^2_{cv} shows the predictive accuracy of the developed QSAR equation. One of the most important descriptor in the correlation equation in **Table 1** is the average structural information content of the second order which reflect the number of different structural fragments in the compounds and may therefore be related to the biological activities. Meanwhile, the minimum net atomic charge for a N atom descriptor, describe the availability of a nitrogen lone electron pair for intermolecular hydrogen bonding. However, the remaining two descriptors which are the minimum coulombic interaction for N-H bond and average bond order of oxygen atom can be related to the intramolecular interactions influencing the biological activity through conformational stability. The application of QSAR in combination with the calculation of the quantum chemical descriptors clearly show the role of these molecular interactions in influencing the biological activity of seconucleoside against the mouse adenosarcoma of the colon 15A (MAC 15A). The large values of ^2SIC and q_N will lead to decrease in the anticancer activity, whereas larger values of q_{NH} and b_{ox} will lead to increase in the activity of these molecules, due to their negative contributions. Hence in designing more active molecules, care should be taken to increase the interactions involving –NH and -O groups. Any factor that increases the number of fragments or the negative charge on the nitrogen atom should be avoided, as these changes would lead to the decrease in the anticancer activity of the compounds.

Hence it can be seen that, in addition to contributing the understanding about the role of these interactions in influencing

the biological activity, this approach can be used to predict the anticancer activity of new seconucleoside nitrosourea analogs which will immensely reduce the experimental efforts required to synthesize the drug candidate.
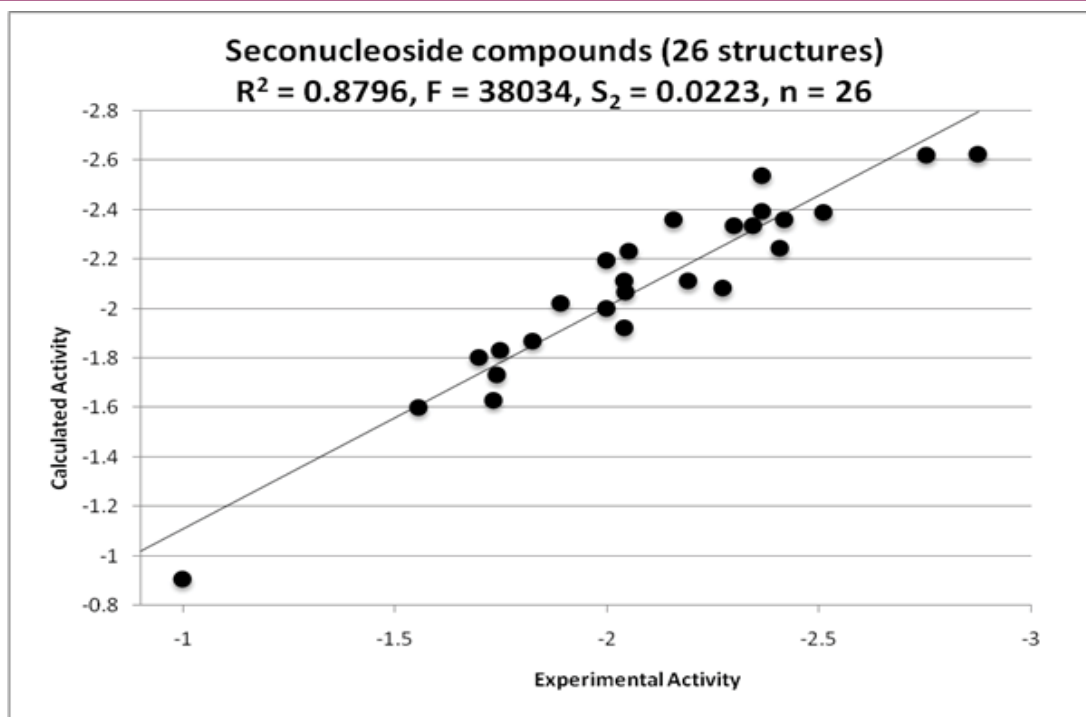
Acknowledgements

Figure 2 Comparison of calculated versus experimental activity (%ILS) using Equation (4).

Table 1 The parameters for the best nonlinear four descriptor QSAR (R^2=0.8796, R^2_{cv}=0.8521, F=38034, S_2=0.0223, n=26).

No	X	Δx	t-test	Descriptor
0	5.1200	3.9048e+00	1.3112	Intercept
1	2.6825	3.7071e-01	7.2362	Average structural information contents of the second order (^2SIC)
2	-2.9245	4.5338e-01	-6.4504	Minimum coulombic interaction for N-H bond, (q_{NH})
3	-2.8587	7.2741e-01	-3.9301	Average bond order of oxygen atom (b_{OX})
4	1.1557	3.9011e-00	2.9176	Minimum net atomic charge for a N atom (q_N)

References

1   Hansch C, Leo A, Schmidt C, Jow PY, Montgomery JY (1980) Antitumor structure activity relations. Nitrosoureas vs. L1210 leukemia. Med Chem 23: 1095-1101.

2   Winkler DA (2001) The role of quantitative structure-activity relationship (QSAR) in biomolecular discovery. Briefing in bioinformatics 3: 73-86.

3   Ericson L, Jawoska J, Worth AP, Cronin MTD, McDowell RM, et al. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression based QSAR. Env health perspective 111: 1361-1375.

4   Fergusan AM, Heritage T, Jonathon P, Pack SE, Philips L, et al. (1997) EVA: A new theoretically based molecular descriptor for use in QSAR/QSPR analysis. J CAMD 11: 143-152.

5   MC, Double JA, McCormick JE, McElhinney RS, Radacic M, et al. (1993) Nucleoside analogues: 13. The effect on anti-tumour activity of varying the uracil 5-substituent and the point of attachment (N1 or N3) of the uracil moiety in seco-nucleoside nitrosoureas. Anticancer Drug Des 8: 115-128.

6   McElhinney RS, McCormick JE, Bibby MC, Double JA, Atassi G, et al. (1989) Nucleoside analogues: 8. Some isomers of B.3839, the original 5-fluorouracil/nitrosourea molecular combination, and their effect on colon, breast and lung tumours in mice. Anticancer Drug Des 4: 1-20.

7   McElhinney RS, McCormick JE, Bibby MC, Double JA, Atassi G, et al. (1989) Nucleoside analogues: 9. Seco-nucleoside analogues of some 5-fluorouracil/nitrosourea molecular combinations having uracil as base: Synthesis and anti-tumour activity. Anticancer Drug Des 4: 191-207.

8   Gnewuch CT, Sosnovsky G (1997) A critical appraisal of the evolution of N-Nitrosoureas as anticancer agent. Chem Rev 97: 829-1013.

9   ACD/ChemSketch (2015) Advanced Chemistry Development Inc. Accessed from: http://www.acdlabs.com

10   CAChe WorkSystem Pro, Version 6.1.1, Fujitsu Limited, Tokyo, Japan.

11   Stewart JJP, Frank J (1990) MOPAC6. Seiler Research Laboratory, United States Air Force Academy, Colorado, USA.

12   Rocha GB, Freire RO, Simas AM, Stewart JJP (2006) RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. J Comput Chem 27: 1101-1111.

13   CODESSA (1995) Semichem: 7204 Summit, Shawnee, Kansas, USA.

14   Katritzky AR, Lobanov VS (1995) QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. Chem Soc Rev 279-287.

15   Karelson M, Lobanov VS (1996) Quantum chemical descriptors in QSAR/QSPR studies. Chem Rev 96: 1027-1043.

16   Zupan J, Gasteiger J (1991) Neural networks - a new method for solving chemical problems or just a passing phase. Anal Chim Acta 248: 1-30.

17   Zefirov NS, Kirpichenok MA, Izmailov FF, Trofimov MI (1987) Scheme for the calculation of the electronegativities of atoms in a molecule in the framework of Sanderson's principle. Dokl Akad Nauk SSSR 296: 883-887.

18   Kirpichenok MA, Zefirov NS (1987) Electronegativity and molecular geometry. I. General principles of the method and analysis of the effect of short-range electrostatic interactions on bond lengths in organic molecules. Zh Org Khim 2: 673-691.