Chemical Data Collections 7–8 (2017) 8–50

Contents lists available at ScienceDirect

Chemical Data Collections

journal homepage: www.elsevier.com/locate/cdc

Data article

Quantitative structure-activity and toxicity relationship study of CCRF-CEM and RPMI 8402 cell line apoptosis with some anticancer compounds

David Ebuka Arthur*, Adamu Uzairu, Paul Mamza, Abechi Eyeji Stephen, Gideon Adamu Shallangwa

Department of Chemistry, Ahmadu Bello University Zaria, Kaduna State, Nigeria

ARTICLE INFO

Article history:
Received 28 October 2016
Revised 13 December 2016
Accepted 20 December 2016
Available online 3 January 2017

Keywords:
GA-MLR
Apoptosis
Molecular descriptors
Williams plot
Cell-lines

ABSTRACT

This research entails the use of 112 anticancer compounds from NCI database to develop QSAR and QSTR models that could be used to predict the activity and toxicity of newly designed compounds which are within the applicability domain of the model. The QSAR/QSTR models of CCRF-CEM and RPMI 8402 cell lines were validated using OECD metric standard for external and internal validation, after which the data were subjected to leverage applicability domain test to identify the chemical space within which the models are most effective. The mean effect of the molecular descriptors in the models were calculated to ascertain the descriptors with the most significant effect on the activities and toxicities of the data set, the result suggests that an increase in the number of double bonds and rings system in the molecules will improve their biological activities and greatly decrease their toxicity at the same time.

* Corresponding author.
E-mail address: hanslibs@myway.com (D.E. Arthur).

http://dx.doi.org/10.1016/j.cdc.2016.12.002
2405-8300/

Specifications Table

Subject area	*Computational chemistry*
Compounds	Aminopterin, camptothecin derivatives and colchicine analogues. Other compounds such as dolastine, glycinate, chlorozotocin as well as other heterocyclic compounds were also in the data set.
Data category	*Data pre-treatment computational simulations, QSAR model, Molecular descriptors contribution.*
Data acquisition format	*Computational studies, Data optimization and Descriptors calculation.*
Data type	*Raw*
Procedure	*Geometric optimization, Molecular descriptor calculation, Data normalization, Dataset division, QSAR model development and validation, Applicability domain of the model.*
Data accessibility	*The computational and statistical data are with this article, while the* raw data collected from the literature (NCI database), which is openly available to the general public on the DTP web site (http://dtp.cancer.gov/mtargets/mt_index.html), while the anti-cancer screening method and assay used to measure the biological activities are reported in the DTP-NCI web site (http://dtp.nci.nih.gov/branches/btb/ivclsp.html).

1. Rationale

Cancer is one of the main cause of death worldwide. Cancer deaths associated with Leukaemia can be credited to a significant % of cancer deaths (26% cases) whereby 30% of this case eventually leads to death [1]. Since 1950 the treatment of leukaemia has made a significant improvement on the activities of the drugs, death rates for infant cancer has fallen by more than 50%. The global survival rate for acute myeloid leukaemia has also improved from 38% to 65% indicating an impressive growth with the search form or effective drugs [2]. Although the accomplishment of new drugs and treatments have limitations related to their side effects and the progress of developed drug resistance [3]. Active new therapeutic agents with novel modes of action and fewer side effects are still needed [4].

Through the past few decades QSAR methods has proven to be an impressive asset in drug design and development [5,6] which computes numerous quantitative descriptors using information extracted from the molecular structure [7,8,6]. Genetic Algorithm multiple linear regression has effectively interrelated and predicted diverse biological properties with molecular finger print [9]. The present study reports the QSAR modelling of biological data obtained at the National Cancer Institute, Bethesda, USA using the leukaemia CCRF-CEM and RPMI 8402 cell lines according to standard OECD procedures [10,11]. CCRF-CEM and RPMI 8402, which originated from the peripheral blood of patients with acute lymphocytic leukaemia, were reported to have T-cell characteristics. However, the present work is focused on the preparation and anti-tumour bioactivity screening of compounds predicted from a QSAR model and presented as a medicinal chemistry study targeting the discovery of anti-tumour active hits.

Procedure

2. Materials and methods

2.1. Experimental dataset

In this study, a dataset of 112 compounds was used to model the relationship between the chemical fingerprints of the compounds and their anticancer activities on leukaemia cell lines. The chemical structures of the data set, NSC and CAS number were taken from the drug discovery and development arm of the National Cancer Institute (NCI). Eligible compounds were determined by reviewing and curating the raw data collected from the literature (NCI database), which is openly available to the general public on the DTP web site (http://dtp.cancer.gov/mtargets/mt_index.html), while the anti-cancer screening method and assay used to measure the biological activities are reported in the DTP-NCI web site (http://dtp.nci.nih.gov/branches/btb/ivclsp.html). The data contains aminopterin and camptothecin derivatives, colchicine analogues and so on. The anticancer activity results are expressed as GI_{50}, which is the concentration for 50% of maximal inhibition of cancer cell proliferation, and LC_{50} which is the 50% lethal concentration of the drug on the cell lines.