

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED KHIDER BISKRA

Faculté des sciences exactes et
des sciences de la nature et de la vie

Département des sciences de la matière

THESE

Présentée par

Samia BOUDERGUA

En vue de l'obtention du diplôme de :

DOCTORAT EN SCIENCES

Option :

Chimie informatique et pharmaceutique

Intitulée:

*Virtual screening and QSAR modeling for antioxidant
activity of benzofurans and flavonoids*

Soutenue le : 29/11/2020

Devant la commission d'examen :

Mme. Dalal HARKATI	MC/A	Université de Biskra	Présidente
Mr. Salah BELAIDI	Professeur	Université de Biskra	Directeur de thèse
Mr. Lotfi BELKHIRI	Professeur	Université de Constantine 1	Examineur
Mr. Lazhar BOUHLALEG	MC/A	Université de Batna 2	Examineur

To all those I love...

To all those I respect...

Acknowledgements

First of All, I have to thank ALLAH, the almighty for everything I have and I could do...

I would like to thank my Supervisor Pr. Salah Belaidi for accepting to supervise my doctoral research and for his precious advices...

I express my deep gratitude to my dissertation committee: Miss. Dalal Harkati, MC/A at the University of Biskra, as a president of the jury, Mr. Lotfi Belkhiri, Professors at the University of Constantine, and Mr. Lazhar Bouchlaleg, MC/A at the University of Batna 1, for accepting to examine my work.

I would like to acknowledge my colleagues at Khemis Miliana University for their technical support, particularly Miss. Aouda Bounif and Miss. Sanaa Zaoui.

I am grateful to Pr. Samir Chtita, Imane Benbrahim, and Toufik Salah for their scientific aid.

I would also thank Pr. Majdi Hochlaf and Miss. Mebarka Alloui for their valuable contribution in this work.

All my family members ...

All my teachers especially my teacher at the primary school

Mr. Abdelkader Mahmoudi...

THANK YOU...

Table of Contents

Acknowledgements	
List of tables	
List of figures	
List of Abbreviations	
Nomenclatures	
General introduction	-1-
References	-3-
Chapter 1. Free Radicals and antioxidants	
1.1 Introduction	-4-
1.2 Free radicals and oxidative stress	-4-
1.2.1 Origin of free radicals	-5-
1.2.2 Oxidative stress effect on health	-7-
1.3 Antioxidants	-9-
1.3.1 Definition	-9-
1.3.2 Origin	-10-
1.3.3 Types	-10-
1.4 Structure- Antioxidant activity relationship	-14-
1.5 Antioxidant activity measurement using DPPH method	-15-
1.6 Uses of antioxidants	-16-
1.7 Conclusion	-17-
References	-18-
Chapter 2. Computational and simulation techniques	
2.1 Introduction	-23-

Table of Contents

2.2 QSAR techniques	-23-
2.2.1 Descriptors generation	-24-
2.2.2 Modeling methods	-27-
2.2.3 Multivariate techniques	-33-
2.3 Drug-Likeness Screening	-33-
2.4 Conclusion	-35-
References	-36-
 Chapter 3. QSAR and drug-likeness study for antioxidant activity of benzofurans	
3.1 Introduction	-39-
3.2 Computational details	-40-
3.2.1 Furan geometry	-40-
3.2.2 Calculation of molecular descriptors	-40-
3.2.3 Data set	-41-
3.3 Results and discussion	-42-
3.3.1 Equilibrium geometry of furan	-42-
3.3.2 Vibrational analysis	-44-
3.3.3 Data set for analysis	-46-
3.3.4 QSAR study	-47-
3.3.5 Drug-likeness screening	-48-
3.4 Conclusion	-50-
References	-51-
 Chapter 4. QSAR study for flavonoids set	
4.1 Introduction	-53-
4.2 Descriptors calculation	-54-

Table of Contents

4.3 Results and discussion	-57-
4.3.1 Descriptors generation	-57-
4.3.2 Principal component analysis and hierarchical clustering	-60-
4.3.3 QSAR modeling using Gaussian Process	-63-
4.4 Structure-activity relationship of the studied flavonoids	-64-
4.5 Conclusion	-65-
References	-66-
General conclusion	-68-
References	-70-
Appendix	
Abstract	

List of Tables

Table 1.1. ROS Characteristics	-6-
Table 1.2. Principal sources of antioxidants	-10-
Table 1.3. Endogenous antioxidants	-11-
Table 1.4. Exogenous antioxidants	-12-
Table 3.1. Chemical structures and experimental activity of the molecules under study	-41-
Table 3.2. Bond lengths and valence angles of furan	-42-
Table 3.3. Anharmonic vibrational frequencies (in cm^{-1}) of furan and of benzofuran as computed using B3LYP/6-31 G (d,p)	-44-
Table 3.4. Values of molecular descriptors used in the QSAR study	-46-
Table 3.5. Drug-likeness parameters and lipophilicity indices of furan derivatives	-49-
Table 4.1. Chemical structures and experimental activity of the data set	-55-
Table 4.2. Bond lengths (in Å) and valence angles (in degree, °) of quercetin	-58-
Table 4.3. Values of molecular descriptors used in the statistical analysis Principal components	-59-
Table 4.4. Principal components	-60-

List of Figures

Figure 1.1. Mitochondrial respiratory chain	-6-
Figure 1.2. Structural features for high antioxidant capacity	-14-
Figure 1.3. DPPH assay	-15-
Figure 1.4. Relation between free radicals and antioxidants	-17-
Figure 2.1. Drug discovery process	-23-
Figure 2.2. QSAR pipeline	-24-
Figure 2.3. Schematic of biological neuron	-29-
Figure 2.4. Single artificial neuron	-29-
Figure 2.5. Preliminary QSAR and Drug-Likeness study	-35-
Figure 3.1. 3D conformations of furan and benzofurans.	-39-
Figure 3.2. 3D MESP for furan (left) and for benzofuran (right)	-43-
Figure 3.3. Structure of ANN	-47-
Figure 3.4. Correlation of experimental and predicted IC ₅₀ as calculated by ANN	-47-
Figure 4.1. Principal flavonoids sub-classes	-54-
Figure 4.2. Quercetin structure	-57-
Figure 4.3. Correlation circle	-61-
Figure 4.4. Hierarchical clustering	-62-
Figure 4.5. Score plot	-62-
Figure 4.6. Comparative bar chart between observed (obs) and predicted (pred) pIC ₅₀ of the test set	-63-
Figure 4.7. Effect of HD on the antioxidant capacity	-64-

List of Abbreviations

ADMET: Absorption, Distribution, Metabolism, Excretion, and Toxicity

AM1: Austin Model 1

ANN: Artificial Neural Network

ATP: Adenosine Triphosphate

A β : β -Amyloid

B3LYP: Becke, three-parameter, Lee-Yang-Parr

CAT: Catalase

CCSD: Coupled Cluster Single-Double

DFT: Density Functional Theory

DNA: Deoxyribonucleic Acid

DPPH: 2, 2-Diphenyl-1-Picrylhydrazyl

EDTA: Ethylene Diaminetetraacetic

GP: Gaussian Process

GPX: Glutathione Peroxidase

HF: Hartree-Fock

HOMO: Highest Occupied Molecular Orbital

IR: Infrared

LUMO: Lowest Unoccupied Molecular Orbital

MESP: Molecular Electrostatic Potential

MM: Molecular Mechanics

MNDO: Modified Neglect of Diatomic Overlap

List of Abbreviations

NADPH: Nicotinamide Adenine Dinucleotide Phosphate

NDDO: Neglect of Diatomic Differential Overlap

NOX: NADPH Oxidase

PCA: Principal Component Analysis

PG: Propyl Gallate

PM3: Parameterization Method 3

QSAR/QSPR: Quantitative Structure Activity/Property Relationship

RNS: Reactive Nitrogen Species

ROS: Reactive Oxygen Species

SAR: Structure/ Activity Relationship

SOD: Superoxide Dismutase

TBHQ: Tertiary Butylhydroquinone.

UV: Ultraviolet

VPT2: Second Order Perturbation Theory.

Nomenclatures

ARE: Average Relative Error

B: Bias

DM: Dipole Moment (debye)

EHOMO/ ELUMO: Energies of the HOMO and LUMO (eV)

f(x): Objective function

f'(x): f(x) Derivative

f''(x): f(x) Second Derivative

HA: Number of Hydrogen bond Acceptors

HD: Number of Hydrogen bond Donors

HE: Hydration Energy (kcal/mol)

HF: Heat of Formation (kcal/mol)

\hat{H} : Hamiltonian operator

I: Inputs

IC50: Half maximal Inhibitory Concentration (μM)

J(x): Jacobian matrix

k(x, x'): Covariance matrix

l: Hyper-parameters called the length scale

LE: Ligand Efficiency

LELP: Ligand-Efficiency dependent Lipophilicity

LipE: Ligand Lipophilic Efficiency

log P: Octanol-Water partition coefficient

MW: Molar Weight (amu)

Nomenclatures

N_H : Number of heavy atoms

O: Output

P: Polarizability (\AA^3)

$r(\mathbf{x})$: Residual function

RB: Number of Rotatable Bonds

R^2 : Squared correlation coefficient for the training set

R_{pred}^2 : Squared correlation coefficient for the test set

S: Surface area (\AA^2)

SSE: Sum Squares of Error

TPSA: Topological Polar Surface Area (\AA^2)

T : Transposition.

V: Volume (\AA^3)

W: Weights

x: Input of a function f

y: Output of a function f

y_* : Value to estimate at x_*

\bar{y}_* : Mean of y_*

Z: Network input to neuron

$\alpha(\mathbf{Z})$: Activation function

ϵ : Noise term of a function

θ : Hyper-parameter in the covariance matrix

σ^2 : Hyper-parameter called the signal variance

$\phi(\mathbf{x})$: Proposed function in ANN modeling

ψ : Wave function

$\delta(\mathbf{x}, \mathbf{x}')$: Kronecker delta function

Nomenclatures

$\nabla f(\mathbf{x})$: Gradient of $f(\mathbf{x})$

$\nabla^2 f(\mathbf{x})$: Hessian of $f(\mathbf{x})$.

General introduction

General introduction

Humanity has witnessed a huge leap in technologic progress. People have benefited of the improvement of life quality in all the fields, yet they become more exposed to new dangers as oxidative stress caused by free radicals. These latter are mainly originated from pollution, smoking, and stress...etc. They can lead to ageing and age-related illnesses [1] like Alzheimer and cardiovascular diseases. In the aim to overcome the detrimental effects of oxidative stress, it is recommended to use the antioxidants. They act with several mechanisms in scavenging the free radicals.

In the case of the pandemic COVID 19 appeared recently in all over the world [2], the treatment suggested to combat the virus comprises glutathione, vitamin C and zinc which are among the strongest antioxidants. They contribute in supporting the immune function. It is also shown that vitamin C may protect lungs from the infections caused by the oxidative stress. Moreover, antioxidants are actually added to drug formulations to enhance their activities. They are even used in the anticancer medical protocol in the aim to boost the anticancer drugs activity. Antioxidants are also primordial for drugs preservation for longer durations [3].

The class of antioxidants is vast; it is clustered according to several criteria such as, action mechanism and origin. The benzofurans and the flavonoids which belong to polyphenols [4] are known to have a high antioxidant activity. Besides, they are abundant in nature, so, they may be a part of our daily diet. For their crucial role, the benzofurans and the flavonoids are largely studied. In order to rationalize their studies, many statistical techniques are used. Quantitative Structure-Activity

Relationship (QSAR) [5] is a powerful tool helping in evaluating the correlation between chemical structures and activities. It may be established within different methods, for instance, Gaussian processes and artificial neural networks. In combination with other multivariate and simulation methods, the data sets may be overly characterized, especially with the integration of the docking [6] and the ADMET in researches [7].

In this context, our present work aims to explore two antioxidant data sets of benzofurans and flavonoids. Therefore, we start with a general introduction of this topic. Then, the core of our work is divided into four chapters; the first one concerns generalities about the antioxidant activity. It presents their types, sources and action mechanisms. Next, we highlight the different computational methods used in QSAR and drug-likeness studies in chapter 2. We give explanations and theories for each technique used in our work such as, DFT, Gaussian process and neural networks. The third chapter focuses on the QSAR modeling and drug-likeness screening of a set of fifteen benzofurans [8, 9], (This work is published in: Journal of Molecular Structure, 1189, pp 307- 314, 2019). The last chapter emphasizes on a QSAR study of some flavonoids [10, 11] by means of Gaussian process infrequently used in this approach, in addition to a clustering study of our data set.

Finally, we finish with a general conclusion and perspectives.

References

- [1] J. Luo, K. Mills, S. le Cessie, R. Noordam, D. van Heemst, *Ageing Res. Rev.* 57 (2020) 1-11.
- [2] R.I. Horowitz, P.R. Freeman, *Med. Hypotheses.* 143 (2020) 1.
- [3] G.T. Yiang, P.L. Chou, Y.T. Hung, J.N. Chen, W.J. Chang, Y.L. Yu, C.W. Wei, *Oncol. Rep.* 32 (2014) 1057-1063.
- [4] N.T. Dat, P.T.X. Binh, L.T.P. Quynh, C.V. Minh, H.T. Huong, J.J. Lee, *Fitoterapia.* 81 (2010) 1224-1227.
- [5] A. Tropsha, *QSAR Modeling and QSAR Based Virtual Screening, Complexity and Challenges of Modern*, in R. Meyers (eds): *Encyclopedia of Complexity and Systems Science.* Springer, New York, 2009.
- [6] L.D. Biasi, R. Fino, R. Parisi, L. Sessa, G. Cattaneo, A. De Santis, P.Iannelli, S. Piotto, *Novel Algorithm for Efficient Distribution of Molecular Docking Calculations*, in F. Rossi et al. (eds): *WIVACE 2015*, Springer, Switzerland, 2016.
- [7] C. Baskaran, M. Ramachandran, *Asian Pac. J. Trop. Dis.* 2 (2012) S734-S738.
- [8] S. Boudergua, M. Alloui, S. Belaidi, M. Mogren Al Mogren, U. A. Abd Ellatif Ibrahim, M. Hochlaf, *J. Mol. Struct.* 1189 (2019) 307-314.
- [9] J.F. Hsieh, W.J. Lin, K.F. Huang, J.H. Liao, M.J. Don, C.C. Shen, Y.J. Shiao, W. T. Li, *Eur. J. Med. Chem.* 93 (2015) 446.
- [10] M. Okawa, J. Kinjo, T. Nohara, M. Ono, *Biol. Pharm. Bull.* 24 (2001) 1202-1205.
- [11] J. Hu, W. Ma, N. Li, K.J. Wang, *J. Mex. Chem. Soc.* 61 (2017) 282-289.

Chapter 1. Free Radicals and Antioxidants

Chapter 1. Free radicals and Antioxidants

1.1 Introduction

Life is impossible without oxygen. Its presence is fundamental for the metabolism of living cells. It is used to metabolizing fats, proteins, and carbohydrates to get energy, so it is primordial for our life, our development as well as our ability of adaptation. However, oxygen is a highly reactive atom; it leads to the production of free radical molecules particularly reactive oxygen species (ROS) which may be at the origin of toxicity, acidity, deterioration, and degeneration. These latter may be detrimental for body cells.

As a remediation, antioxidants can interfere so as to counterbalance the oxidant effects and to remantain the redox equilibrium [1-3]. They can play the role of prophylactic agents [4].

This first chapter will focus on the oxidative stress and its reparation by the several types of antioxidants.

1.2 Free radicals and oxidative stress

Free radicals have unpaired electrons that make them highly reactive and detrimental for lipids, proteins, and DNA. Free radicals impact on cellular membranes by lipid peroxidation and on proteins and nucleic acids by oxidative damage. The high production of free radicals accompanied with their low scavenging activity leads to an oxidative stress [5].

Autooxidation can deteriorate the lipid based food; it causes a loss in the nutritional values, and the production of some undesired derivatives [6]. Furthermore,

free-radical-initiated autooxidation of cellular membrane lipids is accepted to be inductive of many diseases as cancer...etc [7-9].

Oxidative stress was defined by Halliwell and Gutteridge (1985) [10] as the defense incapacity of organs and cells against the aggressive free radicals derived from oxygen.

Furthermore, oxidative stress is known to be an imbalanced equilibrium between the rates of production and elimination of free radicals (oxidants) [11].

1.2.1 Origin of free radicals

ROS are devising into two classes according to their structure. They include nonradical species as hydrogen peroxide (H_2O_2), hypochlorous acid (HOCl), singlet oxygen ($^1\text{O}_2$), and free radicals such as superoxide anion radical (O_2^-), and hydroxyl radical (HO^\cdot)... [12-14].

The free radicals half-life is about 10^{-9} to 10^{-6} s [15].

We can also distinguish the endogenous ROS, which are produced as natural by cellular metabolism (oxidative metabolism) such as the mitochondrial electron transport chain and NADPH oxidase activity. Besides, the extracellular ROS (environmental stressors) generated by UV light and other ionizing radiation, bugs, xenobiotic, and pollutants [16, 17].

❖ Mitochondrial respiratory chain: After the free radicals had been discovered in the biologic system; Harman et al. (1956) evoked for the first time, in 1956, the hypothesis of the damage caused by these radicals. Moreover, it was found that they are leading to the phenomenon of aging.

For the majority of living beings, oxygen (O_2) is crucial for the production of energy by the oxidative phosphorylation reaction generating ATP. This reaction is occurred by the electron transport chain in the internal mitochondrial membrane.

During a normal metabolic chain, the tetravalent reduction of oxygen to water (Figure 1.1) creates primary ROS through many stages. Hydroxyl radical (OH[·]) is considered as the most deleterious one. The degradation of these radicals is controlled by the antioxidants [18].

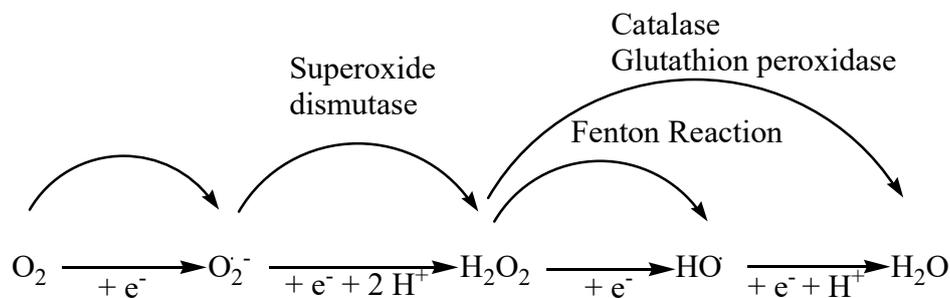


Figure 1.1. Mitochondrial respiratory chain [18].

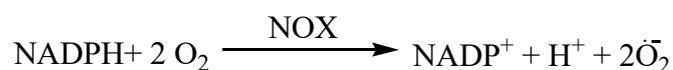
Globally, the oxygen is submitted to quadruple reduction and protonation to form water in the cells. The characteristics of each radical are illustrated on the table 1.1.

Table 1.1. ROS Characteristics [15].

Symbol	Name	Characteristics
O ₂	oxygen	-Stable, -weak oxidant ability.
O ₂ ^{·-}	superoxide anion radical	-Weak reactive radical, -toxic,
¹ O ₂	singlet oxygen	-Very reactive, -can initiate the lipoperoxidation.
H ₂ O ₂	hydrogen peroxide	-Stable and diffusible, -feebly toxic, - antiseptic.
HO [·]	hydroxyl radical	-Very reactive, -principal initiator of the lipoperoxidation, -alters proteins and DNA.

Nevertheless, there is another category of radicals called secondary radicals including: peroxy radical (ROO \cdot), hydroperoxide (ROOH), and the alkoxy radical (RO \cdot).

❖ NADPH oxidase: The majority of cells are able to produce the superoxide radical (O $_2^{\cdot-}$) throughout the membrane NADPH oxidase (NOX) activity. The NADPH oxidase contributes in the defensive system and in the cellular signaling in the phagocytic and non phagocytic cells respectively [18].



We take as an e

xample, the oral cavity which is very susceptible to the oxidative damage. Żukowski et al., 2018 [19] reported that the factors generating ROS in this place may be classified into periodontal inflammation. xenobiotics (ethanol, cigarette smoke, drugs), unhealthy food with high-fat and protein levels, repeated dental treatment using different rays and materials (ozone, non-thermal plasma, laser light, ultraviolet light, fluorides, dental composites, fixed orthodontic appliances, and titanium fixations).

1.2.2 Oxidative stress effect on health

At the physiological concentrations, free radicals are primordial for several biological functions such as playing the role of cell signaling molecules and controlling cell viability, migration, and differentiation. Moreover, free radicals contribute in the defensive system through combating the pathological agents and inactivating them [20- 22]. However, an oxidative-antioxidant imbalance and an oxidative stress may be appeared when the levels of ROS exceed the physiological concentrations [23].

Oxidation process caused by reactive oxygen species such as $O_2^{\bullet-}$, OH^{\bullet} , and lipid peroxyl radicals ROO^{\bullet} can provoke an oxidative stress leading to a considerable damage of cellular proteins, nucleic acids, and lipids. This damage is implicated in the apparition of many diseases [17, 24].

Consequently, free radicals damage can cause proteins denaturation, DNA mutation, and binding to unsaturated lipid membrane leading to lose of fluidity [17].

Our body is able to elaborate its system of protection against the oxidative stress, yet this system may be overwhelmed by the excess generation of free radicals [8].

The oxidative stress is generally considered as one of the major causes of the age-associated diseases such as cardiovascular diseases, cancers, Parkinson's and Alzheimer's diseases...etc [2, 25]:

➤ Cardiovascular diseases: Cardiovascular dysfunction has several causes including hypercholesterolaemia, hypertension, smoking, diabetes, poor diet, stress and physical inactivity. One of the main of them is the atherosclerosis. This latter is initiated by oxidation of human low-density lipoproteins which is mediated by ROS and reactive nitrogen species (RNS) [9].

➤ Cancer: In the case of cancer, ROS proceed as secondary messengers in the intracellular signaling cascades. This act contributes to induce and sustain the oncogenic phenotype of cancer cells. On the contrary, ROS can stimulate the cellular apoptosis and play the role of antitumorigenic species.

A balance between ROS production and scavenging usually leads to a cellular redox homeostasis [26, 27].

➤ Alzheimer's disease: The oxidative stress alters strongly the brain more than any other organs. That is seen when mitochondrial dysfunction, increased metal levels, inflammation, and β -amyloid ($A\beta$) peptides oxidize the neurons components

(lipids, proteins, and nucleic acids). ROS lead to the promoting of A β deposition, the hyperphosphorylation, and the subsequent loss of synapses and neurons in the development of this disease [28].

➤ Skin aging: UV radiation is made up of UVA and UVB. UVA rays are responsible of oxidative process, for instance, lipid peroxidation, modification of structural proteins and DNA damage. These effects may be blocked or repaired by counterbalance the oxidative stress [29]...

1.3 Antioxidants

1.3.1 Definition

An antioxidant is a reducing agent, yet a reducing agent is not necessarily an antioxidant. Hence the antioxidant is able to neutralize an oxidizing agent [30-31].

Even if it is present with low concentrations, an antioxidant is defined as a substance which can considerably decelerate or prevent the process of oxidation by breaking the chain of radical reactions [10].

The principal roles of antioxidants can be cited as following:

- ✓ Quenching the free radicals to end their chain reaction by donating a hydrogen atom or transferring them an electron.
- ✓ Oxidation reactions are generally catalyzed by metals as iron and copper, hence chelating these metals may stop the free radicals generation.
- ✓ Inactivating the free radical generating enzymes such as xanthine oxidase, lipoxygenase, protein kinase C, mitochondrial succinoxidase, and NADPH oxidase...
- ✓ The body has its proper defensive system against oxidation reactions; nevertheless it sometimes needs to stimulate the internal antioxidant enzymes by rich antioxidant diet [17, 32].

1.3.2 Origin

Healthy food can supply the body with the necessary amounts of antioxidants. It provides them even in cheap dishes.

Some of the origins of antioxidants are illustrated on Table 1.2.

Table 1.2. Principal sources of antioxidants [3, 33- 40].

Antioxidant	Origin
Vitamin E	Germ of cereal seeds, polyunsaturated vegetables oils, eggs, walnut...
Vitamin C	Lemon, orange, cantaloup, honeydew melon, cherries, kiwi, tomatoes, cabbage, brussels, cauliflower, peas, red and green peppers, potatoes...
Se	Fish, meat, poultry, eggs, vegetables planted in selenium rich soil...
Fe	Beef, chicken, liver, fish, tofu, spinach and other leafy vegetables...
Cu	Nuts, seeds, seafood, meat and grains...
Mn	Vegetables, rice, nuts, whole grains, sea foods, seeds, chocolates, as well as teas...
Zn	Milk, green vegetables, complete bread, meat, fish...
β- carotene	Orange and dark green fruits and vegetables.
Polyphenols	Chocolate, green tea, vinegar, grapes, berries, red onion, red cabbage, honey...

1.3.3 Types

There is a variety of criteria how to classify the antioxidants. The most frequent classification is based on their origin and their action mechanism.

1.3.3.1 Classification of the antioxidants according to their origin

a) Endogenous antioxidants

The body is the first barrier against the oxidation processes; it relies on two endogenous defensive systems:

- **Enzymatic antioxidants:** The main antioxidants enzymes are: superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPX)...
- **Non enzymatic antioxidants:** They are divided into proteins and non proteins antioxidants (Table 1.3).

Table 1.3. Endogenous antioxidants [34, 41- 43].

Class	Subclass	Antioxidant	Activity	
Endogenous antioxidants	Enzymatic antioxidants	Superoxide dismutase	$2O_2^- + 2H^+ \xrightarrow{SOD} H_2O_2 + O_2$	
		Catalase	$2H_2O_2 \xrightarrow{CAT} 2H_2O + O_2$	
		Glutathione peroxidase	$2GSH + H_2O_2 \xrightarrow{GPX} GSSG + 2H_2O$ GSH: Reduced glutathione. GSSG: Oxidized glutathione.	
	Non enzymatic antioxidants	Proteins	Glutathione	It protects the body against the xenobiotic metabolism. Also, it contributes in the regeneration of the vitamin E.
		metal-binding proteins	-Albumin, -cerulplasmin, -hepatoglobin, -ferritin...	-These proteins are linked to Copper or iron; they control the production of metal catalyzed free radicals.
		Non proteins	-Bilirubin	It is considered as the most effective antioxidant in the lipid peroxidation.
			-Uric acid	It is a good scavenger of singlet oxygen
-Coenzyme Q	-It can regenerate the vitamin E, -it breaks the chain of free radicals reactions.			

b) Exogenous antioxidants

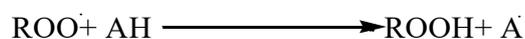
Table 1.4. Exogenous antioxidants [2, 42, 44- 47].

Class	S.Class	Antioxidant	Activity	
Exogenous antioxidants	Natural antioxidants	Vitamins: E, C, A...	-Vitamin E is the most efficient antioxidant of the membrane lipid peroxidation, -vitamin C contributes to regenerate the vitamin E.	
		Micro-nutrients cofactors: Se, Fe, Cu, Mn, Zn...	-Improve the enzymes catalytic activity.	
		Carotenoids: Beta carotene...	-Protect the lipid- rich tissues against oxidation, -have a synergistic protection in addition with vitamins.	
		Polyphenols: flavonoids, polyphenolic acids...	-Scavenge the free radicals, -sequester the antioxidant agents, -modify the cell signaling...	
	Synthetic antioxidants	Drugs	Probucol	Avoids the low density lipoprotein oxidation.
			N-acetylsysteine	Regenerates the glutathione.
		Additives	-Butylated hydroxytoluene - Butylated hydroxyanisole	-Prevent lipid oxidation, -have a synergistic activity.
			Propyl gallate (PG)	Protects oils and fats against oxidation.
			Ethylene diaminetetraacetic (EDTA)	Prevents rancidity and food color changes caused by oxygen and mineral traces (Fe, Cu, Ni...).
			Tertiary butylhydroquinone (TBHQ)	Stabilizes or inhibits the auto-polymerization of organic peroxides.

Besides the body's own antioxidants; the dietary can enhance the fight against the oxidative stress using natural or synthetic antioxidants. These molecules contribute in fighting oxidation processes [25] as well as they serve to preserve food from oxidation and make longer its use duration. Some of the exogenous antioxidants are illustrated on the Table 1.4.

1.3.3.2 Classification of the antioxidants according to their Action mechanism

a) **Type I:** The antioxidants of this category are capable of breaking the chain of free radicals reactions through donating a hydrogen atom or an electron.



ROO[·] : Free radical,

AH: Antioxidant,

A[·] : Antioxidant radical [48, 49].

Consequently, the antioxidants inhibit the propagation of the oxidation chain and scavenge the free radicals.

Vitamins C and A, albumin, bilirubin, and flavonoids act as chain breaking antioxidants [50].

b) **Type II:** The antioxidants that belong to type II have a preventive action [51-53]. They may act as following:

- Chelators of the metals catalyzing free radicals generation (EDTA, citric acid...).

- Quench the ROS (enzyme SOD...).

- Decompose the free radical into non radical compounds (enzyme CAT...)

[50].

c) **Type III:** Some physical parameters can avoid the oxidation in the case of food preservation, for instance vacuum and obscurity [54].

1.4 Structure- Antioxidant activity relationship

Even there are some contradictions concerning the structure- activity relationship (SAR) of the antioxidants, it is believed that the strongest factors are the number of hydroxyl groups and their positions, as well as the conjugation between fractions [55]. This conjugation creates a resonance effect leading to the stability of the antioxidant radical.

For the flavonoids, one of the most influencing parameters is the hydroxylation on B- ring (Figure 1.2) due to the delocalization of the unpaired electron in the aroxyl radical structure. That may contribute in the stabilization of the flavonoids. In addition, a high radical scavenging capacity is proportionally correlated to the high electron-donating ability, and so to low electron-oxidation peak potentials [56].

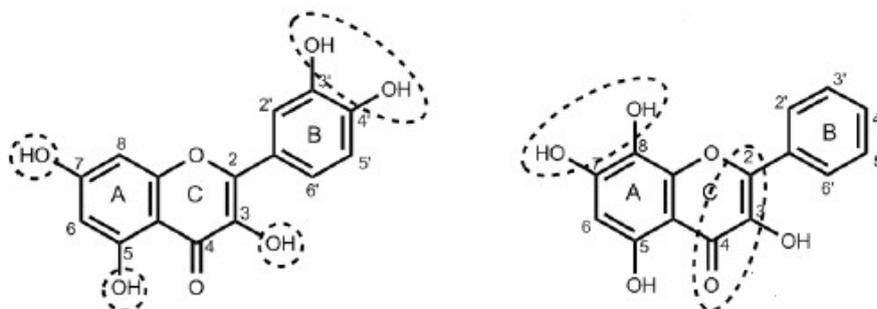


Figure 1.2. Structural features for high antioxidant capacity [17, 55, 57].

When a flavonoid donates hydrogen from a hydroxyl group, it becomes a relatively stable flavonoid phenoxyl radical. Next, this latter may react with a second radical ($RO\cdot$) to give stable quinone [17].

Nevertheless, the presence of glycosides decreases the antioxidant capacity, probably because of the steric hindrance effect [58].

Similarly to the flavonoids, the effect of the number and the position of hydroxyl groups, as well as the extension of conjugation is obvious on the antioxidant capacity of benzofurans, such as the case of arylidene [59]. It is also proved that the mechanism of donating hydrogen is crucial in explaining antioxidant capacity. Moreover, tautomerism contributes in effecting the antioxidant efficiency. It has a role in releasing hydrogen atoms [60]. Whereas, the presence of electron attractor, such as halogens, decreases the ability of free radical scavenging [59].

1.5 Antioxidant activity measurement using DPPH method

DPPH (2, 2-diphenyl-1-picrylhydrazyl) scavenging method is a spectrophotometric assay. It is based on the decolorization of the purple-colored stable radical (DPPH[•]). After has been reduced by an antioxidant, the DPPH becomes non-radical (Figure 1.3). The maximum absorbance is at 517 nm.

Experimentally, the strength of the antioxidant is estimated by its concentration able to reduce 50% of the initial quantity of DPPH.

This test is advantageous since it is simple, quick, and useful for hydrophilic and lipophilic antioxidant compounds [61].

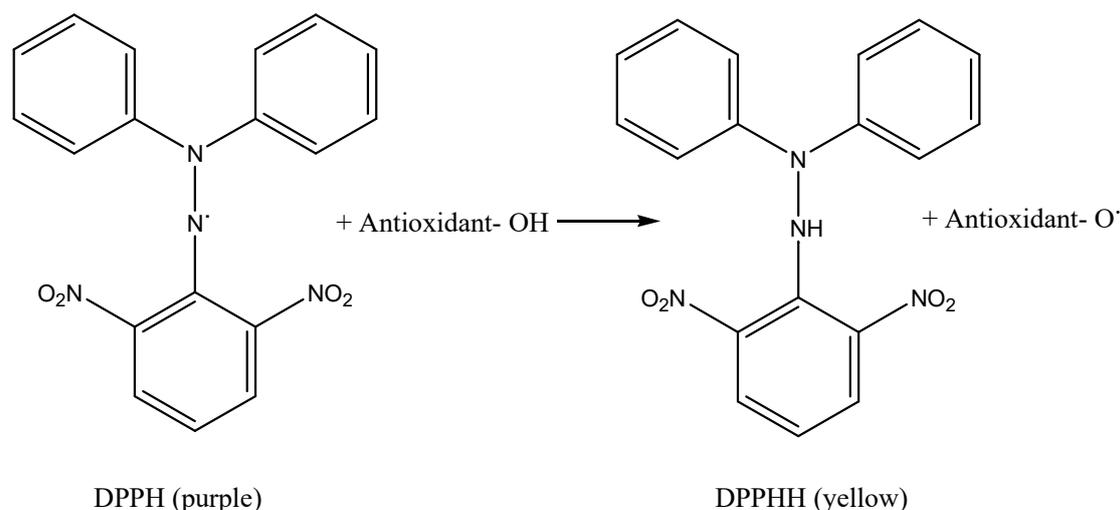


Figure 1.3. DPPH assay [62].

1.6 Uses of antioxidants

The antioxidants are largely used as additives in different fields, principally in food industries, as well as pharmaceutical and cosmetic products.

➤ **Food industries:** Lipid oxidation leading to rancidity and other toxic reactions is one the common reasons of food deterioration. Hence, the antioxidants added with convenient amounts may reduce this effect. They may be natural from herbs and spices, for instance rosemary, thyme, pepper, and clove...Besides, synthetic antioxidants may be used to improve the shelf life of food, such as butylated hydroxytoluene and butylated hydroxyanisole [52].

➤ **Pharmaceutical products:** The aging process accelerated by the oxidative stress may cause many diseases as cancer, diabetes and nervous pathologies. Antioxidants may delay this process by the removal of the free radicals accumulated in vivo and in vitro [63, 64].

➤ **Cosmetic products:** Antioxidants in cosmetics may reinforce the antioxidation capacity of the skin which becomes weak with aging. The most common antioxidants currently used are: ginkgo biloba extracts, resveratrol, α -arbutin, vitamin E, and polysaccharides [63]. They are delivered in different galenic forms: liposoms, nanoparticles, emulsions...[65].

Thus, adding antioxidants to these formulations has a double role. They are benefic to health by combating the oxidative stress. Also, they are a key preservative in the final products where they contribute to prevent lipid oxidation and lengthen the shelf life of food and drugs.

1.7 Conclusion

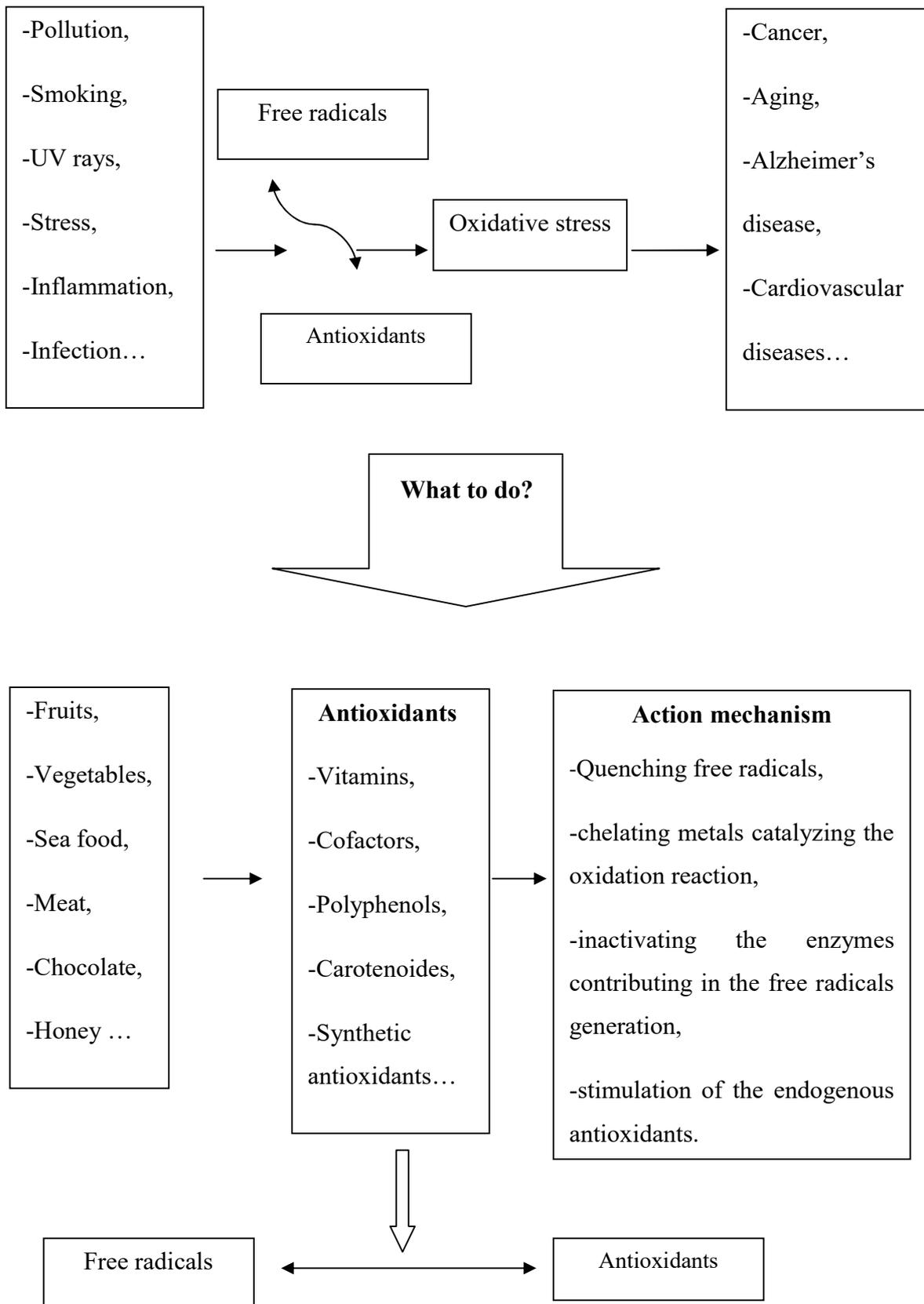


Figure 1.4. Relation between free radicals and antioxidants.

References

- [1] S. Kumar, A.K. Pandey, *Sci. World J.* ID 162750 (2013) 1-16.
- [2] M. Percival, *Clin. Nut. Insights.* 1 (1998) 1.
- [3] C.K. Ramonatxo, *Nutr. Clin. et Metab.* 20 (2006) 165-177.
- [4] L.V.B. Hoelz, B.A.C. Horta, J.Q. Araújo, M.G. Albuquerque, R.B. de Alencastro, J. F.M. da Silva, *J. Chem. Pharm. Res.* 2 (2010) 291-306.
- [5] S.B. Nimse, D. Pal, *Rsc Advances.* 5 (2015) 27986–28006.
- [6] X.C. Weng, Y. Huang, *Grasas Aceites.* 65 (2014) 2.
- [7] A. Saija, M. Scalese, M. Lanza, D. Marzullo, F. Bonina, F. Castelli, *Free Radic. Biol. Med.* 19 (1995) 481-486.
- [8] B. Poljsak, U. Glavan, R. Dahmane, *Int J Canc Prev.* 4 (2011) 193-210.
- [9] T. Bahorun, M.A. Soobrattee, V. Luximon-Ramma, O.I. Aruoma, *Internet J. Medical Update.* 1 (2006) 25-35.
- [10] B. Halliwell, J.M.C. Gutteridge, *Free radicals in biology and medicine.* Clarendon Press, Oxford, 1989.
- [11] V. Rani, G. Deep, R. K. Singh, K. Palle, U. C.S. Yadav, *Life Sciences,* 148 (2016) 183-193.
- [12] I. Gülçin, V. Mshvildadze, A. Gepdiremen, R. Elias, *Phytomedicine,* 13 (2006) 343-351.
- [13] M.M. Ibrahim, G.A.M. Mersal, A.M.M. Ramadan, S.Y. Shaban, M.A. Mohamed, S. Al-Juaid, *J. Mol. Struct.* 1137 (2017) 742-755.
- [14] C.A. Prauchner, *Burns.* 43 (2017) 471-485.
- [15] F. Tessier, P. Marconnet, *J. Sports Sci.* 10 (1995) 2-5.

- [16] A.F. Naeimi, M. Alizadeh, Trends Food Sci Technol. 70 (2017) 34-44.
- [17] S.D.S. Banjarnahor, N. Artanti, Med J Indones. 23 (2014) 239-240.
- [18] C. Migdal, M. Serres, Med Sci. 27 (2011) 405-411.
- [19] P. Żukowski, M. Maciejczyk, D. Waszkiel, Arch. Oral Biol. 92 (2018) 8-17.
- [20] J. Li, O. Wuliji, W. Li, Z.G. Jiang, H.A. Ghanbari, Int. J. Mol. Sci. 14 (2013) 24438-24475.
- [21] J. M. Lu, P.H. Lin, Q. Yao, C. Chen, J Cell Mol Med. 14 (2010) 840-860.
- [22] S. Catarzi, C. Romagnoli, G. Marcucci, F. Favilli, T. Iantomasi, M.T. Vincenzini, Biochim. Biophys. Acta. 1810 (2011) 446-456.
- [23] A. Rahal, A. Kumar, V. Singh, Biomed Res. Int. ID 761264 (2014) 1-19.
- [24] K.E. Heim, A.R. Tagliaferro, D.J. Bobilya, J. Nutr. Biochem. 13 (2002) 572-584.
- [25] S. Das, I. Mitra, S. Batuta, M. Niharul, A. Kunal Roy, N. Ara Begum, Bioorg. Med. Chem. Lett. 24 (2014) 5050.
- [26] H.G. Wittgen, L.C. van Kempen, Melanoma Res. 17 (2007) 400-409.
- [27] P.D. Ray, B. Huang, Y. Tsuji, Cell Signal. 24 (2012) 981-990.
- [28] Z. Chen, C. Zhong, Neurosci Bull. 30 (2014) 271-281.
- [29] M.V. De Gálvez, Actas Dermosifiliogr. 101 (2010) 197.
- [30] R.L. Prior, G. Cao, Free Radic Biol Med. 27 (1999) 1173-1181.
- [31] C. El Kalamouni, Caractérisations chimiques et biologiques d'extraits de plantes aromatiques oubliées de Midi-Pyrénées, PHD Thesis, Sciences of Agrosources, National Polytechnic Institute of Toulouse, France, 2010.
- [32] D. Procházková, I. Bousová, N. Wilhelmová, Fitoterapia. 82(2011) 513-523.

- [33] Y. Hu, Z.J. Pan, W. Liao, J. Li, P. Gruget, D.D. Kitts, X. Lu, *Food Chem.* 202 (2016) 254-261.
- [34] E.M. Atta, N.H. Mohamed, A.A.M. Abdelgawad, *Eur. Chem. Bull.* 6 (2017) 368-370.
- [35] S. Pouil, F. Oberhansli, P. Bustamante, M. Metian, *Chemosphere.* 183 (2017) 503-509.
- [36] K. Shubham, T. Anukiruthika, S. Dutta, A.V. Kashyap, J.A. Moses, C. Anandharamakrishnan, *Trends Food Sci Technol.* 99 (2020) 58-75.
- [37] J. Wang, W. Bi, A. Cheng, D. Freire, P. Vempati, W. Zhao, B. Gong, E.M. Janle, T.Y. Chen, M.G. Ferruzzi, J. Schmeidler, L. Ho, G.M. Pasinetti, *Front Aging Neurosci.* 6 (2014) 1-10.
- [38] S.A. Souyoul, K.P. Saussy, M.P. Lupo, *Dermatol Ther.* 8 (2018) 5-16.
- [39] T.V. Peres, M.R.C. Schettinger, P. Chen, F. Carvalho, D.S. Avila, A.B. Bowman, M. Aschner, *BMC Pharmacol. Toxicol.* 17 (2016) 57.
- [40] M. Zamanian, F. Azizi-Soleiman, *PharmaNutrition.* 11 (2020) 1-4.
- [41] P. Chakraborty, S. Kumar, D. Dutta, V. Gupta, *Res J Pharm Technol.* 2 (2009) 238-244.
- [42] P. Mamta, K. Misra, G.S. Dhillon, S.K. Brar, M. Verma, *Antioxidants*, in: *Biotransformation of Waste Biomass into High Value Biochemicals.* Quebec, 2014.
- [43] G.G. Duthie, K.M. Brown, *Reducing the Risk of Cardiovascular Disease*, in: *Functional Foods*, ed. Goldberg, I. Chapman and Hall. New York, 1994.

- [44] J.L. Baldim, B.G.V. de Alcântara, O.D.S. Domingos, M.G. Soares, I.S. Caldas, R.D. Novaes, T.B. Oliveira, J.H.G. Lago, D.A. Chagas-Paul, *Oxid Med Cell Longev.* 2017 (2017) 2.
- [45] L. Ortega Ramírez, I. Rodriguez García, J. Leyva, M.R. Cruz-Valenzuela, B.A. Silva-Espinoza, G.A. Gonzalez-Aguilar, M.W. Siddiqui, J.F. Ayala-Zavala, *J Food Sci.* 79 (2014) R129-R137.
- [46] A. Yadav, R. Kumar, A. Yadav, J.P. Mishra, S. Srivatva, S. Prabha, *Res. Environ. Life Sci.* 9 (2016) 1328-1331.
- [47] C. Jacobsen, M.B. Let, N.S. Nielsen, A.S. Meyer, *Trends Food Sci Tech.* 19 (2008) 76-93.
- [48] C. Viglianisi, S. Menichetti, *Antioxidants.* 8 (2019) 1-17.
- [49] S.P. Mahantesh, A.K. Gangawane, C.S. Patil, *World Res J Med Aromat Plant.* 1 (2012) 6-10.
- [50] A.M. Papas, *Antioxidant status, diet, nutrition, and health.* CRC Series, Boca Raton, 1998.
- [51] D. De Beer, E. Joubert, W.C.A. Gelderblom, *South African J. Enol. Vitic.* 23 (2002) 48-61.
- [52] M. Kebede, S. Admassu, *Adv Food Technol Nutr Sci Open J.* 5 (2019) 38-40.
- [53] A. Roeding-Penman, M.H. Gordon, *J Am Oil Chem Soc.* 75 (1998) 169-180.
- [54] S. Eymard, *Mise en évidence et suivi de l'oxydation des lipides au cours de la conservation et de la transformation du chinchard (*Trachurus trachurus*) : choix des procédés.* PHD Thesis, Doctoral School-Biochemistry, University of Nantes, France, 2003.

- [55] D. Amić, D. Davidović-Amić, D. Beslo, V. Rastija, B. Lucić, N. Trinajstić, *Curr Med Chem.* 14(2007) 829.
- [56] I. N. Jovanović, A. Miličević, *Arh Hig Rada Toksikol.* 68 (2017) 93.
- [57] G.B. Bubols, R. Vianna Dda, A. Medina-Reimon, G. von Poser, R.M. Lamuela-Raventos, V.L. Eifler-Lima, *Mini Rev Med Chem.* 13(2013) 318-334.
- [58] J. Hu, W. Ma, N. Li, K.J. Wang, *J. Mex. Chem. Soc.* 61 (2017) 282-289.
- [59] A. Baldisserotto, M. Demurtas, I. Lampronti, D. Moi, G. Balboni, S. Vertuani, S. Manfredini, V. Onnis, *Eur. J. Med. Chem.* (2018), doi: 10.1016/j.ejmech.2018.07.001.
- [60] M. Abdel-motaal, E.M. Kandeel, M. Abou-Elzahab, F. Elghareeb. *Eur. Sci. J.* 13 (2017) 304.
- [61] D.J. Charles, *Antioxidant Properties of Spices, Herbs and Other Sources*, Springer Science+Business Media, New York, 2013.
- [62] H.Talbi, A. Boumaza, K. El-Mostafa, J. Talbi, A. Hilali, *Journal de JMESCEN.* 6 (2015) 1111-1117.
- [63] R. Song, Q. Wu, L. Zhao, Z. Yun, *E3S Web of Conferences ChinaBiofilms.* 131 (2019) 3.
- [64] A.A. Hamid, O.O. Aiyelaagbe, L.A. Usman, O.M. Ameen, A. Lawal, *Afr. J. Pure Appl. Chem.* 4 (2010) 142-151.
- [65] L. Montenegro, *J Pharm Pharmacogn Res.* 2 (2014) 73-87.

Chapter 2. Computational and simulation techniques

Chapter 2. Computational and simulation techniques

Chapter 2. Computational and simulation techniques

2.1 Introduction

The ultimate objective of drug companies is drug discovery and development (Figure 2.1). It is a phase which may take many years with a lot of challenges. With recent progresses in bioinformatics as well as computational chemistry and biology, drug discovery is becoming more rational. Nowadays, many researches focus on experimental and theoretical aspects of this discipline.

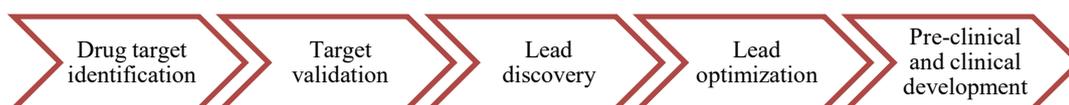


Figure 2.1. Drug discovery process [1].

This chapter will give a general overview on some techniques used in quantitative structure-activity relationships that represent a primordial key in lead discovery. Another tool will also be overly presented; it is drug-likeness screening where we expose the different rules used to determine the drug candidates.

2.2 QSAR techniques

Quantitative Structure Activity/Property Relationship (QSAR/QSPR) approaches are widely studied in chemometrics, pharmacodynamics, pharmacokinetics, and toxicology ...etc [2]. After have been proposed for the first time by Crum-Brown and Fraser in 1868 [3] to study the physiological activity of molecules in function of their composition, QSAR studies have dramatically progressed in term of capabilities of collecting and analyzing all types of data even big data due to the development of

computer sciences and mathematical modeling. For example, Hansch and Fujita [4] introduced the linear free energy relationship model in 1964, however, new approaches are used nowadays like Gaussian processes, neural networks, gene expression programming, project pursuit regression and local lazy regression [2].

The pipeline of QSAR studies starts with data collection and characterization till the design of new compounds candidate to be more active and save on the base of the Structure - Activity Relationship (SAR) and the predictive models (Figure 2.2).

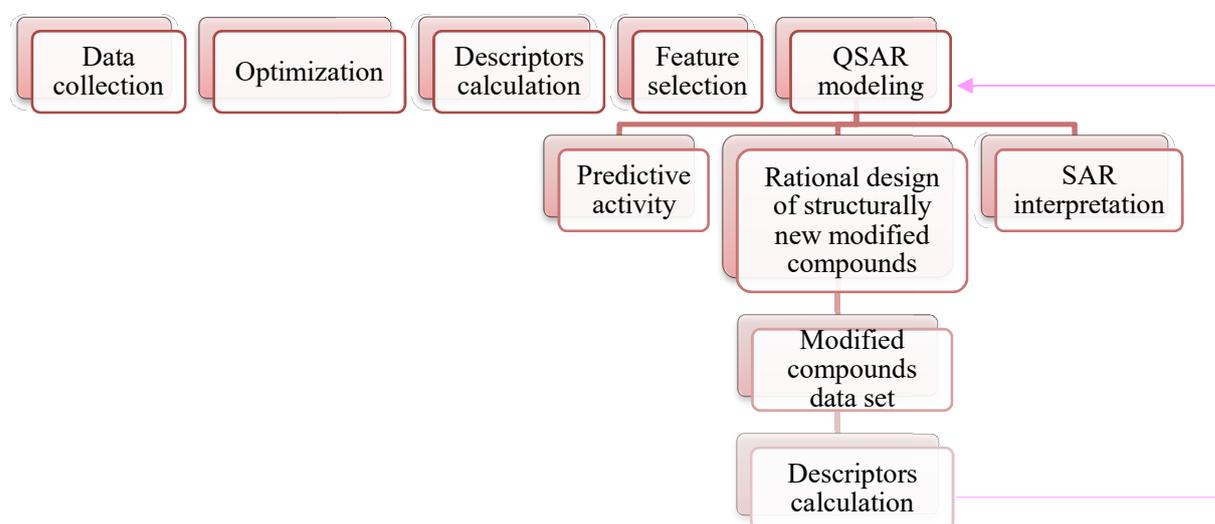


Figure 2.2. QSAR pipeline [5].

2.2.1 Descriptors generation

As presented on Figure 2.2, descriptors generation is a principal feature in QSAR modeling. The descriptors values in addition to the biological activities are the core of the mathematical models. They permit to quantify the relation between the structure and the activity. So, it is needed to have the most accurate values to establish reliable QSAR models. In following some computational methods largely used for descriptors

calculation. In general, these parameters are divided into three categories: topological, electronic, as well as energetic descriptors.

2.2.1.1 Molecular mechanics

Molecular mechanics is founded on the force field; where the variation of the potential energy is correlated with the molecular geometries. In this theory, the molecule seen as a group of atoms in the vacuum is submitted to elastic and harmonic forces. And, just the nucleus is taken in consideration [6].

The energy expression is a simple sum of energies as bond stretching, bond bending, torsions, electrostatic interactions, Van der Waals forces, and hydrogen bonding.

The advantage of molecular mechanics is its application in modeling a large range of molecules even macromolecules, like proteins and segments of DNA. It is also useful for modeling intermolecular forces well. Nevertheless, this method is unable to define many chemical properties as electronic excited states [7].

Example:

The MM+ force field extends the MM2 by supplying additional parameters [8] as the electrostatic term. The advantage of these techniques is that they perform well for a large range of organic molecules [7].

2.2.1.2 Semi-empirical Methods

The longest step for the resolution of Hartree-Fock equation is to calculate the bielectronic integrations. If the base has n dimensions, it will be necessary to carry out n^4 bielectronic integrations. To overcome this weakness, the semi-empiric techniques were introduced so as to simplify the calculation and treat molecules without size restrictions.

Example: Parameterization method 3 (PM3) is one of the Neglect of Diatomic Differential Overlap (NDDO) methods; it is commonly used in organic chemistry to solve the structure and reactivity problems [9].

On average, PM3 predicts energies and bond lengths more accurately than AM1 or MNDO [7].

2.2.1.3 Quantum mechanics

Quantum mechanics explain mathematically the correct behavior of electrons. They aim to study the individual atoms or molecules on the basis of the wave function through resolving the **Schrödinger** equation given as following:

$$\hat{H}\psi = E\psi \quad (2.1)$$

\hat{H} : the Hamiltonian operator,

ψ : the wave function,

E : the energy.

This is a probabilistic description of electrons as waves. So, it can define their locations as probabilities.

In general, the Hamiltonian operator is based on the kinetic energy of the particle within a wave formulation as well as the energy due to Coulombic attraction or repulsion of particles. Nevertheless, this approach is rarely used in the current software. It is more dependent on the Born-Oppenheimer approximation where the Hamiltonian for a molecule with stationary nuclei is given as an algebraic sum of the kinetic energy of just the electrons plus the energies of electrons-nuclei attraction and the electrons repulsion. The fourth term is the repulsion energy between nuclei.

Example: The density functional theory (DFT) is founded on the determination of the energy of a molecule from the electron density instead of a wave function [7]. This method is used in optimizing molecules geometries, studying reactivity, and determining spectroscopic constants...etc. Furthermore, it can establish calculations in fundamental and excited states [10].

This method is advantageous since the integrals for Coulomb repulsion concern only the electron density, given as a three-dimensional function [7].

2.2.2 Modeling methods

2.2.2.1 Gaussian processes

Gaussian process (GP) is a generalization of the Gaussian probability distribution, but for an infinitely numerous possible functions. “The formal definition of GP is that is a set o random variables. Any finite number of the variables has a jointly Gaussian distribution” [11]. In other words, “A Gaussian process generates data located throughout some domain such that any finite subset of the range follows a multivariate Gaussian distribution” [12].

GP is a Bayesian nonparametric modeling technique [13]. It is among the techniques based on the kernel [11].

In Gaussian process regression, we assume that the output y of a function f at an input x can be written as:

$$y=f(x) + \varepsilon \tag{2.2}$$

ε : Noise term.

ε is following a normal distribution as: $\varepsilon \sim N(0, \sigma_n^2)$.

The relation between observations is expressed by the covariance function, $k(x, x')$ which is generally squared exponential:

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^2}{2l^2}\right] + \sigma_n^2 \delta(x, x') \quad (2.3)$$

Where,

x, x' : input points,

σ^2 : a hyper-parameters called the signal variance,

l : a hyper-parameters called the length scale,

$\delta(x, x')$: Kronecker delta function [12].

Thus, to establish a GP regression, we calculate, first, the covariance matrix K , and then we estimate the value of y_* at x_* .

As the GP modeling is based on a multivariate Gaussian distribution, we can write:

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \quad (2.4)$$

$$K_* = [k(x_*, x_1) \quad k(x_*, x_2) \quad \dots \quad k(x_*, x_n)] \quad (2.5)$$

$$K_{**} = k(x_*, x_*) \quad (2.6)$$

^T: Transposition.

Our aim is to find the probability $p(y_*|y)$ which follows a Gaussian distribution:

$$y_*|y \sim N(K_* K^{-1}y, K_{**} - K_* K^{-1}K_*^T) \quad (2.7)$$

The mean of this distribution gives the best estimation of y_* :

$$\bar{y}_* = K_* K^{-1}y \quad (2.8)$$

2.2.2.2 Artificial neural networks

Artificial neural network (ANN) is an imitation of biological neural network. They share the same structure, processing, and learning [11].

A biological neuron works basically as following: Synapses receive impulses, which are processed by the cell body (Soma). Then, the soma sends a response over

an axon to its synapses. Synapses are connected to glands, muscles or other neurons [14].

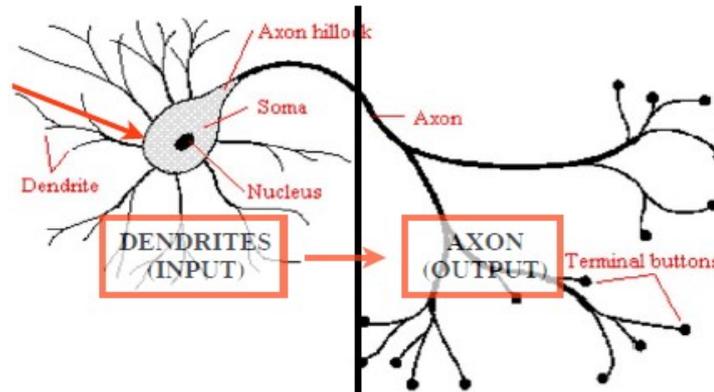


Figure 2.3. Schematic of biological neuron [15].

ANNs are principally arranged of single units known as artificial neurons which are interconnected and organized into layers. Each neuron ends with an output signal based on its activation function [16].

An ANN is composed of three layers: input layer, hidden layer(s), and an output layer [11].

Figure 2.4 shows a single artificial neuron. It is characterized by two parts Z and O. The first linear fragment is a sum of the inputs (I) multiplied by their weights (W) adding to a bias (B).

$$Z = W_1 * I_1 + W_2 * I_2 + W_3 * I_3 + B \quad (2.9)$$

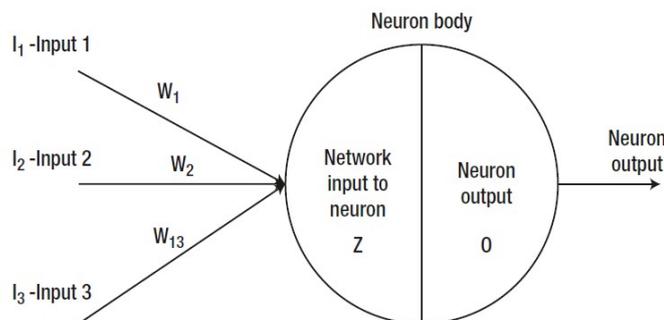


Figure 2.4. Single artificial neuron [14].

Whereas the output O is computed by means of a nonlinear function known as the activation function α :

$$O = \alpha(Z) \quad (2.10)$$

There are some activation function widely used in ANN modeling, for instance sigmoid and tanh activation function.

➤ Sigmoid (logistic) function:

$$\alpha(Z) = \frac{1}{1 + e^{-Z}} \quad (2.11)$$

➤ tanh activation function:

$$\alpha(Z) = \frac{2}{1 + e^{-2Z}} - 1 \quad (2.12)$$

The selection of the suitable activation function is based on giving the best results, in addition to others criterions as interval where it is well-behaved (not saturated) [14].

Instead of fitting data to an equation, neural networks engender hidden models of the relations between inputs and outputs. They are capable to model nonlinear systems, however they have the limitation of overfitting the data [2].

NNs can be distinguished on the base of the direction of excitation propagation within the network as Feed– forward and Recurrent [16].

Multi-layer perceptron and Radial-Basis Function are the two ANNs mainly used in modeling; both are feed-forward networks in which the information has one direction from the input to the output:

Input descriptors → Processing → Output response

The first method is based on a leyered network of interconnected perceptrons. The importance of their inputs is given by their weights which are adjusted in training a

single perceptron. To fulfill a good model, some parameters must be optimized, for instance the number of layers and neurons as well as the neuron transfer function.

Whereas, the radial-basis function neural networks has the following structure: input layer, a single hidden layer and an output layer. Here, the output in the hidden layer is not calculated as the product of the weights by their input values. Each hidden layer neuron is defined by its center and it utilizes a radial basis function as a nonlinear transfer function. The Gaussian function is classically used where the adjusted parameters in the training step are the mean and the width [2, 17].

One of the simplest methods used in ANN is Gauss Newton algorithm. It is an approach in solving nonlinear least squares problems. It is a generalization of Newton's method, for multiple dimensions.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (2.13)$$

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} \quad (2.14)$$

-x: independent variable,

-f(x): objective function,

-f'(x): f(x) derivative,

-f''(x): f(x) second derivative.

In the case of Gauss Newton method, the objective function is the sum squares of error (SSE) which must be minimized.

So, the new objective function is:

$$f(x) = \frac{1}{2} \sum_{i=1}^m (r_i(x))^2 \quad (2.15)$$

-r(x) is the residual function; it is defined as:

$$r_i(x) = \phi(x) - y_i, \quad i = 1, 2, \dots, m \quad (2.16)$$

$-\phi(x)$ is the proposed function and y_i is the observation data.

The gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$ of $f(x)$ can be expressed in terms of the Jacobian as following:

$$\nabla f(x) = J(x)^T * r(x) \quad (2.17)$$

$$\nabla^2 f(x) = J(x)^T * J(x) + \sum_{i=1}^m r_i(x) * \nabla^2 r_i(x) \quad (2.18)$$

$-J(x)^T$ is the transpose of the Jacobian matrix.

Thus, the needed steps to generalize Newton's method for multiple dimensions are:

- ❖ Replace $f'(x)$ (Equation 2.14) with the gradient $\nabla f(x)$.
- ❖ Replace $f''(x)$ (Equation 2.14) with the Hessian $\nabla^2 f(x)$ [18, 19].

2.2.2.3 Statistical analysis

To process to a QSAR study, a collection of experimental data must be firstly done. Then, this data is splitted into training set (model construction) and test set (external validation) [20]. When the model is established, some statistical parameters must be evaluated to validate the model, for instance:

- Squared correlation coefficient for the training set (R^2):

$$R^2 = 1 - \frac{\sum_{m=1}^N (\hat{y}_m - y_m)^2}{\sum_{m=1}^N (y_m - \bar{y})^2} \quad (2.19)$$

- Squared correlation coefficient for the test set [20]:

$$R_{\text{pred}}^2 = 1 - \frac{\sum (\hat{y}_{\text{test}} - y_{\text{test}})^2}{\sum (y_{\text{test}} - \bar{y})^2} \quad (2.20)$$

- Average relative error (ARE) [21]:

$$\text{ARE} = 1 - \frac{\sum_{m=1}^N |\hat{y}_m - y_m| / |y_m|}{N} \quad (2.21)$$

$-y_m$ is the desired output,

$-\hat{y}_m$ is the predicted value by model,

$-\bar{y}$ is the mean of dependant variables (training set),

$-N$ is the number of the molecules in the data set.

2.2.3 Multivariate techniques

2.2.3.1 Principal component analysis

The Principal Component Analysis (PCA) is a statistical test used in the aim to reduce the number of the original variables. It serves to represent a variation present in a data set [22]. The PCA is principally based on finding the eigenvalues and the eigenvectors of the covariance matrix, as well as on maximizing the variance of the linear combinations between the different variables of a data set [23]. The number of the PCs is estimated according to different criterions, such as:

- Select the PCs that represent 80% or 90% of the variation in the data [24].
- Preserve just the PCs whose variances exceed 1. It is Kaiser's rule [25].
- Take the elbow point in the scree plot as the number of PCs [26].

2.2.3.2 Hierarchical clustering

Hierarchical Cluster Analysis aids in simplifying the data by focusing on its clusters and patterns [26]. It collects the samples into nested groups [27] on the base of similarity between them [26].

2.3 Drug-likeness screening

The main objective of this study is to evaluate the oral bioavailability of the data set because in vivo, the physicochemical properties of a drug affect powerfully its pharmacokinetic parameters.

According to the rule of Lipinski [28], a good absorption is attained if:

- the molecular weight is under 500 Da,
- the log P is under 5,
- there are less than 5 H-bond donors,
- there are less than 10 H-bond acceptors.

Veber [29] suggested that a reduced molecular flexibility, as expressed by the number of rotatable bonds (RB), and a low polar surface area (TPSA) may permit to predict a good drug oral bioavailability. He proposed that a high probability of good oral bioavailability is achieved if:

- the polar surface area is equal to or less than 140 \AA^2 ,
- there are 10 or fewer rotatable bonds.

A drug discovery chain is not only based on the optimization of physicochemical parameters, yet it requires evaluating other parameters related to both of toxicity and bioavailability defined as lipophilic efficiency indices: ligand lipophilic efficiency (LipE), ligand efficiency (LE) and ligand-efficiency dependent lipophilicity (LELP) [30]. These factors may contribute significantly to overly estimating the drugs quality at different stages of discovery [31]. LipE is a combination between activity and lipophilicity. It explicates the effect of lipophilicity in binding to a given target. It is also shown that a good, in vivo, performance can be reached when potency increases without increasing logP [32].

It is suggested to target a LipE in a range of 5–7 or even higher [33].

$$\text{LipE} = \text{pIC}_{50} - \log P \quad (2.22)$$

LE is dependant particularly on sizes of small molecules interfering in the affinity with a drug target. Ligand efficiency metrics are primordial for the selection and optimization of fragments, hits, and leads. LE is defined as follows:

$$LE=1.4* pIC50/N_H \quad (2.23)$$

N_H is the number of heavy atoms.

LELP is correlated with both size and lipophilicity [34]. The optimal LELP scores as $-10 < LELP < 10$ [35]. It is given by:

$$LELP= \log P/ LE \quad (2.24)$$

2.4 Conclusion

Drug discovery involves backgrounds in large scientific fields even in its first stages, such as computational chemistry, mathematical modeling and bioinformatics. Our preliminary study concerns tow antioxidant sets since these latter are recently recommended in pharmaceutical, cosmetic and food products [3]. Hence, we have followed the procedure indicated on the Figure 2.5.

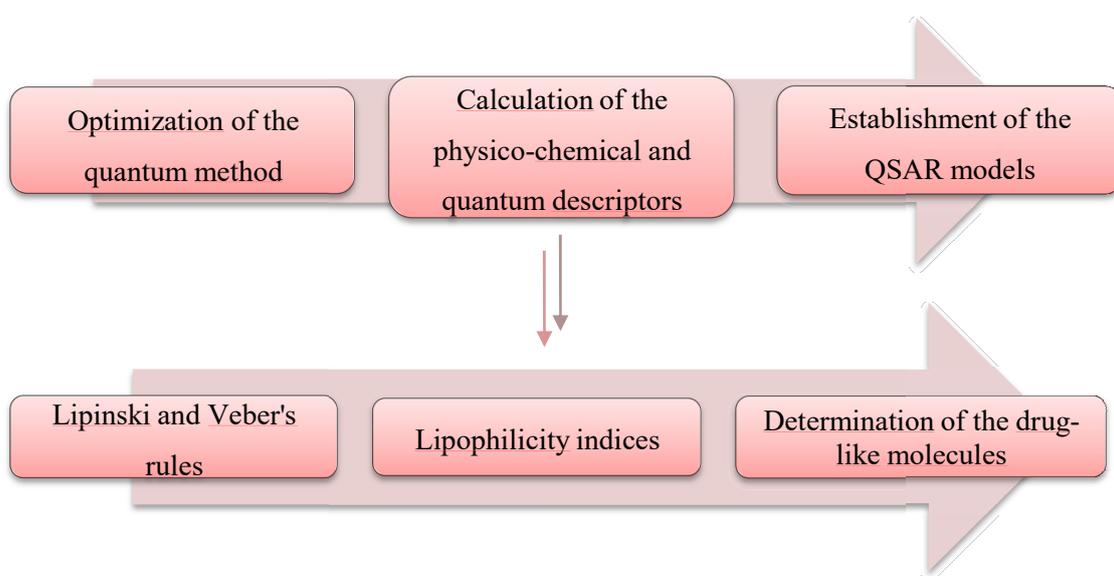


Figure 2.5. Preliminary QSAR and Drug-Likeness study.

References

- [1] N. Prakash, P. Devangi, JAA. 2 (2010) 63.
- [2] P. Liu, W. Long, Int. J. Mol. Sci. 10 (2009) 1978-1998.
- [3] A. Tropsha, Mol. Inf. 29 (2010) 476-488.
- [4] C. Hansch, T. Fujita, J. Am. Chem. Soc. 86 (1964) 1616-1626.
- [5] C. Phanus-umporn, V. Prachayasittikul, C. Nantasenamat, S. Prachayasittikul, V. Prachayasittikul, Excli Journal. 19 (2020) 458-475.
- [6] R. Paugam, Initiation à la modélisation moléculaire, Courses of Chemistry, 1M. Chimie, Orsay, Paris, 2008.
- [7] D.C. Young, Computational chemistry: A practical guide for applying techniques to real-world problems, John Wiley & Sons, Inc. New York, 2001.
- [8] G.B. Jones, B.J. Chapman, Synthesis. 5 (1995) 475-497.
- [9] G. Boucekkine, Méthodes de la chimie quantique, Techniques de l'ingénieur, V1, AF 6050, 2007.
- [10] N. Onofrio, Modélisation de l'interaction d'échange par théorie de la fonctionnelle de la densité couplée au formalisme de la symétrie brisée : Application aux dimères de cuivre, PHD Thesis, Doctoral School of Biochemistry Sciences, Grenoble University, France, 2006.
- [11] L.C. Yee, Y.C. Wei, Current modeling methods used in QSAR/QSPR, Statistical modelling of molecular descriptors in QSAR/QSPR. In Dehmer, K. Varmuza, and D. Bonchev (eds), 2012.

- [12] M. Ebden, Gaussian Processes for Regression: A Quick Introduction, 2008 (arXiv: 1505. 02965 [math.ST]).
- [13] I.E. Poloskov, C. Soize, Appl. Math. Model. 56 (2018) 15-31.
- [14] I. Livshin, Artificial neural networks with Java, Apress, Chicago, 2019.
- [15] P. De Camilli, Neurons and glial cells lecture, 2012 (medcell.med.yale.edu).
- [16] R. Hrabuska, M. Prauzek, M. Venclikova, J. Konecny, IFAC Papers OnLine. 51 (2018) 438-443.
- [17] N.S. Sethi, Int. J. Drug Discov. 3 (2012) 815-836.
- [18] W.H. Lai, S.L. Kek, K.G. Tay, Int. j. innov. sci. eng. technol. 4(2017) 258-262.
- [19] A. Croeze, L. Pittman, W. Reynolds, Lecture: Nonlinear Least-Squares Problems with the Gauss-Newton and Levenberg-Marquardt Methods, Department of Mathematics, Louisiana State University, 2012.
- [20] U. Muhammad, A. Uzairu, D.E. Arthur, J Anal Pharm Res. 7(2018) 240-242.
- [21] Y. Zhang, H. Wang, Z. Yang, J. Li, PLOS ONE. 9 (2014) 4.
- [22] D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, Trends Food Sci. Technol. 72 (2018) 85.
- [23] H. Gatignon, Statistical Analysis of Management Data. Springer Science + Business Media, New York, 2014.
- [24] I.T. Jolliffe, Principal Component Analysis, 2nd Edition, Library of Congress Cataloging-in- Publication Data, Springer –Verlag, New York, 2002.
- [25] H.F. Kaiser, Educ. Psychol. Meas. 20 (1960) 141-151.

- [26] M.M.C. Ferreira, J. Braz. Chem. Soc. 13 (2002) 742-753.
- [27] M. Dash, H. Liu, P. Scheuermann, K.L. Tan, Data Knowl Eng. 44 (2003) 109-138.
- [28] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Adv. Drug Deliv. Rev. 23 (1997) 3.
- [29] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, J. Med. Chem. 45 (2002) 2615- 2623.
- [30] J.A. Arnott, R. Kumar, S.L. Planey, J. Appl. Biopharm. Pharmacokinet. 1 (2013) 31-36.
- [31] A. Zerroug, S. Belaidi, I. Ben Brahim, L. Sinha, S. Chtita, J. King Saud Univ. Sci. 31 (2019) 595-601.
- [32] I. Jabeen, K. Pleban, U. Rinner, P. Chiba, G.F. Ecker, J. Med. Chem. 55 (2012) 3261-3273.
- [33] P.D. Leeson, B. Springthorpe, Nat. Rev. Drug Discov. 6 (2007) 881-890.
- [34] A.L. Hopkins, G.M. Keserü, P.D. Leeson, D.C. Rees, C.H. Reynolds, Nat. Rev. Drug Discov. 13 (2014) 105-121.
- [35] G.M. Keserü, G.M. Makara, Nat. Rev. Drug Discov. 8 (2009) 203-212.

Chapter 3. QSAR
and drug-likeness study

Chapter 3. QSAR and drug-likeness study for antioxidant activity of benzofurans

3.1 Introduction

There is a growing consensus among scientists that a combination of antioxidants, rather than single entities, may be more effective over the long term. Antioxidants may be of great benefit in improving the quality of life by preventing or postponing the onset of degenerative diseases [1]. They are capable of stabilizing or deactivating free radicals before they attack cells. Antioxidants are critical for maintaining optimal cellular and systemic health and well-being [2].

Among compounds that show an antioxidant activity, furan [3] and benzofuran (Figure 3.1) derivatives. They are heterocyclic organic compounds. These derivatives exhibit also high potentialities for use as pharmacological agents. Indeed, these ring systems have emerged as powerful scaffolds for many biological evaluations and play an important role in the design and discovery of new physiological/pharmacologically active molecules.

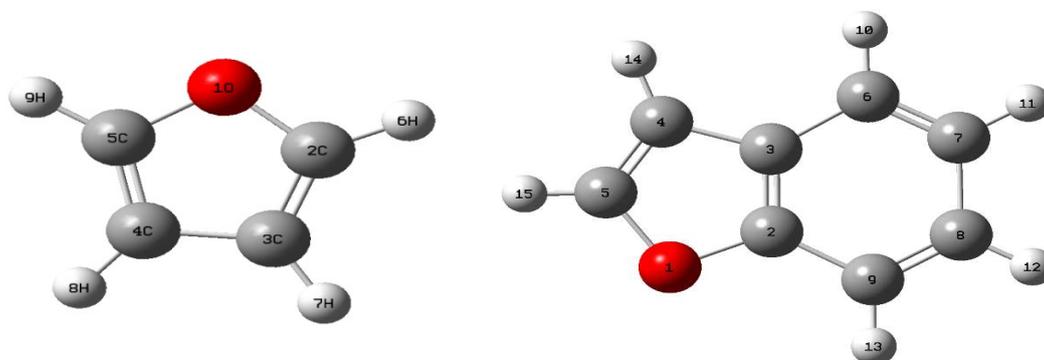


Figure 3.1. 3D conformations of furan and benzofurans.

This chapter [4] aims to derive a quantitative structure- antioxidant relationship model from a data set of fifteen synthetic furan derivatives [3] (Table 3.1) using artificial neural network (ANN) method on JMP 8.0.2 [5]. In the majority of studies, ANNs are used with large sets. We tried here, to take benefit of this technique in this case with just fifteen molecules. Then, Lipinski and Veber rules, and lipophilicity indices are applied to identify “drug-like” compounds.

3.2 Computational details

3.2.1 Furan geometry

We started our investigations by optimizing furan geometry using molecular mechanics, with MM+ force-field, at a gradient of 0.01 kcal/ (Å.mol) [6]. Next, the geometry of furan was reoptimized using DFT/B3LYP, MP2 and CCSD with 6-31G (d,p) and cc-pVDZ basis set by Gaussian 09 program package [7] in order to select the most reliable predictive method comparatively to experiment. Also, 2D and 3D molecular electrostatic potential surface maps (2D MESP/ 3D MESP) are explored.

3.2.2 Calculation of molecular descriptors

Firstly, the geometries of furan derivatives were optimized by molecular mechanics, with MM+ force-field. Then, the QSAR properties module from Hyper Chem 8.08 was used to calculate: molar weight (MW), surface area (S), volume (V), refractivity (R), polarizability (P), octanol-water partition coefficient (log P) and hydration energy (HE).

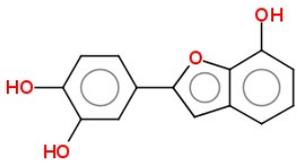
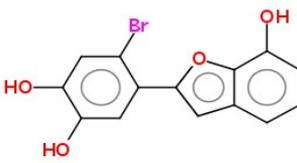
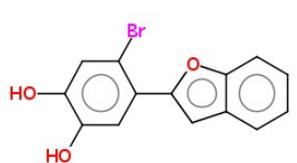
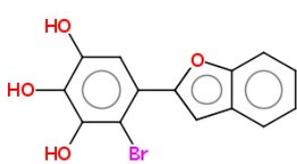
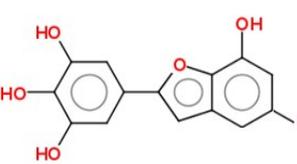
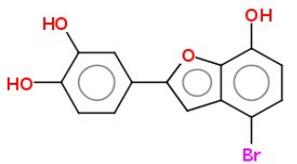
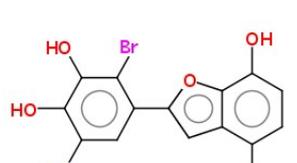
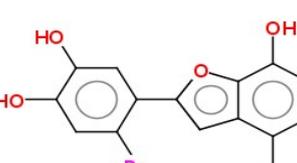
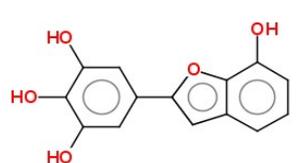
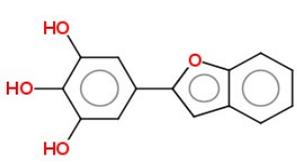
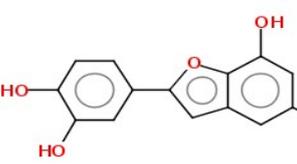
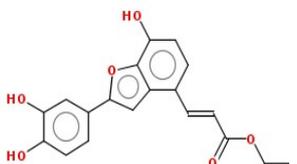
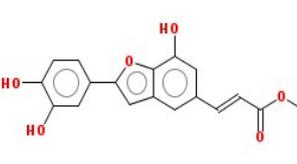
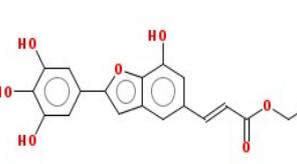
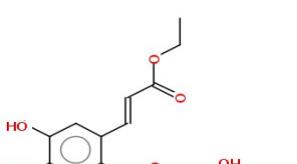
In the next step, quantum computation of the highest occupied molecular orbital energy (EHOMO) and lowest unoccupied molecular orbital energy (ELUMO) were determined using Gaussian 09 at the density functional theory level (DFT) using Becke’s three-parameter Lee- Yang- Parr (B3LYP),with 6-31G (d,p) basis set.

Finally, MarvinSketch 17.1.2 Software [8] was employed to compute: hydrogen bond donors (HD), hydrogen bond acceptors (HA), number of rotatable bonds (RB) and polar surface area (TPSA).

3.2.3 Data set

The studied furan derivatives and their corresponding antioxidant activity using 1, 1- diphenyl-2-picrylhydrazyl (DPPH) were compiled from the literature (Table 3.1).

Table 3.1. Chemical structures of the molecules under study.

N	Structure	IC50	N	Structure	IC50	N	Structure	IC50
1		18.5	2		20.6	3		33.8
4		18.0	5		14.6	6		18.4
7		19.4	8		7.8	9		12.8
10		10.8	11		28.6	12		16.8
13		25.4	14		11.7	15		26.2

All the studied molecules showed a free radical scavenging activity with IC50 values ranged between 7.8 ± 1.6 and 33.8 ± 2.8 μM [3]. 2D structures of furan derivatives were constructed using the Molinspiration Database [9].

3.3 Results and discussion

3.3.1 Equilibrium geometry of furan

The most efficient theoretical method may be selected by comparison with experimental results. A good agreement can be observed between different predicted geometries (bond lengths, valence angles) and experimental data (Table 3.2), mainly for DFT/B3LYP at the basis 6-31G (d,p).

The theoretical dihedral angle values calculated by different methods are either 0° or 180° , which explain that the geometry of furans is plane. From that, the DFT/B3LYP method is found to be more appropriate for further study on furan derivatives, in other parts of this work.

Table 3.2. Bond lengths and valence angles of furan.

Parameters	Experimental [10]	DFT/B3LYP		MP2		CCSD		
		6-31G(d,p)	cc-pVDZ	6-31G (d,p)	cc-pVDZ	6-31G (d,p)	cc-pVDZ	
Bond length (A°)	O1-C2	1.3621	1.3641	1.3635	1.3656	1.3633	1.3673	1.3654
	C2-C3	1.3609	1.3605	1.3643	1.3662	1.3779	1.3580	1.3689
	C3-C4	1.4309	1.4356	1.4381	1.4273	1.4365	1.4403	1.4502
	C2-H6	1.0750	1.0787	1.0857	1.0752	1.0877	1.0752	1.0887
	C3-H7	1.0768	1.0803	1.0870	1.0764	1.0892	1.0765	1.0901
Valence angle ($^\circ$)	O1-C2-C3	110.68	110.53	110.59	110.50	110.78	110.83	111.12
	C2-C3-C4	106.50	106.09	105.99	106.20	105.83	106.00	105.63
	C2-O1-C5	106.55	106.77	106.82	106.60	106.77	106.35	106.49
	O1-C2-H6	115.92	115.78	115.72	115.67	115.77	115.72	115.80
	C2-C3-H7	127.95	126.54	126.55	126.21	126.30	126.54	126.58

Figure 3.2 shows the 3D molecular electrostatic potential surface maps (3D MESP) of furan and of benzofuran. MESP allows understanding various physical and chemical phenomena such as molecular reactivity behavior, intermolecular interactions, molecular recognition, electrophilic reactions, substituent effects and the interactions induced by the reagents, for example between a drug and its cellular receptor [11].

As can be seen, furan and benzofuran present negative electrostatic potentials (red zone) around the oxygen atom due to its high electronegativity. An orange region can be observed around the carbon atoms of the five-ring of furan and the benzene ring of benzofuran. Thus, these parts may be subject to electrophilic attacks. We can see also positive electrostatic potentials (blue zone) around the atoms of hydrogens, which explain that these atomic sites are susceptible for nucleophilic attacks. In sum, furan and benzofuran exhibit common characteristics, which may be helpful for a qualitative understanding of the electrostatic interactions that may take place between reagents or enzyme active sites and the benzofuran derivatives under study.

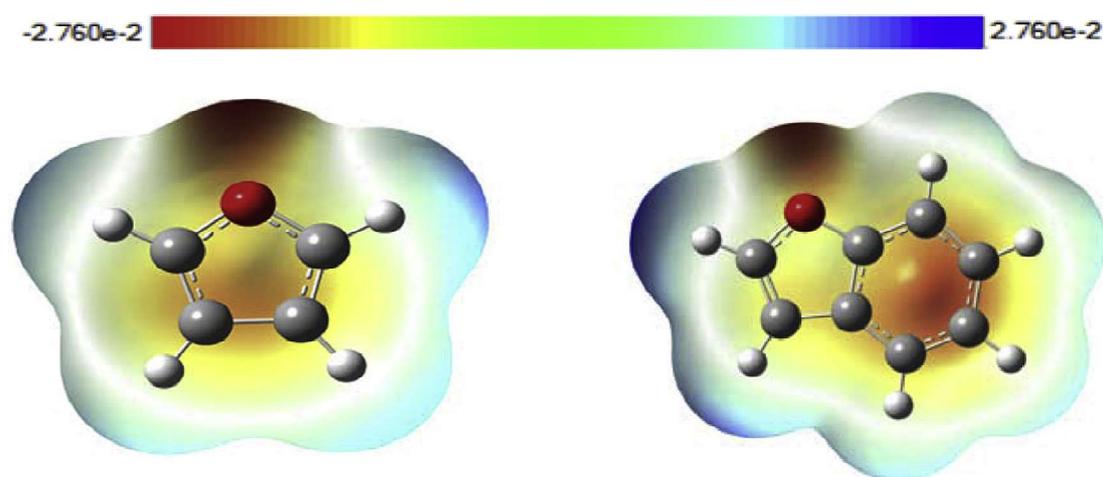


Figure 3.2. 3D MESP for furan (left) and for benzofuran (right). The results are shown by color, from red (most negative) to blue (most positive).

3.3.2 Vibrational analysis

Accurate calculations of anharmonic frequencies of molecular systems allow correct identification of molecules through the assignment of their experimental RAMAN, IR or Far-IR spectra. Such computations are also needed to establish the reliability of the derivatives of the potentials, i.e. force fields, in semi empirical approaches.

Obviously, this is crucial for accurate predictions of thermal contributions to enthalpies and entropies and of the reactive properties of these compounds using such methods. Here, we used second order perturbation theory (VPT2) approach as implemented in Gaussian and MOLPRO [12] to evaluate the anharmonic vibrational frequencies of isolated furan and benzofuran. These data were calculated by means of B3LYP/6-31 G (d,p). The results are listed in Table 3. 3, where they are compared to the respective FT-IR of experimental data.

Table 3.3. Anharmonic vibrational frequencies (in cm^{-1}) of furan [13] and of benzofuran [14] as computed using B3LYP/6- 31 G (d,p).

Furan				Benzofuran			
Mode	Sym.	B3LYP/6-31 (d,p)	Exp.	Mode	Sym.	B3LYP/6-31 (d,p)	Exp.
1	a ₁	3165	3169	1	a'	3165	3158
2	a ₁	3143	3140	2	a'	3126	3124
3	a ₁	1494	1491	3	a'	3085	3094
4	a ₁	1392	1385	4	a'	3057	3077
5	a ₁	1150	1140	5	a'	3065	3067
6	a ₁	1075	1067	6	a'	3029	3047
7	a ₁	1005	995	7	a'	1631	1617
8	a ₁	874	870	8	a'	1601	1595
9	a ₂	865	864	9	a'	1558	1543

10	a ₂	713	722	10	a'	1487	1478
11	a ₂	606	600	11	a'	1461	1457
12	b ₁	3156	3161	12	a'	1373	1346
13	b ₁	3133	3130	13	a'	1344	1329
14	b ₁	1577	1558	14	a'	1274	1264
15	b ₁	1270	1267	15	a'	1260	1253
16	b ₁	1181	1181	16	a'	1181	1180
17	b ₁	1048	1043	17	a'	1164	1161
18	b ₁	878	873	18	a'	1138	1131
19	b ₂	827	838	19	a'	1112	1107
20	b ₂	747	745	20	a'	1044	1036
21	b ₂	612	603	21	a'	1019	1007
				22	a'	902	900
				23	a'	855	855
				24	a'	770	768
				25	a'	614	612
				26	a'	544	539
				27	a'	404	400
				28	a''	966	970
				29	a''	928	929
				30	a''	861	861
				31	a''	848	848
				32	a''	770	763
				33	a''	752	746
				34	a''	737	732
				35	a''	590	584
				36	a''	576	569
				37	a''	425	418
				38	a''	251	246
				39	a''	218	211

The benzofuran is a planar asymmetric rotor with C_s symmetry. The vibrations separate into 27 a' (in-plane) and 12 a'' (out-of plane) fundamentals [15]. Furan is also a planar, near oblate symmetric top rotor. It possesses 8 a₁, 3 a₂, 7 b₁, and 3 b₂

vibrations [16]. In brief, the comparison of the B3LYP/6-31G (d,p) computed and measured fundamentals for furan and benzofurans shows differences of less than 10 cm^{-1} between both sets of data. Again, this confirms the good performance of B3LYP/6-31G (d,p) chosen theoretical level for the derivation of properties of larger benzofuran derivatives.

3.3.3 Data set for analysis

In this work, 9 descriptors were chosen to describe the structure of the furan derivatives. Table 3.4 shows the values of the calculated descriptors used in order to establish the quantitative structure- antioxidant activity relationship.

QSAR was carried out on a series of 15 furan derivatives, the data set was randomly divided into two sets: a training set (12 compounds) and a testing set (3 compounds: 6, 12 and 15) at a ratio of 80:20.

Table 3.4. Values of molecular descriptors used in the QSAR study (*: Test set).

Comp.	MW (amu)	S (\AA^2)	V (\AA^3)	R (\AA^3)	P (\AA^3)	logP	HE (kcal/mol)	EHOMO (eV)	ELUMO (eV)	IC50 (μM)
1	242.23	419.57	685.41	74.61	25.35	-2.44	-23.21	-0.238	0.025	18.5
2	321.13	440.83	732.74	82.15	27.97	-2.39	-22.50	-0.240	0.008	20.6
3	305.13	433.63	711.92	80.54	27.33	-1.36	-16.04	-0.236	0.012	33.8
4	321.13	439.52	729.18	82.15	27.97	-2.39	-20.06	-0.235	0.013	18.0
5	337.13	458.90	765.50	83.75	28.61	-3.41	-27.98	-0.254	0.015	14.6
6*	321.13	447.86	741.16	82.15	27.97	-2.39	-22.68	-0.250	0.018	18.4
7	416.02	474.28	802.32	91.29	31.23	-3.36	-25.56	-0.257	0.004	19.4
8	400.02	471.11	790.38	89.68	30.60	-2.33	-22.03	-0.258	0.005	7.80
9	258.23	426.19	703.13	76.22	25.98	-3.46	-28.55	-0.237	0.023	12.8
10	242.23	418.17	682.36	74.61	25.35	-2.44	-22.56	-0.234	0.026	10.8
11	321.13	450.12	745.24	82.15	27.97	-2.39	-22.53	-0.251	0.017	28.6
12*	340.33	569.21	958.60	100.37	35.05	-2.27	-22.81	-0.237	0.017	16.8
13	340.33	582.08	965.45	100.37	35.05	-2.27	-22.70	-0.239	0.017	25.4
14	356.33	590.10	984.78	101.98	35.69	-3.29	-28.71	-0.240	0.018	11.7
15*	340.33	542.93	941.88	100.37	35.05	-2.27	-22.09	-0.241	0.016	26.2

3.3.4 QSAR study

ANN contained 6 inputs corresponding to the 6 descriptors selected from the correlation matrix, 2 hidden neurons, and 1 output neuron which is the IC₅₀ (Figure 3.3). The number of artificial neurons in the hidden layer was adjusted experimentally [17], two neurons in the hidden layer permitted to attain the best correlation between experimental and predicted data.

The ANN was trained using Gauss Newton method.

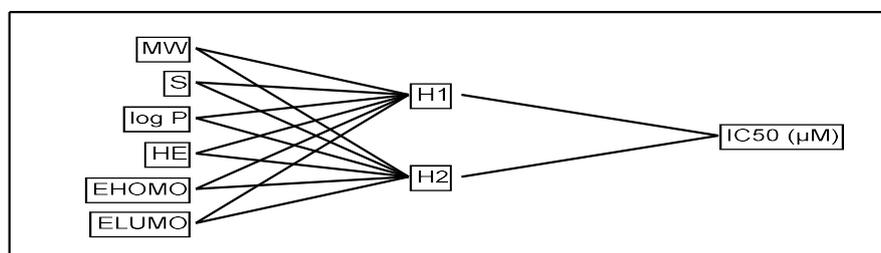


Figure 3.3. Structure of ANN.

A good correlation of experimental and predicted IC₅₀ calculated by ANN is shown in Figure 3.4., and illustrated by R² and ARE values calculated as given by equations 2.19 and 2.21 (R²=0.99 and ARE= 0.99). So, Even for just 12 molecules, the ANN model established a good internal correlation.

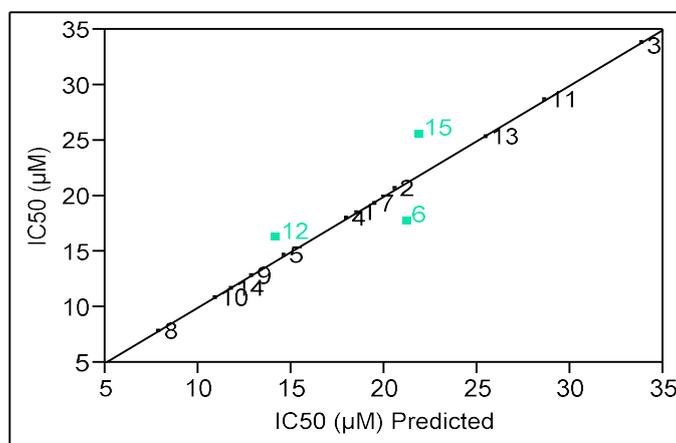


Figure 3.4. Correlation of experimental and predicted IC₅₀ as calculated by ANN.

It is known that the test set must comprise at least five samples to consider the prediction statistics as reliable [18]. In spite of the small number of the test set in our case with just three compounds, we calculate the external correlation coefficient (R^2_{pred}) and ARE_{test} according to the equations 2.20 and 2.21 respectively. That may be considered as only a preliminary prediction for testing new molecules. R^2_{pred} equals about 0.6 which is the acceptable inferior limit in external validation [19] and ARE_{test} is 0.84.

3.3.5 Drug-likeness screening

Table 3.5 presents the drug-likeness parameters and lipophilicity indices of the benzofuran derivatives under investigation. We give also the criteria to satisfy the lipophilicity indices and the Lipinski and Veber rules. As can be seen there, all compounds satisfy the rules, except compounds 5, 7 and 14 which show a slight out of range LELP (<-10) [20]. This table shows also that these compounds have molecular weights of less than 500 Da. Moreover, Table 3.5 gives the values of log P, which is a parameter used to estimate molecular hydrophobicity. For all studied compounds, log P is less than zero, so these molecules possess high water solubility and, in principle, poor membrane permeation capacity, which may be partly remedied by the small molecular weight. Hydrogen bonds increase solubility in water and must be broken for the compound to permeate into and through the lipid bilayer membrane. Thus, an increasing number of hydrogen bonds reduces partitioning from the aqueous phase into the lipid bilayer membrane for permeation by passive diffusion [21].

Table 3.5. Drug-likeness parameters and lipophilicity indices of furan derivatives.

Comp.	Lipinski rules				Veber rules		Lipophilicity indices	
	MW (amu) <500	logP <5	HD <5	HA <10	TPSA (Å ²) <140	RB <10	LipE >5	LLEP -10 < LLEP < 10
1	242.23	-2.44	3	3	73.82	1	7.17	-6.63
2	321.13	-2.39	3	3	73.82	1	7.08	-6.92
3	305.13	-1.36	2	2	53.6	1	5.83	-3.91
4	321.13	-2.39	3	3	73.82	1	7.13	-6.84
5	337.13	-3.41	4	4	94.05	1	8.24	-10.07
6	321.13	-2.39	3	3	73.82	1	7.12	-6.85
7	416.02	-3.36	4	4	94.05	1	8.07	-10.70
8	400.02	-2.33	3	3	73.82	1	7.44	-6.52
9	258.23	-3.46	4	4	94.05	1	8.35	-9.60
10	242.23	-2.44	3	3	73.82	1	7.41	-6.32
11	321.13	-2.39	3	3	73.82	1	6.93	-7.14
12	340.33	-2.27	3	4	100.13	5	7.04	-8.49
13	340.33	-2.27	3	4	100.13	5	6.86	-8.82
14	356.33	-3.29	4	5	120.36	5	8.22	-12.39
15	340.33	-2.27	3	4	100.13	5	6.85	-8.85

We see that all the screened molecules have a number of hydrogen bonds in the appropriate interval of Lipinski. In addition, all benzofuran derivatives show TPSA values in the range of 53-120 Å², which permit them to correlate very well with the

human intestinal absorption, Caco-2 monolayer's permeability, and blood-brain barrier penetration [22]. We can note also that compounds 12, 13, 14 and 15 are likely to be the most flexible with a number of rotatable bonds of 5. In the studied set, LipE takes values between 5.83 and 8.35.

Thus, this set would have a high quality as lead compounds, provided that LELP is situated in the suggested range $-10 < \text{LELP} < 10$ to guarantee that molecules reach both size and lipophilicity criteria, which is the case for all the studied compounds except three molecules (5, 7 and 14) having LELP inferior than the minimum limit (-10.07, -10.70, -12.39).

In general, according to the results listed in Table 3.5, we can estimate that all benzofuran derivatives satisfy the Lipinski and Veber rules, so they would not have problems with oral bioavailability. 80% of compounds reach an LELP in the suggested range $-10 < \text{LELP} < 10$.

3.4 Conclusion

A QSAR study was performed on 15 benzofuran derivatives. QSAR models were obtained using artificial neural network technique. ANN gave a good agreement with experimental values even for our small data set. Calculation of molar weight, octanol-water partition coefficient, number of hydrogen bond donors and hydrogen bond acceptors revealed that all studied molecules obey the Lipinski's rule of five. These molecules satisfy also Veber rules with $\text{RB} < 10$ and $\text{TPSA} < 140$.

Depending on LELP calculated values, all studied set except three compounds (5, 7 and 14) could have promising results in the clinic.

References

- [1] M.N. Alam, N.J. Bristi, M. Rafiquzzaman, Saudi Pharm. J. 21 (2013) 144.
- [2] M. Percival, Clin. Nut. Insights. 1 (1998) 1.
- [3] J.F. Hsieh, W.J. Lin, K.F. Huang, J.H. Liao, M.J. Don, C.C. Shen, Y.J. Shiao, W. T. Li, Eur. J. Med. Chem. 93 (2015) 446.
- [4] S. Boudergua, M. Alloui, S. Belaidi, M. Mogren Al Mogren, U.A. Abd Ellatif Ibrahim, M. Hochlaf, J. Mol. Struct. 1189 (2019) 307-314.
- [5] JMP 8.0.2, SAS Institute Inc., 2009.
- [6] Hyper Chem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2008).
- [7] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, H. Nakatsuji, Gaussian 09, Revision E. 01, 2009, Gaussian. Inc., Wallingford CT.
- [8] MarvinSketch17.1.2, Chemaxon (2017), (<http://www.chemaxon.com>).
- [9] Database. <http://www.molinspiration.com>.
- [10] B. Bak, D. Christensen, W.B. Dixon, L. Hansen-Nygaard, J. Rastrup-Andersen, M. Schottlander, J. Mol. Spectrosc. 9 (1962) 124.
- [11] M. Alloui, S. Belaidi, H. Othmani, N. E. Jaidane, M. Hochlaf, Chem. Phys. Lett. 696 (2018) 70-78.
- [12] H.J. Werner, P.J. Knowles, MOLPRO, Ab Initio Programs Package, 2015. <http://www.molpro.net>.
- [13] A. Mellouki, J. Li-evin, M. Herman, Chem. Phys. 271 (2001) 239.

- [14] W.B. Collier, T.D. Klots, *Spectrochim. Acta, Part A* 51 (1995) 1255.
- [15] T.D. Klots, W.B. Collier, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 51 (1995) 1291- 1316.
- [16] T.D. Klots, R.D. Chirico, W.V. Steele, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 50 (1994) 765- 795.
- [17] M. Larif, A. Adad, R. Hmammouchi, A. I. Taghki, A. Soulaymani, A. Elmidaoui, M. Bouachrine, T. Lakhlifi, *Arab. J. Chem.* 10 (2017) 948.
- [18] A. Golbraikh, A. Tropsha, *J. Comput.-Aided Molec. Des.* 16 (2002) 357.
- [19] A. Tropsha, *Mol. Inf.* 29 (2010) 485.
- [20] G.M. Keserü, G.M. Makara, *Nat. Rev. Drug Discov.* 8 (2009) 203-212.
- [21] E.H. Kerns, L. Di, *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*, Academic Press, USA, 2008.
- [22] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* 43 (2000) 3714-3717.

Chapter 4. QSAR study for antioxidant activity
of flavonoids

Chapter 4: Quantitative structure- antioxidant activity relationship study for flavonoids set

4.1 Introduction

Cultivars and chemical pesticides are two potential ways of combating plant pathogens. However, plants can develop their own defense system so as to resist to biotic and abiotic stresses through the synthesis of a set of secondary metabolites as the flavonoids [1]. The structure of this category is based on a flavan nucleus [2]. The flavonoids are classified according to the pyran ring oxidation and saturation [2, 3] (Figure 4.1). This kind of components exists largely in some foods such as tea, fruits, cocoa, coffee, nuts and vegetables [4]. Honey is also known by its high amounts of flavonoid glycosides [5].

The flavonoids find their applications in various medical fields. They are used as antioxidant [6], antitrypanosomal, antileishmanial [7], anti-inflammatory [8], anticancer [9], as well as antiaging compounds [10]...Since oxidative stresses are strongly related to several diseases [11], the majority of researches emphasize on the flavonoids antioxidant activity and its effect on human health [12].

This part aims to elaborate a statistical study of flavonoids set using principal component analysis and hierarchical clustering. The quantitative structure-antioxidant activity relationship was also carried out by means of Gaussian process. The antioxidant activities of the studied set were obtained from literature; they are measured by the DPPH radical scavenging method [13, 14].

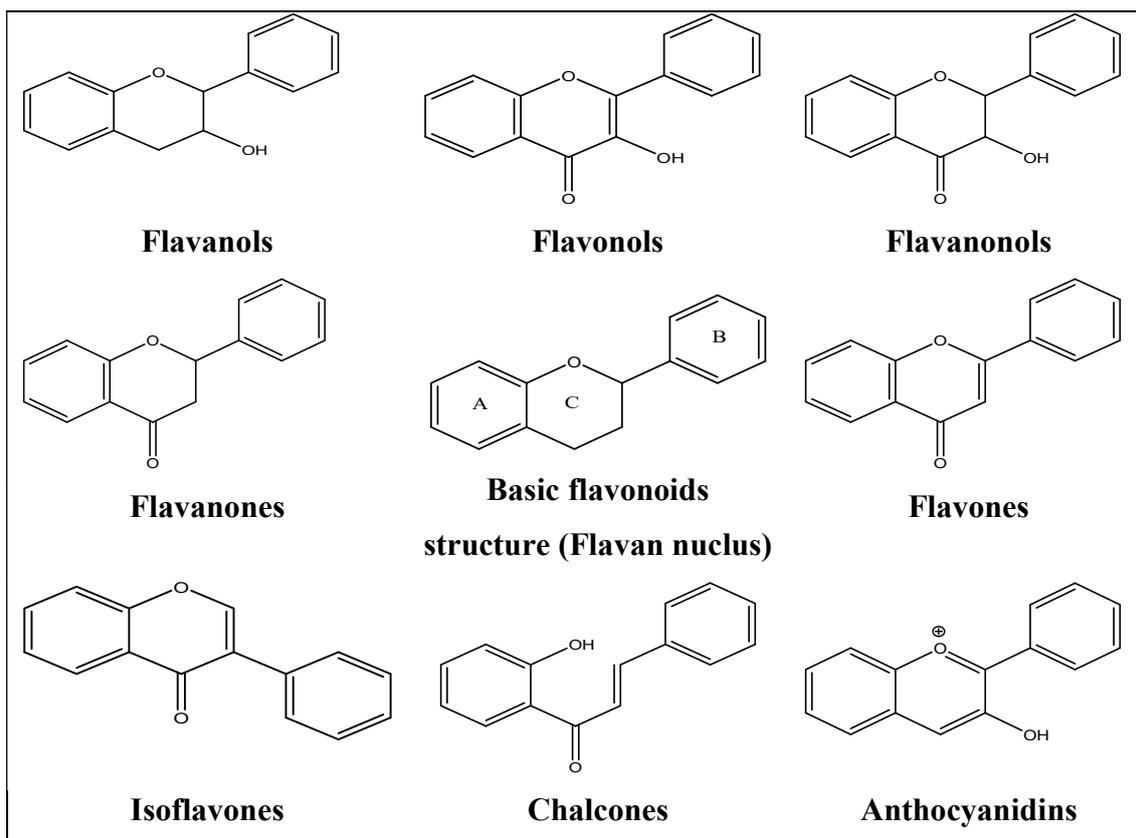


Figure 4.1. Principal flavonoids sub-classes [2, 3].

4.2 Descriptors calculation

Firstly, the molecular structures of the data set were optimized by the molecular mechanics force field (MM+), and then these geometries were reoptimized by means of the semi-empirical method PM3 in HyperChem 8.0.8 software [15]. Exploring this latter, the following molecular descriptors were computed: molar weight, volume, polarizability, octanol-water partition coefficient, hydration energy, as well as the heat of formation (HF).

Secondly, a comparative estimation of the bonds length and the angles of the quercetin, taken as an analogue of the data set, was done in order to scrutinize the goodness of the quantum technique that serves in the electronic descriptors

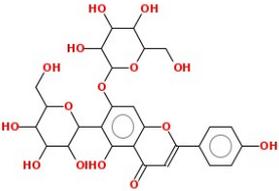
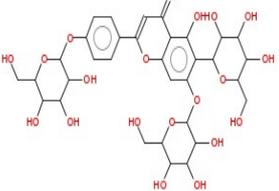
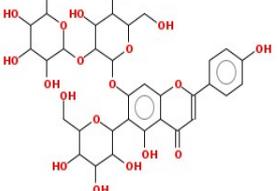
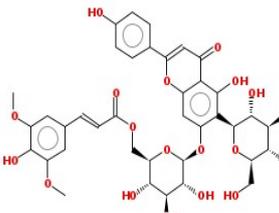
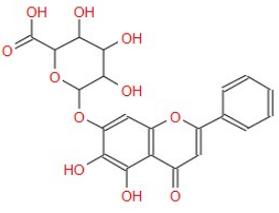
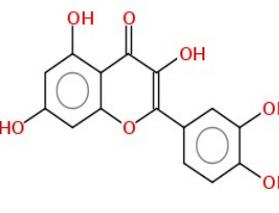
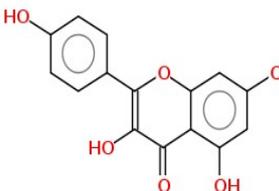
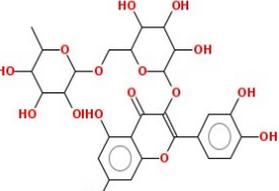
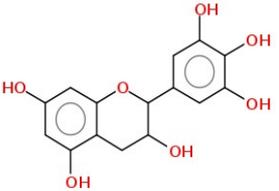
generation. To reach this objective, two techniques were used: DFT/B3LYP (6-31G) and HF (6-31G).

To determine the optimized level of the highest occupied molecular orbital energy, the lowest unoccupied molecular orbital energy and the dipole moment (DM) of the data set, we have used the Gaussian 09 program [16].

Finally, MarvinSketch [17] served to calculate: hydrogen bond donors, hydrogen bond acceptors, number of rotatable bonds and topological polar surface area.

The chemical structures of the data set under study are represented on Table 4.1. They were obtained from the database molinspiration [18].

Table 4.1. Chemical structures and experimental activity of the data set [13, 14].

Molecule	pIC50	Molecule	pIC50	Molecule	pIC50
	3.676		3.536		3.524
1		2		3	
	4.676		4.810		5.051
4		5		6	
	4.385		4.955		5.161
7		8		9	

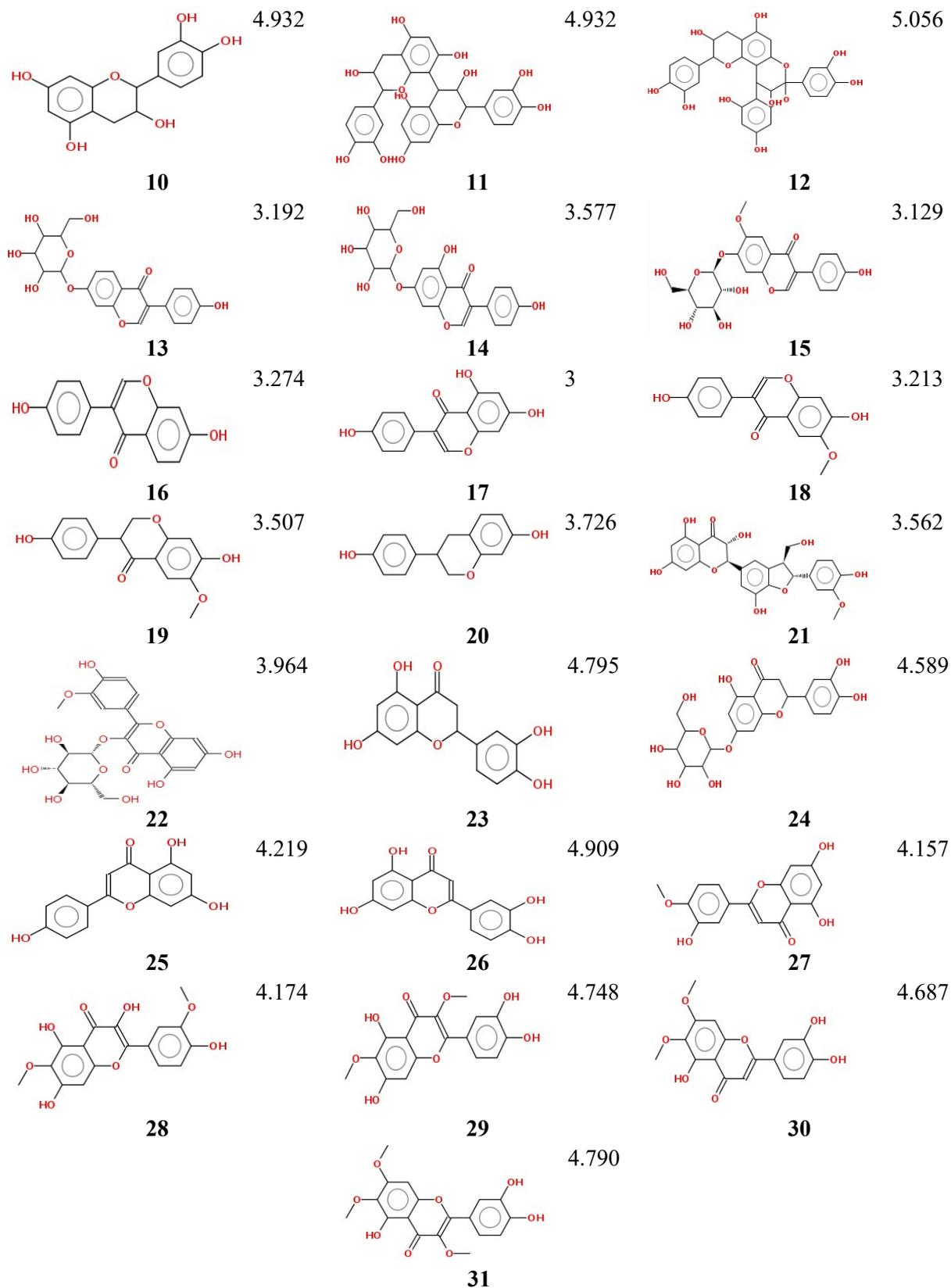


Table 4.2. Bond lengths (in Å) and valence angles (in degree, °) of quercetin [20].

Bond length	Exp.	DFT B3LYP	HF	Angle	Exp.	DFT B3LYP	HF
		(6-31G)	(6-31G)			(6-31G)	(6-31G)
(1-2)	1.36	1.40	1.38	(1-2-3)	121.26	118.17	118.21
(1-6)	1.37	1.39	1.36	(1-2-11)	110.97	112.54	112.13
(2-3)	1.36	1.37	1.34	(3-2-11)	127.75	129.29	129.66
(2-11)	1.48	1.46	1.46	(2-3-18)	122.25	122.86	122.38
(3-4)	1.45	1.44	1.45	(4-3-18)	117.34	114.93	115.60
(3-18)	1.35	1.38	1.37	(3-4-5)	116.79	117.68	117.28
(4-5)	1.42	1.43	1.44	(5-4-17)	122.82	123.79	123.64
(4-17)	1.27	1.29	1.25	(4-5-10)	122.83	121.99	122.76
(5-6)	1.39	1.41	1.40	(6-5-10)	116.96	118.43	118.12
(5-10)	1.42	1.42	1.41	(1-6-7)	116.79	117.82	117.96
(6-7)	1.40	1.39	1.38	(5-6-7)	122.79	122.44	122.70
(7-8)	1.39	1.40	1.39	(6-7-8)	117.08	117.53	117.47
(7-23)	1.02	1.68	1.07	(6-7-8)	117.08	117.53	117.47
(8-9)	1.40	1.40	1.39	(8-7-23)	122.71	120.79	120.92
(8-20)	1.36	1.38	1.36	(7-8-20)	117.66	116.07	116.36
(9-10)	1.36	1.39	1.38	(9-8-20)	119.84	121.70	121.53
(9-24)	1.01	1.08	1.07	(8-9-24)	120.44	121.82	121.88
(10-19)	1.38	1.36	1.35	(5-10-9)	122.14	120.03	120.26
(11-12)	1.39	1.41	1.40	(5-10-19)	118.45	120.42	121.67
(11-16)	1.40	1.42	1.40	(9-10-19)	119.37	119.55	118.07
(12-13)	1.39	1.40	1.39	(2-11-12)	121.34	121.81	122.19
(12-25)	1.00	1.08	1.07	(2-11-16)	119.34	119.38	118.97
(13-14)	1.39	1.39	1.38	(11-12-13)	121.02	120.54	120.42
(13-26)	1.01	1.08	1.07	(11-12-25)	119.35	119.53	119.98
(14-15)	1.38	1.41	1.39	(13-12-25)	119.62	119.93	119.60
(14-22)	1.40	1.38	1.36	(12-13-14)	118.84	120.30	120.31
(15-16)	1.40	1.38	0.37	(15-14-22)	118.33	120.24	120.68
(15-21)	1.37	1.40	0.38	(14-15-21)	118.52	113.88	114.66
(16-27)	1.01	1.08	1.07	(16-15-21)	121.40	125.01	124.37
(18-28)	0.91	0.99	0.96	(3-18-28)	110.29	106.28	110.96
(19-29)	0.95	1.00	0.96	(10-19-29)	101.93	109.35	113.86
(20-30)	0.91	0.98	0.95	(8-20-30)	113.42	112.13	115.44
(21-31)	0.99	0.97	0.95	(15-21-31)	112.59	112.83	115.54
(22-32)	0.98	0.98	0.95	(14-22-32)	101.87	109.68	113.10

Table 4.3. Values of molecular descriptors used in the statistical analysis: (V, Å³), (HE, kcal/mol), (log P), (P, Å³), (MW, amu), (HF, kcal/mol), (EHOMO, eV), (ELUMO, eV), (DM, Debye), (RB), (HD), (HA), and (TPSA, Å²).

N	V	HE	log P	Pol	MW	HF	EHOMO	ELUMO	DM	RB	HD	HA	TPSA
1	1442.92	-47.61	-5.72	54.11	594.53	-553.82	-0.27308	-0.04987	7.80	6	10	15	256.29
2	1801.56	-57.78	-7.18	67.53	756.67	-759.21	-0.26527	-0.04146	6.21	9	13	20	335.44
3	1743.13	-51.53	-6.15	66.89	740.67	-710.11	-0.22100	0.00175	8.27	8	12	19	315.21
4	1817.18	-43.72	-7.04	74.75	800.72	-655.39	-0.14423	-0.14108	14.77	12	10	18	301.05
5	1112.70	-34.05	-3.44	40.77	446.37	-376.39	-0.22794	-0.04651	7.55	4	6	11	183.21
6	755.02	-32.68	-4.01	28.54	302.24	-225.02	-0.25302	-0.04606	4.26	1	5	7	127.45
7	737.08	-27.04	-2.99	27.90	286.24	-181.97	-0.25454	-0.04463	3.76	1	4	6	107.22
8	1480.17	-53.11	-5.91	54.75	610.53	-594.10	-0.26526	-0.03515	5.46	6	10	16	265.52
9	796.96	-35.31	-4.14	29.28	306.27	-251.56	-0.24548	0.01728	4.00	1	6	7	130.61
10	779.09	-31.10	-3.12	28.65	290.27	-209.06	-0.25234	0.00795	3.20	1	5	6	110.38
11	1393.96	-52.08	-6.97	56.52	578.53	-406.53	-0.23910	-0.00041	6.39	3	10	12	220.76
12	1353.53	-50.33	-6.27	55.75	576.51	-396.04	-0.24712	0.00521	4.17	2	9	12	209.76
13	1077.51	-31.80	-2.49	40.05	416.38	-298.59	-0.24421	-0.01044	6.91	4	5	9	145.91
14	1096.72	-35.65	-3.52	40.69	432.38	-340.25	-0.23790	-0.00567	4.37	4	6	10	166.14
15	1154.98	-32.24	-3.49	42.52	446.41	-332.21	-0.24770	-0.01768	3.30	5	5	10	155.14
16	695.89	-19.27	-1.03	26.63	254.24	-91.86	-0.24242	-0.00687	5.33	1	2	4	66.76
17	709.90	-24.68	-2.05	27.27	270.24	-141.38	-0.24776	-0.03413	3.26	1	3	5	86.99
18	773.34	-19.97	-2.02	29.10	284.27	-127.56	-0.24124	-0.00586	1.77	2	2	5	75.99
19	797.68	-19.11	-1.60	29.29	286.28	-147.23	-0.24342	-0.00751	2.77	2	2	5	75.99
20	718.51	-17.37	-0.53	26.74	242.27	-84.41	-0.24625	0.01462	3.54	1	2	3	49.69
21	1206.15	-37.98	-4.95	46.87	482.44	-328.56	-0.24934	-0.05253	3.76	4	6	10	166.14
22	1178.42	-41.82	-5.44	43.80	478.41	-424.64	-0.26574	-0.04158	9.53	5	7	12	195.60
23	746.42	-28.66	-2.59	28.10	288.26	-202.56	-0.25682	-0.03066	3.92	1	4	6	107.22
24	1145.30	-39.96	-4.05	41.52	450.40	-404.53	-0.24626	-0.00343	5.94	4	7	11	186.37
25	721.73	-23.85	-2.09	27.27	270.24	-142.79	-0.26369	-0.02437	5.16	1	3	5	86.99
26	740.32	-29.49	-3.11	27.90	286.24	-185.85	-0.26617	-0.02590	4.86	1	4	6	107.22
27	797.46	-23.32	-3.08	29.74	300.27	-176.55	-0.26297	-0.02516	5.05	2	3	6	96.22
28	886.63	-26.81	-4.97	32.85	346.29	-252.05	-0.25055	-0.04583	3.80	3	4	8	125.68
29	883.77	-29.04	-4.97	32.85	346.29	-249.37	-0.25355	-0.03643	3.74	3	4	8	125.68
30	872.01	-23.24	-4.07	32.21	330.29	-208.74	-0.26586	-0.02684	4.44	3	3	7	105.45
31	942.11	-23.86	-4.94	34.68	360.32	-238.54	-0.25358	-0.03559	3.63	4	3	8	114.68

4.3.2 Principal component analysis and hierarchical clustering

Our simulation includes 31 flavonoids that are characterized by means of 13 descriptors in addition to their biological activities pIC50.

The crucial step in PCA is selecting the most significant components. In order to search for them, we use the cumulative percentage (Cum Percent) of total variation as most evident criterion which must exceed 80% [21]. The importance of the resulted PCs can be visualized through their percentages, where we can estimate that the two first principal components can explain 83.06 % of the total information as shown on Table 4.4. Thus, we can neglect the other components in data representation by PCA which helps us to compress the dataset.

Table 4.4. Principal components.

Number	Eigenvalue	Percent		Cum Percent
1	10.0318	71.656		71.656
2	1.5960	11.400		83.056
3	1.1758	8.399		91.454
4	0.5822	4.158		95.613
5	0.2962	2.116		97.728
6	0.1374	0.981		98.710
7	0.1083	0.774		99.483
8	0.0509	0.363		99.847
9	0.0119	0.085		99.932
10	0.0081	0.058		99.990
11	0.0007	0.005		99.995
12	0.0006	0.004		99.999
13	0.0001	0.001		100.000
14	0.0000	0.000		100.000

In order to understand the correlation between variables, we can perform the loading plot as illustrated by Figure 4.3. Whenever the absolute value of the loading plot is higher, its impact on the PCA- based model is greater [22].

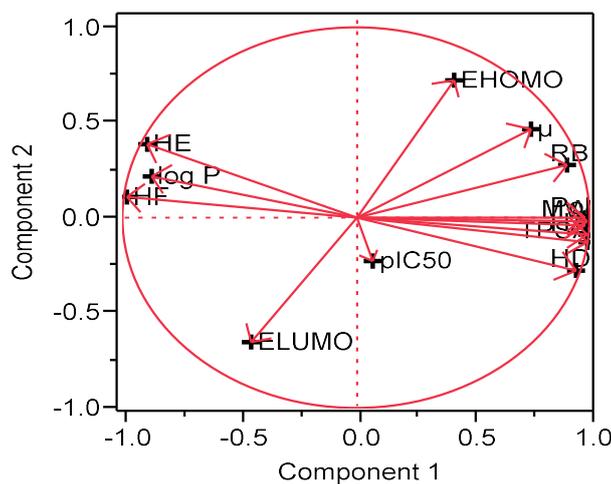


Figure 4.3. Correlation circle.

The correlation circle allows us to study the links between variables. We will therefore seek to find out whether there are groups of variables that are strongly correlated with each other. If such groups exist, then all the variables in a given group will be "synthesizable" by a synthetic variable.

From the above correlation circle, we observe that:

-HD, HA, TPSA, V, Pol, and MW form a subgroup highly correlated because their corresponding vectors are closely superposed.

-The energies HE and HF are negatively correlated with the topologic parameters in the first factorial axis.

-The second factorial axis is predominantly dependent on EHOMO and ELUMO.

Besides, we carry out a hierarchical clustering which is a multivariate technique that leads to assembly similar observations. The clusters share close values according to their descriptors [23].

Figure 4.4 shows that the data set is divided into four different clusters. Obviously, we can distinguish between them as following:

- Group I assembles the molecules: 1, 2, 3, 8, 11, and 12. They are similar in enclosing a glycoside group with a molar weight more than 575 uma.
- Group II contains just the 4th molecule; it is the unique derivative that contains a synapoyl group.
- Group III gathers the derivatives: 5, 13, 14, 15, 21, 22, and 24. This category can be distinguished from the first one by containing a glycoside group, yet with a molar weight less than 500 uma.
- Group IV collects the rest of the set. They are simple flavonoids with a molar weight less than 500 uma.

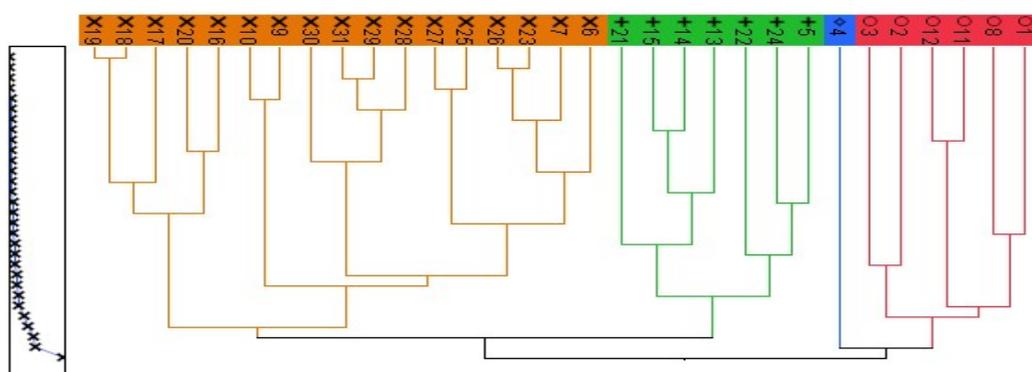


Figure 4.4. Hierarchical clustering.

Figure 4.5 illustrates the distribution of the molecules within the two PC.

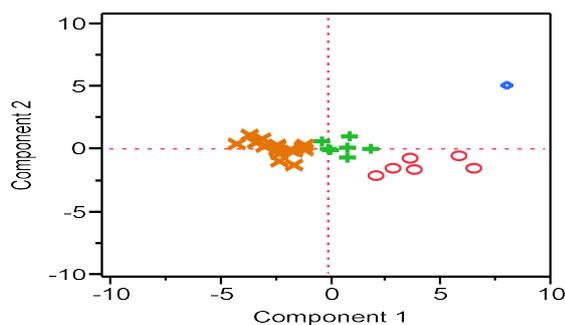


Figure 4.5. Score plot.

4.3.3 QSAR modeling using Gaussian Process

The data set is divided into training and test sets. The test set was randomly selected from the three dominant clusters I, III, and IV with a ratio of about 1:5. The excluded molecules used to validate the model are: 7, 12, 15, 19, and 23.

We can confirm that the EHOMO is the most influencing factor in the antioxidant activity with a total sensitivity of 27.03 %. That is to say that about 27 % of the variation in the activity can be explained by this parameter. Moreover, factors with high θ values have a significant impact on the prediction formula as EHOMO in this case with θ value of 27.86 [24].

Habitually, the Gaussian processes interpolate perfectly the data. They can deal with no-error-term models, where there is a similarity between input and output values [24]. This similarity can be shown by calculating the correlation coefficient (R^2) and the average relative error (ARE) for the training set.

To verify the predictive capacity of the model (Figure 4.6), an external validation was established using the predictive R^2 (R^2_{pred}) and ARE_{pred} on the test set.

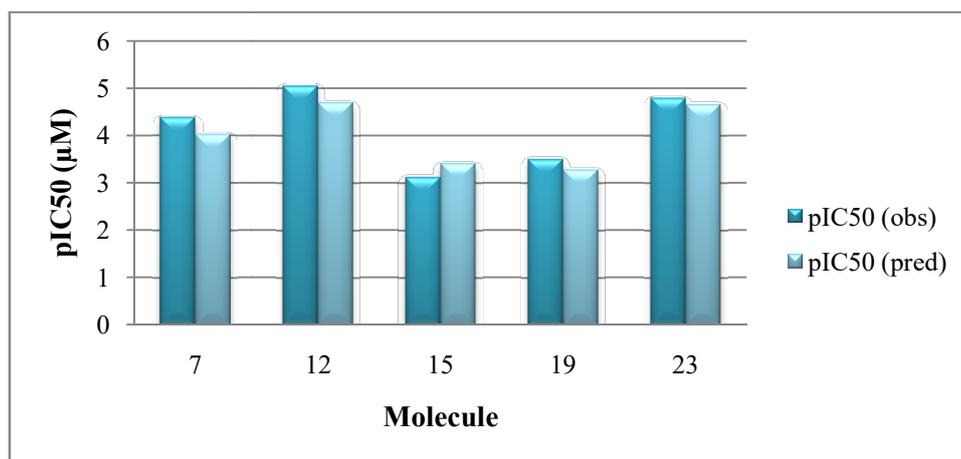


Figure 4.6. Comparative bar chart between observed (obs) and predicted (pred) pIC50 of the test set.

The values of the statistical parameters are given as following:

➤ Training set:

$$R^2=0.99, \text{ and } ARE=0.99.$$

➤ Test set:

$$R^2_{\text{pred}}=0.86, \text{ and } ARE_{\text{pred}} = 0.93.$$

We can confirm that the model has a good predictive capacity with R^2_{pred} of 0.86, which exceeds the inferior limit of 0.6 [25] and ARE_{pred} equals to 0.93.

4.4 Structure-activity relationship of the studied flavonoids

As seen in chapter 1, donating hydrogen is basic in explaining antioxidant capacity of flavonoids. They are stabilized after scavenging free radicals by resonance. Depending on the QSAR study, we can confirm that EHOMO has a high impact on the antioxidant activity since it is directly related to the reactivity of the derivatives. It is also found that the number of OH may have a positive effect on the activity till a limit of 10. After that, we observe an inverse relationship between them. That may be due to the presence of glycosides where we observe, on the Figure 4.7, the presence of the flavonoids that belong to the cluster I (molecules 1, 2, and 3) with a high MW in this part of the graph. Thus, that confirms that the number of OH groups, particularly that belong to glycosides is not enough to increase the antioxidant activity.

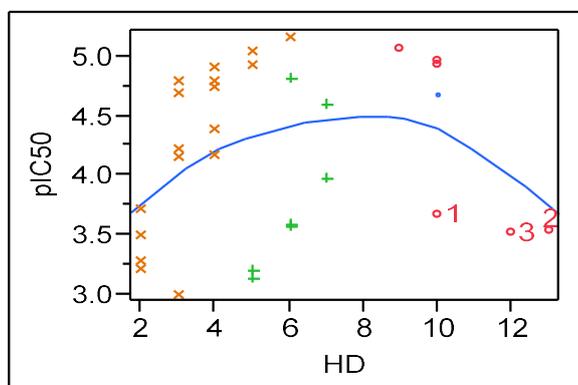


Figure 4.7. Effect of HD on the antioxidant capacity.

4.5 Conclusion

QSAR studies are often established using the common techniques as multiple linear regressions. The objective of this chapter was to carry out a QSAR model using Gaussian process regression which is seldom used in this approach. From our results, we proved that this method is able to predict the antioxidant activity of some flavonoids with a good level. This capacity can be improved by enlarging the training data set.

Depending basically on the molar weight and the presence of glycosides, our set includes four separated clusters.

References

- [1] M. Zaynab, M. Fatima, S. Abbas, Y. Sharif, M. Umair, M.H. Zafar, K. Bahadar, *Microb. Pathog.* 124 (2018) 198.
- [2] A. Dragan, D.A. Dušanka, B. Drago, R. Vesna, L. Bono, T. Nenad, *Curr. Med. Chem.* 14 (2007) 827-845.
- [3] S.D.S. Banjarnahor, N. Artanti, *Med. J. Indones.* 23 (2014) 240.
- [4] K.J. Murphy, K.M. Walker, K.A. Dyer, J. Bryan, *Nutr Res.* 61 (2019) 70.
- [5] P.Truchado, F. Ferreres, F.A. Tomas-Barberan, *J. Chromatogr. A.* 1216 (2009) 7241.
- [6] D. Kruzlicova, M. Danihelova, M. Veverka, *Nova Biotechnol. et Chim.* 11 (2012) 37.
- [7] D. Tasdemir, M. Kaiser, R. Brun, V. Yardley, T.J. Schmidt, F. Tosun, P. Rüedi, *Antimicrob. Agents Chemother.* 50 (2006) 1352-1364.
- [8] X.M. Chen, A.R. Tait, D.D. Kitts, *Food chem.* 218 (2017) 15.
- [9] D. Veeramuthu, W.R. Tharsius Raja, N.A. Al-Dhabi, I. Savarimuthu, *Flavonoids: Anticancer Properties. Biosynthesis to Human Health*, IntechOpen, 2017.
- [10] B.J. Lumbiny, Z. Hui, M.A. Islam, *J. Asiat. Soc. Bangladesh Sci.* 39 (2013) 191.
- [11] S. Das, I. Mitra, S. Batuta, M. Niharul, A. Kunal Roy, N. Ara Begum, *Bioorg. Med. Chem. Lett.* 24 (2014) 5050.
- [12] W.A. Peer, A.S. Murphy, *Flavonoids as signal molecules: Targets of Flavonoid Action. Science of flavonoids*, Erich Grotewold, USA, 2006.
- [13] M. Okawa, J. Kinjo, T. Nohara, M. Ono, *Biol. Pharm. Bull.* 24 (2001) 1202-1205.
- [14] J. Hu, W. Ma, N. Li, K.J. Wang, *J. Mex. Chem. Soc.* 61 (2017) 282-289.

- [15] Hyper Chem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA (2008).
- [16] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, H. Nakatsuji, Gaussian 09, Revision E. 01, 2009, Gaussian. Inc., Wallingford CT.
- [17] MarvinSketch, Chemaxon. (2020). (<http://www.chemaxon.com>)
- [18] Database. <http://www.molinspiration.com>.
- [19] JMP 8.0.2, SAS Institute Inc., 2009.
- [20] M. Rossi, L.F. Rickles, W.A. Halpin, Bioorg. Chem. 14 (1986) 55-69.
- [21] I.T. Jolliffe, Principal Component Analysis, 2nd Edition, Library of Congress Cataloging-in- Publication Data, Springer –Verlag, New York, 2002.
- [22] C. Yoo, M. Shahlaei, Chem. Biol. Drug Des. 91 (2018) 137-152.
- [23] SAS Institute Inc. 2017. JMP® 13 Multivariate Methods, Second Edition. Cary, NC: SAS Institute Inc.
- [24] SAS Institute Inc. 2018. JMP® 14 Predictive and Specialized Modeling. Cary, NC: SAS Institute Inc.
- [25] A. Tropsha, Mol. Inf. 29 (2010) 476-488.

General conclusion

General conclusion

The antioxidants become recently the subject of many researches since they contribute in the well being of humans. They combat the ageing related diseases and scavenge the free radicals originated from the new harmful conditions of life as pollution, smoking, and stress. The antioxidants are also largely used in the formulation of drugs and food. They are even added to the anticancer drugs to boost their activity. In addition, they are vital for preserving medicines and food against oxidation.

This work aimed to establish a QSAR study for two antioxidant sets, benzofurans and flavonoids. Depending on the IC₅₀ values, both of the sets include molecules with high antioxidant capacity, what makes them sources for new derivatives that may be used in drug and food industries.

Neural nets and Gaussian process were used to perform the modeling. These two techniques show a good predictive capacity even with small training sets. The R_{pred}^2 values exceed the inferior limit of 0.6.

The Gaussian process is widely used in physics and computer simulation; nevertheless, it is rarely explored in QSAR approaches. This work and few others [1] confirm that it may be a powerful tool in these studies.

We could also confirm that EHOMO has an important impact on the antioxidant capacity since it is taken in the QSAR models as one of the most efficient factors for both benzofurans and flavonoids. It is related to the reactivity of derivatives which is demanded in the antioxidant activity through donating hydrogen atoms to the free

radicals [2]. Hydroxylation is seen to have a limited effect; it is not sufficient to increase the antioxidant capacity. It is possibly related to the position of OH groups, the conjugation and the MW of the compounds.

Drug-likeness studies were also performed to select the drug-like molecules according to Lipinski and Veber rules, besides the calculation of the lipophilicity indices.

Multivariate techniques are very helpful to simplify the interpretation of data, particularly for large sets. They offer in laps of time wide information about their distribution. In our case, the flavonoids were perfectly clustered into four groups. That may help to establish QSAR studies on distinguished clusters as in the use of Gaussian process since this last is based on similarity between samples.

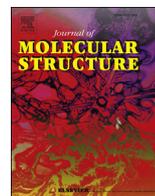
To improve the quality of QSAR models, we can suggest as perspectives for this work:

- Enlarge the data sets.
- Change the ANN and GP parameters.
- Apply separately the models on clusters and compare the results with the whole sets.
- Introduce others descriptors such as charges to improve the quality of correlations between the activity (output) and the different descriptors (inputs).

References

[1] O. Obrezanova, J.M.R. Gola, E.J. Champness, M.D. Segall, *Comput Aided Mol Des.* 22 (2008) 431–440.

[2] C. Viglianisi, S. Menichetti, *Antioxidants.* 8 (2019) 1-17.



QSAR Modeling and Drug-Likeness Screening for Antioxidant Activity of Benzofuran Derivatives

S. Boudergua^a, M. Alloui^{a,b}, S. Belaidi^{a,**}, M. Mogren Al Mogren^{c,***},
U. A. Abd Ellatif Ibrahim^c, M. Hochlaf^{b,*}

^a University of Biskra, Faculty of Sciences, Department of Chemistry, Group of Computational and Pharmaceutical Chemistry, LMCE Laboratory, 07000, Biskra, Algeria

^b Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France

^c Chemistry Department, Faculty of Science, King Saud University, PO Box 2455, Riyadh, 11451, Saudi Arabia

ARTICLE INFO

Article history:

Received 5 February 2019

Received in revised form

20 March 2019

Accepted 1 April 2019

Available online 12 April 2019

Dr. Jean-François Halet for his 60th birthday.

Keywords:

Benzofuran derivatives

Antioxidant

QSAR

MLR

ANN

Drug-likeness

ABSTRACT

In order to explore the relationship between the antioxidant activity and structure of fifteen benzofuran derivatives, we carried out a QSAR study using multiple linear regression (MLR) and artificial neural network (ANN) methods. Six descriptors were used as input data (molar weight, surface area, octanol-water partition coefficient, hydration energy, highest occupied molecular orbital energy and lowest unoccupied molecular orbital energy). The electronic properties were derived at the B3LYP/6–31 G (d,p) level. Benchmarks on furan and benzofuran subunits and their comparison to the experiment showed that this level of theory is good enough. The output data correspond to the antioxidant activity as given by IC₅₀. The predicted properties are in agreement with experimental values. Our study shows that 80% of studied molecules are in accordance with the Lipinski and Veber rules and reach the optimal lipophilicity indices. In addition, statistical analysis reveals that ANN technique with (6-2-1) architecture is more significant than MLR model.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Damage to cells caused by free radicals is believed to play a central role in disease progression [1] including cancer [2], cardiovascular [3] and Alzheimer diseases [4]. The need for antioxidants becomes even more critical with increased exposure to free radicals caused by pollution, cigarette smoke, drugs, illness and stress [1]. Scientists commonly agree that a combination of antioxidants, rather than unique entities, may be more efficient in the end. Antioxidants can be very helpful in improving the quality of life, for instance, by preventing or delaying the onset of degenerative diseases [5]. They are able to stabilize or disable free radicals before they attack the cells. In addition, they are essential for maintaining optimal cellular and system health and well-being [1].

In this context, oxygen heterocycles exhibit diverse biological and pharmacological activities due in part to the similarities with many natural and synthetic molecules with known biological activity. Among these compounds, furan [6] and benzofuran derivatives show efficient antioxidant activity. These derivatives exhibit also high potentialities for use as pharmacological agents. Indeed, these ring systems have emerged as powerful scaffolds for many biological evaluations and play an important role in the design and discovery of new physiological/pharmacologically active molecules. They can arrange to yield potent and selective drugs. We refer to the recent review by Khanam and Shamsuzzaman [7] for a wide presentation of the progress in the numerous pharmacological activities of benzofuran derivatives and of their applications in medicine.

The present theoretical contribution concerns the study of the relationship between antioxidant activity and the structure of fifteen benzofuran derivatives as recently determined by Hsieh et al. [6]. The structures of these compounds are specified in Table 1 as constructed using the Molinspiration Database [8]. Their corresponding antioxidant activity using 1, 1'-diphenyl-2-picrylhydrazyl

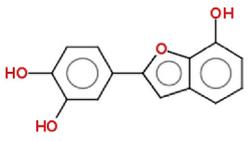
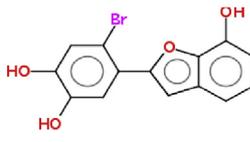
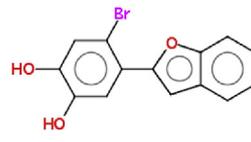
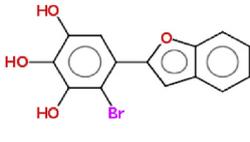
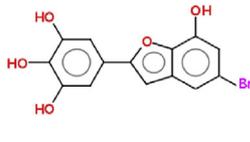
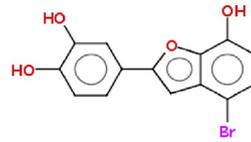
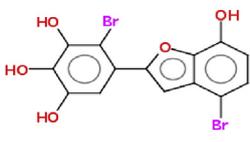
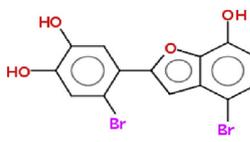
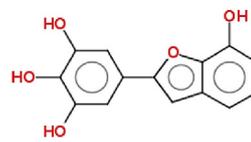
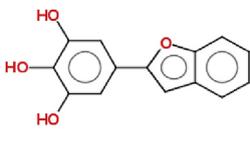
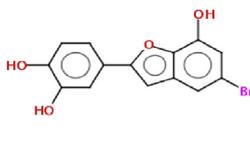
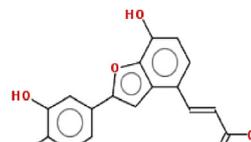
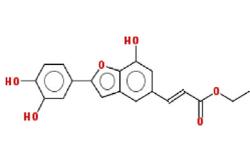
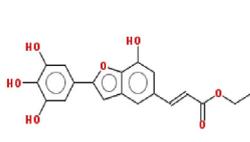
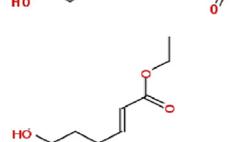
* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: prof.belaidi@gmail.com (S. Belaidi), mmogren@ksu.edu.sa (M.M. Al Mogren), hochlaf@univ-mlv.fr (M. Hochlaf).

Table 1
Chemical structures and experimental activity of the molecules under study [6]. N is the number of the compound as used presently.

N	Structure	IC ₅₀	N	Structure	IC ₅₀	N	Structure	IC ₅₀
1		18.5	2		20.6	3		33.8
4		18.0	5		14.6	6		18.4
7		19.4	8		7.8	9		12.8
10		10.8	11		28.6	12		16.8
13		25.4	14		11.7	15		26.2

(DPPH) is compiled from the literature (Table 1). All these molecules show a free radical scavenging activity with IC₅₀ values ranged between 7.8 ± 1.6 and 33.8 ± 2.8 μ M [6].

Many different chemometric methods, such as multiple linear regression (MLR) [9–11], partial least squares regression (PLS) [12], different types of artificial neural networks (ANN) [13–16], genetic algorithms (GA) [17], and support vector machine (SVM) can be employed to deduce correlation models between the molecular structure and properties [9]. At present, we derive a quantitative structure–activity relationship (QSAR) model using multiple linear regression (MLR) as well as artificial neural network (ANN) methods for the series of benzofuran derivatives of Table 1. Afterwards, Lipinski and Veber rules, and lipophilicity indices are applied to identify “drug-like” compounds. For electronic structure computations, we used Becke’s three-parameter Lee–Yang–Parr (B3LYP) [18] density functional theory (DFT) in conjunction with the 6-31G (d,p) basis set. This level of theory is viewed to be accurate enough after benchmarks on the structure and the vibrational spectroscopy of furan and benzofuran subunits using post Hartree–Fock (MP2, Coupled Clusters) methods and their comparison to experimental (IR, μ w) data.

2. Benchmarks on furan and benzofuran

2.1. Equilibrium geometries of furan and of benzofuran

The most efficient theoretical approach to be used for the larger benzofuran derivatives of interest in the present study may be selected by comparison with experimental results. We started

hence our investigations by optimizing the furan and benzofuran equilibrium structures in order to select the most reliable predictive method comparatively to experiment and with reduced computational cost. Indeed, a compromise between accuracy and computational cost (both CPU and disk occupancy) should be found in order to be able to compute the properties of our series of the benzofuran derivatives.

Table 2 gives the equilibrium structures of furan and of benzofuran as computed at the B3LYP/6-31G (d,p), B3LYP/cc-pVDZ, MP2/6-31G (d,p), MP2/cc-pVDZ, MP2/aug-cc-pVTZ and CCSD/6-31G (d,p) and CCSD/cc-pVDZ levels of theory. These computations are carried out using the GAUSSIAN 09 program package [21]. Table 2 lists also the corresponding experimental geometrical parameters as measured using microwave spectroscopy [19,20]. Since furan and benzofuran are planar, the calculated dihedral angle values are either 0° or 180°.

Close inspection of Table 2 shows that CCSD method in conjunction with either the 6-31G (d,p) or the cc-pVDZ basis sets performs quite well since the differences between the computed and the measured equilibrium distances amount only to few thousandths of Å. The in-plane angles are also close to the experimental ones (within less than 1°). As established in the literature, CCSD approach can be used as a reference method to compare with. Interestingly, the less computationally demanding approaches i.e. B3LYP/6-31G (d,p), B3LYP/cc-pVDZ, MP2/6-31G (d,p), MP2/cc-pVDZ, MP2/aug-cc-pVTZ, perform also quite well compared to CCSD and to experiments since they lead to equilibrium structural parameters which differ by less than few hundredths of Å for the distances and less than 1° for the angles. Especially, a good

Table 2

Bond lengths (in Å) and valence angles (in degree, °) of furan and of benzofuran. See Fig. 1 for the numbering of the atoms. Experimental data for furan and for benzofuran (See NOTE 1) are from Ref. [19] and Ref. [20], respectively.

Furan								
Parameters	Exp.	B3LYP		MP2			CCSD	
		6-31G (d,p)	cc-pVDZ	6-31G (d,p)	cc-pVDZ	aug-cc-pVTZ	6-31G (d,p)	cc-pVDZ
O1–C2	1.3621	1.364	1.363	1.366	1.363	1.360	1.367	1.365
C2–C3	1.3609	1.360	1.364	1.366	1.378	1.365	1.358	1.369
C3–C4	1.4309	1.436	1.438	1.427	1.436	1.426	1.440	1.450
C2–H6	1.0750	1.079	1.086	1.075	1.088	1.075	1.075	1.089
C3–H7	1.0768	1.080	1.087	1.076	1.089	1.076	1.076	1.090
O1–C2–C3	110.68	110.5	110.6	110.5	110.8	110.4	110.8	111.1
C2–C3–C4	106.50	106.1	106.0	106.2	105.8	106.1	106.0	105.6
C2–O1–C5	106.55	106.8	106.8	106.6	106.8	106.9	106.3	106.5
O1–C2–H6	115.92	115.8	115.7	115.7	115.8	115.8	115.7	115.8
C2–C3–H7	127.95	126.5	126.5	126.2	126.3	126.1	126.5	126.6
Benzofuran								
O1–C2	–	1.370	1.370	1.372	1.369	1.365	1.372	1.370
C2–C3	1.409	1.408	1.410	1.407	1.417	1.405	1.400	1.409
C3–C4	1.470	1.444	1.446	1.438	1.446	1.435	1.450	1.458
C4–C5	1.350	1.355	1.359	1.361	1.372	1.360	1.353	1.364
O1–C5	1.362	1.374	1.374	1.375	1.373	1.369	1.377	1.376
C2–C9	1.452	1.390	1.392	1.393	1.401	1.389	1.393	1.402
C3–C6	1.407	1.403	1.406	1.405	1.413	1.402	1.404	1.413
C6–C7	1.397	1.392	1.349	1.390	1.340	1.388	1.390	1.399
C7–C8	1.409	1.408	1.410	1.410	1.419	1.407	1.410	1.419
C8–C9	1.396	1.393	1.396	1.392	1.401	1.390	1.391	1.400
C2–O1–C5	106.7	106.0	105.9	106.0	105.6	105.9	105.8	105.5
O1–C2–C3	110.4	110.3	110.4	110.5	110.8	110.5	110.7	110.9
C2–C3–C4	105.2	105.3	105.3	105.9	105.2	105.4	105.3	105.1
C3–C4–C5	105.4	106.0	106.0	105.6	105.8	105.9	105.6	105.8
O1–C5–C4	112.3	112.4	112.4	112.4	112.6	112.3	112.5	112.8
C2–C3–C6	119.8	118.7	118.9	118.8	119.0	118.8	119.0	119.2
C3–C6–C7	117.6	118.4	118.4	118.3	118.3	118.2	118.3	118.3
C6–C7–C8	121.3	121.4	121.4	121.5	121.4	121.5	121.3	121.2
C7–C8–C9	121.4	121.3	121.2	121.4	121.3	121.4	121.4	121.3

agreement can be observed between the predicted B3LYP/6-31G (d,p) and experimental data. From that, we can conclude that the DFT/B3LYP method is appropriate enough for studies of benzofuran derivatives.

2.2. Vibrational analysis

Accurate calculations of anharmonic frequencies of molecular systems allow correct identification of molecules through the assignment of their experimental RAMAN, IR or Far-IR spectra. Such computations are also needed to establish the reliability of the derivatives of the potentials, i.e. force fields, in semi empirical approaches. Obviously, this is crucial for accurate predictions of thermal contributions to enthalpies and entropies and of the reactive properties of these compounds using such methods. Here, we used second order perturbation theory (VPT2) approach as implemented in GAUSSIAN and MOLPRO [22] to evaluate the anharmonic vibrational frequencies of isolated furan and benzofuran. These data were calculated by means of B3LYP/6-31 G (d,p). The results are listed in Table 3, where they are compared to the respective FT-IR experimental data.

The benzofuran is a planar asymmetric rotor with C_s symmetry. The vibrations separate into 27 a' (in-plane) and 12 a'' (out-of-plane) fundamentals [23]. Furan is also a planar, near oblate symmetric top rotor. It possesses 8 a_1 , 3 a_2 , 7 b_1 , and 3 b_2 vibrations [24] (specified in Table 3). In brief, the comparison of the B3LYP/6-31G (d,p) computed and measured fundamentals for furan and benzofuran shows differences of less than 10 cm^{-1} between both sets of data. Again, this confirms the good performance of B3LYP/6-31G (d,p) chosen theoretical level for the derivation of properties of larger benzofuran derivatives.

2.3. 3D molecular electrostatic potential surface maps (3D MESP) of furan and of benzofuran

Fig. 2 shows the 3D molecular electrostatic potential surface maps (3D MESP) of furan and of benzofuran. MESP allows understanding various physical and chemical phenomena such as molecular reactivity behavior, intermolecular interactions, molecular recognition, electrophilic reactions, substituent effects and the interactions induced by the reagents, for example between a drug and its cellular receptor [11].

As can be seen in Fig. 2, furan and benzofuran present negative electrostatic potentials (red zone) around the oxygen atom due to its high electronegativity. An orange region can be observed around the carbon atoms of the five-ring of furan and the benzene ring of benzofuran. Thus, these parts may be subject to electrophilic attacks. We can see also positive electrostatic potentials (blue zone) around the atoms of hydrogens, which explain that these atomic sites are susceptible for nucleophilic attacks. In sum, furan and benzofuran exhibit common characteristics, which may be helpful for a qualitative understanding of the electrostatic interactions that may take place between reagents or enzyme active sites and the benzofuran derivatives under study.

3. Quantitative structure activity relationships (QSAR) study

When chemical or physical properties and molecular structures are derived from numbers, it is often possible to propose mathematical relations connecting them, which allow making quantitative predictions. The obtained mathematical expressions can then be used as a predictive means of the biological response for similar structures. They are widely used in the pharmaceutical industry to

Table 3
Anharmonic vibrational frequencies (in cm^{-1}) of furan and of benzofuran as computed using B3LYP/6–31 G (d,p). We give also the experimental data.

furan				Benzofuran			
Mode N°	Sym.	B3LYP/6–31 (d,p)	Exp. [25].	Mode N°	Sym.	B3LYP/6–31 (d,p)	Exp [26].
1	a ₁	3165	3169	1	a'	3164.6	3158.0
2	a ₁	3143	3140	2	a'	3126.4	3124.5
3	a ₁	1494	1491	3	a'	3085.2	3094.0
4	a ₁	1392	1385	4	a'	3056.9	3077.0
5	a ₁	1150	1140	5	a'	3064.6	3067.0
6	a ₁	1075	1067	6	a'	3028.6	3047.0
7	a ₁	1005	995	7	a'	1631.2	1617.0
8	a ₁	874	870	8	a'	1601.4	1595.0
9	a ₂	865	864	9	a'	1558.1	1542.8
10	a ₂	713	722	10	a'	1487.1	1478.5
11	a ₂	606	600	11	a'	1461.0	1456.8
12	b ₁	3156	3161	12	a'	1373.0	1346.2
13	b ₁	3133	3130	13	a'	1343.5	1329.0
14	b ₁	1577	1558	14	a'	1274.2	1264.2
15	b ₁	1270	1267	15	a'	1259.6	1252.9
16	b ₁	1181	1181	16	a'	1181.6	1179.9
17	b ₁	1048	1043	17	a'	1164.0	1161.0
18	b ₁	878	873	18	a'	1138.4	1130.8
19	b ₂	827	838	19	a'	1112.4	1107.0
20	b ₂	747	745	20	a'	1044.0	1035.8
21	b ₂	612	603	21	a'	1018.6	1007.0
				22	a'	901.7	900.1
				23	a'	855.4	855.2
				24	a'	770.5	767.6
				25	a'	613.6	611.6
				26	a'	543.7	538.7
				27	a'	404.3	400.3
				28	a''	965.9	970.0
				29	a''	927.6	929.0
				30	a''	861.4	861.1
				31	a''	848.0	848.0
				32	a''	770.4	763.0
				33	a''	751.6	745.6
				34	a''	737.2	731.5
				35	a''	589.6	584.0
				36	a''	575.6	569.0
				37	a''	425.2	417.9
				38	a''	251.3	246.3
				39	a''	217.6	211.2

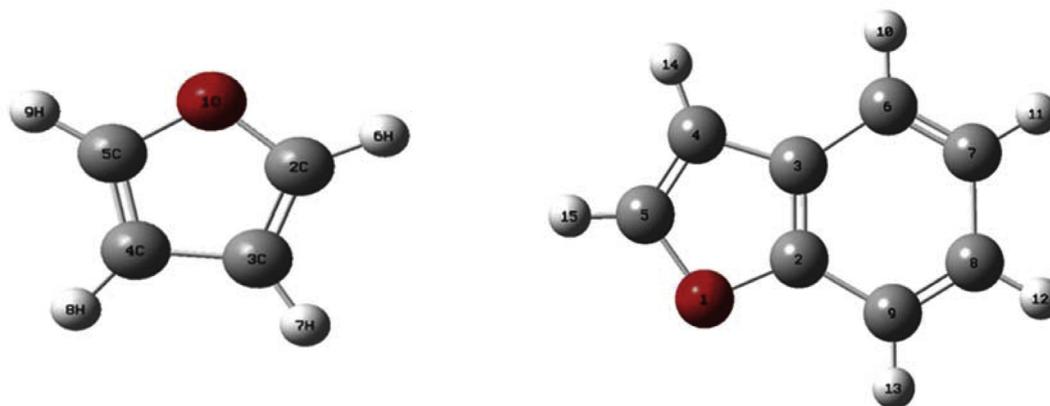


Fig. 1. Structures of furan (left) and of benzofuran (right). We give also the atom numbering.

identify promising compounds, especially at early stages of drug discovery [27].

Relationships between the physicochemical properties of chemical substances and their biological activities can be derived using QSAR (Quantitative Structure-Activity Relationships) concept. These models can also be used to predict the activities of new chemical entities and for their design [28]. Therefore, the

biological activity is quantitatively expressed as the concentration of substance necessary to obtain a certain biological response. For that purpose, multiple linear regression, MLR, and artificial neural networks (ANNs) are used. The accuracy of such models is mainly evaluated by the correlation coefficient R^2 [9] and the average relative error ARE [29] (see NOTE 2). The MLR and ANN models were generated using the software JMP 8.0.2 [30].

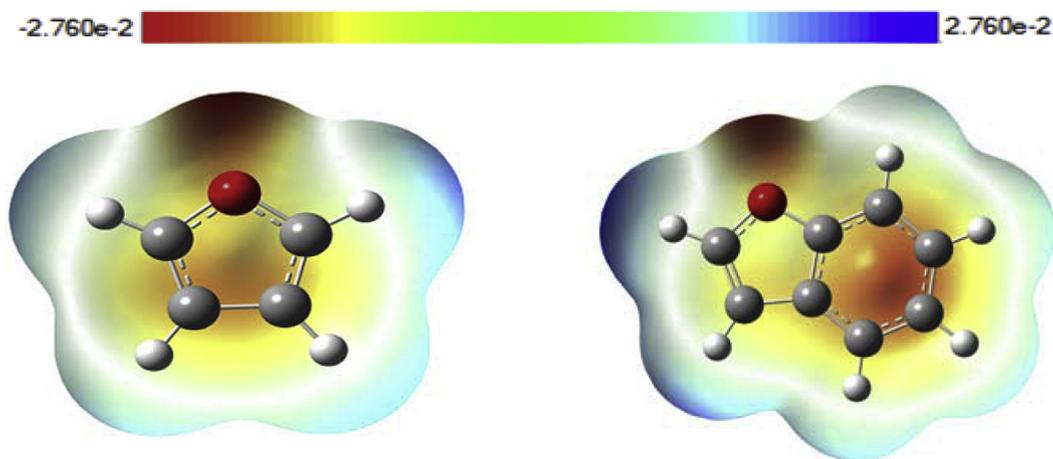


Fig. 2. 3D MESP for furan (left) and for benzofuran (right). The results are shown by color, from red (most negative) to blue (most positive).

The equilibrium geometries and the highest occupied molecular orbital energy (E_{HOMO}) and lowest unoccupied molecular orbital energy (E_{LUMO}) of benzofuran derivatives were determined at the B3LYP/6-31G (d,p) level of theory. We list in Table S1 of the supplementary material the Cartesian coordinates of the optimized benzofuran derivatives equilibrium structures. Then, the QSAR properties module from Hyper Chem 8.08 [33] was used to calculate: molar weight (MW), surface area (S), volume (V), refractivity (R), polarizability (P), octanol-water partition coefficient (log P) and hydration energy (HE). Finally, MarvinSketch17.1.2 Software [34] was employed to compute: hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), number of rotatable bonds (NRB) and polar surface area (PSA).

3.1. Multiple linear regression (MLR)

MLR is one of the earliest and still one of the most commonly used methods for constructing QSAR mathematical models [31] because of its simplicity, transparency, reproducibility, and easy interpretability [32]. At present, QSAR was carried out on a series of fifteen furan derivatives (Table 1). In this work, nine descriptors were chosen to describe the structure of the benzofuran derivatives. Table 4 shows the values of the calculated descriptors used in order to establish the quantitative structure - antioxidant activity. The data set was randomly divided into two sets: a training

set (twelve compounds) and a testing set (three compounds: 6, 12 and 15) at a ratio of 80:20. Correlation matrix between parameters was performed on all nine descriptors. Nevertheless, the analysis revealed six independent descriptors for the development of the model.

The resulting MLR QSAR model is represented by the following equation:

$$\text{IC}_{50} = -212.620 - 0.785 \text{ MW} + 0.397 \text{ S} - 44.368 \log \text{ P} + 9.274 \text{ HE} - 1815.594 E_{\text{HOMO}} - 2746.546 E_{\text{LUMO}} \quad (1)$$

where, IC_{50} is the response or dependent variable. MW, S, log P, HE, E_{HOMO} and E_{LUMO} are descriptors (features or independent variables). Within the regression, the coefficients in front of these descriptors are optimized [32].

For validation of the model, we plot in Fig. 3 the experimental activities against the predicted values as determined by equation (1). We can observe that the predicted IC_{50} values are in an acceptable agreement and regular distribution with experimental ones with $R^2 = 0.64$ and ARE of 0.74.

The positive values of S and HE coefficients indicate their promoter effect on the response, whereas the other factors (MW, log P, E_{HOMO} and E_{LUMO}) exhibit an opposite effect on IC_{50} . QSAR models aim also to define the importance of factor's effects. In our case, the

Table 4

Values of molecular descriptors used in the QSAR study. We give the molar weight (MW, amu), surface area (S, \AA^2), volume (V, \AA^3), refractivity (R, \AA^3), polarizability (P, \AA^3), octanol-water partition coefficient (log P), hydration energy (HE, kcal/mol), energies of the HOMO (E_{HOMO} , eV) and LUMO (E_{LUMO} , eV) and IC_{50} (in μM) (as given in Table 1).

Compound	MW	S	V	R	P	log P	HE	E_{HOMO}	E_{LUMO}	IC_{50}
1	242.23	419.57	685.41	74.61	25.35	-2.44	-23.21	-0.238	0.025	18.5
2	321.13	440.83	732.74	82.15	27.97	-2.39	-22.50	-0.240	0.008	20.6
3	305.13	433.63	711.92	80.54	27.33	-1.36	-16.04	-0.236	0.012	33.8
4	321.13	439.52	729.18	82.15	27.97	-2.39	-20.06	-0.235	0.013	18.0
5	337.13	458.90	765.50	83.75	28.61	-3.41	-27.98	-0.254	0.015	14.6
6*	321.13	447.86	741.16	82.15	27.97	-2.39	-22.68	-0.250	0.018	18.4
7	416.02	474.28	802.32	91.29	31.23	-3.36	-25.56	-0.257	0.004	19.4
8	400.02	471.11	790.38	89.68	30.60	-2.33	-22.03	-0.258	0.005	7.80
9	258.23	426.19	703.13	76.22	25.98	-3.46	-28.55	-0.237	0.023	12.8
10	242.23	418.17	682.36	74.61	25.35	-2.44	-22.56	-0.234	0.026	10.8
11	321.13	450.12	745.24	82.15	27.97	-2.39	-22.53	-0.251	0.017	28.6
12*	340.33	569.21	958.60	100.37	35.05	-2.27	-22.81	-0.237	0.017	16.8
13	340.33	582.08	965.45	100.37	35.05	-2.27	-22.70	-0.239	0.017	25.4
14	356.33	590.10	984.78	101.98	35.69	-3.29	-28.71	-0.240	0.018	11.7
15*	340.33	542.93	941.88	100.37	35.05	-2.27	-22.09	-0.241	0.016	26.2

* corresponds to test molecules.

Table 5
Drug-likeness parameters and lipophilicity indices of benzofuran derivatives. See text and NOTES 3–7 [40–43] for the definition of the parameters given here.

Compound	Lipinski rules				Veber rules		Lipophilicity indices	
	MW (amu)	logP	HBD	HBA	PSA (Å ²)	NRB	LipE	LPLP
	<500	<5	<5	<10	<140	<10	>5	-10 < LPLP < 10
1	242.23	-2.44	3	3	73.82	1	7.17	-6.63
2	321.13	-2.39	3	3	73.82	1	7.08	-6.92
3	305.13	-1.36	2	2	53.6	1	5.83	-3.91
4	321.13	-2.39	3	3	73.82	1	7.13	-6.84
5	337.13	-3.41	4	4	94.05	1	8.24	-10.07
6	321.13	-2.39	3	3	73.82	1	7.12	-6.85
7	416.02	-3.36	4	4	94.05	1	8.07	-10.70
8	400.02	-2.33	3	3	73.82	1	7.44	-6.52
9	258.23	-3.46	4	4	94.05	1	8.35	-9.60
10	242.23	-2.44	3	3	73.82	1	7.41	-6.32
11	321.13	-2.39	3	3	73.82	1	6.93	-7.14
12	340.33	-2.27	3	4	100.13	5	7.04	-8.49
13	340.33	-2.27	3	4	100.13	5	6.86	-8.82
14	356.33	-3.29	4	5	120.36	5	8.22	-12.39
15	340.33	-2.27	3	4	100.13	5	6.85	-8.85

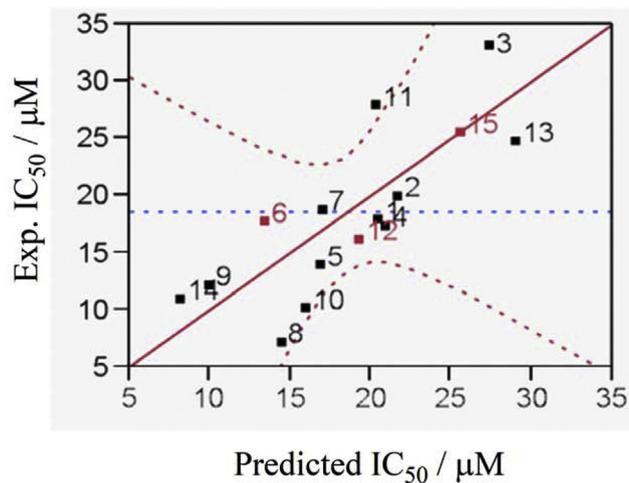


Fig. 3. Correlation of experimental and predicted IC_{50} (in μM) as derived using MLR.

surface is presented as the most influencing parameter on IC_{50} , followed by the molar weight and the hydration energy. The least important factor is log P. As the antioxidant activity is required to be improved, i.e. IC_{50} decreased, it will be recommended to increase MW and reduce S and HE.

3.2. Artificial neural networks

Artificial neural networks (ANNs) models are non-linear models useful to predict the biological activity of large data sets of molecules [15]. In contrast to classical statistical methods such as regression analysis or partial least squares analysis, ANNs enable the investigation of complex and nonlinear relationships. Neural networks are therefore ideally suited for use in drug design and QSAR [35]. They are applied for simulating various non-linear complex systems of pharmaceutical, engineering, psychology and medicinal chemistry domains [36]. For instance, ANN was successfully used for the prediction and synthesis of new organic chemical compounds [16].

In this work, ANN contained six inputs corresponding to the six descriptors selected from the correlation matrix, two hidden neurons, and one output neuron which is IC_{50} (Fig. 4). The number of artificial neurons in the hidden layer was adjusted experimentally

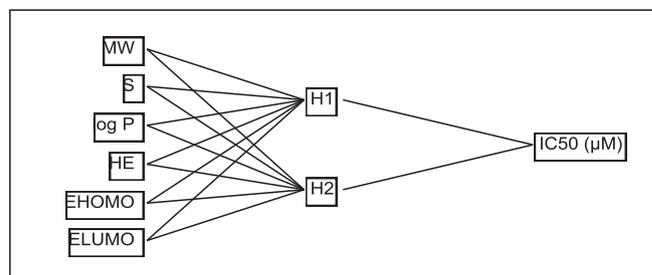


Fig. 4. Structure of ANN.

[37], two neurons in the hidden layer permitted to attain the best correlation between experimental and predicted data. Then, the ANN was trained using Gauss Newton method. A good correlation of experimental and predicted IC_{50} by ANN is found. This is shown in Fig. 5, and illustrated by R^2 and ARE values of $R^2 = 0.99$ and $ARE = 0.99$.

From both results of training and test sets (Fig. 5), we can conclude that the ANN model with (6-2-1) architecture is able to establish a satisfactory relationship between the six descriptors and the antioxidant activity. For instance, all test molecules (6, 12 and 15) are in a good agreement with the two models.

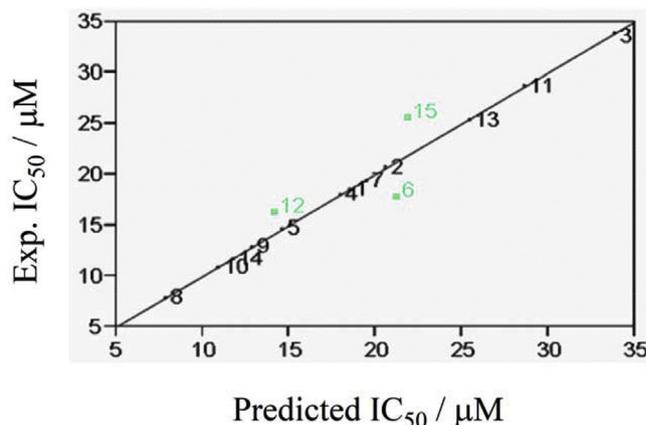


Fig. 5. Correlation of experimental and predicted IC_{50} as calculated by ANN.

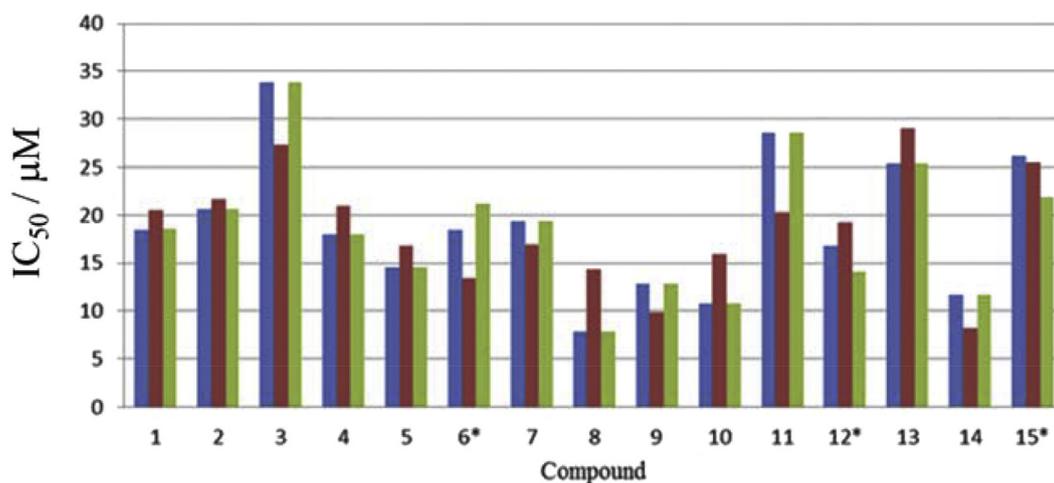


Fig. 6. Comparative bar chart between experimental (in blue) and predicted IC₅₀ (in μM) using MLR (in red) and ANN (in green) methods.

3.3. Drug-likeness screening

Drug-likeness is a qualitative concept used in drug design [38]. It is described to encode the balance among the molecular properties of a compound that influences its pharmacodynamics, pharmacokinetics and ADME (Absorption, Distribution, Metabolism and Excretion) in human body [39]. Several rules have been proposed to evaluate drug-likeness. The most commonly used are Lipinski's rule of five [40] (see NOTE 3), Veber rules [41] (see NOTE 4), and lipophilicity indices [42,43], which are viewed to be quite effective and efficient [44]. These represent guidelines and not absolute values to attest that a compound is a drug-like or not. Indeed, pharmacokinetic parameters are strongly influenced in vivo by the physicochemical properties of a drug [45]. Moreover, successful drug discovery not only requires the optimization of physicochemical parameters, but also more complex parameters related to toxicity and bioavailability defined as lipophilic efficiency indices [46]: ligand lipophilic efficiency LipE (see NOTE 5), ligand efficiency (LE see NOTE 6) and ligand-efficiency dependent lipophilicity LELP (see NOTE 7) [42]. At present, we evaluated the oral bioavailability of the fifteen benzofuran derivatives under study. Our parameters can be used to increase the overall efficiency of drugs at different stages of discovery [45].

Table 5 presents the drug-likeness parameters and lipophilicity indices of the benzofuran derivatives under investigation. We give also the criteria to satisfy the lipophilicity indices and the Lipinski and Veber rules. As can be seen there, all compounds satisfy the rules, except compounds 5, 7 and 14 which show a slight out of range LELP (<-10) [47]. This table shows also that these compounds have molecular weights of less than 500 Da. Moreover, Table 4 gives the values of log P, which is a parameter used to estimate molecular hydrophobicity. For all studied compounds, log P is less than zero, so these molecules possess high water solubility and, in principle, poor membrane permeation capacity, which may be partly remedied by the small molecular weight. Hydrogen bonds increase solubility in water and must be broken for the compound to permeate into and through the lipid bilayer membrane. Thus, an increasing number of hydrogen bonds reduces partitioning from the aqueous phase into the lipid bilayer membrane for permeation by passive diffusion [44]. We see that all the screened molecules have a number of hydrogen bonds in the appropriate interval of Lipinski. In addition, all benzofuran derivatives show PSA values in the range of 53–120 Å², which permit them to correlate very well with the human intestinal absorption, Caco-2 monolayer's

permeability, and blood-brain barrier penetration [48]. We can note also that compounds 12, 13, 14 and 15 are likely to be the most flexible with a number of rotatable bonds of 5.

In the studied set, LipE takes values between 5.83 and 8.35. Thus, this set would have a high quality as lead compounds, provided that LELP is situated in the suggested range $-10 < \text{LELP} < 10$ to guarantee that molecules reach both size and lipophilicity criteria, which is the case for all the studied compounds except three molecules (5, 7 and 14) having LELP inferior than the minimum limit ($-10.07, -10.70, -12.39$).

In general, according to the results listed in Table 5, we can estimate that all benzofuran derivatives satisfy the Lipinski and Veber rules, so they would not have problems with oral bioavailability. 80% of compounds reach an LELP in the suggested range $-10 < \text{LELP} < 10$.

4. Summary and conclusion

A quantitative analysis of the structure-antioxidant activity relationship of fifteen antioxidant benzofuran derivatives is performed. QSAR models were obtained using multiple linear regression (MLR) and artificial neural network (ANN) techniques. Our work shows that this series of compounds obey the Lipinski's rule of five. These molecules satisfy also Veber's rules with NRB < 10 and PSA < 140. According to the calculated LELP values, these derivatives, expect compounds 5, 7 and 14, could have promising results in the clinic tests.

From a mythological point of view, we used first principles approaches to treat the furan and benzofuran subunits. The comparison of the spectroscopic parameters of these entities with experiment showed that the B3LYP/6-31 G (d,p) approach is sufficiently reliable. In addition, its low computational cost allows for investigating larger molecular systems. Afterwards, we showed that both ANN and MLR methods provide similar QSAR model accuracy. For instance, Fig. 6 reveals that MLR and ANN predict almost coinciding IC₅₀s, whereas some systematic deviations can be observed with the measured IC₅₀s. In sum, to improve the QSAR models, research needs to be conducted with more activity data of similar compounds and molecular parameters [49].

Acknowledgements

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for

funding the research through the Research Group Project No. RGP-333.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molstruc.2019.04.004>.

5. Notes

NOTE 1: The experimental values of bond length and angles are estimated from data concerning the 5-membered rings taken from microwave studies of thiophene and furan, and data for the 6-membered ring taken from a crystal study of 3-indolylacetic acid [20].

NOTE 2: The MLR and ANN models were generated using the software JMP 8.0.2 [30]. The accuracy of models is mainly evaluated by the correlation coefficient R^2 and the average relative error ARE. The relations of R^2 [9] and ARE [29] are given

by the following equations: $R^2 = 1 - \frac{\sum_{m=1}^N (\hat{y}_m - y_m)^2}{\sum_{m=1}^N (y_m - \bar{y})^2}$ and

$ARE = 1 - \frac{\sum_{m=1}^N \frac{|\hat{y}_m - y_m|}{y_m}}{N}$, where: y_m is the desired output, \hat{y}_m is the predicted value by model, \bar{y} is the mean of dependent variable, and N is the number of the molecules in the data set.

NOTE 3: According to Lipinski's rule [40], a good absorption is achieved if: (i) molecular weight is under 500 Da; (ii) log P is under 5; (iii) there are less than 5 H-bond donors; (iv) there are less than 10 H-bond acceptors.

NOTE 4: Veber [41] found that reduced molecular flexibility, as measured by the number of rotatable bonds (NRB), and low polar surface area (PSA) or total hydrogen bond count (sum of donors and acceptors) are important predictors of good oral bioavailability, independent of molecular weight. He suggested that a high probability of good oral bioavailability is obtained if: (i) the polar surface area is equal to or less than 140 \AA^2 ; (ii) there are 10 or fewer rotatable bonds.

NOTE 5: $\text{LipE} = \text{pIC}_{50} - \log P$. It combines both potency and lipophilicity. It gives an estimation on the binding of a ligand to a given target. It is suggested to target a LipE in a range of 5–7 or even higher [43]. There is a greater likelihood of achieving good in vivo performance when potency can be increased without increasing logP [42].

NOTE 6: LE is defined as $LE = 1.4 \times \text{pIC}_{50}/N_H$, where N_H is the number of heavy atoms. It gives a measure of the properties of molecules, such as size and lipophilicity, needed to increase their binding affinity to a drug target [42].

NOTE 7: $\text{LELP} = \log P/\text{LE}$. The optimal LELP scores are $-10 < \text{LELP} < 10$ [47].

References

- [1] M. Percival, *Clin. Nut. Insights*. 1 (1998) 1.
- [2] B. Poljsak, U. Glavan, R. Dahmane, *Int. J. Cancer Res. Prev.* 4 (2011) 193.
- [3] T. Bahorun, M.A. Soobrattee, V.L. Ramma, O.I. Arouma, *Internet J. Med. Update* 1 (2006) 25.
- [4] Z. Chen, C. Zhong, *Neurosci. Bull.* 30 (2014) 271.
- [5] Md N. Alam, N.J. Bristi, M. Rafiquzzaman, *Saudi Pharm. J.* 21 (2013) 144.
- [6] J.F. Hsieh, W.J. Lin, K.F. Huang, J.H. Liao, M.J. Don, C.C. Shen, Y.J. Shiao, W.T. Li, *Eur. J. Med. Chem.* 93 (2015) 446.
- [7] H. Khanam, Shamsuzzaman. Bioactive Benzofuran derivatives: a review, *Eur. J. Med. Chem.* 97 (2015) 483–504.
- [8] Database. <http://www.molinspiration.com>.
- [9] R. Darnag, B. Minaoui, M. Fakir, *Arab. J. Chem.* 10 (2017) 600–608.
- [10] K. Dermeche, N. Tchouar, S. Belaidi, T. Salah, J. Bionanoscience 9 (2015) 4.
- [11] M. Alloui, S. Belaidi, H. Othmani, N.E. Jaidane, M. Hochlaf, *Chem. Phys. Lett.* 696 (2018) 70–78.
- [12] P. Xuan, Y. Zhang, T.J. Tzeng, X.F. Wan, F. Luo, *Glycobiology* 22 (2012) 554.
- [13] S. Kothiwale, C. Borza, A. Pozzi, J. Meiler, *Molecules* 22 (2017) 6.
- [14] Z. Hajimahi, A. Ranjbar, A.A. Suratgar, A. Zarghi, *Iran, J. Pharm. Res.* 14 (2014) 73.
- [15] J.M. Díaz-González, F.J. Aguirre-Crespo, X. García-Mera, F.J. Prado-Prado, *MOL2NET*. 1 (2015) 2.
- [16] M. Lazar, A. Mouzdahir, M. Badia, M. Zahouily, *J. Pharm. Res.* 8 (2014) 119.
- [17] E. Pourbasheer, S. Vahdani, D. Malekzadeh, R. Aalizadeh, A. Ebadi, *Iran, J. Pharm. Res.* 16 (2017) 970.
- [18] C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* 37 (1998) 785–789.
- [19] B. Bak, D. Christensen, W.B. Dixon, L. Hansen-Nygaard, J. Rastrup-Andersen, M. Schottlander, *J. Mol. Spectrosc.* 9 (1962) 124.
- [20] A. Hartford, A.R. Muirhead, J.R. Lombardi, *J. Mol. Spectrosc.* 35 (1970) 199–213.
- [21] M.J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, H. H. Nakatsuji, Gaussian 09, Revision E. 01, 2009, Gaussian, Inc., Wallingford CT.
- [22] H.-J. Werner, P.J. Knowles, et al., MOLPRO, Ab Initio Programs Package, 2015. <http://www.molpro.net>.
- [23] T.D. Klots, W.B. Collier, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 51 (1995) 1291–1316.
- [24] T.D. Klots, R.D. Chirico, W.V. Steele, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 50 (1994) 765–795.
- [25] A. Mellouki, J. Liévin, M. Herman, *Chem. Phys.* 271 (2001) 239.
- [26] W.B. Collier, T.D. Klots, *Spectrochim. Acta, Part A* 51 (1995) 1255.
- [27] F. Soualmia, S. Belaidi, H. Belaidi, N. Tchouar, Z. Almi, *J. Bionanoscience* 11 (2017) 584.
- [28] B. Jhanwarb, V. Sharmaa, R.K. Singla, B. Shrivastava, *Pharmacologyonline* 1 (2011) 306.
- [29] Y. Zhang, H. Wang, Z. Yang, J. Li, *PLoS One* 9 (2014) 4.
- [30] JMP 8.0.2, SAS Institute Inc., 2009.
- [31] P. Liu, W. Long, *Int. J. Mol. Sci.* 10 (2009) 1979.
- [32] K. Roy, S. Kar, R.N. Das, *A Primer on QSAR/QSPR Modeling*, Springer Briefs in Molecular Science, USA, 2015.
- [33] Hyper Chem (Molecular Modeling System) Hypercube, Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA, 2008.
- [34] MarvinSketch17.1.2, Chemaxon, 2017. <http://www.chemaxon.com>.
- [35] L. Terfloth, J. Gasteiger, *DDT*. 6 (2001) S102.
- [36] F. Cheng, V. Sutariya, *Clin. Exp. Pharmacol.* 2 (2012) 1.
- [37] M. Larif, A. Adad, R. Hmamouchi, A.I. Taghki, A. Soulaymani, A. Elmidaoui, M. Bouachrine, T. Lakhlifi, *Arab. J. Chem.* 10 (2017) 948.
- [38] M. Joydeep, C. Raja, S. Saikat, V. Sanjay, D. Biplab, T.K. Ravi, *Der. Pharma. Chemic.* 1 (2009) 188–198.
- [39] G. Vistoli, A. Pedretti, B. Testa, *Drug Discov. Today* 13 (2008) 285–294.
- [40] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug Deliv. Rev.* 23 (1997) 3.
- [41] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, *J. Med. Chem.* 45 (2002) 2615–2623.
- [42] A.L. Hopkins, G.M. Keserü, P.D. Leeson, D.C. Rees, C.H. Reynolds, *Nat. Rev. Drug Discov.* 13 (2014) 105–121.
- [43] P.D. Leeson, B. Springthorpe, *Nat. Rev. Drug Discov.* 6 (2007) 881–890.
- [44] E.H. Kerns, L. Di, *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*, Academic Press, Elsevier, USA, 2008.
- [45] A. Zerroug, S. Belaidi, I. Ben Brahim, L. Sinha, S. Chtita, *J. King Saud Univ. Sci.* (2018) in press, <https://doi.org/10.1016/j.jksus.2018.03.024>.
- [46] J.A. Arnott, R. Kumar, S.L. Planey, *J. Appl. Biopharm. Pharmacokinet.* 1 (2013) 31–36.
- [47] G.M. Keserü, G.M. Makara, *Nat. Rev. Drug Discov.* 8 (2009) 203–212.
- [48] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* 43 (2000) 3714–3717.
- [49] J. Luo, J. Hu, L. Fu, C. Liu, X. Jin, *Procedia Eng* 15 (2011) 5162–5163.

Abstract

ملخص: اكتشاف الأدوية عملية تتطلب وقتا و إمكانيات ضخمة من أجل البحث و التطوير. غير أن دراسة العلاقة الكمية بين خصائص المركبات الكيميائية و نشاطها بالإضافة إلى قواعد قابلية الاستعمال الدوائي تساهم بفعالية في اقتراح و اكتشاف مركبات فعالة جديدة.

تعتبر مضادات الأكسدة حاليا من أكثر المواد المدروسة و المستعملة في الصناعات الدوائية و الغذائية نظرا لفعاليتها ضد تأثيرات الشيوخوخة، حيث تثبط الجذور الحرة المسببة للإجهاد التأكسدي.

في هذا الإطار، قمنا في هذا العمل بدراسة العلاقة الكمية بين الخصائص الكيميائية و النشاط المضاد للأكسدة لنوعين من المركبات هما benzofurans و flavonoids. تمت الدراسة بطرق الشبكات العصبية و Gaussian process القليلة الاستعمال في هذا المجال. معامل التنبؤ كان مقبولا و فاق قيمة 0.6.

قابلية الاستعمال الدوائي القائمة على قاعد فابر و ليبينسكي بالإضافة الى معايير الدسامة سمحت بتعيين المركبات الخاضعة لخصائص الأدوية.

الكلمات الجوهرية : مضادات الأكسدة، benzofurans، flavonoids، العلاقة الكمية بين الخصائص و النشاط، الشبكات العصبية، Gaussian process، قابلية الاستعمال الدوائي.

Abstract: Drug discovery takes many years and requires big budgets for research and development. QSAR in addition to drug likeness studies contribute strongly to predict and discover new active molecules.

Nowadays, the antioxidants are among the most studied and used molecules in drugs and food industries because of their anti-ageing effects. They scavenge free radicals causing oxidative stress.

In this work, we perform a QSAR modeling of the antioxidant activity for two sets of benzofurans and flavonoids by artificial neural networks and Gaussian process seldom used in this approach. Their predictability coefficient was acceptable with a value that exceeds 0.6.

Drug likeness studies based on Lipinski and Veber rules, besides the lipophilicity indices permitted to define the drug like molecules.

Keywords: Antioxidants, Benzofurans, Flavonoids, QSAR, ANN, Gaussian process, Drug likeness.

Résumé: La découverte de nouveaux médicaments demande des années et des grands budgets pour la recherche et le développement. Néanmoins, QSAR et les études de ressemblance aux médicaments contribuent fortement dans la prédiction et la découverte de nouvelles molécules actives.

Les antioxydants sont actuellement parmi les produits les plus étudiés et utilisés dans l'industrie pharmaceutique et alimentaire vue leurs effets anti-âge. Ils inhibent les radicaux libres qui initient les stress oxydatifs.

Dans ce travail, on établit des modèles QSAR pour l'activité antioxydante de deux séries de benzofuranes et de flavonoïdes par les techniques de réseaux de neurones artificiels et de processus gaussien rarement utilisé dans cette approche. Les coefficients de prédiction se trouvent acceptables avec une valeur qui dépasse 0.6.

L'étude de ressemblance aux médicaments, basée sur les règles de Lipinski et Veber ainsi que les indices de lipophilicité, a permis de définir les molécules candidates d'être des médicaments.

Mots clés: Antioxydants, Benzofuranes, Flavonoïdes, QSAR, ANN, Processus gaussien, Ressemblance aux médicaments.