Scientific Research Publishing

Computational Chemistry, 2022, 10, 186-202
https://www.scirp.org/journal/cc
ISSN Online: 2332-5984
ISSN Print: 2332-5968

Prediction of Anti-Inflammatory Activity of a Series of Pyrimidine Derivatives, by Multiple Linear Regression and Artificial Neural Networks

Yafigui Traoré[1,2*], Jean Missa Ehouman[1], Mamadou Guy-Richard Koné[1,3], Donourou Diabaté[2], Nahossé Ziao[1,3]

[1]Laboratory of Thermodynamics and Environmental Physico-Chemistry, Nangui Abrogoua University, Abidjan, Ivory Coast
[2]Training and Research Unit Biological Sciences, University Peleforo Gon Coulibaly, Korhogo, Ivory Coast
[3]Ivorian Research Group in Disease Modeling (GIR2M), Abidjan, Ivory Coast
Email: *tyafiguis09@gmail.com

How to cite this paper: Traoré, Y., Ehouman, J.M., Koné, M.G.-R., Diabaté, D. and Ziao, N. (2022) Prediction of Anti-Inflammatory Activity of a Series of Pyrimidine Derivatives, by Multiple Linear Regression and Artificial Neural Networks. *Computational Chemistry*, **10**, 186-202.
https://doi.org/10.4236/cc.2022.104009

Received: August 26, 2022
Accepted: October 25, 2022
Published: October 28, 2022

Abstract

Anti-inflammatory activity of a series of tri-substituted pyrimidine derivatives was predicted using two Quantitative Structure-Activity Relationship models. These relationships were developed from molecular descriptors calculated using the DFT quantum chemistry method using the B3LYP/6-31G(d,p) level of theory and molecular lipophilicity. Thus, the four descriptors which are the dipole moment μ_D, the energy of the highest occupied molecular orbital E_{HOMO}, the isotropic polarizability α and the ACD/logP lipophilicity were selected for this purpose. The Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) models are respectively accredited with the following statistical indicators: $R^2 = 91.28\%$, $R^2_{aj} = 89.11\%$, RMCE = 0.2831, $R^2_{ext} = 86.50\%$ and $R^2 = 98.22\%$, $R^2_{aj} = 97.75\%$, RMCE = 0.1131, $R^2_{ext} = 98.54\%$. The results obtained with the artificial neural network are better than those of the multiple linear regression. However, these results show that the two models developed have very good predictive performance of anti-inflammatory activity. These two models can therefore be used to predict anti-inflammatory activity of new similar pyrimidine derivatives.

Keywords

Anti-Inflammatory Activity, Multiple Linear Regression, Artificial Neural Network, QSAR

1. Introduction

Inflammation is a local response of the organism to an agression of exogenous or endogenous origin [1]. It aims to circumscribe and repair this aggression and involves a series of events that are characterized by a combination of redness, heat, edema and pain [2]. Like pain, inflammation changes behavior. It can then lead to the loss of jobs and even the marginalization of the patient by relatives. It therefore has a very significant social and economic cost [3]. To take part in the treatment of inflammation, several varieties of drugs are available such as aspirin and Non-Steroidal Anti-Inflammatory Drugs (NSAIDs). But many of these varieties have adverse side effects for the body [4]. This is why researchers continue to mobilize in order to find new effective molecules with fewer side effects. Pyrimidine derivatives are a promising avenue. Evidenced by the many studies on series of molecules comprising the core of the pyrimidine and which have analgesic and anti-inflammatory properties [5] [6] [7]. A fairly large number of these studies relate to tri-substituted derivatives of pyrimidine [8] [9]. The results obtained are very encouraging and various substituents are tested. To participate in this dynamic research, Quantitative Structure Activity-Relationship (QSAR) models of anti-inflammatory activity, developed for other organic compounds, are available [10] [11]. But few models relate to the anti-inflammatory activity of pyrimidine derivatives. A QSAR model is an alternative and complementary solution to traditional methods for investigating a biological activity [12]. This approach is increasingly used to reduce the excessive number of experiments, which are sometimes long, dangerous and costly in terms of time and finance [13]. The model establishes a quantitative relationship between biological activity and molecular descriptors. Most models use multiple linear regression. But sometimes linear models are not sufficient to explain all sources of variability due to the complex nature of the relationships between molecular structure and activity [14]. Therefore, nonlinear modeling approaches are used to develop statistically significant and predictive QSAR models [15]. The aim of this work is to develop QSAR models of the anti-inflammatory activity of a series of tri-substituted pyrimidine derivatives using molecular descriptors.

2. Materials and Method

2.1. Computational Theory Level

The quantum descriptor calculation program used in this work is Gaussian 09 [16] with its graphical interface GaussView05. The optimization and the calculation of the frequencies of the molecules were carried out using the Density Functional Theory (DFT) method with the B3LYP functional. The B3LYP functional is a hybrid functional that combines Becke's third parametrization for the exchange energy and the Lee, Yang and Parr functional for the correlation energy [17]. This functional has shown its efficiency for the calculation of many molecular properties [18]. The basis retained is the split-valence and double-dzeta 6-31G(d,p). This basis is sufficiently extensive and the consideration of polariza-

tion functions is important for the explanation of dipole and multipole moments. The B3LYP/6-31G(d,p) level of theory was used to determine the quantum molecular descriptors. The logP molecular lipophilicity of the derivatives was estimated using the ChemSketch software from ACD/Lab [19].

2.2. Molecular Descriptors Used

Some theoretical descriptors have been characterized in order to develop our QSAR model. In particular, the dipole moment μ_D, the energy of the highest occupied molecular orbital E_{HOMO}, the isotropic polarizability α and the logP molecular lipophilicity.

The dipole moment related to the charge distribution is a parameter that relies on the existence of electrostatic dipoles. It is an overall distribution of electric charges in a molecular system, such that the barycenter of the positive charges does not coincide with that of the negative charges. The dipole moment is a vector quantity. The dipole moment makes it possible to describe the global polarity as well as the existence of interaction of molecular systems such as Van der Waals forces, and also to predict their solubility in polar solvents. The dipole moment is an important property that gives an idea of the reactivity of the molecule [20]. It also indicates the stability of a molecule in water. Thus, a strong dipole moment will reflect low solubility in organic solvents and high solubility in water [21].

The highest occupied molecular orbital (HOMO) plays a fundamental role in the qualitative interpretation of chemical reactivity [22]. It is considered the outer orbital containing electrons and it tends to behave as an electron donor.

Another parameter studied is the isotropic polarizability α. It is the ease of a building to deform under the action of an electric field [23]. It is defined by the following relationship [24]:

$$\alpha = \frac{1}{3}\left(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}\right) \tag{1}$$

Finally, the last descriptor evaluated is molecular lipophilicity, which is very important. It is intimately linked to the notion of partition of a molecule between an aqueous phase and a lipid phase [25]. We now know that this capacity for partitioning of a molecule between two phases partly conditions its biological properties such as transport, passage through membranes, bioavailability (distribution and accumulation), affinity for a receptor, protein binding, pharmacological activity, toxicity, accumulation in aquatic organisms, etc. [19].

2.3. Quantitative Structure Activity-Relationship (QSAR)

The objective of a QSAR study is to establish a mathematical relationship between molecular properties called descriptors and a given biological activity, for a series of similar compounds [26]. The equation of such a relationship, when validated, makes it possible to determine the values of the parameters which correspond to optimal activity and to predict the most promising molecular

structure which should be synthesized and tested in the laboratory [27]. It can also be used for the prediction of the properties of molecules already synthesized or not for which the biological activities are not available. The development of a QSAR model must then follow a rigorous scheme in order to achieve a reliable and quantitative result. Thus, the development of a QSAR model 1) begins with the selection of reliable experimental data [28], 2) the calculation of molecular descriptors, as many as possible, 3) the selection of independent and relevant descriptors [29], 4) setting up the QSAR relationship with the selected descriptors using data analysis tools and 5) validating the model developed [30]. Various internal validation criteria exist such as internal correlation coefficient R^2, adjusted correlation coefficient R_{aj}^2, standard deviation RMCE, Fisher coefficient F [30], cross-validation Q_{cv}^2 [31], randomization [32] and also external validation criteria such as R_{ext}^2 and standard deviation RMSEP for the test set, the criteria of Golbraikh and Tropsha [33] as well as those of Roy *et al.* r_m^2, $r_m'^2$, Δr_m^2, $r_m^2(\mathrm{LOO})$ and $r_m^2(\mathrm{overall})$ [34]. These various criteria make it possible to establish the significance, robustness and reliability of the model developed. XLSTAT 2014 and EXCEL 2013 softwares were used to develop the QSAR models and to perform the various calculations.

2.4. Multiple Linear Regression (MLR)

Multiple linear regression is the statistical tool which consists in modeling, using a multiple linear combination, a dependent quantitative variable Y by several independent quantitative explanatory variables X_i ($i=1,\cdots,p$), according to the Equation (2) [35].

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\cdots+\beta_pX_p+\varepsilon \tag{2}$$

where β_i are the regression coefficients and ε is the model error. These coefficients β_i and the variance σ^2 are estimated by minimizing the least squares criterion. The analysis of variance, which is generally done using an ANOVA table, provides access to various model validation parameters such as R^2, R_{aj}^2, RMCE and F defined below [36]:

$$R^2=\frac{\mathrm{SSM}}{\mathrm{SST}}=1-\frac{\mathrm{SSR}}{\mathrm{SST}} \tag{3}$$

$$R_{aj}^2=1-\frac{(n-1)(1-R^2)}{n-p} \tag{4}$$

$$\sigma=\mathrm{SD}=\mathrm{RMCE}=\sqrt{\frac{\mathrm{SSR}}{n-p-1}} \tag{5}$$

$$F=\frac{n-p-1}{n}\frac{\mathrm{SSM}}{\mathrm{SSR}} \tag{6}$$

with:

$$\mathrm{SSM}=\sum_{i=1}^{n}\left(Y_{ipred}-\overline{Y}_{TS}\right)^2 \tag{7}$$

$$\mathrm{SST} = \sum_{i=1}^{n} \left(Y_{iexp} - \overline{Y}_{TS} \right)^2 \tag{8}$$

$$\mathrm{SSR} = \sum_{i=1}^{n} \left(Y_{iexp} - Y_{ipred} \right)^2 \tag{9}$$

$$\mathrm{SST} = \mathrm{SSM} + \mathrm{SSR} \tag{10}$$

n is the number of molecules in the training set (TS) and p the number of descriptors in the model. Y_{iexp} and Y_{ipred} are the experimental and predicted values of the dependent variable Y_i for molecule i; $\overline{Y}_{TS}$ is the mean value of the dependent variable for the training set.

2.5. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a biologically inspired computer algorithm designed to work in the same way as the human brain processes information [37]. It consists of a number of processing elements (or cells) which represent artificial neurons. Each neuron has an input, weights (w_i) associated with each input, a transfer function (f) and an output (a) [38] (see **Figure 1(a)**), which can then branch out to feed a variable number of other neurons [39]. The neurons are interconnected to form the artificial neural network with variable coefficients or weights and are organized into layers: input layer, hidden layers and output layer [40] (see **Figure 1(b)**).

Artificial neural networks have shown great efficiency in modeling nonlinear relationships [15]. The algorithm of multilayer neural networks (or Multilayer Perceptrons) with backpropagation remains the most productive model at the application level and the most widely used [41]. The MATLAB 2017a program was used to build the artificial neural networks of this work.

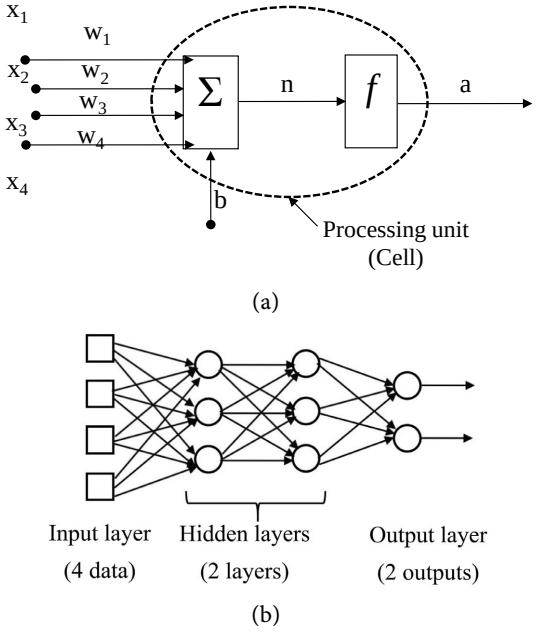


Figure 1. (a) Single neuron with 4 inputs and one output and (b) multilayer perceptron.

3. Results and Discussion

3.1. Analysis of Molecular Descriptors

The work of Vishal *et al.* [42] and Yejella *et al.* [43] provided twenty-eight tri-substituted pyrimidine derivatives with anti-inflammatory activity expressed as a percentage. The general structure of these molecules is as shown in **Figure 2**. The designation codes of the derivatives, the substituents and the percentages of inhibition of inflammation (PI), are collated in **Table 1**.

The results of the calculations of the various molecular descriptors, namely, μ_D, E_{HOMO}, α and logP, for the 28 molecules, are collated in **Table 2**. This table also contains the anti-inflammatory activity expressed by logAI for each derivative of the series. Indeed, the values of the percentages of inhibition of inflammation (PI), were transformed into decimal logarithms logAI according to the expression (12). These new values are collated in **Table 2**.

The decimal logarithm of the anti-inflammatory activity logAI [44] represents the magnitude to be explained in this study. This quantity takes into account both the experimental dose D (10 or 100 mg/kg) of the molecule injected into

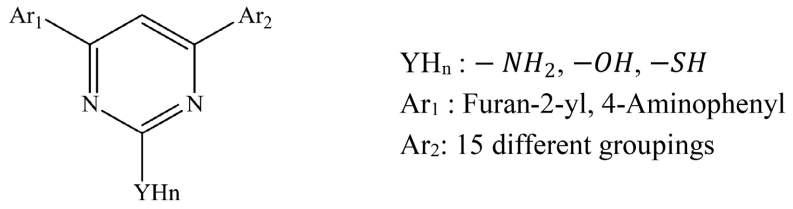


Figure 2. General structure of pyrimidine derivatives.

Table 1. Designation codes, substituents Ar_1, Ar_2 and YH_n and percentages of inhibition of inflammation (PI) of the 28 Pyrimidine derivatives.

CODE	Ar_1	Ar_2	YH_n	PI	CODE	Ar_1	Ar_2	YH_n	PI
DP01	C_4H_3O	4-C_6H_4Br	OH	53.86	DP15	C_4H_3O	4-C_6H_4-CH_3	NH_2	64.10
DP02	C_4H_3O	4-C_6H_4Br	SH	43.82	DP16	C_6H_4-NH_2	4-C_6H_4-Cl	NH_2	82.54
DP03	C_4H_3O	4-C_6H_4Br	NH_2	48.16	DP17	C_6H_4-NH_2	1.4-C_6H_4-Cl_2	NH_2	87.23
DP04	C_4H_3O	4-C_6H_4Cl	OH	49.93	DP18	C_6H_4-NH_2	4-C_6H_4-F	NH_2	83.46
DP05	C_4H_3O	4-C_6H_4Cl	SH	41.08	DP19	C_6H_4-NH_2	3-C_6H_4-Br	NH_2	86.99
DP06	C_4H_3O	4-C_6H_4Cl	NH_2	46.53	DP20	C_6H_4-NH_2	4-C_6H_4-F	NH_2	83.47
DP07	C_4H_3O	4-C_6H_4F	OH	51.43	DP21	C_6H_4-NH_2	4-C_6H_4-OCH_3	NH_2	85.60
DP08	C_4H_3O	4-C_6H_4F	SH	54.28	DP22	C_6H_4-NH_2	3.4-C_6H_4-$(OCH_3)_2$	NH_2	82.13
DP09	C_4H_3O	4-C_6H_4F	NH_2	43.40	DP23	C_6H_4-NH_2	1.3.4-C_6H_4-$(OCH_3)_3$	NH_2	82.82
DP10	C_4H_3O	4-C_6H_4-NH_2	OH	89.71	DP24	C_6H_4-NH_2	4-C_6H_4-C_6H_3	NH_2	83.50
DP11	C_4H_3O	4-C_6H_4-NH_2	SH	69.71	DP25	C_6H_4-NH_2	9-$C_{14}H_9$	NH_2	91.26
DP12	C_4H_3O	4-C_6H_4-NH_2	NH_2	90.21	DP26	C_6H_4-NH_2	2-NC_5H_4	NH_2	83.53
DP13	C_4H_3O	4-C_6H_4-CH_3	OH	68.54	DP27	C_6H_4-NH_2	4-NC_5H_4	NH_2	85.12
DP14	C_4H_3O	4-C_6H_4-CH_3	SH	37.54	DP28	C_6H_4-NH_2	3-NC_5H_4	NH_2	83.96

Table 2. Anti-inflammatory activities (logAI), dipole moment μ_D(Debye), energies of the highest occupied molecular orbital E_{HOMO}(eV), isotropic molecular polarizability α(Bohr^3) and logP lipophilicity of tri-substituted pyrimidine derivatives.

Derivatives	logAI	μ_D	E_{HOMO}	α	logP	Derivatives	logAI	μ_D	E_{HOMO}	α	logP
DP01	0.567	4.252	−6.208	201.710	2.440	DP15	0.640	0.833	−5.731	196.720	3.530
DP02	0.413	4.278	−6.227	219.400	2.790	DP16	2.146	4.716	−5.473	231.730	3.240
DP03	0.466	2.342	−5.875	205.490	3.980	DP17	2.353	3.841	−5.467	239.920	3.740
DP04	0.433	4.321	−6.216	192.590	2.100	DP18	2.150	4.018	−5.428	216.110	2.610
DP05	0.303	4.329	−6.234	210.220	2.450	DP19	2.357	3.349	−5.463	234.920	3.350
DP06	0.373	2.430	−5.885	196.330	3.640	DP20	2.169	1.597	−5.302	239.440	2.490
DP07	0.433	3.928	−6.154	177.320	1.470	DP21	2.282	2.554	−5.358	252.320	2.160
DP08	0.509	4.092	−6.175	194.720	1.820	DP22	2.209	3.554	−5.342	270.040	1.710
DP09	0.291	1.791	−5.820	181.230	3.010	DP23	2.124	2.739	−5.346	231.160	3.130
DP10	1.484	3.538	−5.476	198.280	0.330	DP24	2.189	2.982	−4.966	265.160	2.780
DP11	1.596	4.353	−5.558	216.090	0.680	DP25	2.578	3.433	−5.097	313.030	4.640
DP12	1.615	2.558	−5.326	199.000	1.870	DP26	2.125	1.958	−5.310	211.860	1.330
DP13	0.408	2.944	−6.037	194.070	1.990	DP27	2.178	5.441	−5.529	208.820	1.290
DP14	0.504	3.435	−6.073	210.520	2.340	DP28	2.139	5.040	−5.478	209.650	1.360

the animal, the molar mass M of this injected molecule as well as the physiological response of the animal expressed as a percentage of inhibition (PI) of inflammation [45] according to the expression (12).

$$\log \mathrm{AI} = \log \frac{\mathrm{PI}}{100 - \mathrm{PI}} - \log \frac{D}{M} \tag{12}$$

The analysis of the results of this table reveals that these descriptors vary from one derivative to another and thus depend on the substituents attached to the nucleus of the pyrimidine.

3.2. Statistical Analysis of Data

We seek to build mathematical models capable of explaining and predicting anti-inflammatory activity based on descriptors of free molecules.

For the QSAR model to be simple and understandable, the descriptors used must be meaningful and interpretable [46]. The selection of candidate descriptors for the model is a crucial step and the quality of the model will depend on their relevance because they must provide information that can explain the response (biological activity). To this end, the processing of the descriptors was carried out on the one hand using the one-factor variances and on the other hand using the Pearson correlation coefficients. The analysis of variances makes it possible to eliminate constant or little varied descriptors [28] [47]. The correlation coefficients of the descriptors are calculated taking into account the biological activity expressed by logAI. This consists in bringing the descriptors strongly correlated with each other to the one which is most strongly correlated

with the biological activity. Indeed, descriptors strongly correlated between them are redundant, because they have the same information [48]. These two methods, the results of which are presented in **Table 3** and **Table 4**, show that the four descriptors which are the dipole moment μ_D, the energy of the highest occupied molecular orbital E_{HOMO}, the isotropic polarizability α and the lipophilicity logP, vary well from one derivative to another and that they are linearly independent.

3.3. Prediction of Anti-Inflammatory Activity by Multiple Linear Regression (MLR)

To build and test the multiple linear regression (MLR) model, the initial set of 28 molecules was subdivided into a training set (75%) and a test set (25%) [33] using hierarchical ascending clustering (CAH) [49]. The Euclidean distance between the observations, in the space defined by the descriptors, was retained as the dissimilarity criterion and Ward's method as the aggregation criterion [49]. The multiple linear regression method applied to the training set, using the four descriptors, gave the Equation (13) below:

$$\begin{aligned}\text{Pred}_{\text{MLR}}\left(\log \text{AI}\right) = 11.1714 + 0.2822 * \mu_D + 2.0471 * E_{\text{HOMO}} \\ + 0.0034 * \alpha + 0.0772 * \text{logP}\end{aligned} \tag{13}$$

The statistical indicators of this model are:

$N = 21$, $n = 7$, $R^2 = 91.28\%$, $R^2_{aj} = 89.11\%$, $\text{RMCE} = 0.2831$, $F = 41.8959$, $p < 0.0001$, $Q^2_{cv}\left(\text{LOO}\right) = 84.26\%$, Randomization $R^2_p = 0.7160$, $R^2_{ext} = 86.50\%$, $r^2_m = 0.7978$, $r'^2_m = 0.7928$, $\Delta r^2_m = 0.005$

All these statistical indicators are strongly different from the defined limit values [28] [34]. They thus show that the developed MLR model explains the anti-inflammatory activity of this series of pyrimidine derivatives in a statistically significant and satisfactory manner. This model can thus be considered robust and stable. The predicted values for each set are recorded in **Table 5** as well as the residuals between experimental and predicted values.

Table 3. Variances of anti-inflammatory activity (logAI) and various descriptors.

Grandeurs	logAI	μ_D	E_{HOMO}	α	logP
Variance	0.753	1.218	0.145	885.21	1.034

Table 4. Correlation matrix of the 4 calculated descriptors and the logAI anti-inflammatory activity.

	logAI	μ_D	E_{HOMO}	α	logP
logAI	1				
μ_D	0.09	1			
E_{HOMO}	0.90	−0.22	1		
α	0.71	0.01	0.67	1	
logP	0.01	−0.33	0.07	0.38	1

Table 5. Experimental logAI values, predicted values and residuals e(MLR) of the multiple linear regression model.

TRAINING			
Derivatives	logAI	$Pred_{MLR}$(logAI)	e(MLR)
DP01	0.5668	0.4731	0.0937
DP02	0.4132	0.5235	−0.1103
DP03	0.4664	0.7458	−0.2794
DP05	0.3028	0.4688	−0.1661
DP06	0.3727	0.6972	−0.3245
DP07	0.4332	0.3424	0.0908
DP08	0.5092	0.4264	0.0828
DP10	1.4838	1.5966	−0.1128
DP12	1.6146	1.7483	−0.1337
DP13	0.4076	0.3962	0.0114
DP14	0.5037	0.5384	−0.0346
DP15	0.6403	0.5545	0.0859
DP17	2.3530	2.0932	0.2598
DP18	2.1503	2.0612	0.0891
DP19	2.3567	1.9162	0.4405
DP20	2.1688	1.6999	0.4689
DP22	2.2091	2.2025	0.0066
DP23	2.1242	1.9556	0.1686
DP24	2.1887	2.8790	−0.6902
DP27	2.1776	2.1320	0.0456
DP28	2.1390	2.1310	0.0080
TEST			
DP04	0.4334	0.4209	0.0125
DP09	0.2914	0.5542	−0.2628
DP11	1.5957	1.7407	−0.1450
DP16	2.1460	2.2623	−0.1163
DP21	2.2822	1.8686	0.4136
DP25	2.5777	3.0684	−0.4908
DP26	2.1253	1.6098	0.5155

The analysis of these results shows that, for the two sets, the absolute values of the residuals range from 0.01 to 0.69 with a mean absolute difference (RMCE) of 0.27 defined for the two sets. This result confirms that the predicted values are close to the experimental values overall. The MLR model therefore has a good predictive performance of the anti-inflammatory activity of this series of derivatives.

3.4. Prediction of Anti-Inflammatory Activity by Artificial Neural Networks (ANN)

A feed-forward Backpropagation neural network (a multilayer perceptron) [40] was used, with four inputs corresponding to the four descriptors (μ_D, E_{HOMO}, α and logP), ten hidden layers and one output. The 28 derivatives are randomly divided into three subsets. Training (70%) (20 molecules), validation (15%) (4 molecules) and test (15%) (4 molecules). The training set adjusts the connection weights and model-fitting biases. The validation set verifies the performance of the model throughout the training process and stops training to avoid over-training [38]. The activation functions are the hyperbolic tangent function for the hidden and output layers and the Levenberg-Marquard function for the training set. The performance of the developed model was evaluated by the residual e (ANN) between the predicted and experimental values for each value of logAI. The predicted and experimental values as well as the residuals are presented in **Table 6**.

These results indicate that, for all three sets, the absolute values of the residuals range from 0.00 to 0.29 with a mean absolute deviation (RMCE) of 0.11. This confirms that the predicted values are very close to the experimental values. The ANN model therefore has a very good predictive performance of the anti-inflammatory activity of this series of derivatives.

3.5. Comparison of the Two Established Models

Table 7 brings together the values predicted by each of the two models as well as the residuals and the experimental values for the 28 derivatives studied.

The statistical parameters of the two models are collected in **Table 8**. These parameters show that the two models can predict the anti-inflammatory activity of this series of pyrimidine derivatives, in a statistically significant and satisfactory way. But the results obtained with the model of artificial neural networks are better than those of multiple linear regression. This demonstrates that the model obtained with artificial neural networks has a better predictive capacity of anti-inflammatory activity than that obtained by multiple linear regression. **Figure 3** shows the fit of predicted values and experimental values for the two models. We can see a better match between the values predicted by the artificial neural networks and the experimental values.

Table 6. Experimental values, predicted values of logAI and e(ANN) residuals of the artificial neural network model.

TRAINING			
Derivatives	logAI	$Pred_{ANN}$(logAI)	e(ANN)
DP02	0.4132	0.3976	0.0156
DP03	0.4664	0.4716	−0.0052
DP05	0.3028	0.4259	−0.1231
DP06	0.3727	0.3974	−0.0247

Continued

DP08	0.5092	0.5092	0.0000
DP11	1.5957	1.6404	−0.0447
DP12	1.6146	1.6358	−0.0212
DP13	0.4076	0.4153	−0.0077
DP14	0.5037	0.4135	0.0902
DP15	0.6403	0.4434	0.1969
DP18	2.1503	2.1969	−0.0466
DP19	2.3567	2.2501	0.1066
DP20	2.1688	2.1872	−0.0184
DP21	2.2822	2.2725	0.0097
DP22	2.2091	2.1071	0.1020
DP23	2.1242	2.2650	−0.1408
DP25	2.5777	2.2884	0.2893
DP26	2.1253	1.9419	0.1834
DP27	2.1776	2.0162	0.1614
DP28	2.1390	2.0904	0.0486
	VALIDATION		
DP01	0.5668	0.4425	0.1243
DP09	0.2914	0.3759	−0.0845
DP10	1.4838	1.3969	0.0869
DP16	2.1460	2.3098	−0.1638
	TEST		
DP04	0.4334	0.4938	−0.0604
DP07	0.4332	0.5406	−0.1074
DP17	2.3530	2.3101	0.0429
DP24	2.1887	2.3704	−0.1817

Table 7. Experimental logAI values, predicted values $Pred_{MLR}$(logAI), $Pred_{ANN}$(logAI), residuals e(MLR) and e(ANN) of the MLR and ANN models for the 28 derivatives studied.

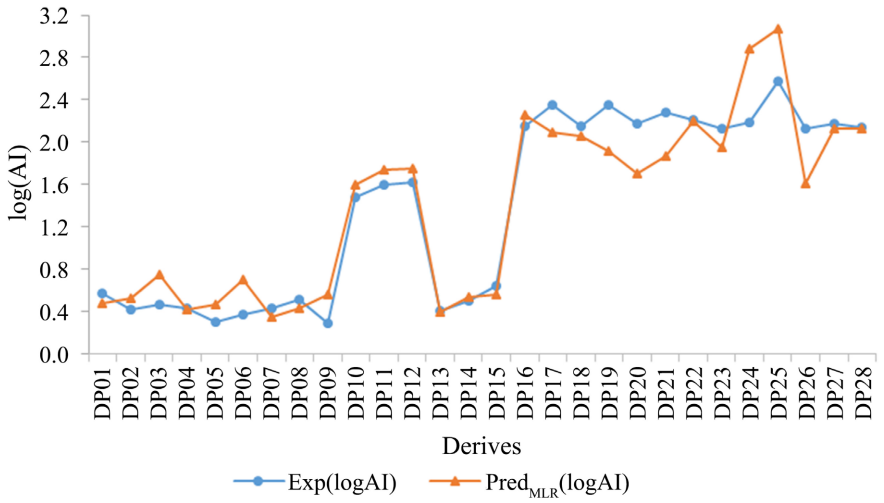
		MLR		ANN	
Derivatives	logAI	$Pred_{MLR}$(logAI)	e(MLR)	$Pred_{ANN}$(logAI)	e(ANN)
DP01	0.5668	0.4731	0.0937	0.4425	0.1243
DP02	0.4132	0.5235	−0.1103	0.3976	0.0156
DP03	0.4664	0.7458	−0.2794	0.4716	−0.0052
DP04	0.4334	0.4209	0.0125	0.4938	−0.0604
DP05	0.3028	0.4688	−0.1661	0.4259	−0.1231
DP06	0.3727	0.6972	−0.3245	0.3974	−0.0247
DP07	0.4332	0.3424	0.0908	0.5406	−0.1074
DP08	0.5092	0.4264	0.0828	0.5092	0.0000
DP09	0.2914	0.5542	−0.2628	0.3759	−0.0845

Continued

DP10	1.4838	1.5966	−0.1128	1.3969	0.0869
DP11	1.5957	1.7407	−0.1450	1.6404	−0.0447
DP12	1.6146	1.7483	−0.1337	1.6358	−0.0212
DP13	0.4076	0.3962	0.0114	0.4153	−0.0077
DP14	0.5037	0.5384	−0.0346	0.4135	0.0902
DP15	0.6403	0.5545	0.0859	0.4434	0.1969
DP16	2.1460	2.2623	−0.1163	2.3098	−0.1638
DP17	2.3530	2.0932	0.2598	2.3101	0.0429
DP18	2.1503	2.0612	0.0891	2.1969	−0.0466
DP19	2.3567	1.9162	0.4405	2.2501	0.1066
DP20	2.1688	1.6999	0.4689	2.1872	−0.0184
DP21	2.2822	1.8686	0.4136	2.2725	0.0097
DP22	2.2091	2.2025	0.0066	2.1071	0.1020
DP23	2.1242	1.9556	0.1686	2.2650	−0.1408
DP24	2.1887	2.8790	−0.6902	2.3704	−0.1817
DP25	2.5777	3.0684	−0.4908	2.2884	0.2893
DP26	2.1253	1.6098	0.5155	1.9419	0.1834
DP27	2.1776	2.1320	0.0456	2.0162	0.1614
DP28	2.1390	2.1310	0.0080	2.0904	0.0486
RMSE(RMCE)			**0.2715**		**0.1138**

Table 8. Statistical parameters R^2, R^2_{aj}, MCE and R^2_{ext} of each of the two models.

Model	R^2	R^2_{aj}	MCE	R^2_{ext}
MLR	91.28	89.11	0.0802	0.8650
ANN	98.22	97.75	0.0128	0.9854



(a)

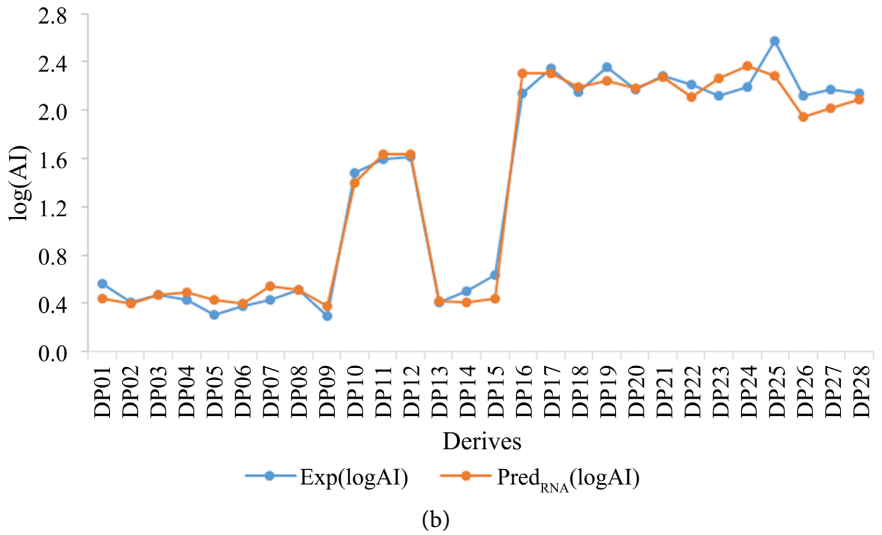


Figure 3. Similarity between predicted and experimental values of logAI. (a) Similarity Between Exp(logAI) and Pred$_{MLR}$logAI; (b) Similarity Between Exp(logAI) et Pred$_{RNA}$logAI.

4. Conclusion

This work allowed us to build two models for predicting the anti-inflammatory activity of a series of tri-substituted derivatives of pyrimidine using quantum descriptors such as the dipole moment μ_D, the energy of the highest occupied molecular orbital E_{HOMO}, isotropic polarizability α and molecular lipophilicity logP. Multiple linear regression (MLR) and artificial neural networks (ANN) methods were used to develop these models. The multiple linear regression model has obtained the following statistical parameters: $R^2 = 91.28\%$, $R^2_{aj} = 89.11\%$, RMCE = 0.2831, $R^2_{ext} = 86.50\%$ while that of the artificial neural networks has the following values: $R^2 = 98.22\%$, $R^2_{aj} = 97.75\%$, RMCE = 0.1131, $R^2_{ext} = 98.54\%$. The results obtained with RNA are better than those obtained with RLM. However, the statistical parameters show that the two models have a very good predictive performance of anti-inflammatory activity. In short, the two models developed make it possible to explain the anti-inflammatory activity of this series of pyrimidine derivatives, in a statistically significant and satisfactory manner. They can be considered sturdy and stable. In perspective, these two models can be used to predict the anti-inflammatory activity of new pyrimidine derivatives for which no experiment has yet been carried out in this direction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Fokunang, C.N., Tembe, E., Frederick, K. and Ngadjui, B. (2018) Overview of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) in Resource Limited Countries.

MOJ Toxicology, **4**, 5-13. https://doi.org/10.15406/mojt.2018.04.00081

[2] Yaro, B., *et al.* (2015) Activités anti-inflammatoire, analgésique et antioxydante de l'extrait aqueux des tiges feuillées de Saba senegalensis Pichon (Apocynaceae) Anti-Inflammatory, Analgesic and Antioxidant Activities of an Aqueous Extract of *Saba senegalensis* Pichon Stems with. *Phytothérapie*, **14**, 213-219. https://doi.org/10.1007/s10298-015-0992-5

[3] Martel, M.O., Finan, P.H., Dolman, A.J., Edwards, R.R., Wasan, A.D. and Robert, N. (2016) Self-Reports of Medication Side Effects and Pain-Related Activity Interference in Patients with Chronic Pain: A Longitudinal Cohort Study. *Pain*, **156**, 1092-1100. https://doi.org/10.1097/j.pain.0000000000000154

[4] Hanna, M.M. (2012) New pyrimido[5,4-e]pyrrolo[1,2-c]pyrimidines: Synthesis, 2D-QSAR, Anti-Inflammatory, Analgesic and Ulcerogenicity Studies. *European Journal of Medicinal Chemistry*, **55**, 12-22. https://doi.org/10.1016/j.ejmech.2012.06.048

[5] Shefali, A. and Vijay, S. (2011) QSAR Study on Some Newly Synthesized Pyrimidobenzimidazole Derivatives as Analgesic Agents. *International Journal of Pharmacy and Pharmaceutical Sciences*, **3**, 457-461.

[6] Abu-hashem, A.A. and Youssef, M.M. (2011) Synthesis of New Visnagen and Khellin Furochromone Pyrimidine Derivatives and Their Anti-Inflammatory and Analgesic Activity. *Molecules*, **16**, 1956-1972. https://doi.org/10.3390/molecules16031956

[7] Kumar, V., Sharma, S. and Husain, A. (2015) Synthesis and *in Vivo* Anti-Inflammatory and Analgesic Activities of Oxadiazoles Clubbed with Benzothiazole Nucleus. *International Current Pharmaceutical Journal*, **4**, 457-461. https://doi.org/10.3329/icpj.v4i12.25597

[8] Giles, D., Roopa, K., Sheeba, F.R., Gurubasavarajaswamy, P.M., Divakar, G. and Vidhya, T. (2012) Synthesis Pharmacological Evaluation and Docking Studies of Pyrimidine Derivatives. *European Journal of Medicinal Chemistry*, **58**, 478-484. https://doi.org/10.1016/j.ejmech.2012.09.050

[9] Nassar, E., El-badry, Y.A. and El Kazaz, H. (2016) Synthesis, *in Vivo* Anti-Inflammatory, and *in Vitro* Antimicrobial Activity of New 5-Benzofuranyl Fused Pyrimidines. *Chemical and Pharmaceutical Bulletin*, **64**, 558-563. https://doi.org/10.1248/cpb.c15-00922

[10] Khedekar, P. (2015) QSAR Studies of 6-Aminopyrimidin-4-One Derivatives as Anti-Inflammatory Agent. *International Journal of Pharmacy & Biomedical Research*, **2**, 13-16.

[11] Kawade, D.P., Khedekar, P.B. and Bhusari, K.P. (2011) QSAR Studies of 6-Aminopyrimidin-4-One Derivatives as Anti-Inflammatory Agent. *International Journal of Pharmacy & Biomedical Research*, **2**, 13-16.

[12] Traoré, Y., Koné, M.G.-R., Ouattara, O. and Ziao, N. (2018) QSAR Approach to Estimating the Analgesic Activity of a Series of Tri-Substituted Pyrimidine Derivatives. *Journal of Computational Chemistry & Molecular Modelling*, **3**, 221-234. https://doi.org/10.25177/JCCMM.3.1.2

[13] Rekka, E.A. and Kourounakis, P.N. (2008) Chemistry and Molecular Aspects of Drug Design and Action. Taylor & Francis Group, Abingdon-on-Thames. https://doi.org/10.1201/9781420008272

[14] Zuvela, P., David, J. and Wong, M.W. (2018) Interpretation of ANN-Based QSAR Models for Prediction of Antioxidant Activity of Flavonoids. *Journal of Computational Chemistry*, **39**, 953-963. https://doi.org/10.1002/jcc.25168

[15] Mitchell, J.B. (2014) Machine Learning Methods in Chemoinformatics. *Wiley Interdisciplinary Reviews*: *Computational Molecular Science*, **4**, 468-481. https://doi.org/10.1002/wcms.1183

[16] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., *et al.* (2009) Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford.

[17] Erickson, C.B., Ankenman, B.E. and Sanchez, S.M. (2018) Comparison of Gaussian Process Modeling Software. *European Journal of Operational Research*, **266**, 179-192. https://doi.org/10.1016/j.ejor.2017.10.002

[18] Gayathri, R. and Arivazhagan, M. (2014) Molecular Structure, Vibrational Investigation of 2-chloro-*α*-*α*-*α*-trifluoro-3,5-dinitrotoluene Using DFT (LSDA, B3LYP, B3PW91 and MPW1PW91) and UV-Vis Absorption Spectra in Organic Solvents: A IEF-PCM/TD-DFT Study. *Spectrochimica Acta*, *Part A*: *Molecular and Biomolecular Spectroscopy*, **123**, 309-326. https://doi.org/10.1016/j.saa.2013.12.012

[19] Arnott, J.A., Kumar, R. and Planey, S.L. (2013) Lipophilicity Indices for Drug Development. *Journal of Applied Biopharmaceutics and Pharmacokinetics*, **1**, 31-36.

[20] Costa, C. and Silva, B. (2017) Analysis and Validation of Dipole Moment Calculations in Chemistry Teaching. *The Electronic Journal of Chemistry*, **9**, 360-368. https://doi.org/10.17807/orbital.v9i5.1020

[21] Peter, E., Teresa, M. and Catanho, J.D.A. (2008) Quantitative Structure-Activity Relationships (QSAR) of 4-Amino-2,6-diarylpyrimidine-5-carbonitriles with Anti-Inflammatory Activity. *Journal of the Brazilian Chemical Society*, **19**, 337-343. https://doi.org/10.1590/S0103-50532008000200021

[22] Lopachin, R.M., Geohagen, B.C. and Nordstroem, L.U. (2019) Mechanisms of Soft and Hard Electrophile Toxicities. *Toxicology*, **418**, 62-69. https://doi.org/10.1016/j.tox.2019.02.005

[23] Sabirov, D.S. (2013) Anisotropy of Polarizability of Fullerene Higher Adducts for Assessing the Efficiency of Their Use in Organic Solar Cells. *The Journal of Physical Chemistry*, **117**, 9148-9153. https://doi.org/10.1021/jp401944x

[24] Prabavathi, N., Nilufer, A., Krishnakumar, V. and Akilandeswari, L. (2012) Molecular and Biomolecular Spectroscopy Spectroscopic, Electronic Structure and Natural Bond Analysis of 2-aminopyrimidine and 4-aminopyrazolo [3, 4-d] pyrimidine: A Comparative Study. *Spectrochimica Acta Part A*: *Molecular and Biomolecular Spectroscopy*, **96**, 226-241. https://doi.org/10.1016/j.saa.2012.05.015

[25] Rutkowska, E. (2013) Review Lipophilicity—Methods of Determination and Its Role in Medicinal Chemistry. *Acta Poloniae Pharmaceutica ñ Drug Research*, **70**, 3-18.

[26] Peter, S.C., *et al.* (2018) Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. In: Ranganathan, S., *et al.*, Eds., *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, Amsterdam, 661-676. https://doi.org/10.1016/B978-0-12-809633-8.20197-0

[27] Wang, T., Wu, M., Lin, J. and Yang, L. (2015) Quantitative Structure-Activity Relationship Promising Advances in Drug Discovery Platforms. *Expert Opinion on Drug Discovery*, **10**, 1283-1300. https://doi.org/10.1517/17460441.2015.1083006

[28] Roy, K. and Kar, S. (2015) A Primer on Qsar/Qspr Modeling. Springer, Berlin, 37-59. https://doi.org/10.1007/978-3-319-17281-1

[29] Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C.P. and Agrawal, R.K. (2011) Validation of QSAR Models—Strategies and Importance. *International Journal of Drug Design & Discovery*, **2**, 511-519.

[30] Khan, A.U. (2016) Descriptors and Their Selection Methods in QSAR Analysis: Paradigm for Drug Design. *Drug Discovery Today*, **21**, 1291-1302. https://doi.org/10.1016/j.drudis.2016.06.013

[31] Qin, L., Zhang, X., Chen, Y., Mo, L., Zeng, H. and Liang, Y. (2017) Predictive QSAR Models for the Toxicity of Disinfection Byproducts. *Molecules*, **22**, Article No. 1671. https://doi.org/10.3390/molecules22101671

[32] Lipiński, P.F.J. and Szurmak, P. (2017) Scramble "N" Gamble: A Tool for Fast and Facile Generation of Random Data for Statistical Evaluation of QSAR Models. *Chemical Papers*, **71**, 2217-2232. https://doi.org/10.1007/s11696-017-0215-7

[33] Martin, T.M., *et al.* (2012) Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *Journal of Chemical Information and Modeling*, **52**, 2570-2578. https://doi.org/10.1021/ci300338w

[34] Roy, P.P., Paul, S., Mitra, I. and Roy, K. (2009) On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules*, **14**, 1660-1701. https://doi.org/10.3390/molecules14051660

[35] Rooki, R., Ardejani, F.D. and Moradzadeh, A. (2014) Hole Cleaning Prediction in Foam Drilling Using Artificial Neural Network and Multiple Linear Regression. *Geomaterials*, **4**, 47-53. https://doi.org/10.4236/gm.2014.41005

[36] Affi, T.S., Ziao, N., Kone, M.G.-R., Bamba, K. and Ouattara, O. (2017) Can Empirical Descriptors Reliably Predict Molecular Lipophilicity ? A QSPR Study Investigation. *International Journal of Engineering Research and Applications*, **7**, 50-56. https://doi.org/10.9790/9622-0705015056

[37] Ge, B., Ishaku, M.M. and Lewu, H.I. (2021) Research on the Effect of Artificial Intelligence Real Estate Forecasting Using Multiple Regression Analysis and Artificial Neural Network: A Case Study of Ghana. *Journal of Computer and Communications*, **9**, 1-14. https://doi.org/10.4236/jcc.2021.910001

[38] Baskina, I.I., Winklerb, D. and Tetko, I.V. (2016) A Renaissance of Neural Networks in Drug Discovery. *Expert Opinion on Drug Discovery*, **11**, 785-795. https://doi.org/10.1080/17460441.2016.1201262

[39] Bahia, I.S.H. (2013) Using Artificial Neural Network Modeling in Forecasting Revenue: Case Study in National Insurance Company/Iraq. *International Journal of Intelligence Science*, **3**, 136-143. https://doi.org/10.4236/ijis.2013.33015

[40] Wesolowski, M. and Suchacz, B. (2012) Artificial Neural Networks: Theoretical Background and Pharmaceutical Applications: A Review. *Journal of AOAC International*, **95**, 652-668. https://doi.org/10.5740/jaoacint.SGE_Wesolowski_ANN

[41] Sutariyaa, V., Grosheva, A., Sadanab, P., Bhatiab, D. and Pathak, Y. (2013) Artificial Neural Network in Drug Delivery and Pharmaceutical Research. *The Open Bioinformatics Journal*, **7**, 49-62. https://doi.org/10.2174/1875036201307010049

[42] Joshi, V.D., Kshirsagar, M.D. and Singhal, S. (2012) Synthesis and Pharmacological Study of Some Novel Pyrimidines. *Der Pharmacia Sinica*, **3**, 343-348.

[43] Yejella, R.P. and Atla, S.R. (2011) A Study of Anti-Inflammatory and Analgesic Activity of New 2,4,6-Trisubstituted Pyrimidines. *Chemical and Pharmaceutical Bulletin* (*Tokyo*), **59**, 1079-1082. https://doi.org/10.1248/cpb.59.1079

[44] Datar, P. (2015) QSAR and Synthesis of a Novel Biphenyl Carboxamide Analogue for Analgesic Activity. *Modern Chemistry and Applications*, **3**, 1-7. https://doi.org/10.4172/2329-6798.1000148

[45] Adeniji, S.E., Uba, S., Uzairu, A. and Arthur, D.E. (2019) A Derived QSAR Model for Predicting Some Compounds as Potent Antagonist against *Mycobacterium tu-*

berculosis: A Theoretical Approach. *Advances in Preventive Medicine*, **2019**, Article ID: 5173786. https://doi.org/10.1155/2019/5173786

[46] Dral, P.O., *et al.* (2016) Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *Journal of Chemical Theory and Computation*, **12**, 1082-1096. https://doi.org/10.1021/acs.jctc.5b01046

[47] Edraki, N., Das, U. and Hemateenejad, B. (2016) Comparative QSAR Analysis of 3,5-bis (Arylidene)-4-Piperidone Derivatives: The Development of Predictive Cytotoxicity Models. *Iranian Journal of Pharmaceutical Research*, **15**, 425-437.

[48] Kausar, S. and Falcao, A.O. (2018) An Automated Framework for QSAR Model Building. *Journal of Cheminformatics*, **10**, Article No. 1. https://doi.org/10.1186/s13321-017-0256-5

[49] Murtagh, F. (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, **31**, 274-295. https://doi.org/10.1007/s00357-014-9161-z