



Democratic and Popular Republic of Algeria
Ministry of Higher Education and Scientific Research
University Mohamed Khider of Biskra



Faculty of Exact Sciences and Science of Nature and Life
Department of Matter Sciences

Ref :

Thesis Presented to obtain the degree of

Doctorate in Chemistry

Option: Pharmaceutical chemistry

Entitled:

**Virtual screening applied in several series of bioactive
heterocyclic molecules**

Presented by:

KHELFA Nedjla

Publicly defended on: 18 / 02 / 2024

In front of the Jury committee composed of:

Mr. DJANI Faical	Professor	University of Mohamed Khider of Biskra	President
Mr. BELAIDI Salah	Professor	University of Mohamed Khider of Biskra	Supervisor
Mr. HOCHLAF Majdi	Professor	University of Gustav Eiffel of Paris	Co-supervisor & invited
Mr. MELKMI Nadjib	Professor	University of Mohamed Khider of Biskra	Examiner
Mr. DIBI Ammar	Professor	University of Hadj Lakhder of Batna1	Examiner

Abstract

Malaria is one of the most widespread and fearsome parasitic disease in the world. *Plasmodium falciparum* is the parasite that causes the most severe type of **malaria**, this species transmits in the human body and attack the important organ; which is the human liver cells, will then reenter the blood stream and begin infecting the red blood cells. The major enzymes in this parasite is *Plasmodium falciparum dihydrofolatereductase (PfdHFR)*, it is responsible for the biosynthesis of essential amino acids and nucleotide bases. The emergence of resistance of this kind of parasite to antimalarial agents had led for several years, still represents a threat, which draws our attention for the adoption of new guidelines for the treatment of **malaria** cases of *P. falciparum* species. For this reason, in this thesis we focus light on the discovery and development of **Sym-triazine** derivatives, which have provided a class of antimalarial drugs highly effective against *PfdHFR*. In this context, it is necessary to focus on **Virtual Screening** computational approaches in the fields of target identification and lead discovery. The aim of our study was to apply this type of in silico methodology, with the aim of modelling and evaluating for screening bioactive molecules derived from Sym-triazine. Our work combined between ligand-based **VS** methods; we based on the QSARs methods Two-dimensional (**2D-QSAR(MLR/ANN)**) and a three-dimensional stereo (**3D-QSAR(PLS)**) which contains effective biological properties. The both of methods coupled with a Virtual Screening examination, by using a technique similarity search. Subsequently, we confirm the powerful and robustness of developed QSAR models by using various statistical OECD principles for the validation; internal and external validation (for training and test set), Y-randomization, and exploit of applicability domain. And structure-based **VS** methods; we concretized on the molecular docking analysis to determine the best interactions of the most active compound or the reference ligand which form stable complexes with the *PfdHFR* enzyme. The final results of our study, these different in silico methods allowed us to identify 9 new derivatives of **Sym-triazine** from both studies, show excellent inhibitory concentration activities against resistant *P. falciparum* bearing the mutant enzymes, making them good candidates for further development as antimalarial drugs.

Keywords: Malaria, *PfdHFR*, Sym-triazine, 2D-QSAR(MLR/ANN), 3D-QSAR(PLS), Virtual Screening.

ملخص

المالاريا من أكثر الأمراض الطفيلية انتشارًا ورعبًا في العالم. حيث نجد أن *Plasmodium falciparum* هو الطفيلي الذي يتسبب في جميع أشكال المالاريا الأكثر خطورة. وذلك عن طريق انتقال هذا النوع في جسم الإنسان ليهاجم العضو المهم؛ وهي خلايا كبد الإنسان على وجه التحديد، ليعود بعدها إلى مجرى الدم ويبدأ في إصابة خلايا الدم الحمراء. إن الإنزيمات الرئيسية في هذا الطفيلي هي *Plasmodium falciparum dihydrofolate reductase (PfdHFR)*، وهي مسؤولة عن التصنيع الحيوي للأحماض الأمينية الأساسية وقواعد النيوكليوتيدات. وعليه فظهور مقاومة هذا النوع من الطفيليات للعوامل المضادة للمالاريا ضل لعدة سنوات، ولا يزال يمثل تهديدًا، الأمر الذي لفت الانتباه نحو اعتماد مبادئ توجيهية جديدة لعلاج حالات المالاريا الخاصة بأنواع *P. falciparum*. وعليه سنسلط في هذه الأطروحة الضوء على اكتشاف وتطوير مشتقات سيم-تريازين، والتي قدمت كفاءة من الأدوية المضادة للمالاريا فعالة للغاية ضد *PfdHFR*. وفي هذا السياق، من الضروري التركيز على الأساليب الحسابية للفحص الافتراضي في مجالات تحديد الأهداف واكتشاف مركبات ذات جودة وفعالية بيولوجية عالية. بحيث يصبح من أهداف دراستنا الحالية هو السعي نحو تطبيق هذا النوع من المنهجية في *In silico*، من خلال نمذجة وتقييم الجزئيات النشطة بيولوجيًا والمشتقة من سيم تريازين. حيث عملنا أولاً على تطبيق أساليب **VS** القائمة على المركب؛ اعتمدنا فيها على طرق **QSARs** ثنائية الأبعاد (**2D-QSAR(MLR/ANN)**) وثلاثية الأبعاد (**3D-QSAR(PLS)**) والتي تحتوي على خصائص بيولوجية فعالة. بحيث تقترن كلتا الطريقتين بالفحص الافتراضي باستخدام تقنية البحث عن التشابه. وبعد ذلك، نؤكد قوة ومثانة نماذج **QSAR** المطورة باستخدام مبادئ **OECD** الإحصائية المختلفة للتحقق من صحتها؛ **internal and external validation** (للمجموعتين **test** و **training**)، **Y-randomization**، و **applicability domain**. وثانياً طبقنا أساليب **VS** القائمة على البنية؛ لقد قمنا بالتحليل **molecular docking** لتحديد أفضل التفاعلات للمركب الأكثر نشاطاً في السلسلة المدروسة أو المركب المرجعي أيهما يشكل روابط مستقرة مع إنزيم *PfdHFR*. النتائج النهائية لدراستنا، من هذه الأساليب المختلفة *In silico*، سمحت لنا بتحديد 9 مشتقات جديدة لـ سيم-تريازين من كلتا الدراستين، وأظهرت أنشطة تركيز مثبطة ممتازة ضد *Plasmodium falciparum* المقاومة التي تحمل الإنزيمات الطافرة، مما يجعلها مرشحة جيدة لمزيد من تطوير الأدوية المضادة للمالاريا.

الكلمات المفتاحية: المالاريا، *PfdHFR*، سيم-تريازين، **2D-QSAR(MLR/ANN)**، **3D-QSAR(PLS)**، الفحص الافتراضي.

Table of content

Acknowledgements	i
List of works	ii
List of Abbreviations	iv
List of Figures	vi
List of Tables	vii
INTRODUCTION	1
<i>Chapter I : Background on Malaria Diseases & Their Inhibitors</i>	6
I.1. Introduction	6
I.2. Malaria disease	7
I.2.1. Epidemiology of malaria	7
I.2.1.1. Epidemiological facts	8
I.2.2. Pathogens and vectors agents	8
I.2.2.1. The pathogen agent	8
I.2.2.2. The vector agent	10
I.2.3. The evolutionary cycle of Plasmodium	10
I.2.3.1. In human body	10
I.2.3.2. In female Anopheles	11
I.2.4. Pathophysiology	12
I.3. The PfDHFR therapeutic target	13
I.3.1 Generality on folates	13
I.3.2. Plasmodium falciparum protein dihydrofolate reductase (PfDHFR)	13
I.4. Antimalarial drugs and resistance	15
I.4.1. Pharmacological classes and mechanisms of action	15
I.4.2. Focus on sym-triazine derivatives	16
I.4.2.1. Physiological and biochemical role of amino-sym-triazine derivatives	17
I.4.2.2. Diamino-sym-triazines as antimalarial drugs	18
<i>References</i>	20
<i>Chapter II: Virtual Screening in Drug Design & Discovery.</i>	24
II.1. Introduction	24

II.2. Virtual screening	25
III.2.1. Ligand-based virtual screening	27
II.2.1.1. Similarity-Based Virtual Screening	27
II.2.1.2. Quantitative structure activity relationship	28
II.2.1.2.1. Generale methodology of a QSAR study	30
A. Biological data collection	30
B. Molecular descriptors	31
C. Development of statistical models	32
D. Validation of QSAR model	34
E. Applicability domain (AD)	36
II.2.1.2.2. 2D and 3D QSAR analysis	37
A. 2D QSAR	37
B. 3D QSAR	37
II.2.1.3. Ligand-based pharmacophore approaches	41
II.2.1.3.1. Elucidation of the pharmacophore	42
II.2.2. Structure-Based Virtuel Screening	42
II.2.2.1. Molecular Docking	43
II.2.2.1.1. Docking process	44
a. Ligand-protein docking	45
b. Protein-protein docking	45
II.2.2.1.2. Search algorithms	45
II.2.2.1.3. Score functions	46
II.2.2.1.4. Analysis of results	48
II.2.2.1.5. Different types of docking	49
<i>a- Rigid docking</i>	49
<i>b- Semi-flexible docking</i>	49
<i>c- Docking flexible</i>	49
II.2.2.2. Molecular dynamics	50
II.2.2.2.1. General principle	50
II.2.2.2.2. Key components of molecular dynamics	51
II.2.2.2.3. Issues and Limitations of Molecular Dynamics	52
References	53
Realized Works	62

<u>Chapter III : In silico-Based Identification of new anti-PfDHFR drug candidates via 1,3,5-triazine derivatives.</u>	63
III.1. Introduction	64
III.2. Materials and methods	65
III.2.1. Computational details	65
III.2.2. Dataset selection	65
III.2.3. QSAR modeling studies	67
III.2.3.1. Statistical analysis and model validation	68
III.2.3.2. Applicability domain approach	68
III.2.4. Drug likeness parameter and lipophilicity indices	69
III.3. Results and discussion	69
III.3.1. Validation method	69
III.3.1.1. Equilibrium geometries of 1,3,5-triazine	69
III.3.1.2. 3D molecular electrostatic potential surface maps (3D MESP) of 1,3,5-triazine	70
III.3.2. Quantitative structure-activity relationship studies	71
III.3.2.1. Multiple linear regression (MLR)	71
III.3.2.2. Artificial Neural Networks (ANN)	75
III.3.3. Design of Novel PfDHFR Inhibitors	77
III.3.4. Drug likeness screening of 1,3,5-triazine derivatives	78
III.3.4.1. ADME study of new designed compounds	81
References	83
<u>Chapter IV : Combined 3D-QSAR, molecular docking, ADMET, and drug likeness scoring of novel Diaminodihydrotriazines as potential antimalarial agents.</u>	88
IV.1. Introduction	89
IV.2. Methodologies	91
IV.2.1. Dataset	91
IV.2.2. 3D-QSAR, In silico pharmacokinetics, ADMET study and Molecular docking analysis	92
IV.3. Results and discussion	94
IV.3.1. Validation of the developed CoMSIA model	94
IV.3.2. Contour Plot Analysis	95

IV.3.2.1. Steric contour map	96
IV.3.2.2. Electrostatic contour map	96
IV.3.2.3. Hydrophobic contour map	97
IV.3.2.4. H-Bonds	97
IV.3.3. In silico pharmacokinetic/ADMET	97
IV.3.4.1. Toxicity	99
IV.3.4.2. Drug-Likeness screening using SWISS/ADME	100
IV.3.4. Docking analysis result	100
IV.3.5. Design of new inhibitors	101
<i>References</i>	103
GENERAL CONCLUSION & PERSPECTIVES	105
APPENDIX	107

To my dear parents

To my dear husband and my daughter

To my sisters

To my brothers

Acknowledgments

"First of all, many thanks to "Allah", the most beneficent, the most merciful".

First of all, I would like to thank my thesis supervisor, Prof. **Salah BELAIDI**, head of the computing and pharmaceutical team at the Laboratory of Molecular Chemistry and Environment (LCME) at the Mohamed Khider University of Biskra, where this work was carried out, for having supervised me in the best possible way over the last four years, I would like to thank you from the bottom of my heart for allowing me to put the beginning of my thesis on hold when it proved necessary.

In particular, I would like to thank Prof. **Djani Faical** for giving me the honor of being the president of the members of my jury, with the examiners Prof. **Melkmi Nadjib**, Prof. **Dibi Ammar** and Prof. **Hoclaf Majdi** for agreeing to read my manuscript and evaluate my work, I would also like to thank them for taking the time to come to Biskra to listen to my presentation.

I would like to express my sincere thanks to Prof. **Majdi Hochlaf** as co-supervisor at the University of Gustav Eiffel in France for your support and your information has been invaluable during the years I have spent in your company. And many thanks to Prof. **Samir Chtita** at the University of Moulay Ismail in Morocco for your scientific help, advice, encouragement, patience and availability.

My sincere thanks also go to Dr. **Soualmia Fatima** at the University of Oran USTO for your scientific help and for all the discussions you had about computational methods. Thank you for your frankness and humour.

I would also like to thank the entire **Bioinformatics PhD team** at Moulay Ismail University, for the quality of the training offered, without which I would certainly not have undertaken to do a thesis, and for the good collaboration of the whole team.

I am very grateful to my parents, my sisters: **Mebarka, Hassina, Khadidja, Karima**, and **Amira**, and my brothers: **Abdelazize, Nadhir** and **Djamel**. It would have been extremely impossible to get to where I am today, with so much encouragement and love, the secret of my pride and the sun that enlightened my life. What a joy that my family has grown over the past four years

It is very difficult for me to talk about you *my dear husband Med Elcherif*, the angel of my life. Words are not enough to express my thanks....Thank you very much for your patience, your sacrifices, your moral support, your valuable advice, your unrivalled kindness and your deep attachment have enabled me to succeed in my studies. May Allah unite our paths for a long and serene life and may this work be the testimony of my gratitude and my sincere and faithful love.

Finally, I also express my sincere thanks to all my friends: **Amina, Fatima, Nadjla, Maroua, Yasmin, Aida, Fifi, Ahlem**...for the patience, support, helpfulness and encouragement in times when I needed it most. May Allah bring you happiness in your lives.

List of Works

International publications:

1. **Khelfa N**, Belaidi S, Zerroug A, Soualmia F, Chtita S (2023) In silico-Based Identification of new anti-PfDHFR drug candidates via 1,3,5-triazine derivatives. *Main Group Chemistry*,22(4): 521-539.
2. **Khelfa N**, Belaidi S, Abchir O, Chtita S, MOGREN AL-MOGREN M, Hochlaf M (2023) Combined 3D-QSAR, Molecular Docking, ADMET, and drug likeness scoring of Novel Diaminodihydrotriazines as Potential Antimalarial Agents (in review).

International conferences:

1. Multivariate analyzes (Drug likeness, SAR, QSAR) of the structure-activity relationships of a series of 1,3,5-triazine derivatives; The 1st international scientific days:"Molecular modeling: challenges of application to the fight against emerging diseases", 25 and 26 December 2020, Casablanca, Morocco.
2. Development and validation of a predictive qsar-based on artificial neural network for effective design of pfdhfr inhibitors; 8th international black sea coastline countries scientific research conference, 29-30 August 2022, Sofia, Bulgaria.
3. Development of QSAR-based (MLR/ANN) predictive models for effective design of PfDHFR inhibitors, 5^{ème} colloque maghrébin sur la chimie hétérocyclique CMCH 5-2022, 13-15 July 2022, Casablanca, Morocco.
4. In silico drug discovery of PfDHFR enzyme inhibitors based on Quantitative Structure-Activity Relationship and Drug-Likeness evaluation, the 1st international seminar on chemical process & environment, 14&15 March 2023, Biskra, Algeria.
5. QSAR modeling study and Drug-Likeness filters applied in potent 1,3,5-triazine derivatives inhibitors as antimalarial, 20&21 October 2023, Istanbul, Turkey.

National conferences:

1. QSAR Modeling and Drug-Likeness filters applied in 1,3,5-triazine derivatives; Journées Nationale sur le SARS-CoV-2 : Bioinformatique & Biotechnologie (JNSARSCoV2BB), 31 March 2021, Boumerdes, ALGERIA.
2. ADME predictions and docking molecular models to predict anti-malaria activity on a series of 1,3,5-triazine derivatives as anti-Pfdhfr agents ; 1st National Seminar on the

evaluation of the biological activities of medicinal plants and molecular docking
(SC.SNV), 22 March 2022, Batna, ALGERIA.

List of Abbreviations

AD: Applicability domain

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

ANN: Artificial Neural Network

BBB: factor of Blood-Brain Barrier permeability

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

CADD: Computer Aided Drug Design

CoMFA: Comparative Molecular field Analysis

CoMSIA: Comparative Molecular Similarity Indices Analysis

CNS: Central Nervous System

CV: Cross-Validation

Cyc: Cycloguanil (antimalarial drug)

CYP: primary cytochromes of the P450 super family

DFT: Density Functional Theory

DHF: Dihydrofolate

DHFR-TS: dihydrofolatereductase-thymidylate synthase

DNA: Deoxyribonucleic Acid

Fu: fraction unbound

HOMO: Highest Occupied Molecular Orbital

IUPAC: International Union of Pure and Applied Chemistry

HBD and HBA: Number of Hydrogen-Bond Donors and Acceptors

HF: Hartree-Fock

HIA: Human Intestinal Absorption

IC50: Half maximal Inhibitory Concentration

IUPAC: International Union of Pure and Applied Chemistry

LBVS: Ligand Based Virtual Screening

Log Kp: Skin permeability

Log P: Partition coefficient water / octanol

Log S: Molecular solubility

LD50: Half maximal Lethal Dosage

LOO: Leave One Out

LUMO: Lowest Unoccupied Molecular Orbital

MLR: Multiple Linear Regressions

MR: Molar Refractivity

MSE: Mean Squared Error

MW: Molecular Weight

NADP: Nictinamide Adenine Dinucleotide Phosphate

NMR: Nuclear Magnetic Resonance

NRB: Number of Rotatable Bonds

OECD: Organization for Economic Cooperation and Development

ONC: Optimal Number of Components

PDB: Protein Data Bank

PfDHFR: *Plasmodium falciparum* dihydrofolate reductase

PLS: Partial least squares regression

PM3: Parametric Method 3

PRESS: Predictive Residual Sum of the Squares

PSA: Polar Surface Area

Pyr: Pyrimethamine (antimalarial drug)

QSAR: Quantitative structure-activity relationship

RMSEC: Root-Mean-Squar-Error of Prediction of training set

RMSEP: Root-Mean-Squar-Error of Prediction of test set

RNA: Ribonucleic Acid

SBVS: Structure-Based Virtual Screening

SDEP: Standard Deviation of Error of Prediction

SEP: Standard Error of Prediction

SPILF: Society of 'Pathologie Infectieuse de Langue Française'

THF: Tetrahydrofolate

VD: Volume of Distribution

WHO: World Health Organization

List of Figures

Chapter I

Figure I.1: Incidence of Malaria in the world in 2023.

Figure I.2: Taxonomic classification of some Plasmodium species.

Figure I.3: The Plasmodium spp. life cycle. [12]

Figure I.4: Chemical structures of the folate family.

Figure I.5: Structure of the PfDHFR enzyme (PDB:1J3J).

Figure I.6: Chemical structures of "HMM" anticancer amino-sym-triazines.

Figure I.7: Structure of pyrazolo [1,5-a] -1,3,5-triazines.

Figure I.8: Chemical structures of diaminodihydrotriazines derivatives.

Figure I.9: Structure of 4,6-diamino-2,2-dimethyl-1,2-dihydro-1,3,5-triazine derivatives.

Chapter II

Figure II.1: Traditional drug research and development procedure[2].

Figure II.2: Schematic illustration of ligand-based and structure-based approaches in VSDD.

Figure II.3: Layout of Quantitative Structure-Activity Relationship Analysis [21]

Figure II.4: Neural network architecture.

Figure II.5: Flowchart to construct 3D-QSAR model.

Figure II.6: The steric and electrostatic field in a CoMFA grid [76]

Figure II.7: Contour maps and molecular alignment in a CoMSIA grid.

Figure II.8: Overview of pharmacophore mapping.

Figure II.9: Outline of process involved in Molecular Docking.

Figure II.10: Outline of process involved in Molecular Dynamic.

Chapter III

Figure III.1: In-silico studies involved in testing of 1,3,5-triazine derivatives.

Figure III.2: 3D MESP of 1,3,5-triazine. The results are color-coded, from red (most negative) to blue (most positive).

Figure III.3: Correlations of experimental versus predicted pIC₅₀ values using MLR.

Figure III.4: MLR model's applicability domain plot. The vertical dashed line represents the warning leverage ($h^*=0.782$), whereas the horizontal lines denote ± 3 .

Figure III.5: Architecture of ANN.

Figure III.6: Correlations of experimental versus predicted pIC₅₀ values using ANN.

Figure III.7: Compound 10.

Figure III.8: Permeability and clearance patterns in vitro for MW and logD.

Chapter IV

Figure IV.1: Chemical structures of diaminodihydrotriazines.

Figure IV.2: Computational modeling techniques involved in our study.

Figure IV.3: 3. 3-D QSAR Model building flowchart.

Figure IV.4: 3D-structure of Plasmodium falciparum DHFR enzyme (code PDB :1j3k).

Figure IV.5: Experimental pIC₅₀ vs. Plot of the predicted pIC₅₀ according to the CoMSIA model.

Figure IV.6: Fields contributions of CoMSIA model with the most active chemical (compound 8). (a): Steric field. (b): Electronegative (red, negative) and electropositive (blue, positive) contour maps. (c): CoMSIA hydrophobic contour maps (yellow, favored area; green, unfavored area). (d): H-bond acceptor and (e): H-bond donor.

Figure IV.7: Chemical structures of most potent compound (i.e. compound 8, left) and of the reference ligand (right).

Figure IV.8: Various interactions formed between the compound 8 (upper trace) and co-ligand (lower trace) with the active site of PfDHFR.

List of Tables

Chapter I

Table I.1. Taxonomic classification of *Plasmodium*.

Table I.2. Antimalarial drugs: pharmacological classes, modes of action, half-life and genes associated with resistance [36].

Chapter II

Table II.1. Classification of QSAR methodologies based on dimensionality. [21]

Table II.2. Types of biological data used in QSAR analysis.

Table II.3. Popularly known molecular descriptors dependent on various dimensions [38].

Table II.4. Statistical parameters for fitting and cross-validation [51].

Table II.5. Statistical parameters for external validation[51].

Chapter III

Table III.1. Optimized structures of the molecules under study[15].

Table III.2. Bond lengths (in Å) and valence angles (in degree, °) of 1,3,5-triazine.

Experimental data for 1,3,5-triazine are collected from Ref.[37]

Table III.3. Chemical descriptors used in the regression analysis. They correspond to number of rotatable bond on the molecule (**NRB (eV)**), energy of highest occupied molecular orbital (**E_{HOMO} (eV)**), energy of lowest unoccupied molecular orbital (**E_{LUMO} (eV)**), refractive index (**n**) and dipolar moment (**μ(D)**).

Table III.4. MLR statistics of predicted model.

Table III.5. ANN and MLR statistics of predicted models

Table III.6. Values of descriptors and pIC₅₀ for the new designed compounds (derivatives of comp.10) of the Figure III.7.

Table III.7. Drug-likeness parameters and Lipophilicity indices of 1,3,5-triazine derivatives.

Table III.8. Drug-likeness of the new designed compounds and reference compounds.

Chapter IV

Table IV.1. Chemical structures of the diaminodihydrotriazines under investigation and those of pyrimethamine (Pyr) and of cycloguanil (Cyc).

Table IV.2. pIC₅₀ experimental activities, pIC₅₀ predicted activities and residues of the series of diaminodihydrotriazines under study. * denotes test set compounds.

Table IV.3. PLS statistics results and fields contributions of CoMSIA model. **E, S, H, A** and **D** stand for electrostatic field, steric field, hydrophobic field, H-bond acceptor field and H-bonds donor field, respectively.

Table IV.4. Absorption, distribution, metabolism, excretion and toxicity predictions for compound **8** and for the reference compound (Ref) using pkCSM. We give their physiochemical properties as evaluated calculated using SWISS/ADME and their Drug Likeness rules.

Table IV.5. Chemical structures of compound **8** and of the newly designed potent compounds.

General Introduction

Drug discovery process is a wide scientific field; very complex and includes an interdisciplinary effort that faces many challenges these days for designing effective and commercially feasible drug. In the 20th century mankind has obtained the ability to discover highly active, yet small, organic molecules which are used for treatment purposes. However, medicinal chemists have always struggled with the selection of which compounds to

synthesize: a chemist has to choose the compounds to be synthesized from among millions of possible molecules.[1-3] Among them are extremely high development costs, long development times, as well as a low number of new drugs that are approved each year. To solve these problems, new and innovative technologies are needed that make the drug discovery process of small molecules more time and cost-efficient, and which allow to target previously undruggable target classes [2]. The computational chemist is encouraged to develop some kind of computer program capable of automatically evaluating very large libraries of compounds and integrate it into the drug discovery process. This is called “virtual screening” (VS). [3]

Virtual screenings (VS) have an enormous potential for the development of new small-molecule drugs, and are already starting to transform the focal points of early stage drug discovery [2] which are represented in the identification of lead compounds showing pharmacological activity against a biological target and the progressive optimization of the pharmacological properties and potency of these compounds. [4] All of these approaches can be very fast and cost-effective. They can lead to highly potent initial hit compounds, and address even challenging target sites such as protein-protein interactions. VS are procedures, in which a collection of ligands is computationally screened for their ability to bind to a given receptor structure. Receptors are most often proteins in the biomedical sciences but can be any type of biological macromolecule, including RNA or DNA. [2]

In this thesis, we will give an application of virtual screenings methods in order to identification and design of novel chemical entities specifically affecting these targets could lead to better drugs for the treatment of malaria. [5]

Malaria was long considered to be the most common infectious disease caused by *Plasmodium falciparum*. It was the disease most feared by explorers who, lacking a drug to protect or treat them, were reluctant to venture deep into the African continent. It is difficult to know what was meant by the word "fever" until the end of the 19th century. Throughout history, malaria has been confused with fevers of all origins, in particular typhoid and yellow fever [6], which affects or threatens more than half the world's population [7]. It remains one of the world's major public health problems, although significant progress has been made by the World Health Organization[8].

By its impact on populations and the seriousness of its pathology, although effective antimalarial agents have been known for a long time, the alarming spread of drug resistant strains of *Plasmodium falciparum*, which is the most lethal parasite species, undergoes the urgency and continuous need for the discovery of new therapeutics. A major initiative in this direction is to “*P. falciparum* Dihydrofolate Reductase” (*PfDHFR*) enzyme targets that are

critical to the disease process or essential for the survival of the parasite. The widespread occurrence of malaria could be attributed to the development of resistance of the parasite to the available antimalarial drugs such as chloroquine, cycloguanil and pyrimethamine. These drugs have been used clinically in the treatment of malaria for longer period of time. But due to the emergence of drug-resistant parasite in many countries, there is need of new and effective drugs for treatment of malaria. [9]

An effort to combat this disease is inspired by the priorities of studying series of antimalarial compounds based on the sym-triazines and especially on the diaminodihydrotriazines [10], which represent an inexhaustible source of active molecules, and its structural diversity can be the starting point for many advanced research projects in the pharmaceutical field, and developing new drugs.

Our objective of the study is to exploit the different computational tools and methods based on virtual screening to score and filter a set of chemical structures in order to procure new bioactive molecules and to study their ability of interactions with the enzyme Plasmodium falciparum dihydrofolate reductase (*PfDHFR*).

Our main contributions are summed in these essential points, namely:

- Structural and electronic study of the basic nuclei of heterocyclic compounds of 1,3,5-triazine.
- Drug-likeness and pharmacokinetic study using several empirical rules such as Lipinski's rule and MPO methods.
- Establish at the molecular level 2D-QSAR (MLR/ANN) models for some heterocyclic series of *PfDHFR* inhibitors.
- 3D-QSAR study was applied for modeling of the studied molecules.
- Analysis by the molecular docking method of the most active chemical of the series and the reference ligand.

In order to carry out this work properly and to achieve the main objectives, we have organized our thesis into four chapters:

Chapter I- Background on malaria diseases & their inhibitors

In the chapter I we will describe the malaria diseases, their pathogenesis and their treatment. In addition, it contains a description of some their inhibitors resistance.

Chapter II- Virtual Screening in Drug Design & Discovery

The second chapter covers the literature review required for Virtual Screening, and is divided into two parts: Ligand-based (LBVS) and Structure-Based Virtual Screening (SBVS), which have become indispensable for the design of new drug substances. We focus briefly on the objective of its definition, its principle, and its techniques and methods.

↳ Realized Works

In this part will be devoted to the practical implementation and interpretation of the results obtained from our work throughout this project.

Chapter III: In silico-Based Identification of new anti-PfDHFR drug candidates via 1,3,5-triazine derivatives

- The 1st point, a 2 Dimension Quantitative structure-activity relationship (2D-QSAR) models were generated using MLR and ANN methods for series of 28 derivatives of 1,3,5-triazine with the use of 20 molecular descriptors
- In the 2nd point, we will present drug-likeness screening studies of the interest inhibitors of PDHFR enzyme.
- At last, the obtained QSAR models were employed to define biological activities of potentially novel active compounds by means of in silico screening processes.

Chapter IV: Combined 3D-QSAR, molecular docking, ADMET and drug likeness scoring of novel Diaminodihydrotriazines as potential antimalarial agents

- The 1st point, consists 3D-QSAR study using the statically method PLS to determine the best CoMSIA model for a series of 42 diaminodihydrotriazine derivatives.
- In the 2nd point, In silico pharmacokinetic/ADMET studies of the most active compound with the reference ligand derivatives of diaminodihydrotriazine.
- Finally, a molecular docking analysis recognizes which molecule; the most active compound or the reference ligand are possible to interact toward the PfDHFR enzyme. Three new developed ligands can then form novel drugs.

We conclude this thesis with a general conclusion with different perspectives.

References

1. Begum S, Shahidulla SM (2019) Role of Computer Aided Drug Design in Drug Development and Discovery: An Overview. *Int J Res Eng Sci Manag* 2(2):445-450.
2. Gorgulla C (2022) Recent Developments in Structure-Based Virtual Screening Approaches. arXiv preprint arXiv:2211.03208.
3. Pasupa K (2012). The Review of Virtual Screening Techniques. *KMITL Inf Tech J* 1(1).
4. Lionta E, Spyrou G, K Vassilatis D, Cournia Z (2014) Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry* 14(16):1923-1938.
5. Philips RS (2001) Current status of malaria and potential control. *Clin Microbiol Rev* 14:208-226.
6. Mouchet J, Carnevale P, Coosemans M, Julvez J, Manguin S, Richard-Lenoble D (2018) Biodiversité du paludisme dans le monde. Paris: J. Libbey; 2004.
7. Aafje R (2004) Questions about Malaria- WHO/Roll Back Malaria.
8. OMS. Rapport (2011) sur la lutte contre le paludisme dans le monde.
9. Dube PN, Mokale S, Datar P (2014). CoMFA and docking study of 2, N6-disubstituted 1, 2-dihydro-1, 3, 5-triazine-4, 6-diamines as novel PfDHFR enzyme inhibitors for antimalarial activity. *Bulletin of Faculty of Pharmacy, Cairo University* 52(1):125-134.
10. Kamchonwongpaisan S, Charoensetakul N, Srisuwannaket C, Taweechai S, Rattanajak R, Vanichtanankul J, Vitsupakorn D, Arwon U, Thongpanchang C, Tarnchompoo B, Vilaivan T, Yuthavong Y (2020) Flexible diaminodihydrotriazine inhibitors of Plasmodium falciparum dihydrofolatereductase: Binding strengths, modes of binding and their antimalarial activities. *Eur J Med Chem* 195:112263.

Chapter I:

Background on Malaria Diseases & Their Inhibitors

I.1. Introduction

Malaria is one of the oldest diseases known to mankind [1]. The word malaria comes from the Latin *paludis*, meaning swamp. This word clearly reflects the relationship between the disease and the ecology of its vector, a mosquito. In Anglophone countries, the term malaria, which comes from the Italian *mal'aria*, meaning bad air, has been retained and has spread to many foreign countries, despite its inappropriate pathogenic connotations. The clinical

manifestations of malaria have been known since ancient times. Its evolution in time and space follows the history of mankind. Physicians in Vedic and Brahmanic India were already distinguishing characteristic intermittent fevers as far back as 1000 BC (Before Christ). When humans began to settle as communities, favoring the transmission of infections [1,2].

In the middle Ages, a large part of Europe suffered from it. The disease was also rife on the new continent before it was discovered, since it was the Spaniards who learned from the Indians about the febrifuge properties of cinchona bark. It was in 1880 that Laveran, a French military doctor in Algeria, discovered the haematozoan of malaria. And it was in 1898 that Grassi demonstrated that the anopheles is the vector of human malaria. In the 1950s, the malaria "eradication" programme emerged, but this was soon replaced by control strategies (in the sense of combating or controlling) in the early 1970s.

I.2. Malaria disease

I.2.1. Epidemiology of malaria

Malaria is endemic worldwide. It is one of the leading causes of infant mortality in developing countries. It is caused by a specific haematozoan, plasmodium, inoculated by the bite of female mosquitoes belonging to various varieties of Anopheles [3,4]. The distribution of malaria varies enormously from one geographical area to another, and even within the same village. The distribution of malaria is extremely variable from one geographical area to another, and even within the same village. This heterogeneity is influenced by numerous factors such as the vector, the host and the parasite. All these factors are in dynamic relationship with the environmental and socio-economic factors that condition the epidemiology of malaria [5].

According to the WHO, Globally, an estimated 2 billion malaria cases and 11.7 million malaria deaths were averted in the period 2000-2021. Most of the cases (82%) and deaths (95%) averted were in the WHO African Region, followed by the WHO South-East Asia Region (cases 10% and deaths 3%) estimated deaths between 2019 and 2021, there were 63 000 deaths that were due to disruptions to essential malaria services during the COVID-19 pandemic [6].

The situation is currently worsening because Plasmodium has become resistant to quinine and synthetic antimalarial (Figure I.1), which were the basis of both treatment and chemoprophylaxis. Reversing this situation will depend on the creation of new drugs and, above all, the development of a vaccine [3,4].

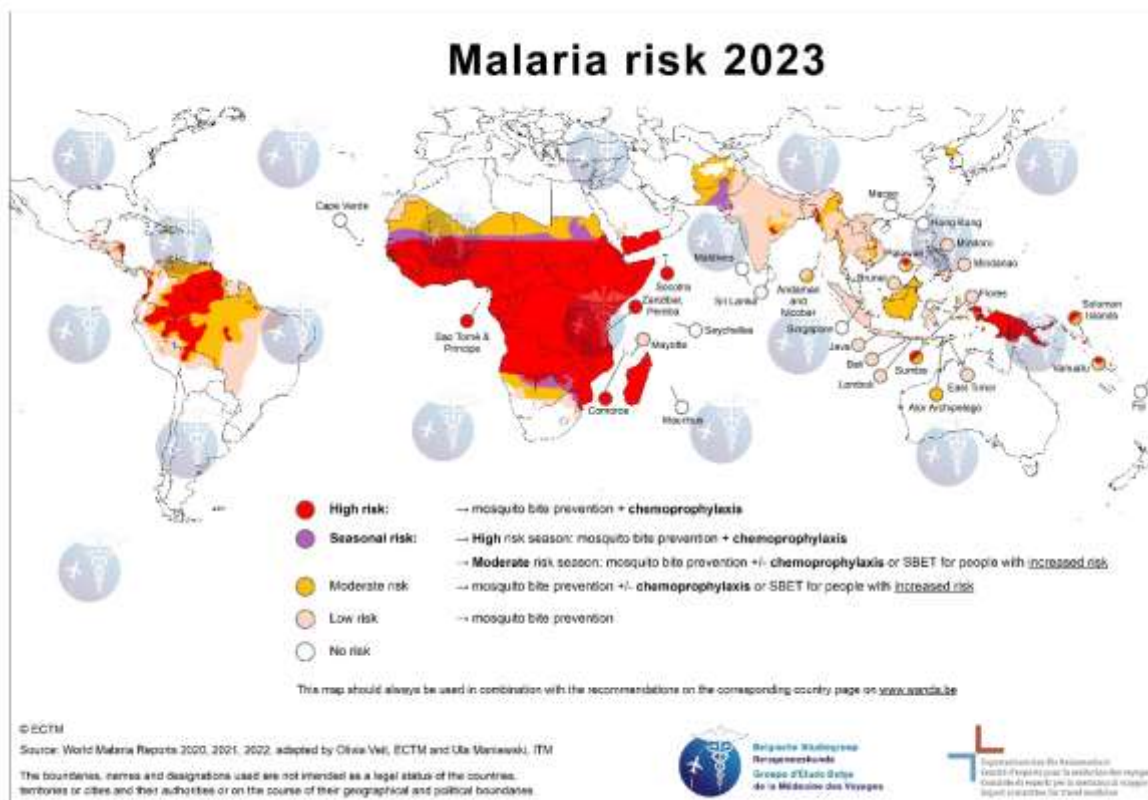


Figure I.1: Incidence of Malaria in the world in 2023.

I.2.1.1. Epidemiological facts

The indices used to classify malaria transmission zones are based on the factors influencing the epidemiology of malaria. The stability index classifies malaria into two zones:

✓Zones with stable or endemic malaria: malaria transmission is long and intense, resulting in a form of premunity that enables individuals to limit serious clinical manifestations in young children.

✓Unstable or epidemic malaria zones: malaria transmission is very short-lived and occurs in epidemic form. This episodic nature of transmission does not allow for the development of premunity. All individuals are at risk of developing the disease [7].

I.2.2. Pathogens and vectors agents

I.2.2.1. The pathogen agent

a. Taxonomy

Table I.1. Taxonomic classification of *Plasmodium*

Kingdom :	Protista
Subregnum :	Protozoa
Phylum :	Apicomplexa

Class :	Sporozoa
Subclass :	Eucoccidia
Ordre :	Haemosporidae
Family :	Plasmodidae
Genre :	<i>Plasmodium</i>
Species :	<i>falciparum, malariae, ovale, vivax, knowlesi</i>

The *Plasmodium* genus comprises 172 species that infect birds, reptiles and mammals. Other genera in the same group include *Hepatocystis*, *Haemoproteus* and *Leucocytozoon*, none of which are infectious to humans. Parasites of humans and primates all belong to either the subgenus *Plasmodium* or the subgenus *Plasmodium* (*Laverania*), while all other species infecting mammals belong to the subgenus *Plasmodium* (*Vinckeia*).

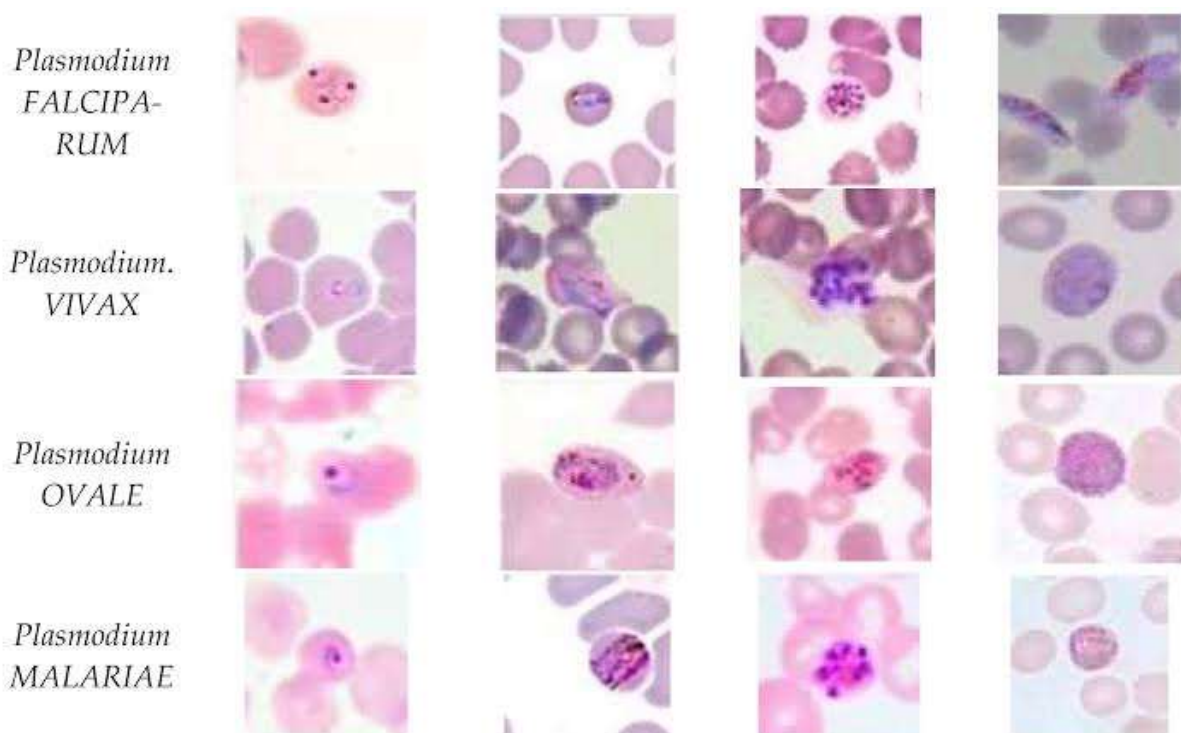


Figure I.2: Taxonomic classification of some *Plasmodium* species.

The various sub-genre (Figure I.2) are distinguished from one another by morphological features and life history characteristics that were used as taxonomic criteria in Garnham's classic classification.[8]

b. Species

Malaria is transmitted by a protozoan of the genus *Plasmodium*. There are many species of *Plasmodium* affecting various animal species, but only five of these species are found in human pathology [9].

These are *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. The five species differ in terms of their biological and clinical characteristics, their geographical distribution and their ability to develop resistance to antimalarial drugs [10].

I.2.2.2. The vector agent

The vector of malaria is the female Anopheles. It is a haematophagous insect of the order Diptera nematocera, family Culicidae and genus Anopheles. More than 500 species of Anopheles have been described, around fifty of which can act as vectors of Plasmodium in humans, and 20 of which are of real epidemiological importance. Numerous climatic and/or environmental factors, whether natural or due to human activity, can modify the distribution of Anopheles in a given region and therefore influence the transmission of Plasmodium. The female anopheles only bites after sunset, with peak activity between 11pm and 6am; this explains the use of mosquito nets for personal prevention [10].

I.2.3. The evolutionary cycle of Plasmodium

The cycle takes place successively in human body (asexual phase in the intermediate host) and in the female anopheles (sexual phase in the definitive host) [11] (Figure I.3).

I.2.3.1. In human body

The cycle is divided into two phases:

- The hepatic or pre-erythrocytic or exo-erythrocytic phase, which corresponds to the incubation phase, clinically asymptomatic;
- The blood or erythrocyte phase, which corresponds to the clinical phase of the disease.

a. The liver phase

The sporozoites inoculated by the female Anopheles during its blood meal remain in the skin, lymph and blood for a maximum of thirty minutes. Many are destroyed by macrophages, but some manage to reach the hepatocytes[11]. They transform into pre-erythrocytic schizontes or "blue bodies" (multinucleated forms) which, after a few days of maturation, burst and release thousands of merozoites into the blood (10,000 to 30,000 depending on the species).

Hepatic schizogony is unique in the cycle, as the liver cell can only be infected by sporozoites [11].

In *P.vivax* and *P.ovale* infections, delayed hepatic schizogony (hypnozoites) can lead to the release of merozoites into the blood several months after the mosquito bite, thus explaining the late relapses observed with these two species [11].

b. The erythrocyte phase

This part of the cycle corresponds to the clinical phase. Merozoites released during the hepatic phase enter the red blood cells. The penetration of the merozoite into the erythrocyte and its maturation into a trophozoite and then a schizont depends on the species and leads to the destruction of the host red blood cell and the release of 8 to 32 new merozoites[11].

I.2.3.2. In female Anopheles

Gametocytes ingested by the mosquito during a blood meal on an infected subject are transformed into male and female gametes, which fuse into a free, mobile egg called an ookinete [18]. This ookinete leaves the lumen of the digestive tract, attaches to the outer wall of the stomach and transforms into an oocyst. Parasitic cells multiply within this oocyst, producing hundreds of sporozoites which, once the oocyst has burst, migrate to the mosquito's salivary glands [11]. These sporozoites are the infectious forms ready to be inoculated with the mosquito's saliva during a blood meal on a vertebrate host.

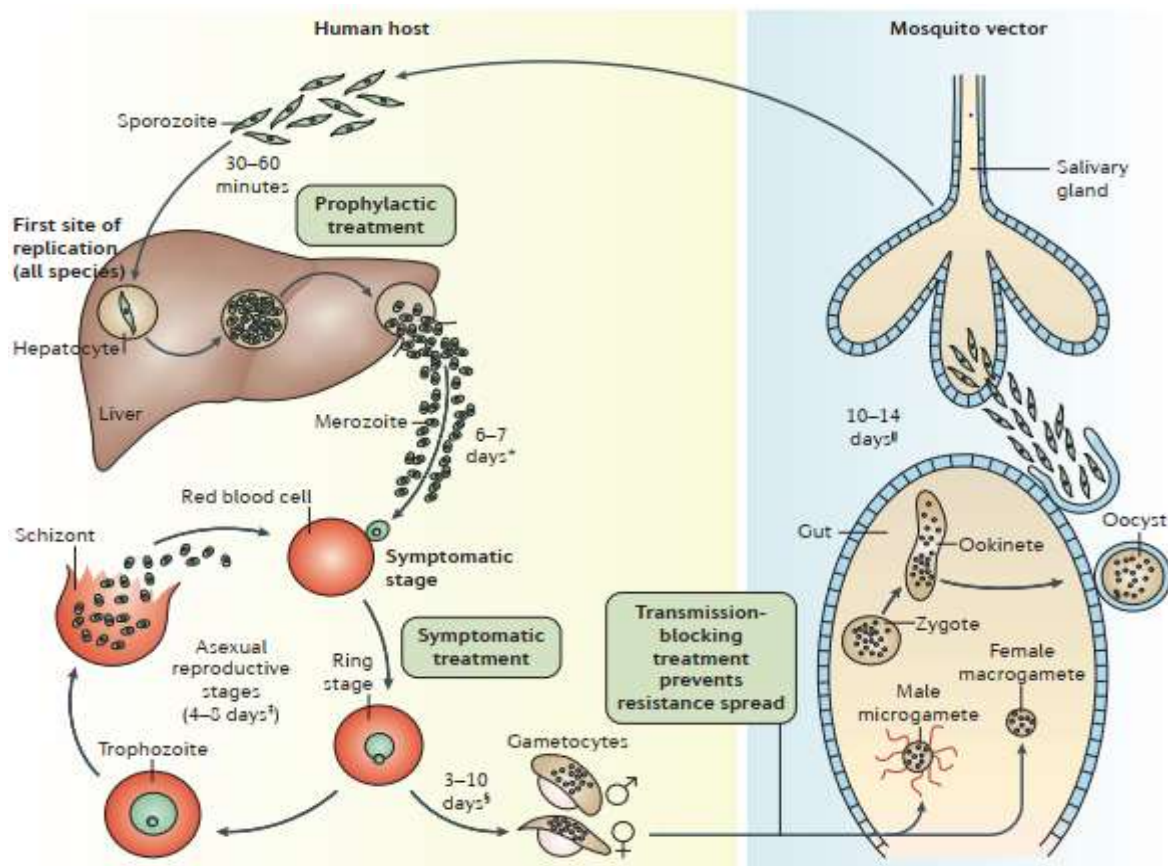


Figure I.3: The *Plasmodium* spp. life cycle. [12]

I.2.4. Pathophysiology [13]

The pathophysiology of malaria is highly complex. For all plasmodial species, the clinical manifestations are observed during endo-erythrocytic multiplication and their severity depends on the plasmodial species, the parasite density and the degree of immunity of the host. In simple malaria attacks, the parasites are much diluted and cannot be detected under the microscope at the start of the blood infection.

Their number increases and the clinical threshold is marked by the periodicity of febrile crises after several schizogonic cycles.

The fever is triggered by the release of malar pigment or haemozoin when the parasitized red blood cells burst. Haemozoin, a pyrogenic substance, acts on the bulbar thermoregulatory centers. Initially, the fever is fairly continuous because the endo-erythrocytic cycle is still poorly synchronized. When the cycle becomes synchronized, the release of haemozoin is regularly repeated, giving the fever a periodic appearance characteristic of a malarial attack.

Anaemia results from the bursting of parasitized red blood cells. Hepatomegaly and splenomegaly are due to hyperactivity of the monocyte-macrophage system responsible for clearing malarial pigment and erythrocyte debris.

I.2.5. Symptoms

The intra-erythrocytic phase of the parasite cycle is responsible for all the symptoms. Fever is the most frequent symptom of malaria, but there are no signs of malaria. Pathognomonic of the infection. Malaria attacks can therefore be confused with the following infections viruses such as influenza, particularly during epidemic periods for respiratory viruses. Medical staff need to be vigilant and made aware that any fever on returning from a malaria-endemic area is malaria until proven otherwise. The fever is then rhythmic, occurring every 48 to 72 hours depending on the parasite species and the length of its intra-erythrocytic development cycle (48 hours for *Plasmodium ovale* spp). In addition to fever, patients frequently present with headache, arthralgia/myalgia and general ill health. Digestive symptoms are less frequent, often combining nausea and vomiting with diarrhoea, anorexia and abdominal pain. This is known as febrile gastric embarrassment. Together, these symptoms define uncomplicated malaria, which accounts for around 85% of malaria attacks in France [14]. Splenomegaly and hepatomegaly may also be observed, mainly in cases of *Plasmodium falciparum* malaria.

However, some patients may present with more severe symptoms, requiring treatment in intensive care units (ICUs) or continuous care units (CCUs). The World Health Organization

(WHO) has defined a set of clinicobiological criteria for defining severe malaria attacks [15], adapted by the Société de Pathologie Infectieuse de Langue Française (SPILF) [16].

The majority of severe malaria attacks are associated with *Plasmodium falciparum*, notably due to its capacity for cytoadherence (*i.e.* the ability of infected red blood cells to adhere to vascular endothelia) and sequestration in deep organs, mediated by Variant Surface Antigen [17,18]. However, *Plasmodium vivax* is also an important agent of severe malaria [19-21]. More than twenty cases of severe malaria caused by *Plasmodium ovale* spp have been described in the literature [22-26].

I.3. The PfDHFR therapeutic target

I.3.1 Generality on folates

Folates (pteroylglutamates), based on the structure of folic acid, belong to a family of B vitamins (Figure I.4) and are essential components of cell growth and proliferation. Members of this family differ from each other in several ways, namely the redox state of the pyrazine ring (X), substitution at the N5 and N10 positions (Y) and the presence of additional glutamate residues linked to the γ -carboxyl group of the only glutamate radical intrinsic to the folate structure (Z).[27, 28]

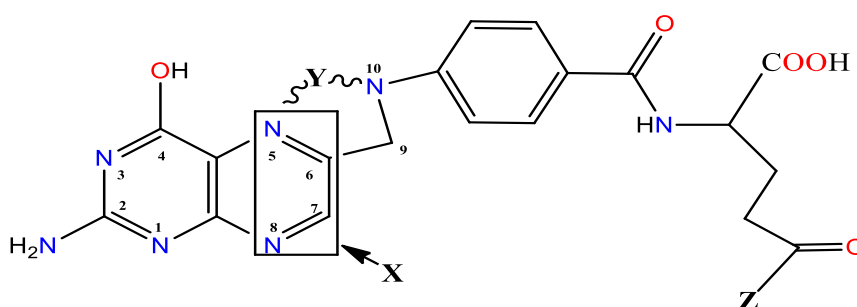


Figure I.4: Chemical structures of the folate family.

Folate cofactor plays a major role in the general synthesis of amino acids and nucleotide bases of DNA and RNA in the malaria parasite. Therefore, folate synthesis pathway has been regarded as an important target for antimalarial drugs, namely dihydrofolate reductase inhibitors.[29]

I.3.2. *Plasmodium falciparum* protein dihydrofolate reductase (PfDHFR)

Among the members of the folate family is the *Plasmodium falciparum* dihydrofolate reductase DHFR (Figure I.5), which is one of the well-established therapeutic targets in malaria-affected *P. falciparum*. It is a bifunctional enzyme known as dihydrofolate reductase thymidylate synthase (DHFR-TS). [30]

It is an essential enzyme that is also ubiquitous in most living organisms[31], being a key enzyme responsible for the reproduction cycle of the Plasmodium parasite [33,33].

The first study of the bifunctional PfDHFR-TS enzyme was carried out in 1984. [34] This study demonstrated that the PfDHFR-TS protein is a bifunctional enzyme with a dimer composed of two subunits of identical size, DHFR and TS. Subsequently, Bzik [35], using cloning and isolation of the pfDHFR-TS gene, performed a sequence analysis of this protein. It was shown that the residual composition of each monomer was linked by a polypeptide junction chain. In which DHFR and TS exist in a single chain in *p.falciparum* whereas they are separated in humans [36].

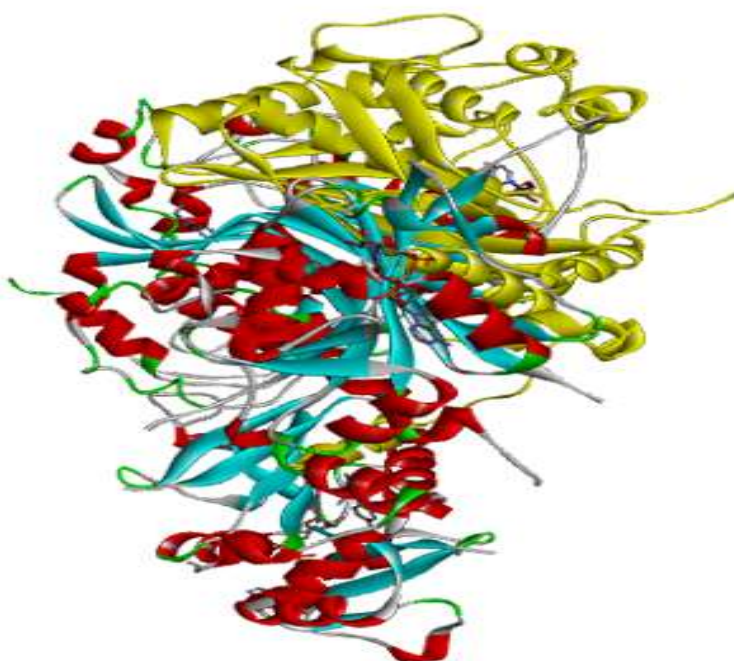


Figure I.5: Structure of the PfDHFR enzyme (PDB:1J3J).

It catalyses the NADPH-dependent reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) using NADPH as a cofactor, which plays a crucial role in many biochemical processes such as folate metabolism and DNA synthesis. Thus, inhibition of DHFR causes an interruption in DNA formation, which eventually leads to the death of parasitic cells. [30]



I.4. Antimalarial drugs and resistance

I.4.1. Pharmacological classes and mechanisms of action

Today, more than fifteen compounds are used against malaria. Their modes of action are shown in the table below (Table I.2), along with their half-life (not including metabolites) and the genes involved in resistance to *P. falciparum*[37].

Table I.2. Antimalarial drugs: pharmacological classes, modes of action, half-life and genes associated with resistance [37].

Chemical class	Molecule	Pharmacological class	mode of action	Half-life duration *	Resistance genes **
Natural or Hemisynthesised antibiotics					
amino-alcohol quinine		Erythrocytic schizonticide	inhibition of heme detoxification in the digestive vacuole	5 to 6pm	Pfnhe1 Pfmpr
	sesquiterpene lactone artémisinine and derivatives	Erythrocyte schizonticide + gametocyte action	alkylation of haemoglobin metabolites, production of free radicals	1 to 4 hours depending on derivatives	Pfk13
Synthetic antimalarial drugs					
4-aminoquinoline					
	amodiaquine chloroquine piperazine	Erythrocytic schizonticide	inhibition of heme detoxification in the digestive vacuole	1 to 6 hours 2 to 3 days 22 days	Pfmdr1 Pfmdr1 Pfcrt Pfmrt
8-amino-quinoline					
	primaquine tafenoquine	schizonticide érythrocytaire + intrahépatique + gamétocydiq	interference with the functioning of plasmodial DNA undetermined	3 to 6 hours 14 days	

Amino alcohol				
Halofantrine lumefantrine mefloquine	erythrocytic schizonticide	inhibition of heme detoxification in the digestive vacuole	2 to 3 days 3 to 6 days 15-22 days	Pfmdr1 Pfmdr1
antifolates and antifolinics				
Sulfadoxine Pyrimethamine Proguanil	Erythrocyte schizonticides	inhibition of nucleic acid biosynthesis	7 to 9 days 2 to 4 days 8 to 24 hours	Pfdhfr Pfdhps Pfmpr
Hydroxynaphtoquinone				
Atovaquone	erythrocytic and intrahepatic schizonticides	inhibits electron transport in the mitochondria and therefore ATP synthesis	2 to 3 days	Pfcytb
Antibiotics				
Cyclines				
Doxycycline	erythrocytic and intrahepatic schizonticides	binding to the 30S subunit of the ribosome	6 to 10 pm	PftetQ

* not including active metabolites

** genes associated with resistance in *Plasmodium falciparum*

I.4.2. Focus on sym-triazine derivatives

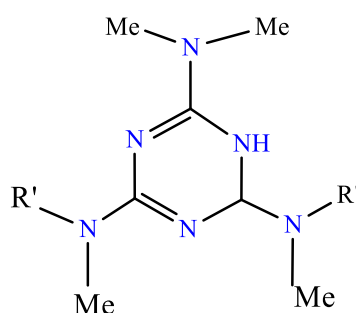
Heterocycles containing a nitrogen atom are among the most common elements in medicinal chemistry, and are found in biomolecules in the DNA chain in purine and pyrimidine bases, as well as in many natural products [38]. These compounds include sym-triazines, which are among the oldest nitrogen-containing organic heterocycles [39].

Sophisticated derivatives of free sym-triazine can be easily prepared from the cheap and readily available 2,4,6-trichloro-1,3,5-triazine (cyanuric chloride) 1 [40- 42]. Replacement of chloride ions in cyanuric chloride give several variants of 1,3,5-triazine derivatives one of them is the amino-sym-triazines, which were investigated as biologically active small molecules [43- 47].

I.4.2.1. Physiological and biochemical role of amino-sym-triazine derivatives

The remarkable development of amino-sym-triazines for various diseases in a very short space of time proves their importance for chemical research. Amino-sym-triazine occupies a prominent position and possesses a wide range of biological activities (Figure I.4). amino-symtriazine is presented in many potent biologically active molecules with promising biological potential, making it an attractive support for the design and development of new drugs. The broad spectrum of biological activities of this moiety has attracted attention in the field of chemistry and especially medicinal chemistry[48].

From a therapeutic point of view, there are a large number of active ingredients containing amino-sym-triazines with diverse therapeutic activities in a wide range of fields [49]. Since the 1970s, several studies have been carried out on the antitumour activity of 2,4,6- tris(N,N-dialkylamino)-1,3,5-triazines. One of these analogues, hexamethylmelamine (HMM) (Figure I.6), is effective against lung, breast and ovarian cancers. Some structural analogues of HMM have been prepared and tested [50].



R=R'=Me, HMM

Figure I.6: Chemical structures of "HMM" anticancer amino-sym-triazines

Many other therapeutic agents have fused diamino-sym-triazines into their chemical structures. In the 1980s, 7-methyl-pyrazolo [1,5-a] -1,3,5-triazine-2,4-diamine was developed as a new bronchodilator and anti-allergy compound[51]. Research interests in pyrazolo [1,5-a] -1,3,5-triazines (Figure II.7) have focused on corticotropin-releasing factor (CRF) receptor blocking activity[52,53]

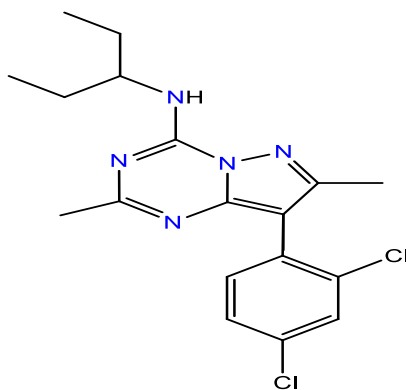


Figure I.7: Structure of pyrazolo [1,5-a] -1,3,5-triazines.

I.4.2.2. Diamino-sym-triazines as antimalarial drugs

Sustained efforts are being made to design and develop a potent PfDHFR inhibitor for malaria control. It is therefore of interest to screen PfDHFR with, 1,3,5-triazine derivatives[54]. Current therapeutic approaches for treating malaria infection include a number of different antimalarial drugs such as antifolates [55]: 4,6-diamino-2,2-dimethyl-1,2-dihydro-1,3,5-triazine (Baker triazines) derived from diaminodihydrotriazines which is displayed in figure I.8.

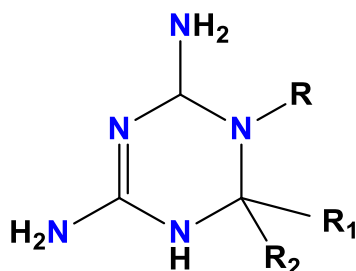


Figure I.8: Chemical structures of diaminodihydrotriazines derivatives.

Its derivatives are becoming increasingly important as pharmaceutical products. Many of these compounds are inhibitors of *P. falciparum* dihydrofolate reductase. Baker triazineantifol was currently in clinical trials as a drug for cancer chemotherapy [56,57].

Resistance to DHFR inhibitors is conferred by single mutations in the gene encoding the respective enzyme, resulting in substitutions in the amino acid chain [58].

But new antimalarial treatments should have new mechanisms of action that are effective against existing multi-resistant strains. In addition, the interruption of parasite transmission, which could contribute to the eradication of malaria, should be exploited by the next generation of antimalarial drugs [59].

These PfDHFR inhibitors (Figure I.9) have been recognized for their therapeutic value as antiprotozoal agents.

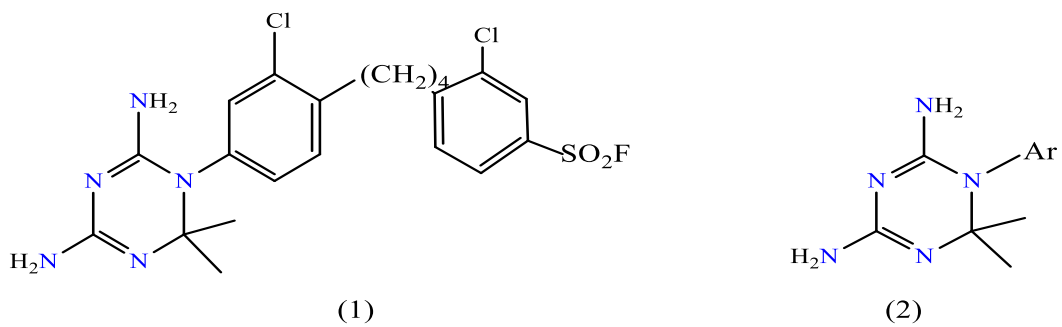


Figure I.9: Structure of 4,6-diamino-2,2-dimethyl-1,2-dihydro-1,3,5-triazine derivatives.

References

1. Kwiatkowski DP (2005) How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *Am J Hum Genet* 77(2):171-192.
2. Diamond J (2011) *Le troisième chimpanzé. Essai sur l'évolution et l'avenir de l'animal humain*. Folio essais. Gallimard.
3. De Lamare G (2000) *Dictionnaire des termes de médecine*, 26^{ème} édition.
4. Gentilini M (1993) *Médecine tropicale*, Flammarion, France, 5^{ème} édition.
5. ANOFEL. *Paludisme*. Association Française des Enseignants de Parasitologie et Mycologie, Paludisme (2014).
6. World Health Organization (2022). WHO European regional obesity report 2022. World Health Organization. Regional Office for Europe.
7. Macdonald G (1957) *The Epidemiology and Control of Malaria*. Oxford University Press, London, p201.
8. Garnham PC (1966) Immunity against the different stages of malaria parasites. *Bull Soc Pathol Exot Filiales* 59(4):549-557.
9. Support de cours de paludisme (version PDF) (2010-2011) Université Médicale Virtuelle Francophone.
10. Robert V, Boudin C (2002) Biologie de la transmission homme-moustique du plasmodium. *Parasitologie*. Manuscrit n°2454a.
11. Tanrikulu Y, Schneider G (2008) Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nature Rev Drug Discov* 7(8), 667-677.
12. Ambroise T (1991) P : Paludisme : Physiopathologie-réceptivité-résistance innée Edition ellipses/AUPELF 60-65.
13. Kendjo E, Houzé S, Mouri O, Taieb A, Gay F, Jauréguiberry S, et al. (2019) Epidemiologic Trends in Malaria Incidence Among Travelers Returning to Metropolitan France, (1996-2016). *JAMA Netw Open* 2(4):e191691.
14. Severe malaria. (2014) *Trop Med Int Health*. 1:7-131.
15. Bouchaud O, Bruneel F, Caumes E, Houzé S, Imbert P, Pradines B, et al. (2020) Management and prevention of imported malaria. 2018 update of the 2007 French clinical guidelines. *Médecine et Maladies Infectieuses* 50(2):161-93.
16. Wahlgren M, Goel S, Akhouri RR (2017) Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. *Nat Rev Microbio* 15(8):479-91.
17. Kraemer SM, Smith JD (2006) A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol* 9(4):374-80.

18. Naing C, Whittaker MA, Nyunt Wai V, Mak JW (2014) Is Plasmodium vivax malaria a severe malaria?: a systematic review and meta-analysis. *PLoS Negl Trop Dis* 8(8):3071.
19. Mathews SE, Bhagwati MM, Agnihotri V (2019) Clinical spectrum of Plasmodium vivax infection, from benign to severe malaria: A tertiary care prospective study in adults from Delhi, India. *Trop Parasitol*. Dec 9(2):88–92.
20. Guedes KS, Sanchez BAM, Gomes LT, Fontes CJF (2019) Aspartate aminotransferase-to-platelet ratio index (APRI): A potential marker for diagnosis in patients at risk of severe malaria caused by Plasmodium vivax. *PLoS ONE* 14(11):e0224877.
21. Singh R, Jain V, Singh PP, Bharti PK, Thomas T, Basak S, et al. (2013) First report of detection and molecular confirmation of Plasmodium ovale from severe malaria cases in central India. *Trop Med Int Health* Nov 18(11):1416–20.
22. Strydom K-A, Ismail F, Freaun J (2014) Plasmodium ovale: a case of not-so-benign tertian malaria. *Malar J* 13:85.
23. Tomar LR, Giri S, Baudhdh NK, Jhamb R (2015) Complicated malaria: a rare presentation of Plasmodium ovale. *Trop Doct*. 45(2):140–2.
24. D'Abramo A, Gebremeskel Tekle S, Iannetta M, Scorzolini L, Oliva A, Paglia MG, et al. (2018) Severe Plasmodium ovale malaria complicated by acute respiratory distress syndrome in a young Caucasian man. *Malar J* 17(1):139.
25. Groger M, Fischer HS, Veletzky L, Lalremruata A, Ramharter M (2017) A systematic review of the clinical presentation, treatment and relapse characteristics of human Plasmodium ovale malaria. *Malar J* 11;16(1):112.
26. Baugh CM, Krumdieck CL, Ann NY (1971) *Acad. Sci.*, 186: 7.
27. McGuire J (2003) *J Curr Pharm Des* 9 : 2593-2613.
28. Wazhurst DC (2002) Resistance to antifolates in *P. falciparum*, the causative agent of tropical malaria. *Sci prog* 85: 89-111.
29. Tibon NS, Hee Ng C, Cheong SL (2020) *Eur. J. Med. Chem* 188:111983. <https://doi.org/10.1016/j.ejmech.2019.111983>
30. Blakley RL (1984) (Blakley, R.L. and Benkovic, S.J. eds.), John Wiley & Sons, Inc. USA, 1: 191-253.
31. Yuthavong Y, Vilaivan T, Chareonsethakul N, Kamchonwongpaisan S, Sirawaraporn W, Quarrell R, Lowe G (2000) *J Med Chem* 43: 2738-2744.
32. Vanichtanankul J, Taweechai S, Uttamapinant C, Chitnumsub P, Vilaivan T, Yuthavong Y, Kamchonwongpaisan S (2012) *Antimicrob Agents Chemother* 56: 3928-3935.

33. Garrett CE, Coderre JA, Meek TD, Garvey EP, Claman DM, Beverley S M, Santi DV (1984) *Mol Biochem Parasitol* 11: 257-265.
34. Li DJ, Bzik W, Horii T, Inselburg, J (1987) *Proc Natl Acad Sci* 84: 8360-8364.
35. Delfino R, Santos-filho O, Figueroa-Villar J (2002) *J Braz Chem Soc* 13:727-741.
36. National College of Medical Pharmacology, France.
37. AMIRA A (2015) Doctoral thesis of University Badji Mokhtar, Annaba, Algeria.
38. Blaney JM, Hansch C, Silipo C, Vittoria A (1984) *Chem Rev* 84: 333-407.
39. Koç ZE (2011) Complexes of iron (III) and chromium (III) salen and salo-phen Schiff bases with bridging 1, 3, 5-triazine derived multidirectional ligands. *J Heterocycl Chem* 48(4):769–775.
40. Carofiglio T, Varotto A, Tonellato U (2004) One-pot synthesis of cyanuric acid-bridged porphyrin-porphyrin dyads. *J Org Chem* 69(23):8121–8124.
41. Mooibroek TJ, Gamez P (2007) The s-triazine ring, a remarkable unit to generate supramolecular interactions. *Inorg Chim Acta* 360(1):381–404.
42. Porter JR, Archibald SC, Brown JA, Childs K, Critchley D, Head JC et al (2002) Discovery and evaluation of N-(triazin-1, 3, 5-yl) phenylalanine derivatives as VLA-4 integrin antagonists. *Biorg Med Chem Lett* 12(12):1591–1594.
43. Mylari BL, Withbroe GJ, Beebe DA, Brackett NS, Conn EL, Coutcher JB et al (2003) Design and synthesis of a novel family of triazine-based inhibitors of sorbitol dehydrogenase with oral activity: 1-{4-[3R, 5S-dimethyl-4-(4-methyl-[1, 3, 5] triazin-2-yl)-piperazin-1-yl]-[1, 3, 5] triazin-2-yl}-(R) ethanol. *Biorg Med Chem* 11(19):4179–4188.
44. Henke BR, Consler TG, Go N, Hale RL, Hohman DR, Jones SA et al (2002) A new series of estrogen receptor modulators that display selectivity for estrogen receptor β . *J Med Chem* 45(25):5492–5505.
45. Klenke B, Stewart M, Barrett MP, Brun R, Gilbert IH (2001) Synthesis and biological evaluation of s-triazine substituted polyamines as potential new anti-trypanosomal drugs. *J Med Chem* 44(21):3440–3452.
46. D'Atri G, Gomasasca P, Resnati G, Tronconi G, Scolastico C, Sirtori CR (1984) Novel pyrimidine and 1, 3, 5-triazine hypolipemic agents. *J Med Chem* 27(12):1621–1629
47. Singla P, Luxami V, Paul K (2015) *Eur J Med Chem* 102: 39-57.
48. Spaltenstein A, Almond MR, Bock WJ, Cleary DG, Furfine ES, Hazen RJ, Kazmierski, WM, Salituro FG, Tung RD, Wright LL (2000) *Biorg Med Chem Lett* 10:1159.25.
49. Foster BJ, Harding BJ, Leyland-Jones B, Hoth D (1986) *Cancer Treat. Rev* 38: 197-217.

50. Guillaume M, Lakatos C (1983) Prostaglandins Series 3: 332-7.
51. Gilligan PJ, Robertson DW, Zaczek R (2000) J Med Chem 43: 1641-1660.
52. Gilligan PJ, Folmer BK, Hartz RA, Koch S, Nanda KK, Andreuski S, Fitzgerald L, Miller K, Marshall W (2003) J Bioorg Med Chem 11: 4093-4102
53. Singh IV, Mishra S (2018) Molecular Docking Analysis of Pyrimethamine Derivatives with Plasmodium falciparum Dihydrofolate Reductase. Bioinformation 14(5): 232–235.
54. Saifi M, Beg T, Harrath A, Altayalan F, Al Quraishy S (2013) African J Pharm Pharmacol 7: 148-156.
55. Baker BR, Ashton WT (1973) J Med Chem 16: 209-214.
56. Kim KH, Dietrich SW, Hansch C, Dolnick BJ, Bertino JR (1980) J Med Chem 23, 1248-1251.
57. Cortese JF, Plowe CV (1998) Mol Biochem Parasitol 94: 205-214.
58. Talisuna OA, Bloland P, D'Alessandro U (2004) History, dynamics and public health importance malaria parasite. Clin Microbiol Rev 17:235-54.
59. Sharma YD (2005) Genetic alteration in drug resistance markers of P. falciparum. Ind J Med Res 121:13-22.

Chapter II: Virtual Screening in Drug Design & Discovery

II.1. Introduction

Traditional Approach of Drug Design and Discovery

In the field of therapeutics and medicine a drug discovery is an integrated process by which new drug candidate are discovered. Traditionally, new drug molecules were discovered through identifying the active ingredient from traditional remedies or by serendipitous discovery. As time went and a several number of therapeutically applicable molecules were discovered and

a library of molecule with same or different activity. Later chemical libraries with identical or similar parent scaffold of synthetic small molecules, natural products (plants, marine, animals) or extracts were randomly or desirably screened against to specific cells or a whole organism to identify the desirable therapeutic activity/effect known as classical pharmacology. Synthesizing list/derivatives of compound and patenting them is an essential phenomenon to secure the pharmaceutical industry data in preventing the scientific and economical loss.[1]

For an individual drug candidate a pharmaceutical industry synthesize >1000 structural derivative and depending on the initial screening results protection of such data is carried out, which significantly increases the cost of drug discovery procedure in its initial phases which is directly proportionate to the number of molecule you design, synthesize and test [1]. Les différentes étapes sont illustrées schématiquement dans la Figure II.1. Au cours des phases successives, des milliers de molécules doivent être triées et sélectionnées, afin d'obtenir un nombre très limité de candidats.

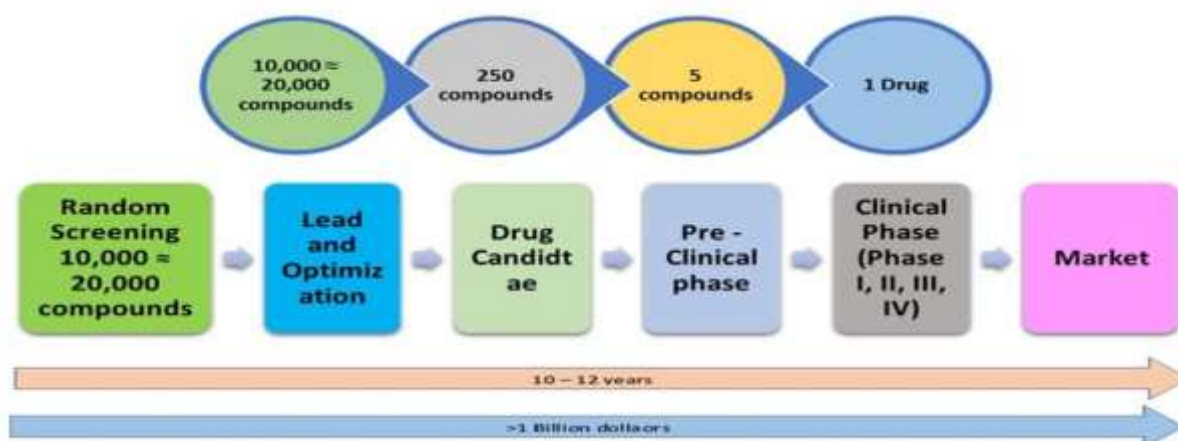


Figure II.1: Traditional drug research and development procedure [2].

In general, it is estimated that the medication research and development process takes 10-12 years and costs more than \$1 billion in total. As a result, computer assisted drug design (CADD) is extensively employed as a novel drug design [2]. In recent times virtual experimentation in CADD has become known as an innovative way of testing high performance especially in terms of low cost and the chances of obtaining the most suitable novel beaten by a large filter of compound libraries[3,4].

II.2. Virtual screening

In biomedical arena, the process of drug development and discovery is very challenging, expensive and time consuming [5]. Regulatory agencies as well as pharmaceutical industry are actively involved in development of computational tools that will improve effectiveness and

efficiency of drug discovery and development process, decrease use of animals, and increase predictability. It is expected that the power of computer aided drug design also known as *in silico* screening will grow as the technology continues to evolve [1,6].

Virtual screening is a set of computational methods or *in silico* analogues of biological screening [7]. It has been used as the most useful tool now in the day to find the most interesting bioactive compounds [4]. The aim of VS is used to discover new drug candidates from 3-dimensional chemical structure databases with the help of information about the target protein or known active myths.[4,5]

It is intended to reduce the size of chemical space and thereby allow focus on more promising candidates for lead discovery and optimization. The goal is to enrich set of molecules with desirable properties (active, drug-like, leadlike) and eliminate compounds with undesirable properties (inactive, reactive, toxic...etc). The rapid growth of virtual screening is evidenced by increase in the number of citations matching keywords “virtual screening”. [5]

Major types of approaches in Virtual Screening Drug Design (VSDD) [7]:

There are types of approaches for drug design through VSDD is the following (Figure II.2):

- Ligand-based virtual screening (LBVS),
- Structure-based virtual screening (SBVS).

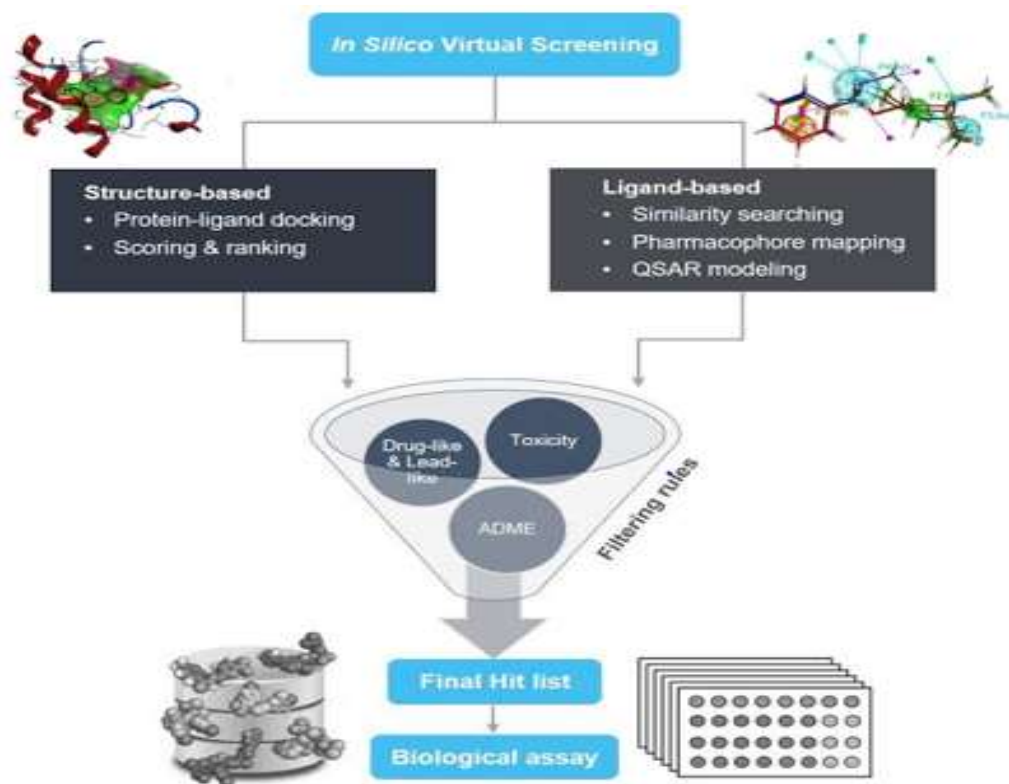


Figure II.2: Schematic illustration of ligand-based and structure-based approaches in VSDD.

II.2.1. Ligand-based virtual screening:

When at least one ligand of the target under investigation is known, ligand-based virtual screening can be implemented. The basic principle common to all ligand-based methods is that similar molecules will tend to exhibit similar activity profiles [8]. The similarity of molecules can be measured by looking for common properties, which are used as similarity descriptors. Depending on the number of reference ligands for the target and the type of descriptors, different methods can be employed: similarity search, screening using QSAR and pharmacophore methods.

II.2.1.1. Similarity-Based Virtual Screening

Similarity-based virtual screening and candidate ranking are considered to be one of the most powerful tools in medicinal chemistry[9,10] and have been successfully applied in a number of cases.[11] Similarity searching [7] is used for finding those compounds which are most similar to a query compound in a database. This involves comparing the query compound with every compound in the database in turn and returns a ranked list of all the compounds that are judged to be similar to the query. Similarity searching in chemical databases was first introduced in the mid-1980s [12, 13]. The rationale for similarity searching is the “similar property principal” which states that structurally similar molecules will exhibit similar properties and biological activity [14].

Similarity searching programs can generally be categorized into 2D and 3D similarity according to whether 3D conformation information is considered:

- The most commonly used similarity method is based on 2D fingerprints [7] and there are numerous studies and reviews of similarity coefficients [15, 16]. Similarity coefficients can be classified into three major classes namely: association coefficients, correlation coefficients, and distance coefficients [16]. 2D similarity methods are efficient for quickly profiling neighboring compounds. However, it may to some extent provide different hits for the same queries as different 2D similarity definitions target different aspects of the information. This method also tends to discover close structural analogues instead of novel scaffold hits[11,17].

Apart from 2D fingerprint-based methods, similarity matching using algorithm is also used for graphical descriptors that can compare objects represented as a graph. Recently algorithm is able to perform tens of thousands of comparisons in a very short time. [18] Evaluated both graph-based and fingerprint-based measures of structural similarity. The results show that, in VS, there is no statistically significant difference in the number of active molecules retrieved by graph-based and fingerprint-based approaches. Another

type of similarity search system is text-based molecular description. Recently, [19] introduced a new algorithm into QSAR model, LINGO, based on the fragmentation of SMILES strings into overlapping substrings of a defined size.[7]

- However, 3D similarity methods [11] typically consider multiple aspects of the 3D conformation, including pharmacophores, molecular shapes, and molecular fields. 3D methods can be conveniently used to accomplish scaffold hopping to identify novel compounds. Based on the pharmacophore matching approach, which was used as the engine of the previously mentioned PharmMapper Server[20], a method named SHAFTS (SHApe-FeaTure Similarity) has been developed for rapid 3D molecular similarity calculation. This method adopts hybrid similarity metrics of molecular shape and colored (or labeled) chemistry groups annotated by pharmacophore features for 3D calculation and ranking in order to integrate the strength of both pharmacophore matching and volumetric similarity approaches. The triplet hashing method is used to enumerate fast molecular alignment poses. The hybrid similarity consists of shape-densities overlaps and pharmacophore feature fit values and is used to score and rank alignment modes[17].

II.2.1.2. Quantitative structure activity relationship

Among the virtual screening approaches, *Quantitative structure activity relationship* (QSAR) is the most powerful method due to its high and fast throughput [21]. The process by which a chemical structure is correlated with a specific effect such as biological activity or chemical reactivity, toxicity, or other kinds of activities on their molecular characteristics [22,23]. In 1868, A. Crum-Brown and T.R Fraser-formulated a suggestion that physiological activity of molecules depends on their constitution [21]:

$$\text{Activity} = F(\text{Structure}) \quad \text{Eq II.1}$$

The QSAR analysis includes all statistical methods by which biological activities (most often expressed as logarithms of equipotential molar activities) are linked with structural elements (Free Wilson analysis), physico-chemical properties (Hansch analysis) or various field-related parameters helping to describe the structure (3D QSAR). Different types of tools can be used: multi-linear regressions (MLR) [24], partial least squares (PLS) regressions [25], neural networks [26-28].

QSAR methodologies based on dimensionality (Table II.1) [21] have the potential of decreasing substantially the time and effort required for the discovery of new medicines. A major step in constructing the QSAR models is to find a set of molecular descriptors that represents variations of the structural properties of the molecule. The QSAR analysis employs

statistical methods to derive quantitative mathematical relationship between chemical structure and biological activity. The process of QSAR modeling can be divided into three stages: development, model validation and application.[29]

Table II.1. Classification of QSAR methodologies based on dimensionality. [21]

1D-QSAR	Molecular representations and molecular fragments i.e., pKa, log P with biological activity.
2D-QSAR	Contains topological information i.e., physicochemical properties with biological activity.
3D-QSAR	Correlation of various 3D properties which surrounds the molecule.
4D-QSAR	Ligand receptor interactions of the drug molecule with the 3D properties.
5D-QSAR	Representing different induced-fit models in 4D-QSAR.
6D-QSAR	Incorporating different salvation models in 5D-QSAR.

➤ The objective

The aim of QSAR method is therefore to analyze the structural data in order to detect the determining factors for the measured property.

The information extracted from QSAR study results can be used to obtain a better understanding of molecular structures and probably the mode of action at the molecular level. This information can then be used to predict the structural properties and biological activities of new compounds, as well as to design new structures (Figure II.3) [30].

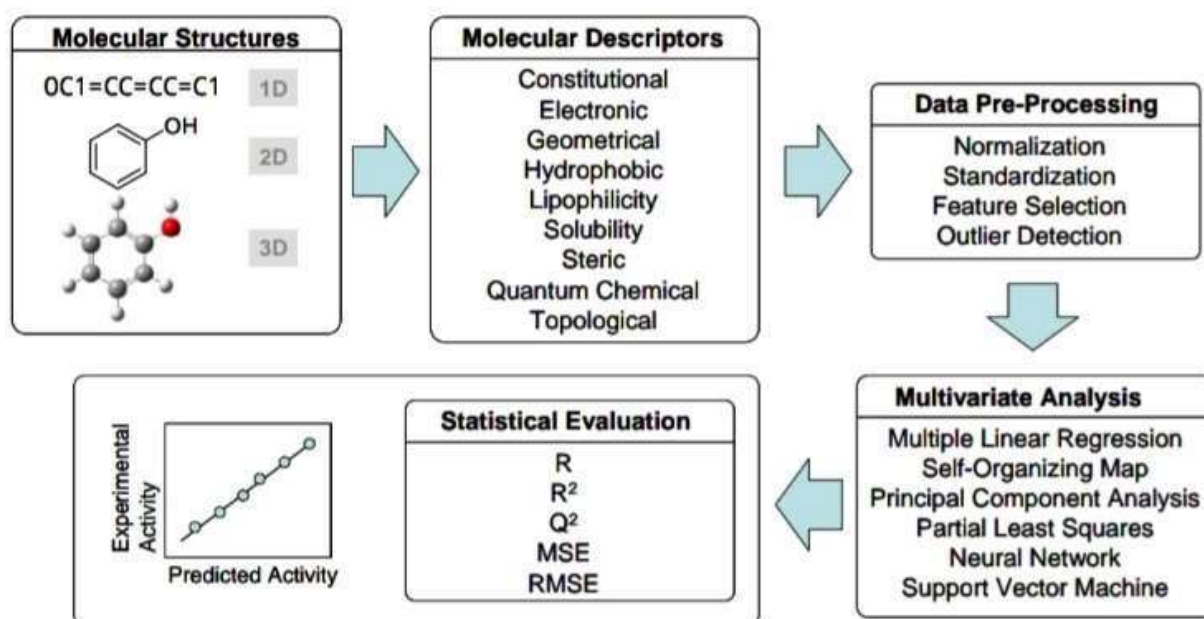


Figure II.3: Layout of Quantitative Structure-Activity Relationship Analysis [21].

II.2.1.2.1. General methodology of QSAR study

A. Biological data collection

By its very construction, a QSAR model is highly dependent on experimental reference data. The choice of database is therefore a critical point in its development. In most cases, experimental data are taken from the literature. They are usually expressed on a logarithmic scale, due to the linear relationship between response and dose logarithm in the central region of the log dose-response curve. Inverse logarithms of activity ($\log 1/C$) are also used to obtain higher mathematical values when structures are biologically highly effective. Examples of biochemical or biological data used in QSAR analysis are described in Table II.2 [31].

Table II.2. Types of biological data used in QSAR analysis.

Source of activity	Biological parameters
<p>1. Isolated receptors</p> <p>Speed constant</p> <p>Michaelis-Menten constant</p> <p>Inhibition constant</p>	<p>$\text{Log } k$</p> <p>$\text{Log } 1/K_m$</p> <p>$\text{Log } 1/K_i$</p>
<p>2. Cellular systems</p> <p>Inhibition constant</p> <p>Cross-resistance</p> <p>In vitro biological data</p> <p>Gene mutation</p>	<p>$\text{Log } 1/IC_{50}$</p> <p>$\text{Log } CR$</p> <p>$\text{Log } 1/C$</p> <p>$\text{Log } TA_{98}$</p>
<p>3. In vivo systems</p> <p>Bioconcentration factor</p> <p>In vivo reaction rates</p> <p>Pharmacodynamic rates</p>	<p>$\text{Log } BCF$</p> <p>$\text{Log } I$ (induction)</p> <p>$\text{Log } T$ (clairance totale)</p>

A database should come from the same analysis protocol, and care should be taken to avoid inter-laboratory variability, to be of high quality of data. Any bad data points will tend to corrupt the correct correlation of structure and activity. [32] Data should be composed of experimental data that are as reliable as possible, since error bars on them will propagate into the final model, as the latter's parameters are adjusted against them. It is therefore important to choose data with low uncertainties in order to limit experimental error bars. Indeed, a model cannot be statistically more robust than the theoretical data from which it was developed.

As far as possible, the database should also be characteristic of the range of values that biological activity may encounter, since it is partly on this criterion that the applicability domain of the final model will be determined. As a general rule, the larger the latter, the more predictive models over a wide range of values can be expected.

The rules of thumb for a good QSAR data set are that the dose-response relationship and activity (or affinity) should be reproducible, the activity range should extend two or more orders of magnitude from the least active chemical to the most active in the series, the number of chemicals used to build the QSAR model should be large enough for stability, the activities of the chemicals should be evenly distributed throughout the activity range, and the chemicals chosen for the training set should possess sufficient structural diversity to cover the range of chemistry space associated with the biological activity under study.[33]

B. Molecular descriptors

A descriptor is a numerical or textual value resulting from an operation performed on a certain representation of the molecule to be described. Many types of chemical structure descriptors are available from commercial software. Descriptors can be grouped according to the way they are encoded (textual, numerical or vector representation), the type of information they carry (physico-chemical, topological, pharmacophoric descriptor, etc.), or the dimensionality of the representation of the molecule from which they have been calculated. There are several thousand of them, and Todeschini attempts to draw up an exhaustive inventory in a leading book in the field [34].

Obtaining a statistically robust model depends very much on the ability of the descriptors selected to encode the variation in activity with structure. The more that is known at the molecular level about the biological mechanism of action of chemicals, the better the chemist can choose from the wide variety and types of specific molecular descriptors in terms of their dimensions. [35] In view of this element, Table II.3 offers a valuable example of largely used molecular descriptors depending on dimensions [36]. Commercially available molecular modeling programs often include statistical tools to help assess which descriptors best encode structure-activity variation. [37]

Table II.3. Popularly known molecular descriptors dependent on various dimensions [38].

Dimension of descriptors	Parameters
0D	Constitutional indices, molecular property, atom, and bond count.
1D	Fragment counts, fingerprints.

2D	Topological, structural, physicochemical parameters including thermodynamic descriptors.
3D	Electronic, spatial parameters, MSA parameters, MFA parameters, RSA parameters.
4D	Volsurf, GRID, Raptor, etc. derived descriptors.
5D	These descriptors consider induced-fit parameters and aim to establish a ligand-based virtual or pseudoreceptor model. These can be explained as 4D-QSAR 1 explicit representation of different induced-fit models. Example: flexible-protein docking.
6D	These are derived using the representation of various solvation circumstances along with the information obtained from 5D descriptors. They can be explained as 5D-QSAR 1 simultaneous consideration of different solvation models. Example: Quasar.
7D	They comprise real receptor or target-based receptor model data.

C. Development of statistical models

After collecting and identifying signal-producing descriptors that correlate with the target variable (biological activity), and noise-producing descriptors that don't, the next step is to pick the descriptors to be used in the created model. Statistical analysis is also used to identify which descriptors are correlated with each other, so that only the most important ones are retained, thereby reducing redundancy of information. In general, methods for designing the QSAR model could be divided into two groups: (i) Classical variable selection and (ii) Variable selection by artificial intelligence algorithms [39]. The choice of method depends mainly on the question being asked and the nature of the data to be processed.

➤ Statistical Methods in QSAR

(i). Multiple linear regression (MLR)

Multilinear regression is of fairly restrictive use in QSAR [40], and is the simplest and most widely used method for developing predictive models [41]. It requires a very complete dataset in which all substituent combinations have been tested. In other words, it requires as many experiments as possible variations, which is rarely the case in practice.

A regression analysis based on the assumption that there is a linear relationship between a dependent variable Y that depends linearly on several independent variables X_1, X_2, \dots, X_j is

called multiple linear regression. The multiple linear regression equation is of the form: $Y = f(X_1, X_2, \dots, X_j)$ where $f(X_1, X_2, \dots, X_j)$ is a linear function of X_1, X_2, \dots, X_j [42]. The aim is to obtain a mathematical equation that can take the following form:

$$Y^i = a + \sum_{j=a}^{j=p} X_j^i \quad \text{Eq II.2}$$

Where (b) are the regression coefficients and (a) the regression constant.

(ii). Partial least squares (PLS)

Partial least squares (PLS) regression, a generalization of multiple linear regression, can be used when the number of descriptors is high and they are highly correlated [43, 44]. It uses a linear transformation to find the axes that best represent the data in space. One of the advantages of this regression method lies in its ability to handle large databases with many correlated variables [45].

PLS gives a statistically robust solution even when the independent variables are highly interrelated, or when the independent variables exceed the number of observations. PLS is an iterative regression method that produces its solutions based on the linear transformation of a large number of original descriptors into a small number of new orthogonal terms called latent variables. As such, this method is counted as a standard statistic [46]. The latent variables T (known as X-score) and U (Y-score) are derived from the large collection of descriptors and the responses (biological activity). The obtained latent variable T (X-score) is used to predict the U (Y-score) and, then, the U (Y-score) is used to predict the response (biological activity) [47].

(iii). Artificial neural networks (ANN)

ANNs are useful methods in QSAR studies, and particularly in cases where it is difficult to specify an exact mathematical model to describe a given structure property relationship. [48] The method of artificial neural networks originates from the real neurons found in an animal brain. ANNs are parallel computing systems made up of groups of highly interconnected processing elements called neurons, arranged in a series of layers. Each layer can perform its calculations independently, and can transmit the results to another layer. In this way, the result of the transfer function is communicated to the neurons in the output layer. This is the point at which the results are finally interpreted and presented. [48] Figure 3 shows an example of neural network architecture.

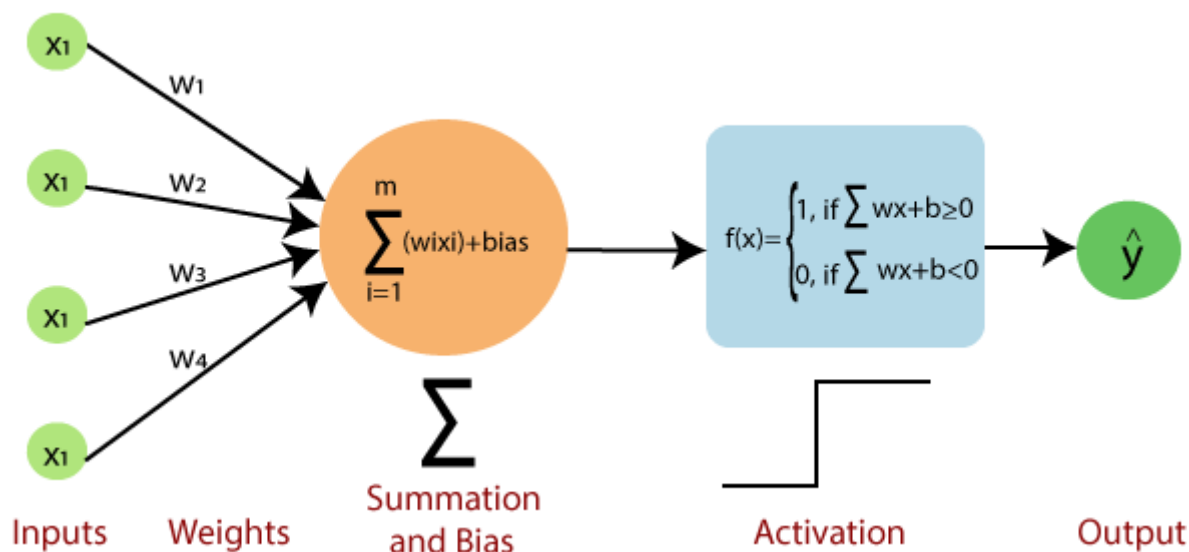


Figure II.4: Neural network architecture.

D. Validation of QSAR model:

In order to assess the importance of QSAR models and, consequently, their ability to predict the activities/properties of other (new) compounds, the validation of QSAR models remains a very sensitive stage in statistical studies. Since a model is the result of a statistical analysis, it must be interpreted and applied within the very precise framework of the domain covered in the analysis [49].

The requisite condition for the validity of the regression model is that the multiple correlation coefficients R^2 and cross validated determination coefficient Q^2 is as close as possible to one and the standard error of the estimation is small; although the former is not essentially a very good predictor of fitness. Apart from the use of fitness parameters to judge the statistical quality of the model, validation of QSAR models is carried out using two major strategies [50]: (i) internal validation using the training set molecules, and (ii) external validation based on the test set compounds by splitting the whole data set into training and test sets.

Table II.4. Statistical parameters for fitting and cross-validation [51].

statistic	Definition	Equations and terms	Threshold
R^2	Coefficient of multiple determination (or correlation)	$R^2 = 1 - \frac{\sum (y_{obs} - y_{cal})^2}{\sum (y_{obs} - \bar{y}_{train})^2}$ <p>y_{train} = mean value of the observed activity of the training set compounds.</p>	$R^2 > 0.6$
R_{adj}^2	Adjusted R^2	$R_{adj}^2 = \frac{\{(n-1) - R\}}{n - p - 1}$ <p>n est le nombre des observations (les molécules) ; p est le nombre de variables indépendantes (les descripteurs) ; R est le coefficient de détermination du modèle.</p>	$R_{adj}^2 > 0.6$

SE	Standard error of estimate	$SE = \sqrt{\frac{\sum(y_{obs} - y_{cal})^2}{(n - p - 1)}}$ <p><i>Y_{obs} and Y_{cal} are the observed (experimental) and estimated scores respectively, while n is the number of compounds and p is the number of descriptors</i></p>	SE should be low for a good model
F	F -value	$F = \frac{\sum(y_{cal} - \bar{y})^2 / (p)}{\sum(y_{obs} - y_{cal})^2 / (n - p - 1)}$ <p><i>Y_{obs} is the observed response, Y_{calc} is the calculated response, n defines the total number of compounds and predictor variables is denoted as p.</i></p>	$F > F$ of fisher table
$PRESS_{CV}$	Predictive residual sum of squares (cross-validation)	$PRESS = \sum (y_{obs} - y_{cal})^2$	$PRESS_{CV}$ should be low for a good model
Q^2_{LOO}	Explained variance in prediction	$Q^2_{LOO} = 1 - \frac{\sum(y_{obs} - y_{pred})^2}{\sum(y_{obs} - \bar{y}_{train})^2}$	$Q^2 > 0.5$
MAE_{CALC}	Mean absolute error in fitting (calculated on training set)	$MAE_{CALC} = \frac{\sum y_{obs} - y_{pred} }{n}$	<ul style="list-style-type: none"> • Good predictions: $MAE \leq 0.1 \times$ training set range, AND $MAE + 3 \times \sigma \leq 0.2 \times$ training set range. • Bad prediction: $MAE > 0.15 \times$ training set range; OR MAE training set rang $MAE + 3 \times \sigma > 0.25 \times$ training set range e.
CCC	Concordance correlation coefficient	$CCC = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$	$CCC < 1$

Besides these techniques, randomization or Y-scrambling and determination of the applicability domain (AD) of the model and selection of outliers are other vital aspects in the course of developing a reliable QSAR model with the spirit of OECD principles.

Table II.5. Statistical parameters for external validation [51].

statistic	Definition	Equations and terms	Threshold
R^2_{pred}	R^2 prediction	$R^2_{adj} = 1 - \frac{\sum(y_{obs} - y_{cal})^2}{\sum(y_{obs} - \bar{y}_{train})^2}$	$R^2_{prd} > 0.6$
Q^2_{F1}	Variance explained in external prediction	$Q^2_{F1} = 1 - \frac{\sum(y_{obs(test)} - y_{cal(test)})^2}{\sum(y_{obs(test)} - \bar{y}_{train})^2}$	$Q^2_{F1} > 0.5$
Q^2_{F2}	Variance explained in	$Q^2_{F2} = 1 - \frac{\sum(y_{obs(test)} - y_{cal(test)})^2}{\sum(y_{obs(test)} - \bar{y}_{test})^2}$	$Q^2_{F2} > 0.5$

	external prediction		
Q_{F3}^2	Variance explained in external prediction	$Q_{F3}^2 = 1 - \frac{[\sum (y_{obs(test)} - y_{cal(test)})^2] / n_{test}}{[\sum (y_{obs(test)} - \bar{y}_{test})^2] / n_{train}}$	$Q_{F3}^2 > 0.5$
$r_{m(rank)}^2$	Closeness between the R^2 and R_0^2 determination coefficients	$r_{m(rank)}^2 = r_{(rank)}^2 \times (1 - \sqrt{r_{(rank)}^2 - r_0^2(rank)})$	
\bar{R}_r	An average of the correlation coefficient for randomized data		$\bar{R}_r < 0.5$
\bar{R}_r^2	An average of the correlation coefficient for randomized data		$\bar{R}_r^2 < 0.5$
\bar{Q}_r^2	An average of leave one out cross for randomized data -validated determination coefficient		$\bar{Q}_r^2 < 0.5$
${}^cR_p^2$	$cRp2 = R^2 \times (1 - \sqrt{ R^2 - \bar{R}_r^2 })$		${}^cR_p^2 > 0.6$
CCC	Concordance correlation coefficient	$CCC = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$	$CCC < 1$

E. Applicability domain (AD):

To evaluate the reliability of any QSAR model and its power to predict new compounds, the domain of applicability must be essentially defined. The applicability domain plays a crucial role for estimating the uncertainty in the prediction of a particular compound based on how similar it is to the compounds employed to construct the QSAR model. The AD is defined as a theoretical region in the chemical space constructed by both the model descriptors and modeled response. Therefore, the prediction of a modeled response using QSAR is applicable only if the compound being predicted falls within the AD of the model as it is unfeasible to predict the whole universe of compounds using a single QSAR model [52,53].

Various methods are in place to assess the AD of QSAR models. From the QSAR publications of the last decade, the most widely used method for estimating interpolation regions is the leverage approach (Williams plot) [54], in which the standardized residuals and the leverage values (h_i) are plotted. It is based on the calculation of the leverage h_i for each compound, for which QSAR model is used to predict its activity:

$$h_i = x_i(X^T X)^{-1} x_i^T \quad \text{Eq II.3}$$

Where x_i is the row vector of the descriptors of compound i and X is the variable matrix deduced from the training set variable values. The index T refers to the matrix/vector transposed. The critical leverage h^* is, generally, fixed at $3(k + 1)/N$, where N is the number of training compounds, and k is the number of model parameters. If the leverage value h of a compound is higher than the critical value (h^*) i.e., $h > h^*$, the prediction of the compound can be considered as not reliable. [55]

II.2.1.2.2. 2D and 3D QSAR analysis

A. 2D QSAR

Classical 2D QSAR analysis (Hansch and Free Wilson analyses) consider only two dimensional (2D) structures. The principle of QSAR methods is, as the name suggests, to establish a mathematical relationship between molecular properties, both electronic and geometric, called descriptors, and a macroscopic observable (e.g. biological activity), for a series of similar chemical compounds, using data analysis methods.

The data used to form the 2D-QSAR equation is represented by a matrix of numbers, with each row representing a compound and each column representing physicochemical properties (descriptors). In 2D-QSAR, there are numerous descriptors. Those most often used are constants logP, MW, RM ...etc. A large number of constant values are collected and a statistical analysis process exploits them to find the relationship between biological data and molecular descriptors [56]

The basic formalism of the QSAR method will result from statistical analyses. The simple mathematical relationship is defined as follows (Eq II.4) [57]:

$$\text{Function} = f(\text{structural molecular or fragment properties}) \quad \text{Eq II.4}$$

To prevent these relationships from being statistically insignificant, or in the event of a one-off error, apply the following approach:

- The ratio of compounds to descriptors must be greater than 5,
- Descriptors must be uncorrelated. The degree of inter-correlation is assessed by the correlation coefficient r . [58]

In addition, the multi-linear regression (MLR) method can be used to solve these problems [24].

B. 3D QSAR

Three-dimensional quantitative structure activity relationships (3D-QSAR) models are useful in the process of new drug design and development as their application helps to reduce the cost and time of the synthesis of medicinally active compounds [59–61].

3D-QSAR are models that establish a relationship between biological activity and structural parameters (molecular descriptors) calculated in three-dimensional space for a group of molecules. 3D QSAR allows the prognosis of activity of structurally varied molecules and also assist in identification of new molecules with enhanced activity [62-67]. The different colored squares generated in 3D QSAR studies provide an idea about essential structural features for better biological activities [68- 70].

The CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis) models are commonly used to study the quantitative structure-activity relationship at the three-dimensional level [71,72].

In these methods, molecule properties are described by different fields. For example:

- The molecular surface accessible to the solvent, connolly or contact surface,
- Electrostatic potential (position of charged groups),
- Participation in hydrogen bonds,
- Molecular lipophilicity potential,
- Molecular orbitals,
- The shape of the molecule.

The principles of these models are based on the use of mesh networks (figure II 5).

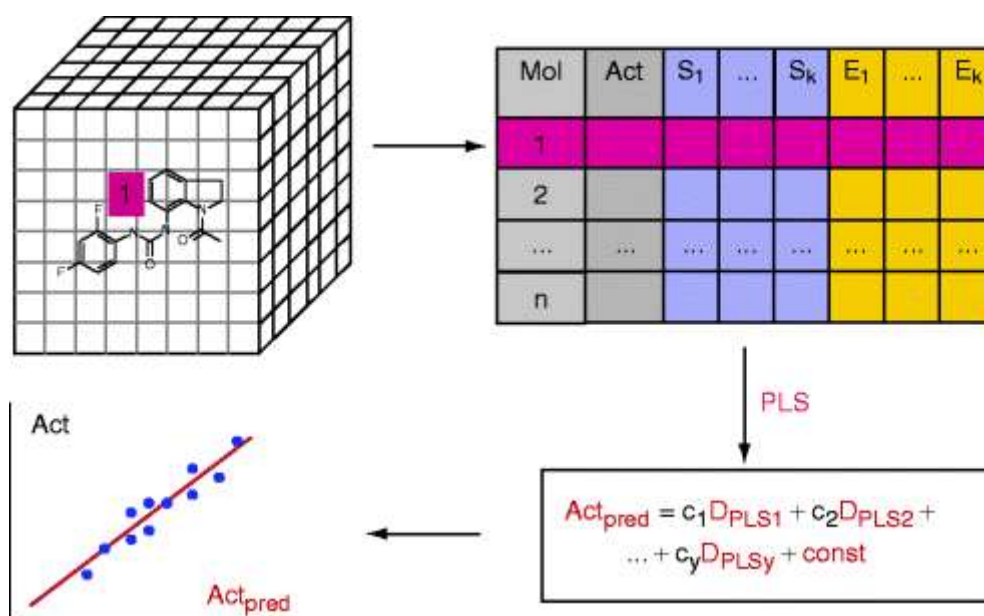


Figure II.5: Flowchart to construct 3D-QSAR model.

➤ Comparative Molecular field Analysis (CoMFA)

The CoMFA is a molecular field-based, alignment-dependent, ligand-based 3D QSAR method which generates a quantitative relationship of molecular structures of a ligand can be predicted from its three dimensions and its biological response [71, 73].

To date, CoMFA is probably the most widely used 3D-QSAR method. A CoMFA study normally begins with traditional pharmacophore modeling to suggest a bioactive conformation for each molecule and ways of superimposing the molecules under study. [71]

The idea behind CoMFA is that differences in a target property, for example, biological activity, are often closely related to equivalent changes in the shapes and strengths of the non-covalent

interaction fields surrounding the molecules employing linear regression methods such as partial least squares (PLS) [25].

Or stated differently, steric and electrostatic fields provide all the information needed to understand the biological properties of a set of compounds [74]. Accordingly, molecules are placed in a cubic grid and the interaction energies between the molecule and a defined probe are calculated for each grid point. [71] Figure II.6 shows the steric and electronic field in a CoMFA grid.

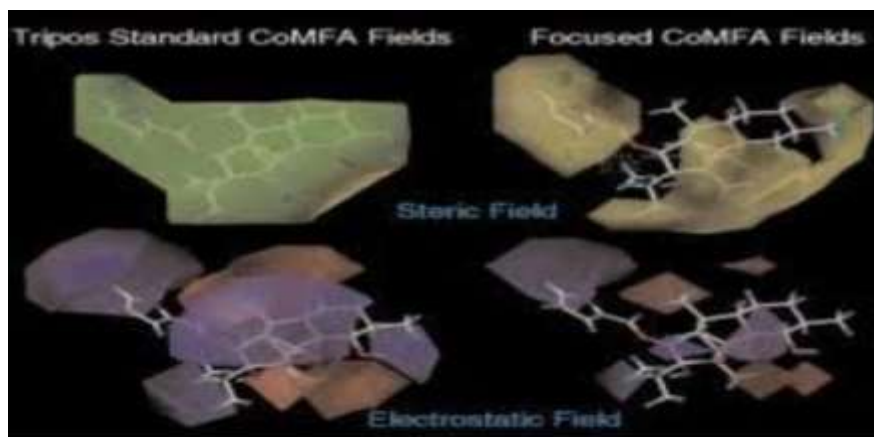


Figure II.6: The steric and electrostatic field in a CoMFA grid [73]

In a CoMFA study, correct alignment of the molecules is essential but often problematic. Optimal alignment can be defined as achieving maximum superposition of the steric and electrostatic fields of a set of molecules. Alignment varies from molecule to molecule, depending on structural similarity or diversity. Consequently, alignment significantly influences model results, and meaningful and relevant results should only be expected for valid alignments. [75]

The CoMFA has the ability to design of new ligands in the structure–activity correlation problems. Along with a good physicochemical property or response.[76]

- **Display and Interpretation of CoMFA-Results [76]**

The results are displayed for a CoMFA model by two ways:

(a) Coefficient contour plots: It portrays vital regions in space around the compounds where specific structural modifications appreciably vary with the response. In CoMFA, two types of contours are shown for each interaction energy field: (i) the positive and (ii) negative contours which are depicted by some specific colors.

(b) Plots from PLS models: Two types of plots are generally created: (i) score plots and (ii) loading/weight plots. The score plots between biological response (Y-scores) and latent variables (X-scores) show relationships between the activity and the structures, whereas plots

of latent variables (X-scores) display the similarity/dissimilarity between the molecules, and their clustering predispositions.

➤ **Comparative Molecular Similarity Indices Analysis (CoMSIA)**

Because of the problems associated with the functional form of the Lennard-Jones potential used in most CoMFA methods, Klebe et al. have developed a CoMFA method based on similarity indices called CoMSIA (comparative analysis of molecular similarity indices). Instead of grid-based fields, CoMSIA is based on similarity indices obtained using a functional form adapted from the SEAL algorithm [77].

Three different indices related to steric, electrostatic and hydrophobic potentials were used in their study of the classical steroidal reference dataset. In CoMSIA, five different similarity fields are calculated at regularly spaced grid points for the aligned molecules [76].

- Steric,
- Electrostatic,
- Hydrophobic,
- Hydrogen bond donor, and
- Hydrogen

Models of comparable statistical quality with regard to the cross-validation of the training set, as well as the predictivities of a test set, were derived using CoMSIA [78].

- **Advantages of CoMSIA [76]**

The CoMSIA technique provides following unique advantages:

- The ‘Gaussian distribution of similarity indices’ overcomes the unanticipated changes in grid-based probe–atom interactions.
- The choice of similarity probe includes steric and electrostatic potential fields as well as hydrogen bonding and hydrophobic fields.
- The effect of the solvent entropic provisions can also be included by employing a hydrophobic probe.
- The CoMSIA contours indicate those areas within the region occupied by the ligands that ‘favor’ or ‘oppose’ the occurrence of a group with a particular physicochemical property or response.

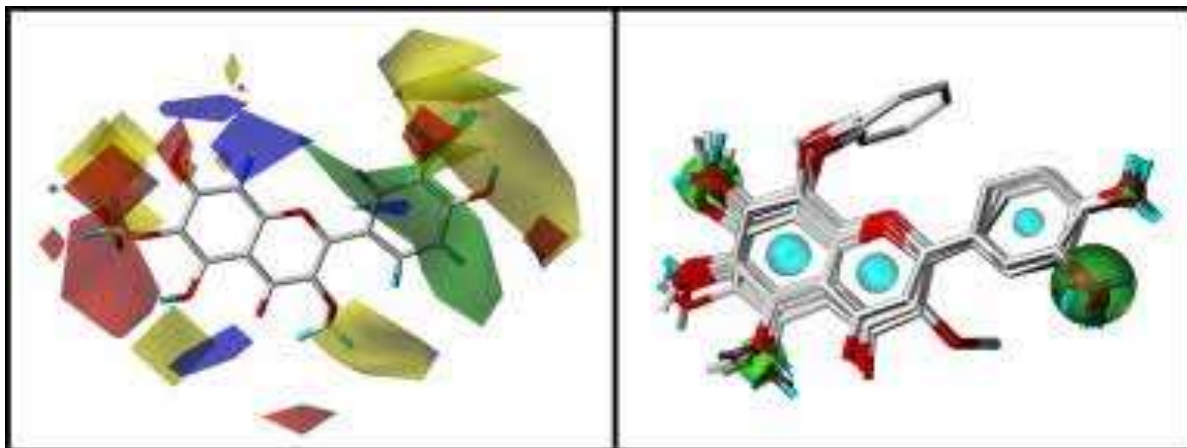


Figure II.7: Contour maps and molecular alignment in a CoMSIA grid.

II.2.1.3. Ligand-based pharmacophore approaches

Pharmacophore approaches have now become one of the main tools used in VSDD [7]. The pharmacophore concept was developed by Ehrlich at the end of the **19th** century [79]. At this time, although the term pharmacophore was not used, Ehrlich developed the idea that certain chemical groups in a molecule are responsible for biological or pharmacological action [80].

According to the official 1998 IUPAC (International Union of Pure and Applied Chemistry) definition, a pharmacophore model is "a set of steric and electronic characteristics necessary to ensure optimal supramolecular interactions between drugs, and a specific biological target to trigger (or block) its biological response" [81].

According to this definition, molecules sharing the same pharmacophore for a given target should therefore bind identically to this receptor and present similar activity profiles. One of the major features of this type of method is that a pharmacophore is defined by mutually complementary pharmacophoric points, which are functional groups rather than groups of atoms. The various pharmacophoric points sought (figure II.8) are hydrogen bond donors and acceptors, positively charged groups which form electrostatic interactions with negatively charged ones and vice versa, and aromatic groups, considered distinctly from the broader class of hydrophobic groups from which they are derived, and both of which are complementary to other hydrophobic groups. [82]

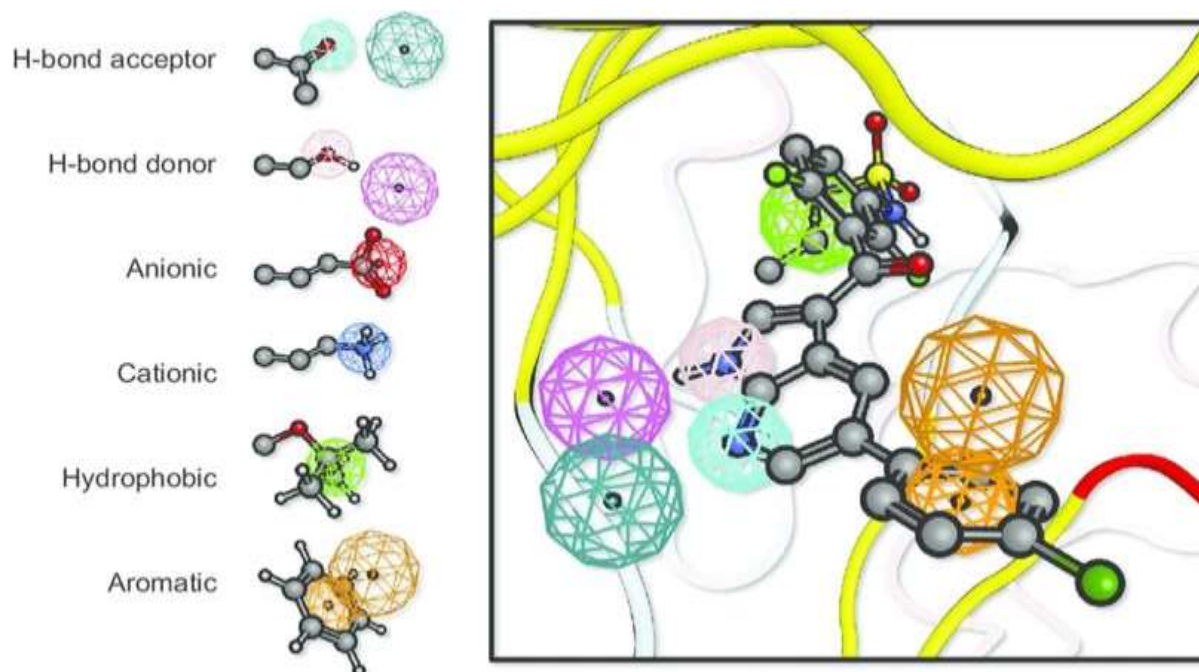


Figure II.8: Overview of pharmacophore mapping.

A pharmacophore is considered to be ligand-based when it is determined from the structure of active reference compounds without knowing or taking into consideration the structure of the receptor. Once the pharmacophore has been obtained, it is used to screen the chemical library for potentially active molecules [83]. Furthermore, it gives good knowledge about molecular interactions of various compounds to their target structure and these features are complimentary to each other in 3-D space. Pharmacophore could be more better though combination with shape and volumes for proper fitting into the site of the receptor because wrong shape prevents fitting of compound into the receptor [84].

II.2.1.3.1. Elucidation of the pharmacophore

The elucidation of a pharmacophore is a complex process divided into several steps [85].

(i) Selection of reference ligands: The reference ligands used to build a pharmacophore are active compounds whose activity on the biological target under study is known and comparable for all.

(ii) Conformational research: The ligands used to create the pharmacophore must be in their bioactive conformations, i.e. the conformation in which they bind to the receptor. However, when this has not been identified, a conformational search must be carried out to include all ligand conformations in the study.

(iii) Determination and representation of pharmacophore points for each ligand: Pharmacophoric points can be of three types: based on atoms, based on topological groups and based on the chemical properties of groups of atoms.

The choice of the type of pharmacophore point is not trivial. Indeed, if the pharmacophoric points of a model are based on atoms or topological groups (for example, an oxygen or a carbonyl), only molecules possessing exactly these atoms or topological groups can be identified as "hits" during screening. Conversely, if the pharmacophore points describe chemical properties, the number of molecules that can satisfy the pharmacophore criteria increases, since different elements can represent the same chemical function (e.g. a nitrogen and an oxygen can both be hydrogen bond acceptors).

(iv) Determining pharmacophore models: Once the pharmacophore points have been defined for each ligand, it is necessary to search for those that are common to the different reference ligands in order to obtain one or more pharmacophore models. To do this, the ligands must be aligned, and algorithms for finding the maximum common substructure (MCS) are generally employed.

(v) Assigning scores to different models and selecting the best pharmacophore(s): In the final step, a score is assigned to the different pharmacophore models, enabling them to be ranked. The various scoring functions used are based on the number and quality of pharmacophore points superimposed between the reference ligands, conformational energy, overlap volume between the different ligands [86], but also the rarity of the pharmacophore.

II.2.2. Structure-Based Virtual Screening

The performance of biochemical processes and cell mechanisms are dependent upon complex and multiple non-covalent intermolecular interactions between proteins and small-molecule modulators. The understanding of the structural and chemical binding properties of important drug targets in biologically relevant pathways allows the design of small molecules capable of regulating or modulating specific target functions in the body that are closely linked to human diseases and disorders, through multiple intermolecular interactions within a well-defined binding pocket [87-90]. It contains important methods which can be used in VS, which are the following:

II.2.2.1. Molecular Docking

Interactions between a protein receptor and its ligand are the basis of most biological mechanisms, so the details of these interactions, at the molecular level, are of great interest and can be studied by docking. [91]

The use of docking methods in the drug design process began over 30 years ago[92]. The main role of this technique is to predict the ability or otherwise of a ligand (substrate, activator or

inhibitor) to bind to the amino acids making up the active site structure of a receptor protein (protein), based on prediction of the conformation and orientation of the molecule on binding to the receptor [93].

Molecular docking takes place in two distinct stages:

- The first step is to position the ligand in the chosen site on the protein.
- The second stage of this method allows the evaluation of energy interactions between ligand and protein. These two steps are different from docking program used [94].

To this end, docking methods combine the use of a search algorithm to generate putative ligand binding modes in the receptor, or "poses", and a scoring function to rank the different poses according to a predicted affinity score [95].

II.2.2.1.1. Docking process

The action of a molecule in a protein is governed by principles of mutual recognition between the molecule and its target. The general characteristics of ligand-protein interactions lie in steric, hydrophobic and polar complementarity between the two structures, and a favorable energy conformation of the ligand that favors good binding affinity [96].

Docking methods seek to assemble proteins to build a complex [97]. The docking process involves the interaction of a small organic molecule with the receptor, usually a protein [98] (figure II.9). The aim is to determine how these molecules will fit together. The problem is therefore to determine the structure of the molecular complex resulting from the association of two molecules of known structure [99].

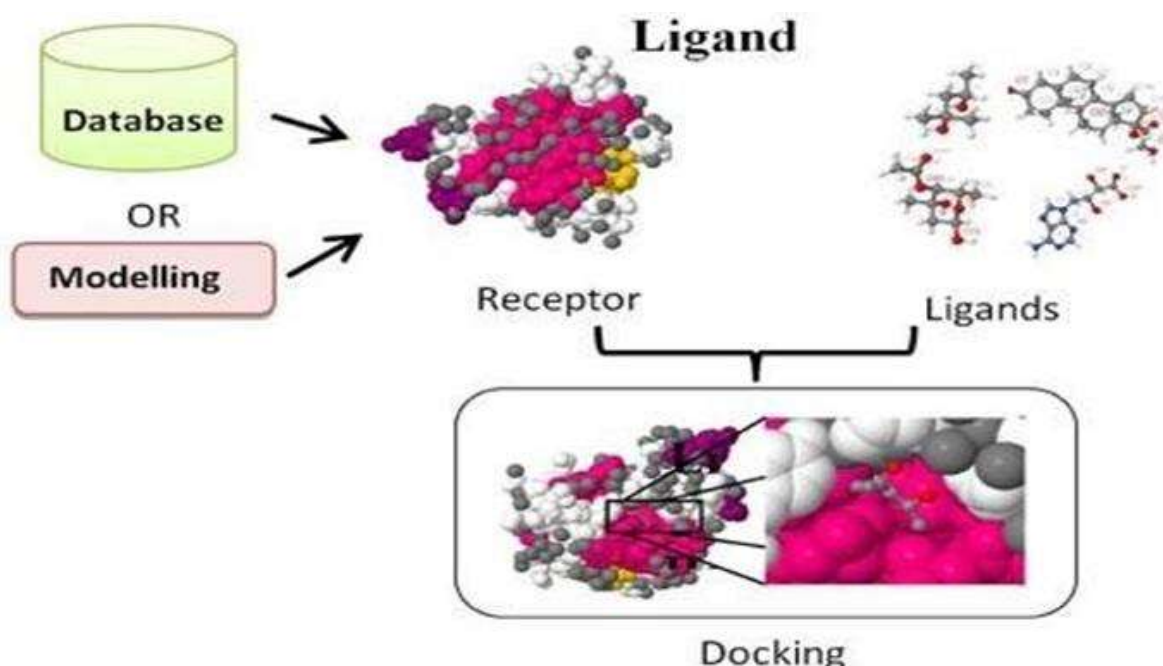


Figure II.9: Outline of process involved in Molecular Docking.

In molecular biology, there are two main docking problems: ligand-protein docking and protein-protein docking.

a. Ligand-protein docking:

This technique requires a large molecule (the protein is also called "the receptor") and a small molecule (the ligand), and is very useful in drug development. The problem to be solved resembles the "key in the lock" situation, when the ligand is docked in the cavity of the protein.

b. Protein-protein docking:

Protein-protein interactions are also extremely important, since they are responsible for many necessary biological functions. Prediction of such interactions is extremely important to the complete understanding of human physiology. Association of two biological macromolecules is a fundamental biological phenomenon and an unsolved theoretical problem. In recent years, several groups have developed a variety of tools in an attempt to solve the so called protein-protein docking problem, that is, the prediction of the geometry of a complex from the atom coordinates of its uncomplexed constituents.[100]

II.2.2.1.2. Search algorithms

Search algorithms for dealing with ligand flexibility can be classified into three broad categories: systematic search algorithms, random or stochastic search algorithms and deterministic or simulation search algorithms[101].

a. Systematic search:

The general principle is to cut the ligand into rigid and flexible fragments. First, one or more fragments that are to be rigid are placed within the active site and thus brought into interaction with the target, then the ligand is reconstructed by placing the flexible fragments in succession while exploiting the angles of twist. [102]

Systematic search algorithms aim to explore all ligand degrees of freedom by rotating all rotatable bonds from 0 to 360° using a chosen incremental step. As a result, the number of conformations generated can be very large. (Equation II.5). This is known as the combinatorial explosion.

$$N_{Conformations} = \prod_{i=1}^N \prod_{j=1}^{n_{inc}} \frac{360}{\theta_{i,j}} \quad \text{Eq II.5}$$

Equation II.5. Calculation of the number of possible conformations ($N_{conformations}$) for a ligand with N the number of rotational bonds, n_{inc} the number of increments and $\theta_{i,j}$ the value of the rotational incremental angle j for bond i . [103]

Two types of methods can be used: exhaustive search methods and incremental reconstruction methods.

b. *Random or stochastic search:*

Random search or stochastic algorithms make random changes in terms of translations, rotations and torsions to generate new ligand conformations. Changes are accepted or rejected using of a probability function [93]. Four main types of algorithm fall into this category class: Monte-Carlo methods, genetic algorithms, algorithms of the taboo search and swarm optimization algorithms [104].

In the most widely used "Monte Carlo" method, the ligand is considered as a whole, and changes are made in translations, rotations and torsions. For each movement, the molecule is minimized and its energy calculated. The conformation obtained by this transformation is tested with a selection criterion based on energy. If this criterion is validated, it will be saved and the program will then generate the next conformation. Iterations continue until the predefined number of conformations has been collected. The main advantage of the Monte Carlo method is that the change can be large enough to allow the ligand to cross energy barriers on the potential energy surface. This is a point not easily reached by simulation methods based on molecular dynamics [105].

c. *Deterministic or simulation research:*

Simulation methods are based on the solution of Newton's equations of motion. They include molecular dynamics techniques and minimization algorithms. The former are never used to generate ligand flexibility, as they require a computation time that is incompatible with the management of molecule databases. The latter, on the other hand, are sometimes used in docking programs, as a complement to another search algorithm, in order to achieve a low-energy conformation. [106]

II.2.2.1.3. Score functions

The score function is a useful numerical datum for quantifying the degree to which a ligand complexes with a receptor. Overall, it is an approximation of the free energy resulting from the transition from the free form of the protein and ligand to the association in complex form. The thermodynamic principle is as follows (Eq. II.6): [107]

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{proteine}} \quad \text{Eq.II.6}$$

Establishing a good score function is a major problem in docking. It often happens that the solution evaluated as the most probable is not the expected native form. This may be due to the fact that the native complex is not necessarily the one with the largest access surface, or the

greatest number of available hydrogen bonds. For this reason, there are different types of score functions, depending on the criteria on which they are based[108]. Chemical affinity can be calculated by the Gibbs free energy ΔG for a temperature T[109,110]:

a. Score functions based on force fields:

Force-field-based score functions use molecular mechanics to calculate the interaction energy of the complex and the internal energy of the ligand. G-Score [111]. In molecular mechanics (MM), a force field corresponds to a set of parameters and functions used to define a system and describe its potential energy landscape. Parameters classically include atomic mass, charge, van der Waals radius and various reference values corresponding to inter-atomic bond lengths and plane and dihedral angles. These parameters are generally derived from experiments or quantum simulations usually performed on small organic molecules. Force field functions correspond to mathematical formalisms that integrate these parameters and are used to calculate various types of potential energy [112].

The main limitations of force-field-based score functions stem from the fact that they were written for gas-phase models and therefore do not contain a solvation effect or entropy term. Extensions including an entropy term for the ligand (in G-Score) and protein-ligand hydrogen bonds (in Gold and Autodock) have recently been added[113].

b. Empirical score functions:

Empirical score functions are used to interpret the interaction energy of a receptor-ligand complex from a summation equation of localized chemical interactions. [111]. This type of function is based on multiple regression to adjust the function coefficients according to the physics of the system. Fitting from a dataset of receptor-ligand complexes with measured affinities[114].

Empirical score functions usually contain terms describing ionic interactions, hydrophobic interactions, hydrogen bridges or bonds and interactions generated by entropy change (entropy penalty). However, these score functions sum up these different terms, weighting them with terms describing the different types of molecular interactions[115]. Most docking software uses this type of function because of its efficiency in terms of speed and precision. However, their main drawback is their strong dependence on calibration parameter data.

c. Knowledge-based scoring functions:

These functions are derived from the analysis of the three-dimensional structures of experimentally determined ligand-protein complexes. Rules defining the preferred geometry of interactions are deduced from these structures by statistical means. [116]

These statistical functions make it possible to establish a correlation between the thermodynamic state of the protein-ligand complex that forms a system and the probability of finding this complex in a given microscopic state. Frequencies are converted into free enthalpy (energy) using a Boltzmann distribution, and the potentials are therefore called Potential of Mean Force (PMF)[116].

d. Consensus score functions:

Numerous scoring functions have been developed, but none is universally applicable. Some will perform well on a set of proteins related to those in the training set used to calibrate their parameters, but will be less suitable for proteins with different physico-chemical properties. In addition, each score function has its own advantages and disadvantages with regard to the model formulated to describe the process of association of a ligand with its receptor. These characteristics nevertheless suggest that several score functions should capture different information. It was on the basis of this idea that the application of consensus scoring was introduced by combining the predictions of multiple score functions [118]. Several strategies varying in the way they combine each score have been undertaken and have shown improvement in the prediction of binding mode, affinity or even the identification of ligands that can effectively bind a receptor in virtual screens [119].

II.2.2.1.4. Analysis of results

Two commonly used criteria were utilized for assessment of docking software accuracy.

(i) Root-mean square deviation (RMSD):

The first way to evaluate quality of a docked pose is to compare its geometry relative to the original experimental structure.

Difference between two conformations (or any three-dimensional structures) is often measured by computing root-mean square deviation (RMSD) [120]. RMSD can be calculated using formula (Equation II.7):

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad \text{Eq II.7}$$

Where N is the total number of atoms in the molecule and δ is a distance between each pair of corresponding atoms.

Concerning current docking software accuracy, the RMSD value of 2 Å is commonly used as a cutoff value. Poses closer to the experimentally determined structure (i.e. with RMSD lower than 2 Å) are generally considered sound.

RMSD for heavy atoms was calculated using *RMSD Tool* plugin implemented in an interactive graphics software *VMD* [121].

(ii) Binding score:

Accuracy of a scoring function was measured by a comparison of predicted binding score of a ligand with experimental value of free energy of binding.

While scoring functions employed in docking software tend to use various approximations to enhance their speed, their accuracy is not on a level of more computationally expensive methods. To provide a context, standard error was estimated to be around 2.5 kcal/mol [122].

In comparing binding energies predicted by docking software with experimentally determined values, two criteria are commonly considered.

First one is comparison of absolute values of energy, therefore accuracy of binding energy prediction. On the other hand, in the field of drug design, researchers are often more interested in comparing inhibitor potency relative to each other. For this purpose, docking software should ideally be able to rank the ligands from the most to the least potent (predict the correct binding trend), even if the absolute values of binding energy are not accurate. Two correlation coefficients are commonly used [123] to quantify relationship between actual and predicted binding trend.

II.2.2.1.5. Different types of docking

a- Rigid docking: The ligand is rigid and the search for the optimum position is limited to positioning. We speak of key-locks, considering only rigid bodies capable of interacting when they present perfect geometric compatibility. This concept was introduced by Emile Fisher in 1894 [124].

b- Semi-flexible docking: the ligand is flexible in order to explore all possible conformations, while the target is kept rigid during assembly. Although this type of docking has the advantage of requiring a relatively short computation time, it does not take into account conformational changes required by the target. Numerous studies using this semi-flexible strategy have led to conclusive results [125,126]. However, proteins remain in constant motion and the major challenge today is to introduce the deformations of the receptor (protein).

c- Docking flexible : When docking methods take ligand flexibility into account, two steps are performed successively throughout the docking process:

- The first step is to explore the conformational space in order to find the bioactive conformations among the proposed conformations.

- In the second stage, a score function evaluates these conformations [127]. There are several types of conformational search algorithms.

II.2.2.2. Molecular dynamics

Molecular dynamics (MD) is a computer simulation technique in which the time evolution of a set of interacting atoms can be followed by integrating their equations of motion.[128,129] The first steps in molecular dynamics were only possible with the arrival of the first computers (1957). But the first real simulations were carried out by Rahman, thanks to his work on the simulation of liquid argon in 1964, with a simulation time of 10-11 s, followed by liquid water [130] in 1971.

Molecular mechanics calculates the position of atoms and molecules in their minimum energy states, at 0°K, i.e. without the slightest vibration or movement. Molecular dynamics attempts to simulate the movement of atoms and molecules over time at temperatures above 0°K.

II.2.2.2.1. General Principle

Molecular dynamics (DM) is a technique for simulating the evolution of several molecules over time [131] (figure II.10). It consists in studying the trajectory of a molecule by applying the laws of classical Newtonian mechanics (Equation II.8), to describe the motion of a molecule as a function of time.[129] In this way, the molecule's trajectory can be described as a function of time.

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2} \quad \text{Eq II.8}$$

Where:

\vec{F}_i : Vector force acting on atom i.

m_i : Masse of atom i.

\vec{a}_i : Acceleration Vector of atom i.

\vec{r}_i : Position of atom i.

The velocities and positions of individual atoms over time can be used to evaluate macroscopic data such as kinetic energy and temperature. It can be used to simulate intramolecular movements, which can then be visualized in real time. These movements correspond to vibrations around a minimum, or to the passage from one minimum to another minimum in energy [130].

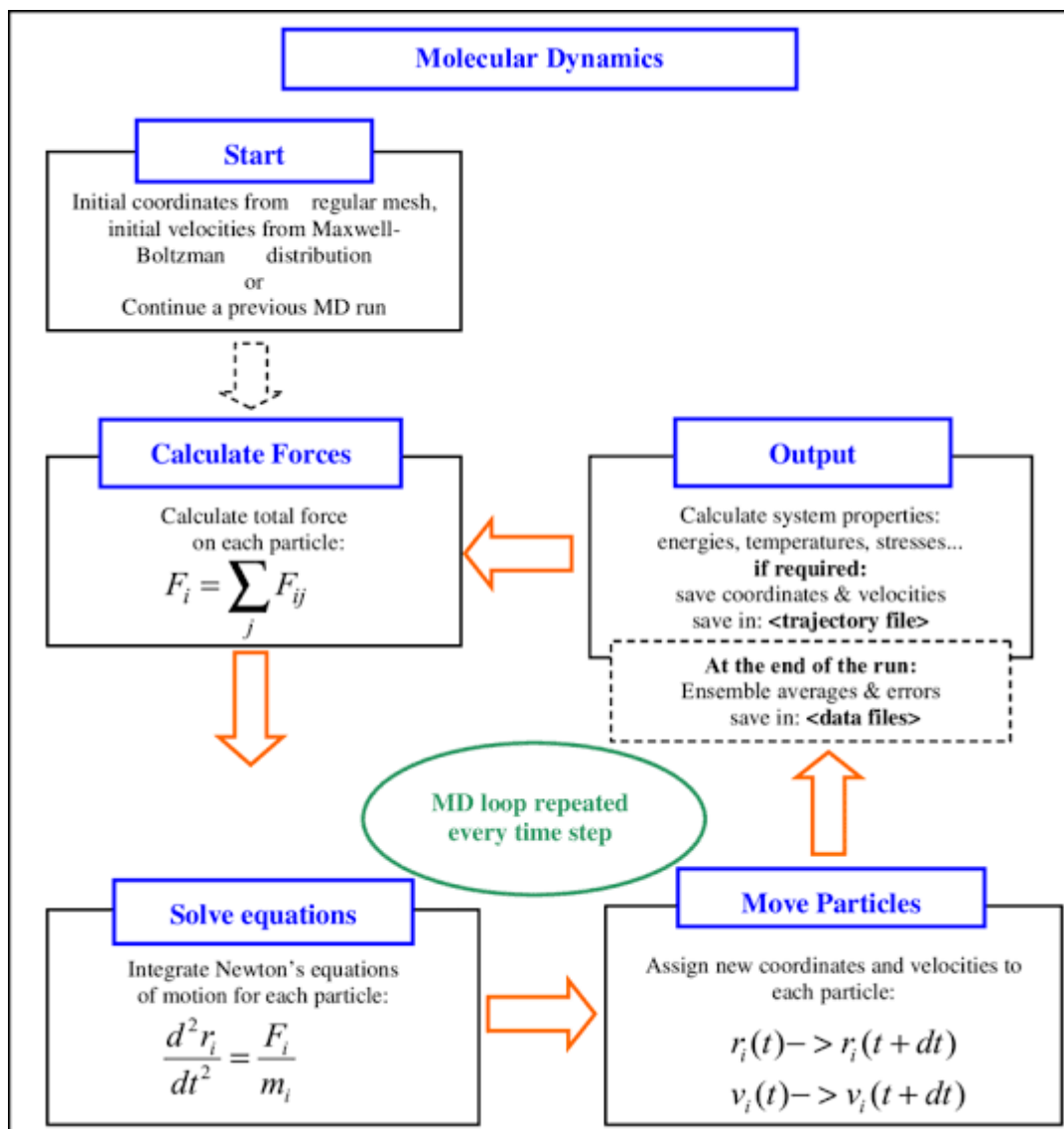


Figure II.10: Outline of process involved in Molecular Dynamic.

However, not all properties and quantities can be directly calculated in an MD simulation and vice versa certain quantities that can be directly estimated in a simulation cannot be tracked in an experiment. A representative example is a simulation of liquid water in which we can measure the coordinates and velocities of each molecule (microscopic properties) at any instance of time. However, there is no experimental method that can produce this kind of information, but rather it will provide us with the averaged properties across a large number of molecules (macroscopic properties). [132]

II.2.2.2.2. Key components of molecular dynamics

There are several key components involved in molecular dynamics [131]:

- **System topology:** defining the set of atoms and their bonds

- **Coordinates and initial velocities of the atoms in the system:** coordinates are usually derived from experimentally determined structures, while initial velocities are often automatically generated.
- **Thermostats/barostats:** to control the temperature and/or pressure of the environment, depending on the thermodynamic set chosen.
- **The force field:** which defines all the forces applied to the system's atoms in the form of potentials. Atoms are usually represented in point form, symbolizing the position of the nucleus. The name "force field" also denotes the parameters associated with the various atoms, bonds and forces (mass, charge, potential constants, etc.).
- **The integrator:** this is the algorithm that generates a new conformation from the coordinates of previous conformations and/or velocities, the force field and any thermostats, barostats and other forces that may have been applied.

II.2.2.2.3. Issues and Limitations of Molecular Dynamics

Molecular dynamics has several issues, some caused by the characteristics of simulated chemical processes, most of them caused by the nature of the method itself.

- In models of computational chemistry, the energy surface maps coordinates of atoms to the energy. This function has many local minima that correspond to more or less stable states of the system. In the interesting processes the system usually crosses the energy barrier and transforms from one state to another. However, crossing the barrier occurs with probability that exponentially relates to its height, i.e. the higher barrier, the less probable crossing. Therefore, classical MD simulations sometimes have to simulate for long simulation time for crossing of the energy barrier to happen. As it can be rather difficult and lengthy to cover and sample whole energy surface, the issue is called the sampling problem [133].
- Needed long simulation times directly lead to long wallclock times as MD solves the initial value problem in a sequence of steps and, moreover, the integration scheme has small time step due to high oscillations of bonds that contain the hydrogen. The evaluation of the potential between atoms in each step remains computationally demanding despite many approximations of long-range interactions. Moreover, evaluation of long-range interactions requires communication between all processors calculating spatially decomposed parts. Many decomposition techniques rather compute the same values on two different processors than send a message which stresses the high temporal cost of communication [134].

References

1. Kapetanovic IM. (2008) Computer-aided drug discovery and development (CADDD): in silico-chemicobiological approach. *Chem Biol Interact* 171: 165- 176.
2. Padole SS, Asnani AJ, Chaple DR, Katre SG (2022) A review of approaches in computer-aided drug design in drug discovery. *GSC Biol Pharma Sci* 19(2), 075-083.
3. López-Vallejo F, Caulfield T, Martínez-Mayorga K, Giulianotti MA, Houghten RA, Nefzi A, Medina-Franco JL (2011) Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Comb. Chem. High Throughput Screen* 14,475–487.
4. Gautam RK, Kamal MA and Mittal P (2023) *Computational Approaches in Drug Discovery, Development and Systems Pharmacology*, Academic Press, Elsevier Inc.
5. Begum S, Shahidulla SM (2019) Role of Computer Aided Drug Design in Drug Development and Discovery: An Overview. *Int J Res Eng Sci Manag* 2(2): 445-450.
6. Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem Biol Interact* 171:165-176.
7. Pasupa, K. (2012). The Review of Virtual Screening Techniques. *KMITL Inf Tech J* 1(1).
8. Johnson MA, Maggiora GM (1990) *Concepts and Applications of Molecular Similarity*. Wiley edition.
9. Muchmore SW, Edmunds JJ, Stewart KD, Hajduk PJ (2010) Cheminformatic tools for medicinal chemists. *J Med Chem* 53:4830–41.
10. Maldonado AG, Doucet JP, Petitjean M, Fan BT (2006) Molecular similarity and diversity in chemoinformatics: from theory to applications. *Mol Divers* 10: 39–79.
11. Ou-Yang SS, Lu JY, Kong X Q, Liang ZJ, Luo C, Jiang H (2012) Computational drug discovery. *Acta Pharmacologica Sinica* 33(9):1131-1140.
12. Carhart R, Smith D, and Venkataraghavan R (1985) “Atom pairs as molecular features in structureactivity studies: definition and applications,” *J Chem Inf Comp Sci* 25(2):64–73.
13. Willett P, Winterman V, Bawden D (1986) “Implementation of nearest-neighbor searching in an online chemical structure search system,” *J Chem Inf Comp Sci* 26(1): 36–41.
14. Johnson MA, Maggiora G. M (1990) Eds., *Concepts and Applications of Molecular Similarity*. New York: John Wiley & Sons.
15. Gower JC (1985) “Measures of similarity, dissimilarity and distance,” *Encyclopedia of statistical sciences* 5:397–405.
16. Ellis D, Furner-Hines J, Willett P (1993) “Measuring the degree of similarity between objects in text retrieval systems,” *Perspec Inf Manag* 3(2):128–149.

17. Liu XF, Jiang HL, Li HL. (2011) SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *J Chem Inf Model* 51:2372–85.
18. Raymond J, Willett P (2002) “Effectiveness of graph-based and fingerprint-based similarity measures for virtual screening of 2D chemical structure databases,” *J Comp Aided Mol Design* 16(1):59– 71.
19. Vidal D, Thormann M, Pons M, (2005) “LINGO – an efficient holographic text based method to calculate biophysical properties and intermolecular similarities,” *J Chem Inf Model* 45(2):386–393.
20. Liu XF, Ouyang SS, Yu BA, Liu YB, Huang K, Gong JY, et al. (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res* 38:W609–14.
21. Sree Mahalakshmi P, Jahnavi Y (2020) A Review on QSAR Studies. *Int J Adv Pharm Biotech* 6(2):19–22.
22. Sippl W, Holtje HD (2000) Structure-based 3D-QSAR—merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods. *J Mol Struct (Theochem)* 503: 31–50. Doi:10.1016/S0166-1280(99)00361-9.
23. Dermeche K, Tchouar N, Belaidi S (2015) *J. Bionanosci* 9:395-400.
24. Ghasemi J, Saaidpour S, Brown SD (2007) *J Mol Struct (Theochem)* 805:27-32.
25. Geladi P, Kowalski BR (1986) *Anal Chim Acta* 185:1-17.
26. Duprat AF, Huynh T, Dreyfus G (1998) *J Chem Inf Comput Sci* 38:586-594.
27. Tetko IV, Villa AEP, Livingstone DJ (1996) *J Chem Inf Comput Sci* 36:794-803.
28. Gasteiger J, Zupan J, (1993) *Angew Chem Int Ed Engl* 32:503-527.
29. Muhammad U, Uzairu A, Ebuka Arthur D (2018) Review on: quantitative structure activity relationship (QSAR) modeling. *J Anal Pharm Res* 7(2):240-242.
30. Waterbeemd H, Rose S (2003) "Quantitative approaches to structure-activity relationships", in Book "Quantitative approaches to structure-activity relationships". Elsevier. 351-367.
31. Almi Z, Belaidi S, Segueni L (2015) *Rev Theor Sci* 3 : 264-272.
32. Blaney JM, Martin EJ (1997) Computational approaches for combinatorial library design and molecular diversity analysis. *Curr opin chem biol* 1(1):54-9.
33. Hopfinger AJ, Duca JS (2000) Extraction of pharmacophore information from high-throughput screens. *Curr opin in biotechno* 11(1):97-103.
34. Todeschini R, Consonni V (2008) *Handbook of Molecular Descriptors*. John Wiley & Sons.
35. Kerassa A, Belaidi S, Harkati D, Lanez T, Prasad O, Sinha L (2016) *Rev Theor Sci* 4:85-96.

36. Belaidi S, Mazri R, Belaidi H, Lanez T, Bouzidi D (2013) *Asian J Chem* 25 : 9241-9245.
37. Todeschini R, Consonni V. (2009) *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references: John Wiley & Sons.*
38. Khan AU (2016) “Descriptors and their selection methods in QSAR analysis : paradigm for drug design,” *Drug Discov Today* 21(8):1291–1302. doi: 10.1016/j.drudis.2016.06.013.
39. Guest S, Section E 2012 “Feature Selection Methods in QSAR Studies,” *J. AOAC Int.*, vol. 95(3):636–651.
40. Rivail JL (1999) *Eléments de chimie quantique à l’usage des chimistes*, 2ième éd., CNRS Edition.
41. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:864-871.
42. Thomas LH (1927) The calculation of atomic fields. *Proc. Camb Phil Soc* 23:542- 548.
43. Puzyn T, Leszczynski J, Cronin MT (2010) *Recent Advances in QSAR Studies: Methods and Applications: Part I Theory of QSAR, Challenges and Advances in Computational Chemistry and Physics*, 8.
44. Tobias RD (2002) *An Introduction to Partial Least Squares Regression*, Statistical Analysis System Institute Inc., Cary, USA.
45. Gauchi JP (1995) *Rev Stat Appl* 43:65-89.
46. Patel HM, Noolvi MN, Sharma P, Jaiswal V, Bansal S, Lohan S, et al. (2014) Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med Chem Res* 23(12):4991-5007.
47. Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression), *Wiley interdisciplinary reviews: computational statistics*, 2(1):97-106.
48. Hoffman BT, Kopajtic T, Katz JL, Newman AH. (2000) 2D QSAR modeling and preliminary database searching for dopamine transporter inhibitors using genetic algorithm variable selection of Molconn Z descriptors. *J med chem* 43(22):4151-9.
49. Tropsha A, Gramatica P, Gombar VK (2003) the importance of Being Earnest: Validation is the Absolute Essential for Successful Application and interpretation of QSPR Models, *QSAR and Combinatorial Sciences* 22(1):69–77.
50. Basak SC, Mills D (2010) “Quantitative structure-activity relationships for cycloguanil analogs as PfDHFR inhibitors using mathematical molecular descriptors,” *SAR QSAR Environ Res* 21:215–229.
51. Luque FJ (2018) *Frontiers in computational chemistry for drug discovery* 23(11): 2872.
52. Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, Oberg T, Todeschini R, Fourches D, Varnek A (2008) *J Chem Inf Model* 48:1733.

53. Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701
54. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) “Applicability Domain for QSAR Models,” *Int. J Quant Struct Relationships* 1:5–63,
55. Ghamali M, Chtita S, Ousaa A, Elidrissi B, Bouachrine M, Lakhlifi T (2017) Review Article on QSAR analysis of the toxicity of phenols and thiophenols using MLR and ANN. *J Taib Univ Sci* 11:1–10
56. Silverman RB (2004) "The Organic Chemistry of Drug Design and Drug Action." 2nd ed, USA: Elsevier.
57. Patel HM, Noolvi MN, Sharma P (2014) “CHEMISTRY Quantitative structure – activity relationship (QSAR) studies as strategic approach in drug discovery,” 4991–5007.
58. Trinajstić N, Nikolić S, Basak SC, Lukovits I, (2001) "Distances indices and their hypercounterparts: Intercorrelation and use in the structure-property modeling." *SAR QSAR Environ Res* 12:31-54.
59. Bhadoriya KS, Sharma MC, Jain SV, Raut GS, Rananaware JR (2013) Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking-based combined in silico rational approach to design potent and novel TRPV1 antagonists. *Med Chem Res* 22(5):2312–27.
60. Amnerkar ND, Bhusari Synthesis KP (2010) Anticonvulsant activity and 3D-QSAR study of some prop-2-eneamido and 1-acetyl-pyrazolin derivatives of aminobenzothiazole. *Eur J Med Chem.* 45(1):149–59.
61. Bhadoriya KS, Sharma MC, Jain SV (2015) Pharmacophore modeling and atom-based 3DQSAR studies on amino derivatives of indole as potent isoprenylcysteine carboxyl methyltransferase (Icmt) inhibitors. *J Mol Struct* 1081:466–76.
62. Mali SN, Chaudhari HK (2018) Computational studies on imidazo [1, 2-a] pyridine-3-carboxamide analogues as antimycobacterial agents: common pharmacophore generation, atom-based 3D-QSAR, molecular dynamics simulation, QikProp, molecular docking and prime MMGBSA approaches. *Open Pharm Sci J* 5(1): 12–23.
63. Bhongade BA, Amnerkar ND, Gadad AK (2020) 3D-QSAR studies on 4, 5-dihydro-1H-pyrazolo [4, 3-h] quinazolines as plk-1, CDK2/A and aur-A serine/threonine kinase inhibitors. *Lett Drug Des* 17(4):388–95.
64. Kesar S, Paliwal S, Sharma S, Mishra P, Chauhan M, Arya R, Madan K, Khan S (2019) in-silico QSAR modelling of predicted rho kinase inhibitors against cardio vascular diseases. *Curr Comput Aided Drug Des* 15(5):421–32.

65. Arya R, Gupta SP, Paliwal S, Sharma S, Madan K, Chauhan M (2019) pharmacophore modeling and docking studies to investigate potential leads for the development of β -secretase APP cleavage enzyme-1 (BACE-1) inhibitors. *Lett Drug Des* 16(7):775–84.
66. Arya R, Gupta SP, Paliwal S, Kesar S, Mishra A, Prabhakar YS (2019) QSAR and molecular modeling studies on a series of pyrrolidine analogs acting as BACE-1 inhibitors. *Lett Drug Des* 16(7):746–60.
67. Asati V, Ghode P, Bajaj S, KJain S, Bharti SK (2019) 3D-QSAR and molecular docking studies on oxadiazole substituted benzimidazole derivatives: validation of experimental inhibitory potencies towards COX-2. *Curr. Comput. Aided Drug Des* 15(4): 277–93.
68. Rajathej DM, Parthasarathy S, Selvaraj S (2019) QSAR analysis of multimodal antidepressants vortioxetine analogs using physicochemical descriptors and MLR modeling. *Curr. Comput. Aided Drug Des* 15(4):294–307.
69. Shirbhate E, Patel P, Patel VK, Veerasamy R, Sharma PC, Rajak H (2020) Searching for potential HDAC2 inhibitors: structure-activity relationship studies on indole-based hydroxamic acids as an anticancer agent. *Lett Drug Des* 17(7):905–17.
70. Ravichandran V, Rohini K, Harish R, Parasuraman SK (2019) Sureshkumar Insights into the key structural features of triazolothienopyrimidines as anti-HIV agents using QSAR, molecular docking, and pharmacophore modeling. *Struct Chem* 30(4): 1471–84.
71. Cramer RD, Patterson DE, Bunce JD (1988) "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins." *J Am Chem Soc* 110: 5959-5967.
72. Xue CB, Zhang L, Luo WC, Xie XY, Jiang L, Xiao T (2007) "3D-QSAR and molecular docking studies of benzaldehyde thiosemicarbazone, benzaldehyde, benzoic acid, and their derivatives as phenoloxidase inhibitors." *Bioorg Med Chem* 15: 2006-2015.
73. Norinder U (1998) Recent progress in CoMFA methodology and related techniques. In: Kubinyi H, Folkers G, Martin YC (eds) 3D QSAR in drug design—recent advances, Kluwer Academic Publishers, New York 3:24–39.
74. Kim KH, Greco G, Novellino E (1998) A critical review of recent CoMFA applications. *Perspectives in drug discovery and design* 12:257-315.
75. Norinder U (1998) Recent progress in CoMFA methodology and related techniques. *Perspectives in drug discovery and design* 12:25-39.
76. Roy K, Kar S, Das RN (2015). A primer on QSAR/QSPR modeling: fundamental concepts. Springer.

77. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J med chem* 37(24):4130-46.
78. Polanski J, Walczak B (2000) The comparative molecular surface analysis (COMSA): a novel tool for molecular design. *Comp & chem* 24(5):615-25.
79. Ehrlich P (1898) Über die Constitution des Diphtheriegiftes. *Deut Med Wochschr* 24:597-600.
80. Guner OF, Bowen JP (2014) Setting the record straight: the origin of the pharmacophore concept. *J Chem Inf Model* 54(5):1269-83.
81. Selassie C, Verma RP (2003) History of quantitative structure–activity relationships. *Burger's Medicinal Chemistry and Drug Discovery*.
82. Horvath D (2008) Topological Pharmacophores, in *Chemoinformatics Approaches to Virtual Screening*, A.T. Varnek, A., Editor. p. 338.
83. Finn, P. W.(1996). Computer-based screening of compound databases for the identification of novel leads. *Drug Disc Today* 1(9):363-370.
84. Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): In silico – chemico - biological approach. *Chemico-Bio Inter*171:165-176.
85. Dror O, Shulman-Peleg A, Nussinov R, et al. (2004) Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem* 11(1):71-90.
86. Gillet VJ (2012) Pharmacophore Models in Drug Design, in *Physico-Chemical and Computational Approaches to Drug Discovery*, J.B. Luque, X., Editor.p. 418.
87. Hopkins AL, Groom CR (2002) The Druggable Genome. *Nat Rev Drug Discov* 1:727-730.
88. Hajduk PJ, Huth JR, Tse C (2005) Predicting Protein Druggability. *Drug Discov Today* 10:1675-1682.
89. Cavasotto CN, Orry AJ (2007) Ligand Docking and Structure-Based Virtual Screening in Drug Discovery. *Curr Top Med Chem* 7:1006-1014.
90. Cardoso CL, Lima VV, Zottis A, Oliva G, Andricopulo AD, Wainer IW, Moaddel R, Cass QB (2006) Development and characterization of an immobilized enzyme reactor (IMER) based on human glyceraldehyde-3-phosphate dehydrogenase for on-line enzymatic studies. *J Chromatogr A* 1120:151-157.
91. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S (2000) GRid-IN dependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J med chem* 43(17):3233-43.

92. Kuntz ID, Blaney JM, Oatley SJ et al. (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161(2):269-88.
93. Kitchen DB, Decornez H, Furr JR et al. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3(11):935-949.
94. Martz F (2007) Développement d'une nouvelle méthode de docking basée sur les mécanismes enzymatiques et guidée par des groupes prosthétiques, thèse de doctorat de l'université paris.
95. Barril X, Soliva R (2006) Molecular modelling. *Mol Biosyst* 2(12):660-681.
96. Sotriffer C, Klebe G, Stahl M, Böhm H-J (2003). Docking and Scoring Functions/Virtual Screening. In *Burger's Medicinal Chemistry and Drug Discovery* (pp. 281–331). Hoboken, NJ, USA: John Wiley & Sons, Inc.
97. Bastard K (2005). Assemblage flexible de macromolécules : la théorie du champ moyen appliquée au remodelage des boucles protéiques. Thèse Doctorat. Université de paris 7.
98. May A, Eisenhardt S, Schmidt-Ehrenberg J, Cordes F (2003). Rigid body docking for Virtual Screening. Konrad-Zuse-Zentrum für Informationstechnik Berlin. Berlin-Dahlem. Retrieved from <https://opus4.kobv.de/opus4-zib/frontdoor/index/index/docId/769>
99. Dréo J, Pétrowsky A, Siarry P, Taillard E (2003). Métaheuristiques pour l'optimisation difficile [recuit simulé, recherche avec tabous, algorithmes évolutionnaires et algorithmes génétiques, colonies de fourmis ...]. Eyrolles. Retrieved from <https://hal.archives-ouvertes.fr/hal-0084302>
100. Vyas V, Jain A, Jain A, Gupta A (2008). Virtual screening: a fast tool for drug design. *Scientia Pharmaceutica* 76(3), 333-360.
101. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 32:335-73.
102. Bouchrit H (2012) Thèse de magister université Mentouri. Algérie. 70.
103. Kitchen DB, Decornez H, Furr J.R et al. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3(11):935-49.
104. Huang, SY, Zou X (2010) Advances and challenges in protein-ligand docking. *Int J Mol Sci* 11(8):3016-34.
105. Xuan YM, Hong XZ, Mihaly M, Meng C (2011) Molecular Docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug* 7(2): 146–157.
106. Vieth M, Hirst JD, Kolinski A, Brooks CL (1998) *J Comput Chem* 19:1612.
107. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Domini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating

- structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33: 889-897.
108. Brut M (2009) Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques : Les Modes Statiques Université Toulouse III - Paul Sabatier.
109. Kollman PA (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93: 2395-2417.
110. Simonson T, Archontis G, Karplus M (2002) Free energy simulations come of age: protein-ligand recognition. *Acc Chem Res* 35: 430-437.
111. Ignarro LJ, Fukuto JM, Griscavage JM, Rogers NE, Byrns RE (1993) Oxidation of nitric oxide in aqueous solution to nitrite but not nitrate: Comparison with enzymatically formed nitric oxide from L-arginine. *Proceedings of the National Academy of Sciences USA* 90:8103-8107.
112. Chevrollier N (2019) Doctoral thesis. Développement et application d'une approche de docking par fragments pour modéliser mes interactions entre protéines et ARN simple-brin. University of Paris-saclay. Paris.
113. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW Taylor(2003) Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics* 52(4): 609– 623.
114. Lafond M (2015) Docking et scoring. Schrödinger.
115. Holloway MK (1995) A priori Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site. *J Med Chem* 38: 305-317.
116. Alban A (2007) Stratégies de docking-scoring assistées par analyse de données. Application au criblage virtuel des cibles thérapeutiques COX-2 et PPAR gamma. Other. Universités d'Orléans. France 191.
117. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 42(25):5100–5109.
118. Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 43(25): 4759–4767.
119. Weisstein EW (2011) Root-mean-square. <http://mathworld.wolfram.com/Root-Mean-Square.html>.
120. Humphrey W, Dalke A, Schulten K (1996)VMD – Visual Molecular Dynamics. *J Mol Graph* 14:33–38.

121. The Scripps Research Institute (2011) AutoDock - AutoDock. <http://autodock.scripps.edu/>.
122. Plewczynski D, Łazniewski M, von Grotthuss M, Rychlewski L, Ginalski K (2011) Votedock: Consensus docking method for prediction of protein-ligand interactions. *J Comp Chem* 32(4):568–581.
123. Fischer E. (1984) *Einfluss*. *Ber* 27:2985-2993.
124. Alvarez JC (2004) *Curr Opin Chem Biol* 8(4):365-370.
125. Ghosh S et al. (2006) *Current Opinion in chemical Biology* 10(3):194-202.
126. Kollman PA, Massova I, Reyes C, Kuhn B, Hou S, Chong L, Lee T, Duan Y, Wang W, Domini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33: 889-897.
127. Heermann DW (1986) 'Computer Simulation Methods in Theoretical Physics', Springer, Berlin.
128. Alder BJ, Wainwright TE (1957) 'Molecular Dynamics by Electronic Computers, Proc. Intern. Symposium on Transport Processes in Statistical Mechanics', Wiley Interscience 97.
129. Weiner SJ, Kollman PA, Nguyent T, Cas DA (1986) An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* 7:230-252.
130. Leach Andrew R. (2001) *Molecular Modelling: principles and applications*. Harlow, UK : Prentice Hall ; 2nd ed.
131. Ganesan A, Coote ML, Barakat K (2017) “Molecular dynamics-driven drug discovery: leaping forward with confidence,” *Drug Discov Today* 22:249–269.
132. Daan F, Berend S. (2001) *Understanding Molecular Simulation*. Academic Press, Inc. 6277 Sea Harbor Drive Orlando, FL, United States.
133. Jensen F (2007) *Introduction to computational chemistry*. John Wiley & Sons Ltd, Great Britain, 2 edition.
134. Bowers KJ, Dror RO, Shaw DE (2005) Overview of neutral territory methods for the parallel evaluation of pairwise particle interactions. *J Phys: Conference Series* 16:300–304.

Realized Works

Chapter III:
***In silico-Based Identification of new
anti-PfDHFR drug candidates via
1,3,5-triazine derivatives***

III.1. Introduction

The *in-silico* prediction (figure III.1) of antimalarial activity has been highlighted as a critical stage in the development of drugs with targeted biological activity. The computer-aided drug discovery (CADD) has proven to be a beneficial method for discovering prospective lead compounds and assisting in the development of new medications for a number of ailments[1]. In medicinal chemistry, computational analyses based on structure-based approaches, such as QSAR(Quantitative Structure Activity Relationship)method and ADME (Absorption, Distribution, Metabolism, and Excretion) proprieties are now commonly employed to assist speed up the drug design process[2-7].

The implementation of these three studies was the subject of this current research. The first section focuses on the development of the best QSAR models using the statically approach Multiple Linear Regression (MLR) and Artificial Neural Networks(ANN), on a set of different molecular characteristics of 1,3,5-triazine derivatives as potential antimalarials. It would reveal fresh knowledge that might be used to design new antimalarials inhibitors with increased potency.

A predictive QSAR model was constructed to be utilized for lead optimization and testing of novel compounds, and an *in-silico* evaluation of drug-likeness features was examined, which gives useful information about the activity of substances in the body that are expected to serve as inhibitors.

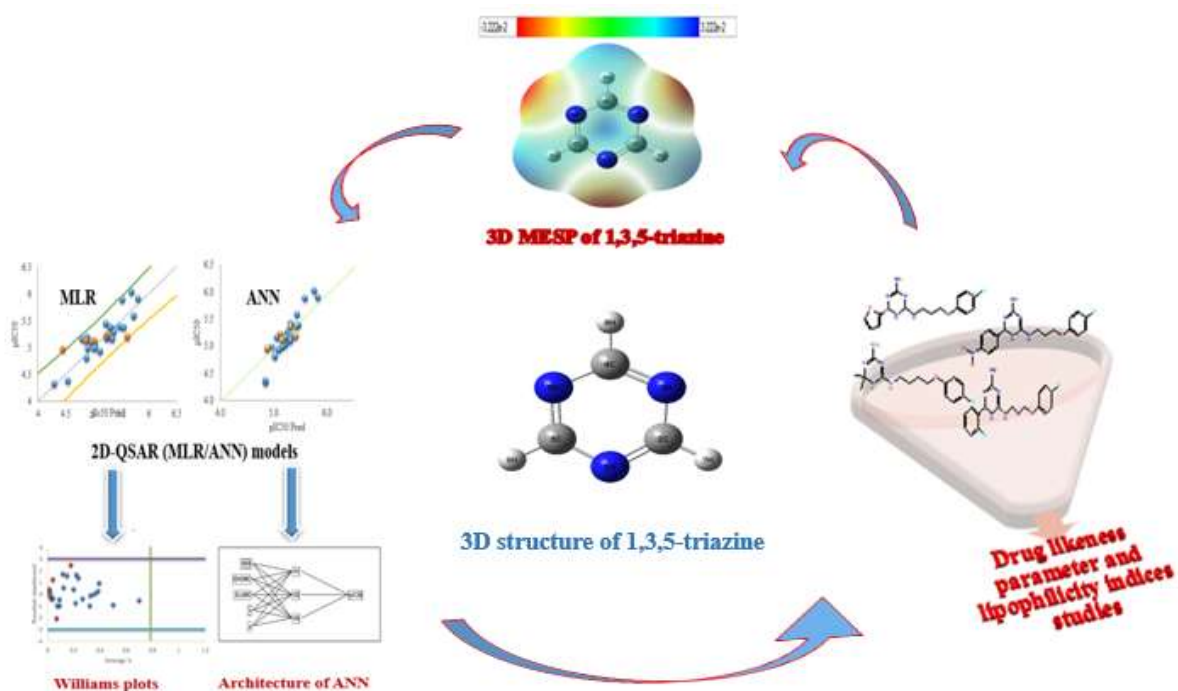


Figure III.1: *In-silico* studies involved in testing of 1,3,5-triazine derivatives.

III.2. Materials and methods

III.2.1. Computational details

Initial calculations were optimized using HyperChem 8.03 software[8]. The geometry of 1,3,5-triazine and its derivatives; were pre-optimized using the MM+ force-field (rms=0,01 Kcal/) in molecular mechanics[9]. The PM3 approach was used to completely re-optimize geometry[10]. In the next step, a parallel study has been made using Gaussian 09 program package, at various computational levels, HF/6-31++G(d,p), HF/6-311++G(d,p), B3LYP/6-31++G(d,p) and B3LYP/6-311++G(d,p)[11].

After that, the different properties of 1,3,5-triazine derivatives were calculated by using MarvinSketch 17.13.0 [12] software, ACD/Chemsketch12.0 [13] and HyperChem software(version 8.0.6) [8]. By means of these software, twenty descriptors are computed and reduced using the stepwise strategy in XLSTAT software[14] to build several QSAR models, just five descriptors of the best QSAR model have been reported in the present work.

III.2.2. Dataset selection

All *in-vitro* IC₅₀ (μM) antimalarial activity data of twenty-eight molecules having similar structures were selected from a series of 1,3,5-triazine-based derivatives expeditiously synthesized and biologically evaluated by Gravestock et al.[15]. The negative logarithm (pIC₅₀ = -log₁₀ (IC₅₀)) was used to convert all of the experimental activity IC₅₀ values for the purpose of providing numerically greater data values, listed in Table III.1, after then, it was employed as the dependent variable in the creation of the QSAR models.

Table III.1. Optimized structures of the molecules under study[15].

N°	Structures	pIC ₅₀ (obs)	N°	Structures	pIC ₅₀ (obs)
1		4.303	15*		5.131
2		5.005	16		5.176
3		4.781	17*		5.338
4		4.954	18		5.100
5		5.075	19		5.176
6*		4.947	20*		5.105
7		5.148	21		5.344

8		5.172	22		5.345
9		5.364	23		4.965
10		6.004	24		5.412
11		5.338	25		4.910
12*		5.100	26		5.143
13		5.176	27		5.567
14*		4.349	28		5.866

(*): Test set compounds. $pIC_{50} = -\log_{10}(IC_{50})$.

III.2.3. QSAR modeling studies

In an attempt to determine the role of structural features of compounds, which appears to have an effect on antimalarial activity, a QSAR models were generated. The field of quantitative structure-activity relationships deals with the development of a predictive models correlating biological activity (pIC_{50}) with the physicochemical properties. Once these are available, by using statistical methods, It is possible to establish this predictive MLR-QSAR and ANN-

QSAR models for a series of biologically active molecules which have shown inhibitory activity against the PfDHFR enzyme[16].

III.2.3.1. Statistical analysis and model validation

To predict the QSAR models, multiple linear regression (MLR) analysis of molecular descriptors was carried out in the present work using the stepwise strategy in XLSTAT software[17]. The MLR method was compared to the artificial neural network (ANN) method, which is another reliable and predictive QSAR model that is ideally suited for treating non-linear correlations between descriptors and activity[18-20]. All the ANN analyses were carried out using JMP 8.0.2. software[21].

Two basic methodologies are used to undertake external and internal validation of models[22]: internal validation utilizing the training set molecules, and external using the test set molecules by partitioning the entire data set into training and test sets; at 80% and 20%, respectively, utilizing the so-called 'Balanced Subset Method' (BSM)[23].

Apart from the use of fitness of several parameters, the statistical qualitative analysis of the QSAR model was validated by using the leave-one-out cross validation method (LOOCV)[24,25]. The best model was chosen in this study using the determination coefficient (R^2) and adjusted determination coefficients (R_{adj}^2), Fisher's value (F-value) should have high values and p-value (including the critical probability : p-value < 0.0001 for descriptors and for the model), predicted residual sum of squares (PRESS) and standard deviation of error of prediction (SDEP), root-mean-square error of prediction of training set ($RMSE_c$) and if the following conditions are satisfied, determination coefficient (R_{test}^2) of external validation, cross validated determination coefficient (Q_{cv}^2), root-mean-square error of prediction ($RMSE_p$) low mean squared error (MSE), and Y-randomization statistical parameters ($R_{(Rand)}^2$, $Q_{(Rand)}^2$ and cR_p^2).

III.2.3.2. Applicability domain approach

Another pivotal issue is the definition of a locale in the compound space containing the structural, physicochemical, or natural properties data set where upon the training set of the created model is through the applicability domain (AD) and for which it is applicable to make predictions for new compounds[26,27]. Even the most comprehensive, significant and validated models cannot reliably predict properties for all existing compounds.

Therefore, the AD of the models must be defined and only predictions for molecules falling in the training set in this domain can be considered acceptable. The method of leverage value h_i for each compound i has a calculated of the QSAR model was represented (Figure III.4)

Therefore, the AD of the models must be defined and only predictions for molecules falling in the training set in this domain can be considered acceptable. The method of leverage value h_i for each compound i has a calculated of the QSAR model was represented (Figure III.4) $h_i = x_i^T (X^T X)^{-1} x_i$ ($i = 1, 2, 3, \dots, n$), where x_i is the query compound's descriptor row-vector while X is the $n \times (k-1)$ matrix of k descriptor values of model for n training set compounds, and the superscript T is the matrix/transpose vector's [28]. And there exists a warning leverage value (h^*), in general, the critical value of the leverage $h^* = 3 \times (k + 1) / N$ [29], where x_i is the query compound's descriptor row-vector while X is the $n \times (k-1)$ matrix of k descriptor values of model for n training set compounds, and the superscript T is the matrix/transpose vector's [28]. And there exists a warning leverage value (h^*), in general, the critical value of the leverage $h^* = 3 \times (k + 1) / N$ [29].

III.2.4. Drug likeness parameter and lipophilicity indices

A drug's physicochemical properties of chemical compounds of series have a substantial impact on its *in vivo* pharmacokinetic characteristics under research with PFDHFR enzyme. The detailed analysis of drug likeness characteristics and lipophilicity indices were carried out by applying the different rules, by Lipinski's rules [30,31], veber's rules [32], lipophilicity indices [33,34], and Golden Triangle tool [35].

III.3. Results and discussion

III.3.1. Validation method

III.3.1.1. Equilibrium geometries of 1,3,5-triazine

The main geometrical parameters of the optimal equilibrium geometry to be employed are the most efficient theoretical strategy for the larger of 1,3,5-triazine (Figure III.1) of interest in the current study perhaps selected by comparison with experimental results.

Our investigations started by performing benchmark studies on 1,3,5-triazine using different theoretical methods (PM3, Ab-initio, DFT) in order to select the most reliable predictive method comparatively to experiment and with reduced computational cost.

Table 1 lists the main geometrical parameters of the optimized equilibrium geometry of 1,3,5-triazine are in accordance with the numbering scheme given in (Figure III.1). Table 1 lists also the corresponding experimental geometrical parameters that have been obtained by X-ray diffraction, which revealed that the molecule had D_{3h} symmetry [36]. Since 1,3,5-triazine are planar, the calculated dihedral angle values are either 0° or 180° .

From the obtained values (Table 1), we can also find that the appropriate method to compute the spectroscopic parameters of the 1,3,5-triazine is the density functional theory (DFT with

B3LYP/6-31++G(d,p)) which will be used to compute the quantum properties of our series of the 1,3,5-triazine derivatives.

Table III.2. Bond lengths (in Å) and valence angles (in degree, °) of 1,3,5-triazine. Experimental data for 1,3,5-triazine are collected from Ref.[37]

	Parameters	EXP	PM3	Ab initio/HF		DFT/B3LYP	
				6-31G++(d,p)	6-311G++(d,p)	6-31G++(d,p)	6-311G++(d,p)
Bond length (Angstrom)	N1-C2	1.338	1.357	1.318	1.317	1.337	1.334
	N1-C6	1.338	1.357	1.318	1.317	1.337	1.334
	C2-N3	1.338	1.357	1.318	1.317	1.337	1.334
	C2-H7	1.084	1.098	1.074	1.075	1.087	1.086
	N3-C4	1.338	1.357	1.318	1.317	1.337	1.334
	C4-N5	1.338	1.357	1.318	1.317	1.337	1.334
	C4-H8	1.084	1.098	1.074	1.075	1.087	1.086
	N5-C6	1.338	1.357	1.318	1.317	1.337	1.334
	C6-H9	1.084	1.098	1.074	1.075	1.087	1.086
Valence angle (°)	C2-N1-C6	113	118.433	114.536	114.475	114.245	114.278
	N1-C2-N3	127	121.567	125.463	125.524	125.755	125.723
	C2-N3-C4	113	118.433	114.536	114.475	114.242	114.275
	N3-C4-N5	127	121.567	125.463	125.524	125.756	125.723
	C4-N5-C6	113	118.433	114.536	114.475	114.245	114.278
	N1-C6-N5	127	121.567	125.463	125.524	125.753	125.721
	C2-N1-C6	113	118.433	114.536	114.475	114.245	114.278

III.3.1.2. 3D molecular electrostatic potential surface maps (3D MESP) of 1,3,5-triazine

The electrostatic potential that is created by a molecule's electron charge density in space expands in the entire space (nuclei considered as point charges). MESP entails comprehending a variety of physical and chemical phenomena, for instance molecular reactivity, molecular recognition, intermolecular contacts, substituent effects, electrophilic reactions, and reagent-induced interactions, such as those between a drug and its cellular receptor[38]. Figure III.2 shows the 3D molecular electrostatic potential surface maps (3DMESP) of 1,3,5-triazine.

As can be seen in Figure III.2, due to its strong electronegativity, 1,3,5-triazine exhibits negative electrostatic potentials (red zone) surrounding the nitrogen atoms (N1, N3, and N5). Additionally, we can observe positive electrostatic potentials (the blue zone) everywhere around the hydrogen and carbon atoms, which explain why these atomic sites are exposed to nucleophilic attack. Carbon atoms attached hydrogen atoms have the most positive charge per atom of hydrogen (dark blue).

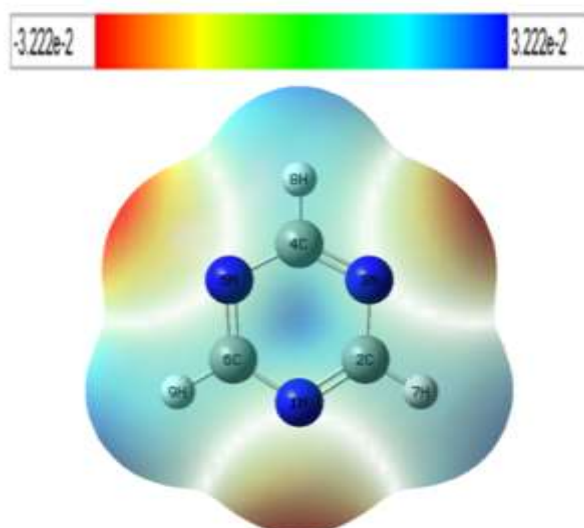


Figure III.2. 3D MESP of 1,3,5-triazine. The results are color-coded, from red (most negative) to blue (most positive).

In sum, 1,3,5-triazine derivatives exhibit a number of properties that may help us better understand the electrostatic interactions that may occur between the 1,3,5-triazine derivatives under study and reagents or enzyme active sites.

III.3.2. Quantitative structure-activity relationship studies

III.3.2.1. Multiple linear regression (MLR)

In the present study we tried to develop the statistical correlation of the best QSAR model that was derived from multi linear regression model generation (MLR).

That the physicochemical descriptors, **NRB** (number of rotatable bond on the molecule), **E_{HOMO}** (energy of highest occupied molecular orbital), **E_{LUMO}** (energy of lowest unoccupied molecular orbital), **n** (refractive index) and **μ** (dipolar moment) of the series of twenty-eight of 1,3,5-triazine derivatives were used as independent variables and were correlated with antimalarial activity (**pIC₅₀**). Additional validation was performed on a data set consisting of 1,3,5-triazine derivatives was randomly divided into two subsets; 23 training and 5 test sets. The values of the descriptors used in MLR analysis are presented in Tables 3. Pearson's correlation matrix has been performed on all descriptors.

Among several MLR equations the best model is expressed by the following relation:

$$pIC50 = -24.602 + 0.265 * NRB - 20.841 * E_{HOMO} + 14.892 * E_{LUMO} + 15.464 * n - 0.224 * \mu \text{ (Eq III.1)}$$

In Figure III.3, We display the experimental activity versus the predicted activity values to further created MLR model's prediction ability.

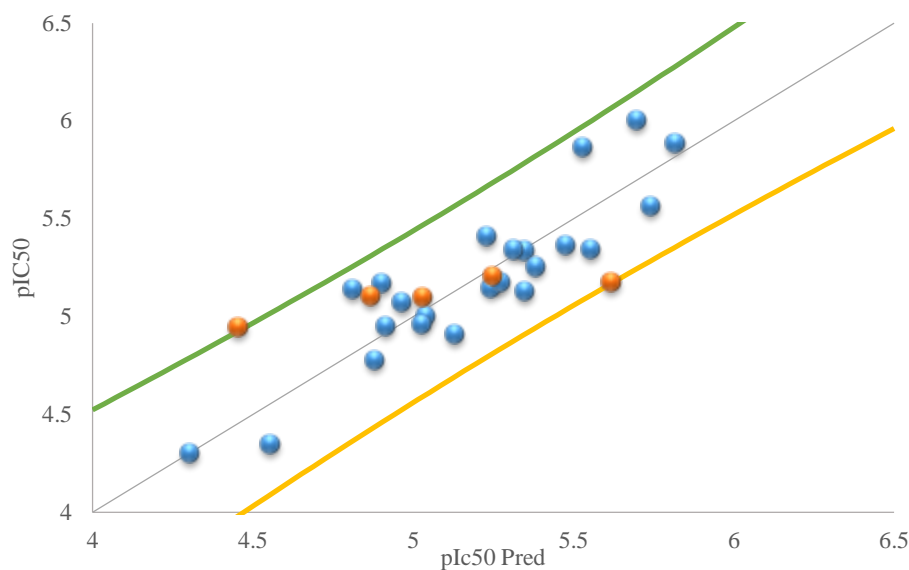


Figure III.3. Correlations of experimental versus predicted pIC_{50} values using MLR.

Once developed, the model must be interpreted by analyzing all the statistical parameters. In the model obtained in equation (1), we note that NRB , E_{LUMO} and n with positive coefficients suggest that biological activity increases with the increase in the values of these descriptors. On the other hand, the negative coefficients of E_{HOMO} and μ suggest the opposite. Thereafter, the molecular descriptors and the values of calculated activities of MLR and ANN are reported in Table III.3.

Table III.3. Chemical descriptors used in the regression analysis. They correspond to number of rotatable bond on the molecule (NRB), energy of highest occupied molecular orbital ($E_{HOMO}(eV)$), energy of lowest unoccupied molecular orbital ($E_{LUMO}(eV)$), refractive index (n) and dipolar moment ($\mu(D)$).

N°	NBR	E_{HOMO}	E_{LUMO}	n	μ	pIC_{50}	MLR	Residues	ANN	Residues
1	2	-0.209	-0.026	1.63	3.67	4.303	4.304	-0.001	4.853	-0.549
2	2	-0.208	-0.026	1.67	3.22	5.005	5.039	-0.034	5.219	-0.214
3	3	-0.214	-0.034	1.63	2.32	4.781	4.881	-0.100	5.031	-0.250
4	5	-0.229	-0.105	1.72	6.87	4.954	4.915	0.039	5.170	-0.216
5	4	-0.220	-0.045	1.67	5.53	5.075	4.967	0.109	5.270	-0.195
6*	5	-0.210	-0.029	1.61	5.11	4.947	4.455	0.492	4.889	0.058
7	6	-0.213	-0.029	1.65	5.75	5.148	5.244	-0.096	5.319	-0.171
8	6	-0.203	-0.024	1.66	7.37	5.172	4.904	0.268	5.314	-0.142
9	6	-0.212	-0.050	1.67	4.56	5.365	5.476	-0.112	5.472	-0.107
10	6	-0.218	-0.028	1.67	5.49	6.004	5.697	0.307	5.771	0.234
11	7	-0.227	-0.103	1.72	7.22	5.131	5.348	-0.216	5.277	-0.146
12*	6	-0.218	-0.031	1.66	5.08	5.177	5.618	-0.442	5.420	-0.243
13	6	-0.211	-0.029	1.65	5.10	5.338	5.349	-0.011	5.343	-0.005
14*	5	-0.210	-0.028	1.65	5.07	5.100	5.031	0.069	5.314	-0.214
15	7	-0.211	-0.028	1.64	5.59	5.177	5.275	-0.098	5.212	-0.035
16	7	-0.181	-0.010	1.64	7.14	4.349	4.555	-0.206	4.860	-0.511
17*	6	-0.217	-0.032	1.64	5.22	5.207	5.250	-0.043	5.121	0.086
18	7	-0.248	-0.047	1.61	5.19	5.257	5.383	-0.126	5.260	-0.003

19	6	-0.214	-0.032	1.68	5.02	5.886	5.818	0.068	5.826	0.061
20*	5	-0.208	-0.040	1.65	4.69	5.106	4.869	0.237	5.159	-0.053
21	6	-0.210	-0.027	1.65	5.26	5.344	5.314	0.029	5.416	-0.072
22	6	-0.221	-0.033	1.66	5.51	5.345	5.555	-0.211	5.314	0.031
23	6	-0.210	-0.027	1.63	5.12	4.965	5.028	-0.063	4.972	-0.007
24	7	-0.225	-0.100	1.67	4.85	5.412	5.231	0.181	5.320	0.092
25	7	-0.206	-0.025	1.62	5.16	4.910	5.131	-0.221	5.096	-0.186
26	6	-0.209	-0.031	1.61	3.91	5.143	4.812	0.331	5.052	0.090
27	7	-0.212	-0.031	1.64	3.76	5.567	5.741	-0.174	5.437	0.130
28	8	-0.212	-0.030	1.63	5.27	5.867	5.529	0.337	5.596	0.270

From the observed and predicted biological activity data of the molecules given in Table III.3, we can notice that there is a strong similarity between the observed and predicted pIC50 values. This can be explained by the low values of the residuals. This means that the QSAR models developed via the MLR and ANN techniques have a strong predictive capacity of the biological activity of the studied molecules according to the selected molecular descriptors (**NBR**, **E_{HOMO}**, **E_{LUMO}**, **n** and **μ**).

In order to detect the absence of the multi-collinearity for the selected descriptors the variance inflation factors (VIF) were calculated [39]. All five descriptors in the MLR-QSAR model have VIF values less than 5 (VIF = 1.716, 1.800, 3.023, 2.572, and 1.855, for **NRB**, **E_{HOMO}**, **E_{LUMO}**, **n** and **μ** respectively), indicating that the multi-collinearity is not present in the MLR model, which is the unique explanation for this condition.

Moreover, the predictive power of the equation of the model is confirmed by metrics of Golbraikh and Tropsha's criteria are listed in Table III.4. The MLR model having $R^2 > 0.6$ for both training and test sets will only be considered for validation [40]. Equation (1) exhibited high values of R^2 and R^2_{adj} , these are essential criteria are confirmed a strong association between the observed activities (pIC50) and the predicted activities (pIC50_{pred}). Our results for these two indices are 0.811 and 0.756, respectively, as shown in Table 4. The small PRESS (0.717) and SDEP (0.176) values indicate the model predictability and lack of over fitting. Value of F-value (14.635) (Table III.4) of MLR model with p-value less than 0.0001 show that the model is statistically significant [41].

This is on the other hand confirmed by metrics for external validation that has also used Golbraikh and Tropsha's criteria to judge the predicted model from a calculation the larger R^2_{test} (0.628) for the test set, Q^2_{cv} (0.578), RMSEP (0.258) and MSE (0.011) values indicate good predictive ability of the MLR model [42].

Table III.4. MLR statistics of predicted model.

Parameter	MLR	Threshold
R^2	0.811	>0.6
R_{adj}^2	0.756	>0.6
RMSE _c	0.195	A low value
PRESS	0.717	A low value
SDEP	0.176	A low value
F-value	14.635	A high value
R_{test}^2	0.628	>0.6
Q_{cv}^2	0.578	>0.5
RMSE _p	0.258	A low value
MSE	0.011	A low value

Furthermore, the robustness of the MLR model was ensuring by applying the Y-randomization test. Mostly, the Y-randomization test is used to test the stability of the predictive power of statistical models. Therefore, in the present study, this test was used to check the stability of the statistically modeled structure-activity relationship. This is to eliminate the probability of generating a QSAR model at random. We performed many Y vector random shuffles. After 100 random tests we have obtained small average values of 0.217 for $R_{(Rand)}^2$ and -0.573 for $Q_{(Rand)}^2$, also the smaller cR_p^2 (0.706 > 0.5) values indicate good predictive ability of the model and demonstrate its high robustness[43].

III.3.2.1.1. QSAR model's applicability domain

A William's plot is used to show the AD of models (Figure 4). According to the "three-sigma rule"[51], the AD is established using Excel 2013 software[44], in this plot inside a square region within the standard deviation x (in this study, $x = 3$). Outliers are molecules with standardized residuals three times higher than the model's standard deviation.

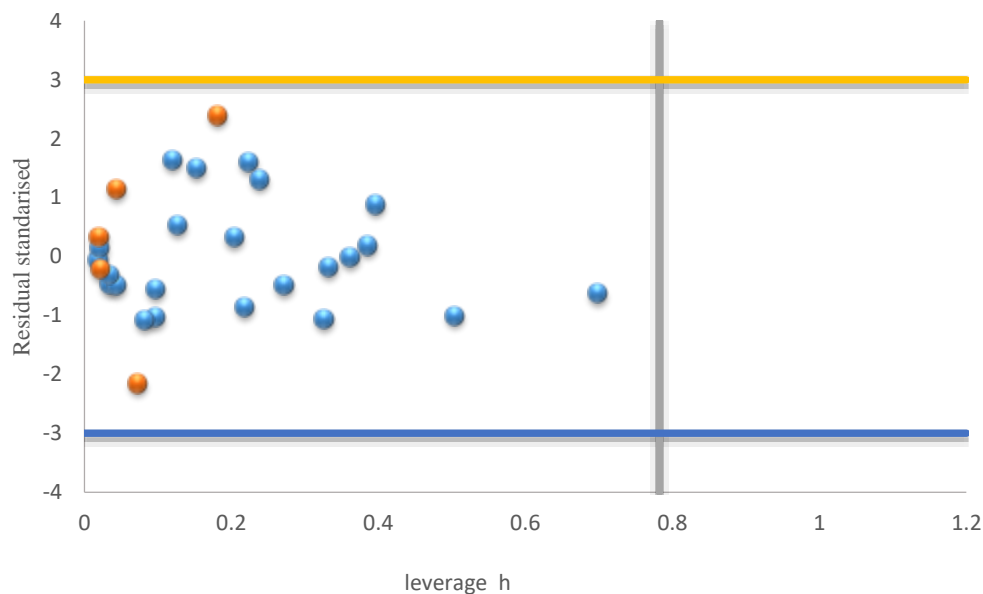


Figure III.4. MLR model's applicability domain plot. The vertical dashed line represents the warning leverage ($h^*=0.782$), whereas the horizontal lines denote ± 3 .

All substances in the dataset fall within the applicability range of the suggested MLR model, as can be seen by carefully examining Figure III.4. The leverage values of the inhibitors are all less than the warning h^* value (0.782), and none of them exhibit standardized residuals that are more than the threshold. As a consequence, the model exhibits the best statistical parameters and strong predictive capabilities, and it can be used in this AD with a high level of confidence.

III.3.2.2. Artificial Neural Networks (ANN)

The existence of a non-linear relationship between pIC_{50} and the five selected descriptors by the MLR model as inputs was studied in the second stage. The number of hidden layers was determined using the value $2n+1$, where n denotes the number of input layers, which plays an important role in establishing the optimum ANN architecture[45].

For pIC_{50} data, the architecture of the chosen ANN model was 5-3-1, with 5 descriptors in the first layer, three neurons in the hidden layer, and one neuron in the output layer after optimization. In this work, the intermediate (hidden) layer is made up of three neurons that form a deep internal pattern that identifies the strongest correlations between expected and experimental data. The output layer is made up of one neuron that returns the value of pIC_{50} (Figure III.5).

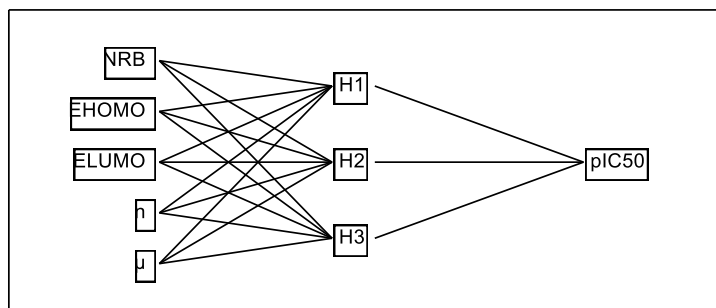


Figure III.5. Architecture of ANN.

The Gauss Newton method was then used to train the ANN[46]. The experimental and predicted pIC₅₀ using ANN are found to be highly correlated. This is indicated in Fig. 6 with 0.987 and 0.841 value of R^2 and Q_{cv}^2 , respectively. We are also obtained 0.968 a high value of R^2_{test} and a low value of MSE (-0.073) for the external validation.

We could infer that the ANN model with (5-3-1) architecture is capable of establishing a suitable link between the five descriptors and antimalarial activity based on both training and test set outcomes (Figure III.5).

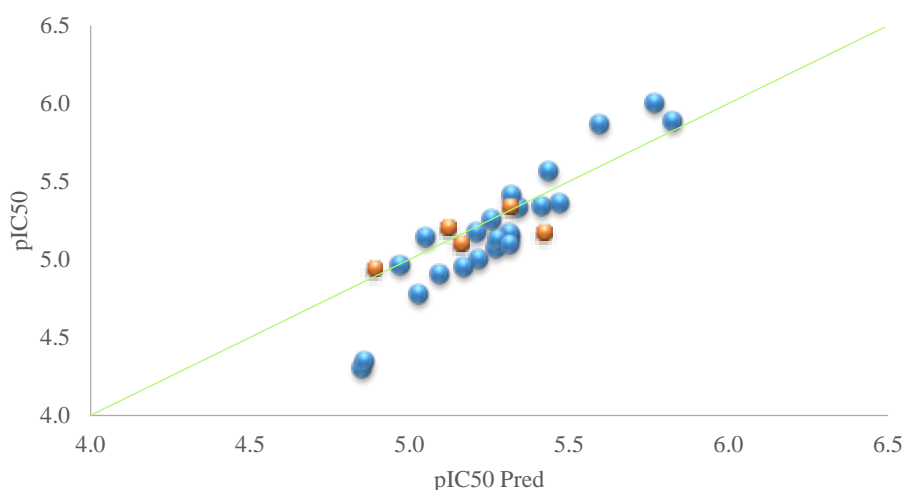


Figure III.6. Correlations of experimental versus predicted pIC₅₀ values using ANN.

A simple comparison of the values of the important statistics of ANN model in Table III.5 with those obtained using the MLR method confirms that ANN outperforms MLR, demonstrating the existence of a non-linear relationship between the pIC₅₀ and the five selected descriptors of the investigated compounds.

Table III.5. ANN and MLR statistics of predicted models.

Parameter	ANN	MLR
R^2	0.987	0.811
R^2_{test}	0.986	0.628
Q^2_{cv}	0.841	0.578
MSE	-0.073	0.011

Thus, therefore, our QSAR models can be successfully applied to predict the anti-PfDHFR activity of this class of molecules.

III.3.3. Design of Novel antimalarials

Using the foregoing results as a guide, we made appropriate substitutions and then proceeded to calculate their activities using the proposed model Eq (III.1). As a result, the proposed model will help us to speed up the time when it comes to synthesizing and assessing the antimalarials activity of 1,3,5-triazine derivatives.

According to the preceding discussions, our MLR model might be used to calculate pIC_{50} pred of various 1,3,5-triazine derivatives as shown in Table III.6 and could contribute to the development of new antimalarials druglike. If we create a new compound with higher values than existing compounds, we may be able to create more active compounds than those now in use. In this manner, we performed structural alteration using compounds with the greatest pIC_{50} values as a template comp.10 (Figure III.7).

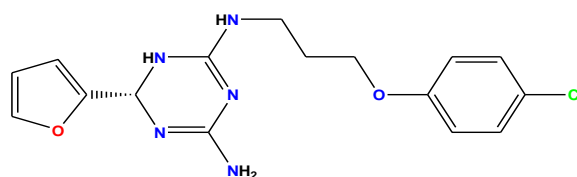
**Figure III.7.** Compound 10.

Table 6 lists the structures of the designed compounds, as well as their parameter values computed using the same procedures and the pIC_{50} values predicted by the MLR model.

Table III.6. Values of descriptors and pIC₅₀ for the new designed compounds (derivatives of comp.10) of the Figure III.7.

N ^o	Comp	NBR	E _{HOMO}	E _{LUMO}	n	μ	pIC ₅₀ pred
C1		7	-0.215	-0.033	1.70	5.99	6.251
C2		7	-0.201	-0.008	1.70	6.18	6.291
C3		7	-0.218	-0.029	1.75	5.35	7.191
C4		7	-0.202	-0.008	1.72	3.73	7.097
C5		7	-0.226	-0.046	1.76	4.91	7.385
C6		7	-0.199	-0.017	1.75	2.85	7.574
Ref		6	-0.218	-0.028	1.67	5.49	6.004

III.3.4. Drug likeness screening of 1,3,5-triazine derivatives

Chemicals' drug-likeness is a qualitative feature [46], beneficial for early-stage drug development. From the standpoint of this concept, it would be ideal to encode the equilibrium between a compound's molecular characteristics that effects its pharmacokinetics and eventually optimizes their absorption, distribution, metabolism and excretion (ADME) in the human body like a medicine.

At present, we should evaluate the oral bioavailability of the twenty-eight 1,3,5-triazine derivatives under study. Continuous, the quickest strategy for appreciating the drug-likeness of a set is to apply “rules”, they have been applied. As first, the most commonly Lipinski's rules are used[30,31].

Continuous, our parameters determined that good absorption or permeation is more likely to occur when: the molecular weight (MW<500da), number of hydrogen bond donors (HBDs <5) (counting the sum of all NH and OH groups), to estimate hydrophobicity of molecules used thepartition coefficient octanol/water (Log p< 5), and the number of hydrogen bond acceptors (HBA<10) are all within a certain range (counting all N and O atoms). In this rule of five-score, there are a total of four violations of Lipinski's rules.

Veber et al. identified the other two descriptors.[32]: number of rotatable bonds (NBR<10) and topological polar surface area (PSA <140 Å²). The TPSA is an important measure for predicting molecular transport properties, especially in the areas of blood-brain barrier (BBB) penetration and intestinal absorption [47,48]. It is well known that molecules with a TPSA of 140 Å² have a great ability to penetrate in an environment that is hydrophobic, like biological membranes. However, this could explain their quick penetration in hydrophilic settings, for instance, the core of transporter proteins[49]. The TPSA values were discovered to be in the range of (74,8 Å²-127,17Å²); these chemicals may be able to penetrate the BBB, resulting in increased bioavailability. TPSA was used to compute the percentage of absorption ((% ABS) in accordance with the equation the equation %ABS = 109 ± 0.345×TPSA[50].All of the compounds had a high %ABS, ranging from 65.126 to 83.194 %, implying that the permeability of their cellular plasmatic membrane is good.

Table III.7. Drug-likeness parameters and Lipophilicity indices of 1,3,5-triazine derivatives.

N°	Lipinski's rules					Veber's rules				Lipophilicity indices			
	MW≤ 500 Da	Log p ≤ 5	HBD ≤ 5	HBA ≤10	Lipin- ski score	NBR ≤10	TPSA ≤ 140 Å ²	Veber score	%ABS	LE	LLE	pIC ₅₀	N _H
1	231.30	2.18	3	5	4	2	74.80	2	83.194	0.354	2.123	4.303	17
2	271.37	3.21	3	5	4	2	74.80	2	83.194	0.350	1.795	5.005	20
3	279.77	2.21	3	6	4	3	74.80	2	83.194	0.352	2.571	4.781	19
4	407.26	1.52	3	9	4	5	117.94	2	68.311	0.257	3.431	4.954	27
5	380.25	1.74	3	8	4	4	74.80	2	83.194	0.284	3.335	5.075	25
6*	309.80	1.42	3	7	4	5	84.03	2	80.01	0.330	3.526	4.947	21
7	357.84	1.77	3	7	4	6	84.03	2	80.01	0.288	3.378	5.148	25
8	392.29	1.55	3	8	4	6	84.03	2	80.01	0.278	3.622	5.172	26
9	426.73	1.33	3	9	4	6	84.03	2	80.01	0.278	4.034	5.364	27
10	347.80	0.04	3	7	4	6	97.17	2	75.476	0.350	5.964	6.004	24
11	402.84	0.96	3	9	4	5	117.94	2	68.311	0.257	4.171	5.131	28
12*	392.29	1.55	3	8	4	6	84.03	2	80.01	0.279	3.626	5.176	26
13	363.89	2.56	3	7	4	6	84.03	2	80.01	0.299	2.778	5.338	25
14*	349.86	2.45	3	7	4	5	84.03	2	80.01	0.297	2.650	5.100	24
15	387.87	0.78	3	8	4	7	93.26	2	76.825	0.268	4.396	5.176	27
16	400.91	0.82	3	8	4	7	87.27	2	78.892	0.217	3.529	4.349	28
17*	375.83	1.17	3	8	4	6	84.03	2	80.01	0.280	4.037	5.207	26
18	425.84	2.34	3	6	4	7	84.03	2	80.01	0.254	2.916	5.256	29
19	373.90	2.12	3	5	4	6	100.1	2	74.465	0.330	3.766	5.886	25
20*	325.86	1.77	3	6	4	5	100.1	2	74.465	0.340	3.336	5.106	21
21	357.84	1.77	3	7	4	6	84.03	2	80.01	0.299	3.574	5.344	25
22	392.29	1.55	3	8	4	6	84.03	2	80.01	0.288	3.795	5.345	26
23	341.39	1.39	3	7	4	6	84.03	2	80.01	0.278	3.575	4.965	25
24	368.40	1.18	3	8	4	7	127.17	2	65.126	0.281	4.232	5.412	27
25	353.42	1	3	7	4	7	93.26	2	76.825	0.264	3.910	4.910	26
26	323.83	1.87	3	7	4	6	84.03	2	80.01	0.327	3.273	5.143	22
27	371.87	2.22	3	7	4	7	84.03	2	80.01	0.300	3.347	5.567	26
28	385.90	2.62	3	7	4	8	84.03	2	80.01	0.304	3.246	5.866	27

The results obtained are shown in Table III.7, they were calculated using HyperChem 8.0.8 (for MW, Log p and N_H) and MarvinSketch 6.2.1 software (for HBD, HBA, NBR and TPSA). As

can be seen there, The Lipinski and Veber criteria are satisfied by all substances, indicating that their theoretical oral bioavailability is optimal. The association between appropriate aqueous solubility and intestinal permeability, as well as these physicochemical molecular properties that represent the first steps in oral bioavailability.

Therefore, for the series of interest, table III.7 shows that the rule of five is 4 and 2 for Veber's score. Indeed, to be marginal for further developments the compounds with Rule of five-scores > 1 are taken into consideration[51]. Overall, our findings show that most compounds defy Lipinski and Veber rules, indicating that all chemical compounds would have no issues with oral bioavailability

We have also defined ligand efficiency (LE), lipophilic ligand efficiency (LLE) and the golden triangle as $LE = 1.4pIC_{50}/NH$, and $LLE = pIC_{50} - LogP$; where NH is the number of heavy atoms [52]. They are described as crucial parameters for drug discovery, furthermore as a means of determining a compound's potency in relation to its molecular weight. LE is influenced by ligand size, with smaller ligands having higher biological efficiency than bigger ligands on average[53,54].

Further, we used LLE to facilitate a deeper comprehension of the affinity of structural alterations in the series, with respect to lipophilia. As a rough guide, LLE values have been stranded in the range 5–7 in drug-like space for medicinal compounds[55]. That compounds with high LE and LLE tries to improve a potency interact with biological targets[56]. In the studied series, the change of LE and LLE during optimization (Table III.7).

Other characteristics that affect ADME and drug-likeness attributes such as molecular weight (MW) and distribution coefficients (logD) were used to illustrate the simultaneous absorption and clearance of optimal medicines using Warring rules and the Golden Triangle tool[35]. The Golden Triangle is a visualization tool developed from in vitro permeability, in vitro clearance and computational data designed to aid medicinal chemists in achieving metabolically stable, permeable and potent drug candidates[35]. Plotting MW vs. logD on estimated octanol: buffer (pH 7.4) and classifying compounds of a series as permeable and stable (pH 7.4).

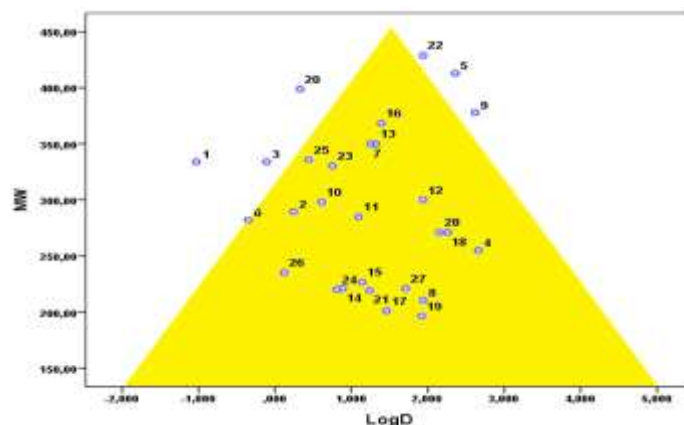


Figure III.8. Permeability and clearance patterns in vitro for MW and logD.

A triangular shape known as the "golden triangle" is formed when the design properties are moved into an area with a baseline of $\log D_{7.4} = -2.0$ to $\log D_{7.4} = 5.0$ at $MW = 200$ Da and a peak at $\log D_{7.4} = 1.0$ to 2.0 and $MW = 450$ Da, this increases the chance of success in maximizing potency, stability, and permeability. According to the fighting rule, with $MW = 414$ Da and $\log D_{7.4} > 1.3$ have a 74% chance of becoming highly permeable. Golden triangle's rules apply to the majority of our substances. These findings should aid in the development of permeability-enhancing chemicals. The metabolic stability and good membrane permeability of compounds found inside the Golden Triangle are more likely to be present than others outside.

According to the Golden Triangle (Figure III.8), the most of compounds under study are located within of it, indicating that these 1,3,5-triazine derivatives have good permeability and clearance [57]. The other compounds are the reverse, consisting of six compounds: 1,3,5,9, 20 and 22.

III.3.4.1. ADME study of new designed compounds

Table III.8 lists the physicochemical parameters of novel designed compounds. All of the proposed compounds' LogP and HBA values which indicate that they have a fair absorbency and resulting in an increase in the electrostatic interactions of the 1,3,5-triazine derivatives with the amino acid residues in the active sites. They were in great accord with the most important rules of drug similarity e.g. Lipinski, Veber and Lipophilicity indices.

Table III.8. Drug-likeness of the new designed compounds and reference compounds.

Lipinski's rules	Veber's rules	Lipophilicity indices
------------------	---------------	-----------------------

Comp	MW ≤ 500Da	Log p ≤ 5	HBD ≤ 5	HBA ≤10	Lipinski score	NBR ≤10	TPSA ≤ 140 Å ²	Veber score	LE	LLE	NH
C 1	361.83	0.19	3	6	4	7	87.94	2	0.350	6.061	25
C 2	361.83	0.19	3	6	4	7	87.94	2	0.352	6.101	25
C 3	363.86	0.38	3	6	4	7	87.94	2	0.419	6.811	24
C 4	343.45	0.76	3	6	4	7	87.94	2	0.414	6.337	24
C 5	382.84	-0.13	3	7	4	7	87.94	2	0.414	7.515	25
C 6	342.46	0.44	3	6	4	7	87.94	2	0.442	7.134	24
Cpd 10	347.80	0.04	3	7	4	6	97.17	2	0.350	5.964	24

The number of rotatable bonds 7(NBR <10), hydrogen bond acceptor 6 or 7 (HBA<10) and octanol/water partition coefficients (Log p<5) are used to forecast a compound's lead-likeness. When compared to the reference compound 10, these compounds showed no discomfort, indicating that they have good drug likeness properties.

References

1. Sharma P, Virmani T (2020) Synthesis, Antimicrobial Evaluation and QSAR Studies of Some Newly Synthesized Imidazole Derivatives. *Int J Adv Sci Technol* 29:9.
2. Ajala A, Uzairu A, Shallangwa, G.A (2023). QSAR, simulation techniques, and DMET/pharmacokinetics assessment of a set of compounds that target MAO-B as anti-Alzheimer agent. *Futur J Pharm Sci* 9, 4. <https://doi.org/10.1186/s43094-022-00452-X2>
3. Fonseca AMD, Neidelenio , Soares NM, Colares RP, Oliveira MMD, Oliveira LS, Marinho GS, de Lima MRP, da Rocha MN, dos Santos HS, Marinho ES (2023) Naphthoquinones biflorin and bis-biflorin (*Capraria biflora*) as possible inhibitors of the fungus *Candida auris* polymerase: molecular docking, molecular dynamics, MM/GBSA calculations and in silico drug-likeness study, *J Biomol Struct Dyn*, DOI: 10.1080/07391102.2022.2163702
4. Ouassaf M, Daoui O, Alam S, Elkhatabi S, Belaidi S, Chtita S (2022) Pharmacophore-based virtual screening, molecular docking, and molecular dynamics studies for the discovery of novel FLT3 inhibitors. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2022.2123403>
5. Fouedjou RT, Daoui O, Nour H, Ayoub M, Fogang HPD, Siddique F, Elkhatabi S, Bakhouch M, Belaidi S, Chtita S (2023) In Silico Approach for Designing Novel SARS-CoV-2 Inhibitors from Medicinal Plants. *Phys. Chem. Res.* <https://doi.org/10.22036/pcr.2022.349693.2138>
6. Abchir O, Daoui O, Belaidi S, Ouassaf M, Qais FA, Elkhatabi S, Belaouad S, Chtita S (2022) Design of novel benzimidazole derivatives as potential α -amylase inhibitors using QSAR, pharmacokinetics, molecular docking, and molecular dynamics simulation studies. *J Mol Model* 28: 1-17.
7. Fouedjou RT, Fogang HPD, Ouassaf M, Daoui O, Elkhatabi S, Bakhouch M, Belaidi S, Chtita S (2022) Targeting the main protease and the spike protein of SARS-COV-2 with naturally occurring compounds from some cameronian medicinal plants: an in-silico study for drug designing. *J Chil Chem Soc*, 67:5602-5614.
8. HyperChem (Molecular Modeling System) Hypercube (2008) Gainesville FL 32601, USA.
9. Belaidi S, Laabassi M, Gree R, Botrel A (2003) Analyse multiconformationnelle des macrolides symétriques de 12 à 28 chaînons basée sur la mécanique moléculaire. *Sci Stud Res* 4:27-38.
10. Stewart JJP (1989) Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem*. <https://doi.org/10.1002/jcc.540100209>

11. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao Y, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant J. C, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R GA, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2010) Gaussian Inc., Wallingford, CT, Gaussian 09.
12. MarvinSketch 17.1.2 (2017) Available on: <http://www.chemaxon.com>
13. ChemSketch, Advanced Chemistry Development Inc, Toronto, Canada. The program (edition 5.x) is available at the ACD at <http://www.acdlabs.com/>
14. XLSTAT, Software. XLSTAT Company (2013) Available on: www.xlstat.com
15. Gravestock D, Rousseau AL, Amanda L, Lourens, Anna CU, Moleele SS, Van Zyl RL, Steenkamp PA (2011) Expeditious synthesis and biological evaluation of novel 2, N6-disubstituted 1, 2-dihydro-1, 3, 5-triazine-4, 6-diamines as potential antimalarials. *Eur J Med Chem* 46:2022-2030
16. Soualmia F, Belaidi S, Tchouar N, Lanez T, Boudergua S (2021) QSAR Studies and Structure Property/Activity Relationships Applied in Pyrazine Derivatives as Antiproliferative Agents Against the BGC823. *Acta Chim Slov* 68:882-895.
17. Wythoff BJ (1993) Backpropagation neural networks: a tutorial. *ChemometrIntell Lab Syst* 18 :115-155.
18. Zupan J, Gasteiger J (1999) *Neural Networks in Chemistry and Drug Design: an Introduction*, Wiley-VCH.
19. JMP 8.0.2, SAS Institute Inc., (2009).
20. Van de Waterbeemd H, Timmerman H, Mannhold R, Krogsgaard-Larsen P (Eds.) (1995) *Chemometric Methods in Molecular Design. Methods and principles in medicinal chemistry*.
21. Rojas C, Duchowicz PR, Tripaldi P, PisDiez R (2015) Quantitative structure–property relationship analysis for the retention index of fragrance-like compounds on a polar stationary phase. *J Chromatogr A* . <https://doi.org/10.1016/j.chroma.2015.10.028>

22. Zhu T, Chen W, Singh RP, Cui Y (2020) Versatile *in silico* modeling of partition coefficients of organic compounds in polydimethylsiloxane using linear and nonlinear methods. *J Hazard Mater*. <https://doi.org/10.1016/j.jhazmat.2020.123012>
23. Golbraikh A, Tropsha A (2002) Beware of q^2 !. *J Mol Graph Model*. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
24. Cheng Z, Chen Q, Pontius FW, Gao X, Tan Y, Ma Y, Shen Z (2020) Two new predictors combined with quantum chemical parameters for the selection of oxidants and degradation of organic contaminants: a QSAR modeling study. *Chemosph*. <https://doi.org/10.1016/j.chemosphere.2019.124928>
25. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) Applicability domain for QSAR models. *Int J Quant Struct Prop Relatsh*. <https://doi.org/10.4018/ijqspr.2016010102>
26. Liu H, Wei M, Yang X, Yin C, He X (2017) Development of TLSE model and QSAR model for predicting partition coefficients of hydrophobic organic chemicals between low density polyethylene film and water. *Sci Total Environ* 574: 1371-1378.
27. Ghamali M, Chtita S, Bouachrine M, Lakhlifi T (2016) Méthodologie générale d'une étude RQSA/RQSP. *Rev Interdiscip* 1:ISSN 2458-7087.
28. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental-and-computational-approaches-to-estimate-solubility-and-permeability-in-drug-discovery. *Adv Drug Deliv Rev*. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
29. Almi Z, Belaidi S, Segueni L (2015) Structural Exploration and Quantitative Structure-Activity Relationships Properties for 1,2,5-Oxadiazole Derivatives. *Rev Theo Sci* 3: 264-272.
30. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J Med Chem* 45: 2615-2623.
31. Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH (2014) The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov*. <https://doi.org/10.1038/nrd4163>
32. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6: 881-890.
33. Johnson TW, Dress KR, Edwards M (2009) Using the Golden Triangle to optimize clearance and oral absorption. *Bioorg Med Chem Lett*. <https://doi.org/10.1016/j.bmcl.2009.08.045>

34. Walker IC, Palmer MH, Ballard CC (1992) The electronic states of the azines. VI. 1,3,5-triazine, studied by VUV absorption, near-threshold electron energy-loss spectroscopy and *ab initio* multi-reference configuration interaction calculations. *Chem Phys*. [https://doi.org/10.1016/0301-0104\(92\)80023-O](https://doi.org/10.1016/0301-0104(92)80023-O)
35. Lancaster JT, Stoicheff BP (1956) High Resolution Raman Spectroscopy Of Gases: VII. Rotational Spectra Of s-Triazine And s-Triazine-d 3. *Canad J Phys* 34: 1016-1021.
36. Asirvatham S, Dhokchawle BV, Tauro SJ (2019) Quantitative structure activity relationships studies of non-steroidal anti-inflammatory drugs: A review. *Arab J Chem* . <https://doi.org/10.1016/j.arabjc.2016.03.002>
37. <http://larmarange.github.io/analyse-R/multicolinearite.html>
38. Agrawal VK, Khadikar PV (2001) QSAR prediction of toxicity of nitrobenzenes. *Bioorg Med Chem*. [https://doi.org/10.1016/S0968-0896\(01\)00211-5](https://doi.org/10.1016/S0968-0896(01)00211-5)
39. Belaidi S, Almi Z, Bouzidi D (2014) Electronic structure and physical-chemistry properties relationship for phenothiazine derivatives by quantum chemical calculations, *J Comput Theor Nanosci* 11: 2481-2488.
40. Almi I, Belaidi S, Zerroug E, Alloui M, Said RB, Linguerri R, Hochlaf M (2020) QSAR investigations and structure-based virtual screening on a series of nitrobenzoxadiazole derivatives targeting human glutathione-S-transferases. *J MolStruct*. <https://doi.org/10.1016/j.molstruc.2020.128015>
41. Gramatica P, Pilutti P, Papa E (2003) *QSAR & Comb Sci* 22: 364-373.
42. Nielsen HR (1987) *Proc. IEEE Int. Conf. Neural Networks III* 10: 11-13.
43. Excel (2013). *Ink, Excel software*, Microsoft 2013.
44. Schneider G, Baringhaus KH (2008) *Molecular Design: Concepts and Applications*.
45. Fairbank M, Alonso E (2012) Efficient Calculation of the Gauss-Newton Approximation of the Hessian Matrix in Neural Networks. *Neural Comp* 24:607–610
46. Schaftenaar G, de Vlieg J (2012) Quantum mechanical polar surface area. *J Comput Aided Mol Des*. <https://doi.org/10.1007/s10822-012-9557-y>
47. Ouassaf M, Belaidi S, Lotfy K, Daoud I, Belaidi H (2018) Molecular Docking Studies and ADMET Properties of New 1.2.3 Triazole Derivatives for Anti-Breast Cancer Activity. *J Bionanosci*. <https://doi.org/10.1166/jbns.2018.1505>
48. Prasanna S, Doerksen RJ (2009) Topological Polar Surface Area: A Useful Descriptor in 2D-QSAR. *Curr Med Chem* 16 : 21-41.

49. Zhao YH, Abraham MH, Le J, Hersey A, Luscombe CN, Beck G, Sherborne B, Cooper I (2002) Rate-Limited Steps of Human Oral Absorption and QSAR Studies. *Pharm Res*. <https://doi.org/10.1023/A:1020444330011>
50. Copeland RA (2013) Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists. John Wiley & Sons, Hoboken, New Jersey, USA
51. Davis A, Ward SE (2014) *The Handbook of Medicinal Chemistry: Principles and Practice*. Royal Society of Chemistry.
52. Edwards PD et al (2007) Application of Fragment-Based Lead Generation to the Discovery of Novel, Cyclic Amidine β -Secretase Inhibitors with Nanomolar Potency, Cellular Activity, and High Ligand Efficiency. *J Med Chem*. <https://doi.org/10.1021/jm070829p>
53. Reynolds CH, Tounge BA, Bembenek SD (2008) Ligand Binding Efficiency: Trends, Physical Basis, and Implications. *J Med Chem*. <https://doi.org/10.1021/jm701255b>
54. Handlon AL, Schaller TL, Leesnitzer LM, Merrihew RV, Poole C, Ulrich JC, Raymond V, Merrihew, Chuck Poole, John C. Ulrich, Joseph Wilson JW, Cadilla R, Turnbull Ph (2016) Optimizing Ligand Efficiency of Selective Androgen Receptor Modulators (SARMs). *ACS Med Chem Lett*. <https://doi.org/10.1021/acsmchemlett.5b00377>
55. Hill RG, Richards D (2021) *Drug Discovery and Development E-Book: Technology in Transition*. Elsevier Health Sciences.
56. Zerroug A, Belaidi S, BenBrahim I, Sinha L, Chtita S (2018) Virtual screening in drug-likeness and structure/activity relationship of pyridazine derivatives as Anti-Alzheimer drugs. *J King Saud Univ-Sci* 31:595-601.

Chapter IV:

Combined 3D-QSAR, molecular docking, ADMET, and drug likeness scoring of novel Diaminodihydrotriazines as potential antimalarial agents.

IV.1. Introduction

In particular, chemoinformatic investigations are in silico techniques with a wide range of uses, including the discovery of new lead chemicals and the improvement of pharmacological activity features of a series of chemical compounds with known pIC₅₀ (biological activity) [1,2]. In this work, we aim biologically evaluate, using in silico approaches, the anti-PfDHFR activity of a series of diaminodihydrotriazine-based PFDHFR inhibitors, which were recently identified by Kamchonwongpaisan et al [3].

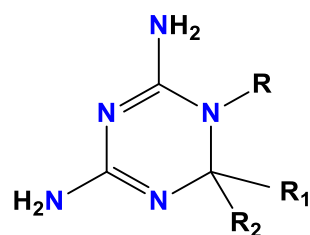
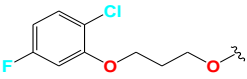
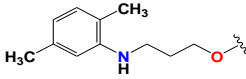
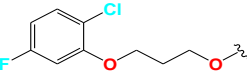
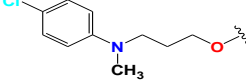
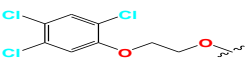
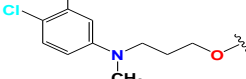
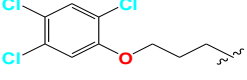

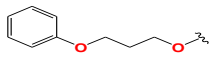
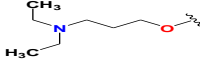
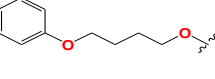
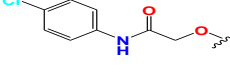
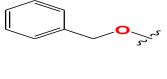
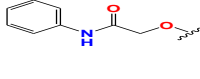
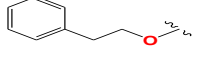
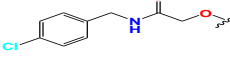
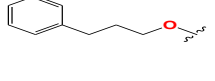
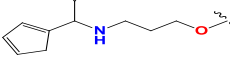
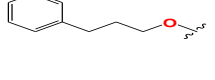
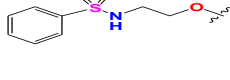
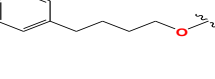
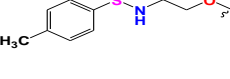
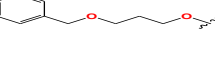
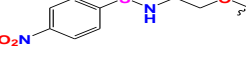
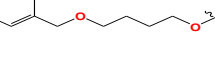
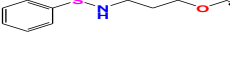
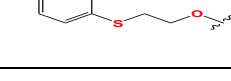
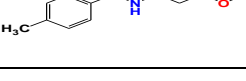


Figure IV.1: Chemical structures of diaminodihydrotriazines.

Table IV.1. Chemical structures of the diaminodihydrotriazines under investigation and those of pyrimethamine (Pyr) and of cycloguanil (Cyc).

Cpd.	R	R ₁	R ₂	Cpd	R	R ₁	R ₂
Pyr		H	CH ₂ CH ₃	Cyc		H	Me
1		H	Me	22		Me	Me
2		H	Ph	23		Me	Me
3		Me	Me	24		Me	Me
4		H	Me	25		Me	Me
5		H	Ph	26		Me	Me
6		Me	Me	27*		Me	Me
7		H	Me	28		Me	Me

8		Me	Me	29		Me	Me
9		H	Me	30		Me	Me
10		Me	Me	31		Me	Me
11		H	Ph	32*		Me	Me
12		Me	Me	33		Me	Me
13*		Me	Me	34		Me	Me
14*		Me	Me	35		Me	Me
15*		H	Ph	36*		Me	Me
16		Me	Me	37*		Me	Me
17		H	Ph	38		Me	Me
18		Me	Me	39*		Me	Me
19		Me	Me	40*		Me	Me
20		Me	Me	41		Me	Me
21*		Me	Me	42		Me	Me

In order to design a series of more efficient potential antimalarial agents, we investigate in-depth these compounds as specified in Table IV.1 and Figure IV.1. We use thus a number of

computational modeling techniques (figure IV.2) such as 3D-QSAR, *in silico* ADMET⁹¹ (Absorption, Distribution, Metabolism, Excretion and Toxicity) modeling profiles and molecular docking methods.

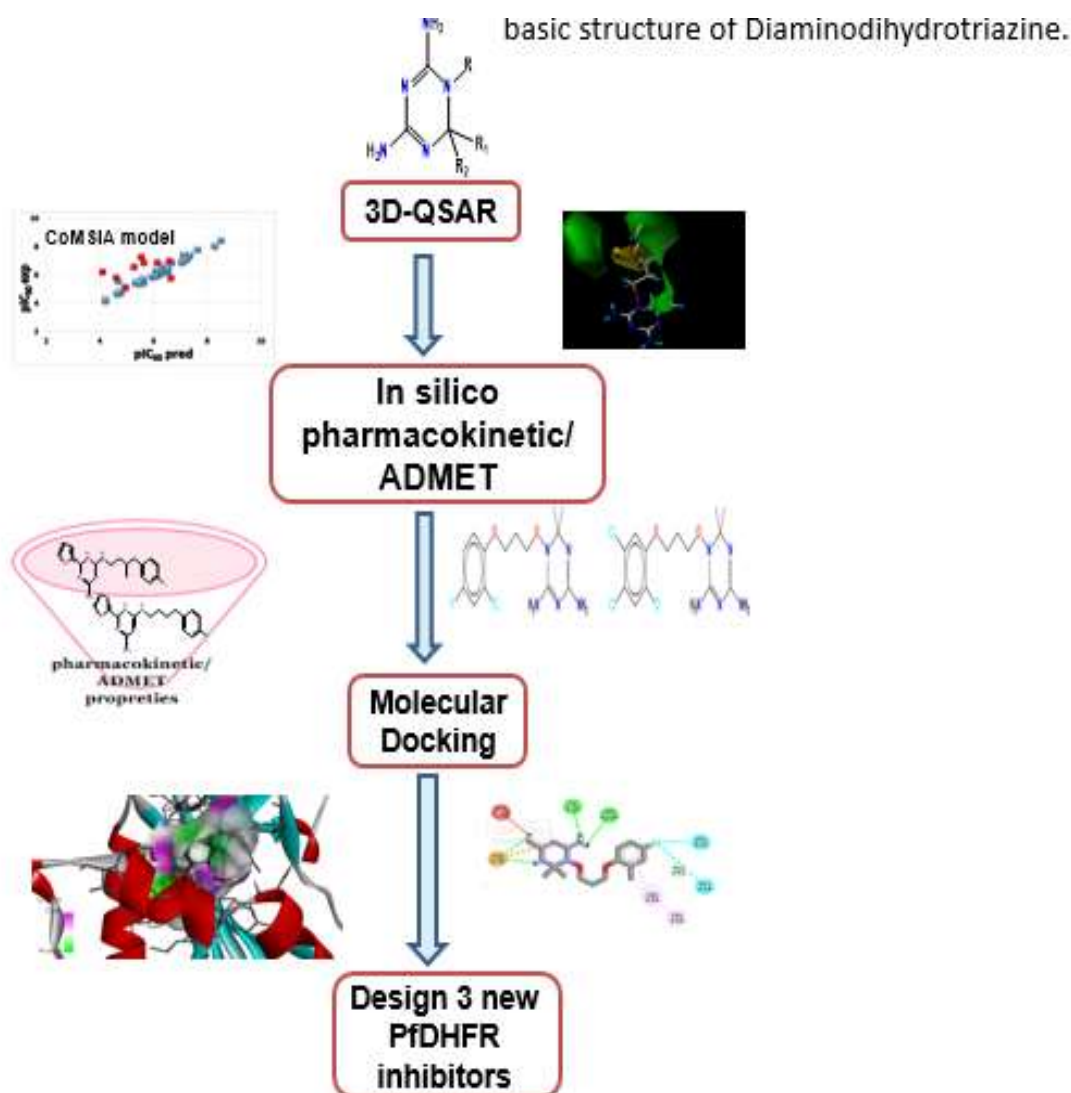


Figure IV.2: Computational modeling techniques involved in our study.

IV.2. Methodologies

IV.2.1. Dataset

The dataset compiled for this study includes 42 diaminodihydrotriazine derivatives as identified by Kamchonwongpaisan et al. [3]. Their chemical structures are shown in Table IV.1. Table IV.2 gives their biological activity values in μM units. The database is randomly divided into two ensembles for the purpose of creating a 3D-QSAR model: a learning set for building the quantitative model (80%) and a test set for assessing the performance of the suggested model (20%).

Table IV.2. pIC₅₀ experimental activities, pIC₅₀ predicted activities and residues of the series of diaminodihydrotriazines under study. * denotes test set compounds.

Comp.	pIC _{50exp}	CoMSIA pIC _{50pred}	Residues	Comp.	pIC _{50exp}	CoMSIA pIC _{50pred}	Residues
1	6.52	6.16	0.36	22	6.59	6.97	-0.38
2	6.32	6.36	-0.04	23	7.36	7.39	-0.03
3	7.64	7.80	-0.16	24	7.11	7.39	-0.28
4	6.09	6.25	-0.16	25	7.19	7.05	0.14
5	5.64	5.44	0.20	26	5.46	5.54	-0.08
6	8.28	8.04	0.24	27*	5.64	6.79	-1.15
7	6.68	6.88	-0.20	28	6.13	6.29	-0.16
8	8.52	8.45	0.07	29	6.46	6.52	-0.05
9	6.54	6.61	-0.07	30	6.38	6.60	-0.22
10	4.94	5.13	-0.19	31	6.54	6.44	0.10
11	5.59	5.71	-0.12	32*	5.54	7.28	-1.74
12	7.28	7.25	0.03	33	6.10	5.89	0.21
13*	6.66	5.78	0.88	34	4.78	4.75	0.03
14*	6.14	6.86	-0.72	35	4.74	4.82	-0.08
15*	4.61	5.81	-1.20	36*	4.96	5.08	-0.12
16	7.05	6.89	0.16	37*	5.27	6.55	-1.28
17	5.32	5.44	-0.12	38	4.11	6.20	-2.09
18	6.23	6.06	0.17	39*	4.21	4.25	-0.04
19	5.92	5.86	0.06	40*	5.46	5.40	0.06
20	4.68	4.57	0.11	41	4.21	4.16	0.05
21*	6.27	6.19	0.08	42	4.63	4.67	-0.04

IV.2.2. 3D-QSAR, *In silico* pharmacokinetics, ADMET study and Molecular docking analysis

In this work we started with the 3 Dimension of Quantitative Structure-Activity Relationship study using the Sybyl-X 2.1.1 software [4, 5] to carry out the 3D-QSAR procedures based on CoMSIA (Tripos Inc., St. Louis, MS, USA). To validate the model, a variety of internal and external validation approaches were used. Cross-validation was utilized to estimate the predictive ability of the CoMSIA model power using the Leave-one-out (LOO) approach. Figure IV.3 resumes all of steps involved in development of our (3D-QSAR) Model:

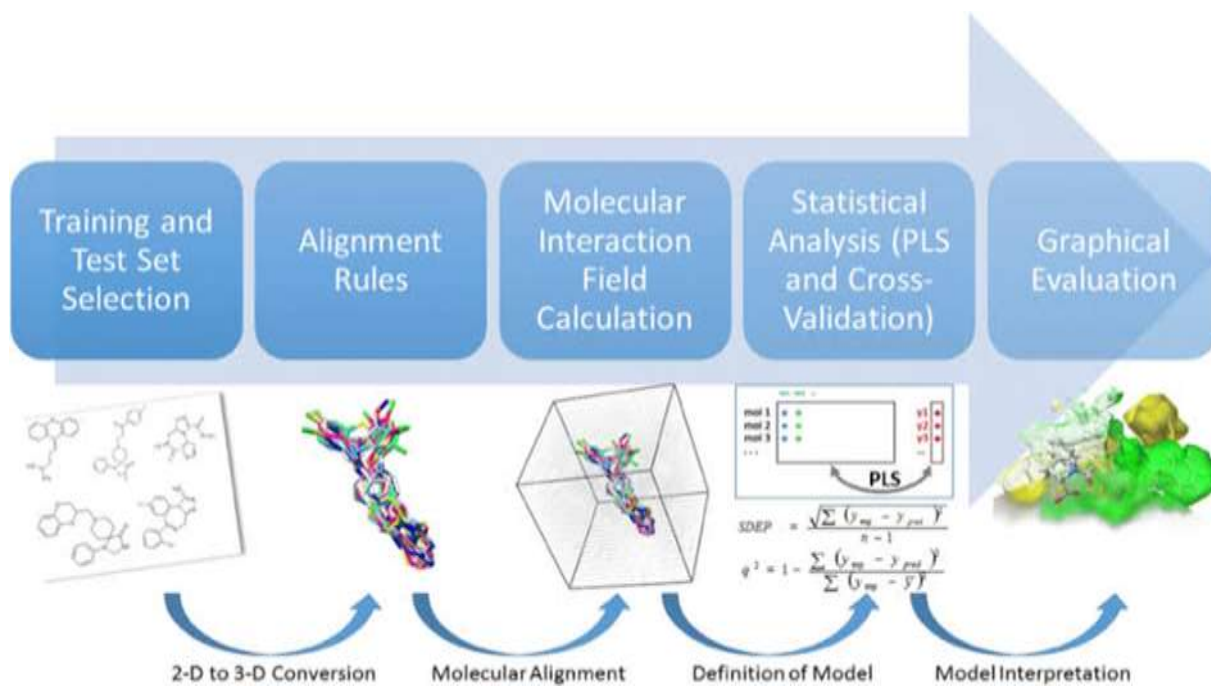


Figure IV.3: 3-D QSAR Model building flowchart.

After that, to predict the ADMET and pharmacokinetics features of the most biologically active compound and the reference ligand, we utilized the web tools at the pkCSM platform [6] and SwissADME server [7]. Finally, we applied molecular docking in the selected compounds to find the binding mode of the active compounds in the active binding region of the *Plasmodium falciparum* DHFR protein (PDB :1j3k)(figure IV.4). PyRx software was used to simulate the docking of molecules [8]. In order to assess the outcomes, BIOVIA discovery Studio 2019 client visualizer was used [9].

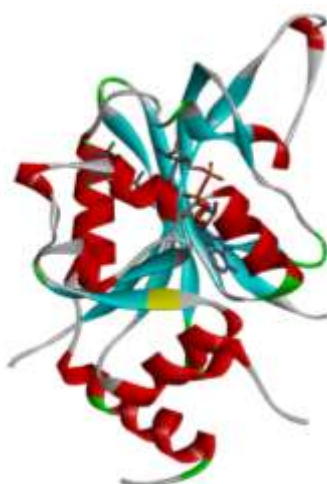


Figure IV.4: 3D-structure of *Plasmodium falciparum* DHFR enzyme (code PDB: 1j3k).

IV.3. Results and discussion

IV.3.1. Validation of the developed CoMSIA model

We performed 3D-QSAR investigations to further explore the antimalarial activity of diaminodihydrotriazine derivatives against *Plasmodium falciparum* DHFR. Thus, we developed a CoMSIA model where we evaluated the electrostatic (E), steric (S), hydrophobic (H), hydrogen bond acceptor (A) and hydrogen bond donor (D) fields [10, 11]. The respective data are given in Table IV.3.

Table IV.3. PLS statistics results and fields contributions of CoMSIA model. **E, S, H, A** and **D** stand for electrostatic field, steric field, hydrophobic field, H-bond acceptor field and H-bonds donor field, respectively.

Model	Q^2	R^2	SEE	F	ONC	R_{pred}^2	Fields contribution				
							S	E	H	D	A
CoMSIA	0.553	0.981	0.163	118.913	5	0.787	0.175	0.212	0.044	0.164	0.405

The table shows that the CoMSIA model based on biological activity (pIC_{50}) values demonstrated strong predictive ability with Q^2 of 0.553 (>0.5), an **ONC** of 5, a **F**-value of 282.251, R^2 of 0.981 (>0.9) and **SEE** of 0.163 (<0.95). For external validation, a set of 9 compounds was used. The corresponding R_{pred}^2 statistical parameter is 0.787. For the training and test set compounds, pIC_{50} 's were predicted. These values are listed in Table IV.2, where they are also compared to the experimental pIC_{50} . Both sets of pIC_{50} agree quite well, which validates the presently established CoMSIA model. Indeed, this model has residual values that range from -2.09 to 0.88.

The experimental and predicted values of pIC_{50} for the test and training sets of the compounds are plotted using linear regression in Figure IV.5. This figure shows that the plot of the experimental pIC_{50} data relate linearly with the predicted pIC_{50} data for the chemicals of both test and training sets. Indeed, a good linear dependence is noticeable which confirms that the proposed model has good predictive power.

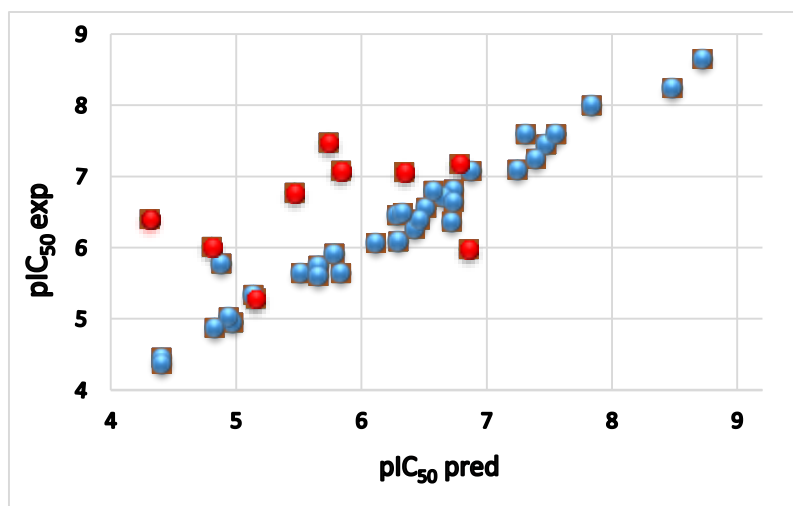
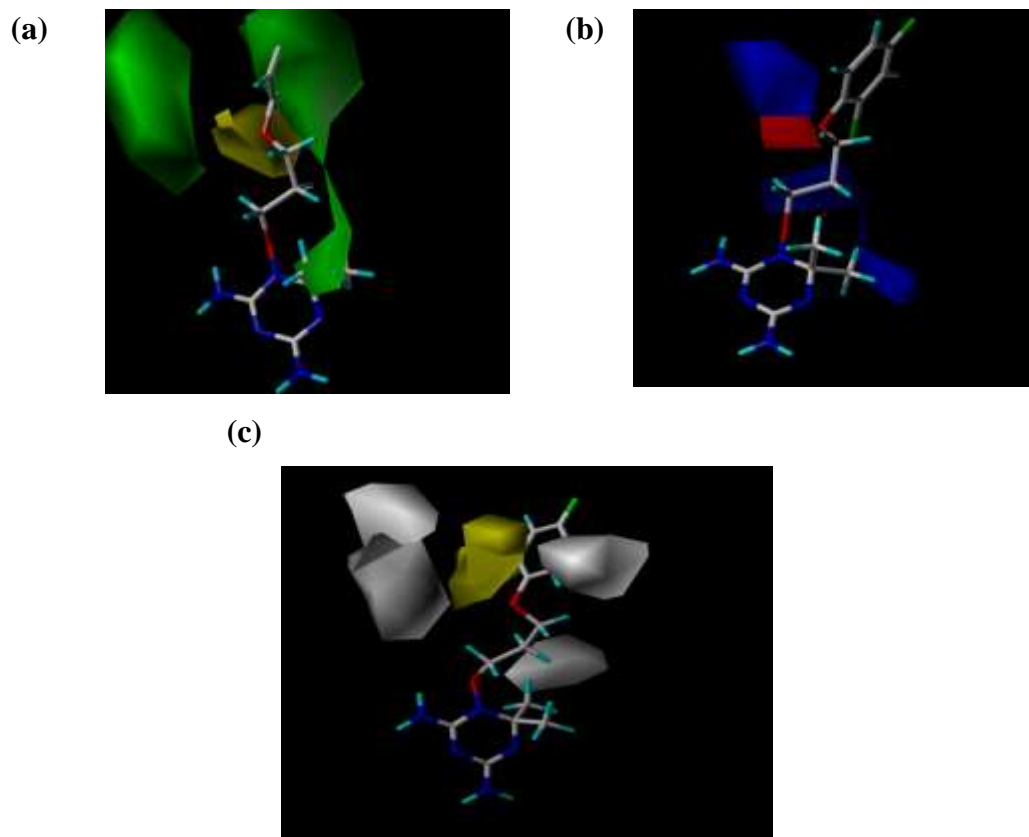


Figure IV.5: Experimental pIC_{50} s. Plot of the predicted pIC_{50} according to the CoMSIA model.

IV.3.2. Contour Plot Analysis

Contour plot analysis was carried out for the most active compound (i.e. compound **8**) at specific spatial areas of the system in order to understand the essential molecular criteria. The map includes descriptions of both positive and negative activity properties, such as (a) steric (b) electrostatic (c) hydrophobic/non-polar, and (c) donor and acceptor hydrogen bonds. Their individual good contributions are shown in Figure IV.6.



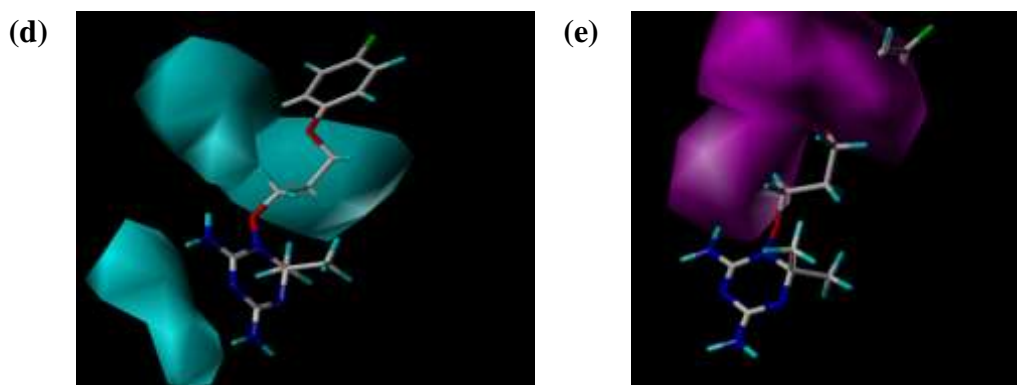


Figure IV.6. Fields contributions of CoMSIA model with the most active chemical (compound **8**). (a): Steric field. (b): Electronegative (red, negative) and electropositive (blue, positive) contour maps. (c): CoMSIA hydrophobic contour maps (yellow, favored area; green, unfavored area). (d): H-bond acceptor and (e): H-bond donor.

IV.3.2.1. Steric contour map

The green contour maps showed in Figure IV.6 (a) of the CoMSIA steric contour map indicate that the bulky groups are advantageous for enhancing activity, whereas the bulky groups are not as preferred in the yellow contour maps. For instance, the phenyl ring of the R group is highlighted by a large green contour map, which implies that inhibitors with large substituents compared to those with no groups or smaller ones, should be more active at this position. This agrees with relatively large pIC_{50} values for compound **2** ($pIC_{50} = 6.319$) and compound **6** ($pIC_{50} = 8.284$), which chemical structures include relatively large steric groups such as O-1,2,4-trichlorobenzene and O-1,2-dichlorobenzene substituted of pyridine, respectively. Whereas compound **15** is lacking such bulky substituents and this results in a drop in its biological activity. Indeed, its pIC_{50} of 4.607 is relatively low. Moreover, Figure IV.6 (a) shows that the O and N atoms of R group have a yellow contour, which indicates that adding bulkier substituents there will greatly reduce the biological activity of the corresponding diaminodihydrotriazine derivatives. This agrees with the small pIC_{50} value (of 4.11) measured for compound **38**. Indeed, this compound is less potent than the other compounds of the series that do not have substitutes attached to these O and N atoms.

IV.3.2.2. Electrostatic contour map

Figure IV.6 (b) displays the electrostatic contour maps as evaluated for compound **8**. The blue and red contours correspond to the electropositive and electronegative charges, respectively. The large red contours in the N, O, and S atoms of the chain of the R group imply that the inhibitory activity of electronegative groups will rise at these positions. This could explain why

compound **3** activity ($pIC_{50} = 7.638$) is higher than that of compound **28** ($pIC_{50} = 6.13$). Besides, ⁹⁷ the activity might be increased by the compound's electropositive group, as indicated by a blue contour area near the C atoms in the chain of the R group. This is in line of the higher activity of compound **14** with $O(CH_2)_2$ group ($pIC_{50} = 6.14$) than that of compound **39** with a sulfonamide group ($pIC_{50} = 4.21$). It may also help to explain why these compounds exhibit less potent PfDHFR inhibiting activity and have electronegative substituents those fall into the unfavorable red area. Indeed, the electronegative groups decrease the activity of the compounds.

IV.3.2.3. Hydrophobic contour map

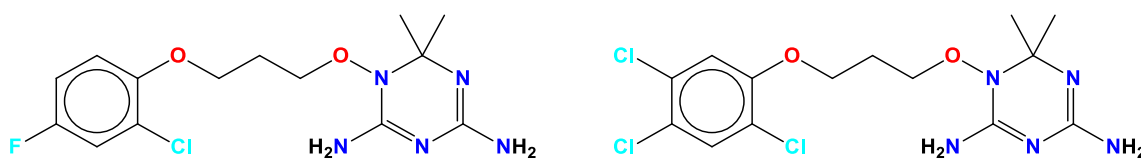
Figure IV.6 (c) displays the hydrophobic field's contour maps. According to the CoMSIA hydrophobic contour plot, the hydrophobic substituent will increase in the yellow zone the inhibitory activity of the compound. Yellow contours around white contours in the region of the benzene rings suggest that the addition of a modest hydrophobic substituent increases the biological activity.

IV.3.2.4. H-Bonds

As shown in Figure IV.7 (d), there are cyan color around positions R. Thus, substitutions at these positions by groups with properties of higher hydrogen bond donor are disadvantageous to the PfDHFR inhibition. This is the case, for instance, of compounds **22** and **27**, which have smaller pIC_{50} than compound **8**. Also, the purple maps of the H-bond acceptors field surrounding the R group in Figure IV.7 (e) imply that the presence of a donor substitution (for example, N, O, or S) at this position may encourage the development of H bond interaction.

IV.3.3. In silico pharmacokinetic/ADMET

The pharmacokinetic properties of the most active compound of the series (i.e. compound **8**) and the co-crystallized compound (denoted hereafter as reference compound, cf. Scheme IV.1), were estimated with the aid of pkCSM. Also, the properties of drug likeness of these compounds were predicted with the aid of SWISS/ADME. The outcomes are shown in Table IV.4, which presents the ADMET predictions and the physiochemical properties and the Drug Likeness rules of these compounds.



Scheme IV.1. Chemical structures of most potent compound (i.e. compound **8**, left) and of the reference ligand (right).

Table IV.4. Absorption, distribution, metabolism, excretion and toxicity predictions for compound **8** and for the reference compound (Ref) using pkCSM. We give their physiochemical properties as evaluated calculated using SWISS/ADME and their Drug Likeness rules.

<i>Absorption and distribution</i>										
Cp.	Water solubility (log mol/L)	Caco2 permeability (log Papp in 10 ⁻⁶ cm/s)	Intestinal absorption (human) (%) (Absorbed)	VDss (human) (log L/kg)	Fraction unbound (human) (Fu)	BBB permeability (log BB)	CNS permeability (log PS)			
8	-2.886	0.902	80.815	-0.054	0.473	-1.6	-3.19			
Ref	-2.820	0.731	79.852	-0.008	0.412	-1.401	-3.103			
<i>Metabolism and excretion</i>										
Cp.	CYP2D6 substrate	CYP3A4 substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	Total Clearance (log ml/min/kg)		
8	No	No	Yes	Yes	No	No	No	0.231		
Ref	No	No	Yes	Yes	No	No	No	0.389		
<i>Toxicity</i>										
Cp.	AMES toxicity	hERG I inhibitor	hERG II inhibitor	Hepatotoxicity	Skin permeation	Max. tolerated dose (human); log (mg/kg/day)				
8	Yes	No	Yes	No	No	0.717				
Ref	No	No	No	Yes	No	0.779				
<i>Physiochemical properties</i>										
Cp.	MW	XlogP	HBA	HBD	TPSA	log S	Fraction Csp3	NRB		
8	343.78	1.61	5	2	98.46	-2.78	0.43	6		
Ref	394.68	2.76	4	2	98.46	-3.81	0.43	6		
<i>Drug Likeness rules</i>										
Cp.	Lipinski rule	Ghose rule	Veber rule	Egan rule	Muegge rule	Bio availability score	Alerts PAINS	Alerts Brenk	Lead likeness	Synthetic Accessibility
8	Yes	Yes	Yes	Yes	Yes	0.55	0	0	0	3.93
Ref	Yes	Yes	Yes	Yes	Yes	0.55	0	1	1	3.98

Absorption and distribution were predicted from many properties. For the absorption, the log S parameter was used to evaluate water solubility of compound **8** and of the reference ligand (Scheme IV.1). We obtained values of -2.820 to -2.886 log mol/L for reference ligand and for compound **8**, respectively. These values indicate that both compounds are soluble in aqueous solutions due to their long R group's chain. For the Caco2 permeability, we have 0.731 to 0.902 log Papp in 10⁻⁶ cm/s for reference ligand and for compound **8**. For the intestinal absorption (% absorbed) (HIA), we found that compound **8** (= 80.815) is slightly better absorbed than reference ligand (= 79.852).

For the evaluation of the distribution of the two compounds, we used the following factors: distribution's volume for human (log L/kg), fraction unbound (Fu), factor of blood-brain barrier permeability (log BB), and CNS permeability (log PS). The volume of distribution of compound **8** as well as its log BB value are larger, in absolute value, than those of the reference

compound. The distribution of medication reserves from the plasma are transferred to the tissue ⁹⁹ [12]. According to compound **8** analyses, this compound has a log VD_{ss} value of 0.054, whereas the reference ligand has a smaller log VD_{ss} value (of 0.008). Thus, the reference ligand is less advantageous than compound **8** in terms of what constitutes an acceptable VD_{ss} value according to Refs. [13-15]. Whereas the fractions unbound of each compound and their CNS permeability values are close to each other (cf. Table IV.4).

The majority of medications in plasma alternate between being unbound and being bound to serum proteins in an equilibrium state. The proportion of drugs in plasma referred to as "fraction unbound" that are not bound to proteins impacts hepatic metabolism and renal glomerular filtration, which in turn influences the distribution volume, total clearance, and effectiveness of medicines [16]. The more tightly a medicine binds to blood proteins, the more difficult it is for it to disperse across cell membranes [17]. In the current study, low values, more particularly ranges between 0.412 and 0.473, were expected for the two chemicals. The BBB is an intricate barrier that isolates the peripheral tissue from the central nervous system (CNS). The BBB regulates the movement of substances, cells, and nutrients from the blood to the brain and from the brain to the blood in order to keep the CNS in a state of homeostasis. Additionally, it helps removing cellular metabolites and toxins from the brain and transports them to the blood [18]. Our compounds are expected to have BBB permeability values between -1.401 and -1.600.

The metabolism was approximated based on the inhibition of the primary cytochromes (CYP) of the P450 super family. The primary mechanism for the interactions of drug-drug based on metabolism, CYP enzyme inhibition, typically entails rivalry between drugs for the same binding site of the enzyme. To metabolizing a number of medications, CYP2C19 is also engaged in the detoxification of possible carcinogens or the bioactivation of various environmental procarcinogens [19]. This is consistent with the outcomes of the metabolism section that these substances are both CYP1A2 and CYP 2C19 inhibitors.

Besides, bioavailability is connected to excretion, which typically occurs as a combination of hepatic and renal clearance. Excretion is crucial for establishing dosage rates to reach steady-state concentrations [18]. The total clearance values of compound **8** and of the reference ligand are 0.231 and 0.389, respectively.

IV.3.3.1. Toxicity

The toxicity of the potent compounds was determined using pkCSM web tool. This allowed to predict mutagenicity and carcinogenicity, AMES toxicity, hepato toxicity, skin permeation and oral rat acute toxicity lethal dosage (LD₅₀) values. In order to create transdermal drug

administration, skin permeability ($\log K_p$) is evaluated. Indeed, skin permeability is an¹⁰⁰ important element in boosting medication effectiveness [20]. Table IV.4 shows that both compounds do not exhibit skin permeability. The hepatotoxicity descriptor implied that the reference chemical might cause hepatotoxicity, whereas compound **8** might not. In the biotransformation and energy exchanges of medicines, xenobiotics, depend heavily on the liver. Damage to the liver constantly interferes with normal metabolism and may even result in liver failure [29]. Compounds **8** and Ref have maximum tolerated doses of 0.717 and 0.779 correspondingly (measured in logarithm value of mg/kg/day for humans).

IV.3.3.2. Drug-Likeness screening using SWISS/ADME

We applied the following filters of drug-likeness: Lipinski, Veber, Ghose, Muegge and Egan rules using the screening with Swiss ADME. The outcomes are given in Table IV.4. This table shows that compound **8** and the reference ligand do not violate any of these rules. Indeed, the physiochemical properties listed in Table IV.4 show that these properties fall within these accepted ranges i.e. $150 \text{ Da} \leq \text{molecular weight (MW)} \leq 500 \text{ Da}$; $-0.7 \leq \text{lipophilicity (XlogP)} \leq 5$; H-bond acceptor (**HBA**) < 10 ; H-bond donor (**HBD**) < 5 ; $20 \leq \text{topology polar surface area (TPSA)} \leq 130$; $-6 \leq \text{molecular solubility (log S)} \leq 0$; $0.25 \leq \text{fraction Csp}^3 \text{ saturation} \leq 1$; and molecular flexibility (**NBR**) < 10 . In sum, these substances have a bioavailability score of 0.55, indicating good bioavailability.

Similarly, compound **8** exhibits no PAINS and no Brenk alerts, indicating that this compound has good medicinal chemistry property, compared to the reference compound which shows one Brenk alerts. Besides, both compounds have synthetic accessibility scores of ~3.9. Therefore, they could be easily synthesizable.

IV.3.4. Docking analysis result

To investigate the interactions of compound **8** and of co-crystallized ligand (i.e. reference ligand) with the active pocket of the PfDHFR protein (PDB ID: 1J3k), we conducted a molecular docking analysis with PyRx software. The compounds' binding locations were visualized using Discovery Studio as displayed in Figure IV.8. This figure shows that there are multiple contact types between compound **8** and the protein 1J3k. For instance, compound **8** was found to have different bond interactions, the important bonds are of hydrogen bonding nature (four conventional interactions between the NH of diaminodihydrotriazine ring and Ile 14, Tyr 170, 2 bonds with ASP 54 amino acid residues) and alkyl and pi-alkyl bonds (with Leu46, Leu40, Val45 and Gly44). Reference compound has however just one H-bond between the O atom of the R group and Val 168, amino acid residue. Besides, it forms alkyl and pi-alkyl

bonds with Gly166, Leu40, Val 195. Besides, we compute a larger binding energy of the best dock score (of -8.4 kcal/mol), in absolute value, for compound **8** compared to the reference ligand (-7.2 kcal/mol). Thus compound **8** binds better than this reference ligand to this enzyme. Accordingly, the reference ligand is less stable than compound **8**. This may justify the relatively large inhibitory activity of this compound against PfDHFR.

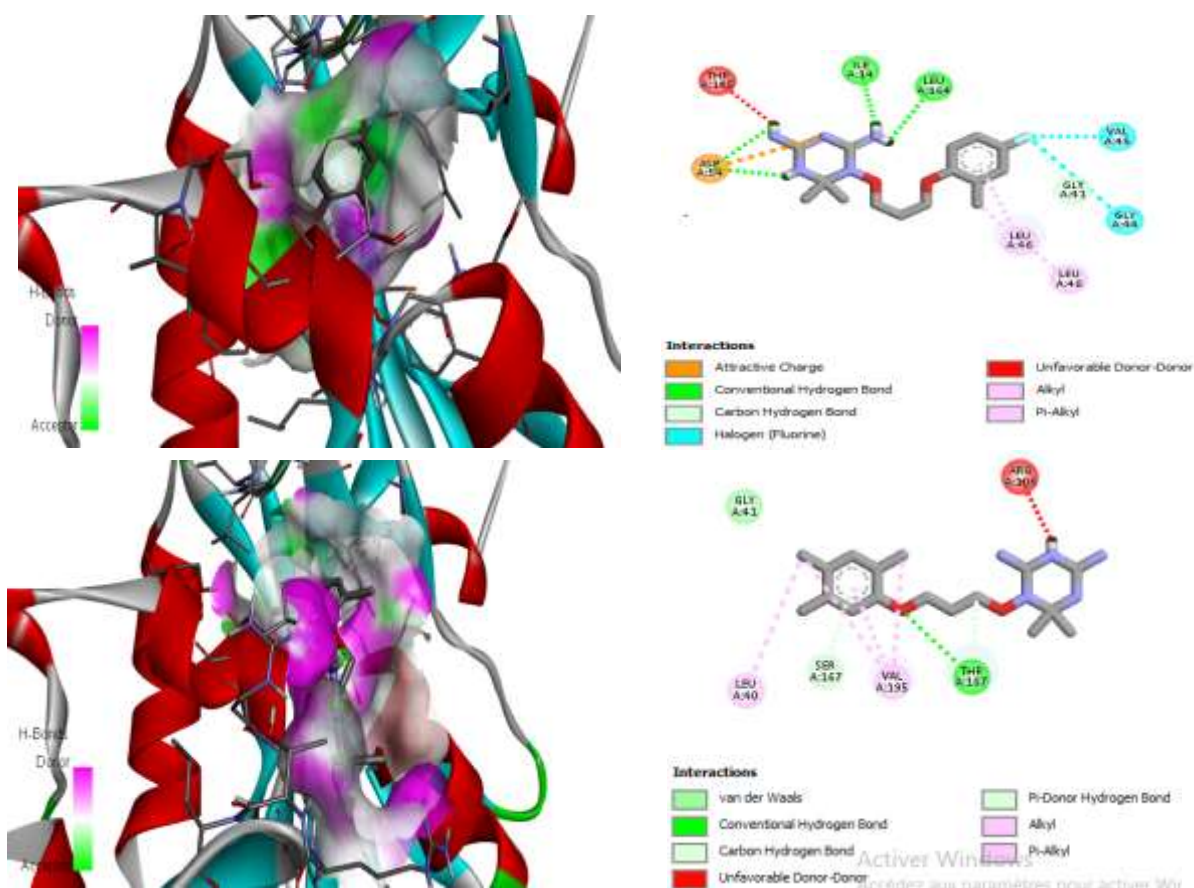


Figure IV.7: Various interactions formed between the compound **8** (upper trace) and co-ligand (lower trace) with the active site of PfDHFR.

IV.3.5. Design of new inhibitors

Based on the present CoMSIA model maps and the impact of the various groups on the inhibitory activity of substituted diaminodihydrotriazine derivatives, we designed new inhibitors. These compounds were determined from different contour maps results, where we tried to change the active groups (R, R₁ and R₂) with other substituents that are more effective and less electronegative than the substituents of compound **8**. Then, their biological activities were evaluated using Sybyl-X 2.1.1 software. Therefore, we designed three novel compounds, which are denoted as compounds **8a**, **8b** and **8c** as shown in Table IV.5. This table gives also the predicted pIC₅₀ of these new designed potent compounds.

Table IV.5. Chemical structures of compound **8** and of the newly designed potent compounds.

Cp.	R1	R2	R	pIC ₅₀ pred
8	Me	Me		8.45
8a	H	Me		8.69
8b	Me	Me		10.01
8c	H	Me		10.69

In comparison to compound **8**, the most active compound in the sequence, compounds **8a**, **8b** and **8c** exhibit large inhibitory activity. Indeed, we evaluate their pIC₅₀ as 8.69, 10.01 and 10.69, which are larger than pIC₅₀ of compound **8** (of 8.45). In particular, compounds **8b** and **8c** are predicted to be highly efficient.

References

1. Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: application to targets and beyond. *Br J Pharmacol* 152:9–20.
2. Tabeshpour J, Sahebkar A, Zirak MR, Zeinali M, Hashemzaei M, Rakhshani S, (2018) Computer-aided drug design and drug pharmacokinetic prediction: a mini-review. *Curr Pharm Des* 24:3014–3019.
3. Kamchonwongpaisan S, Charoensetukul N, Srisuwannaket C, Taweechai S, Rattanajak R, Vanichtanankul J, Vitsupakorn D, Arwon U, Thongpanchang C, Tarnchompoo B, Vilaivan T, Yuthavong Y (2020) Flexible diaminodihydrotriazine inhibitors of Plasmodium falciparum dihydrofolate reductase: Binding strengths, modes of binding and their antimalarial activities. *Eur J Med Chem* 195:112263.
4. www.tripossoftware.com
5. Vanommeslaeghe K, Guvench O, MacKerell AD (2014) Molecular mechanics. *Curr Pharm Des* 20: 3281-3292.
6. Pires DE, Blundell TL, Ascher DB (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J med chem* 58(9): 4066-4072. <http://biosig.unimelb.edu.au/pkcsm/prediction>
7. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports* 7(1): 42717. <https://www.swissadme.ch>
8. Trott O, Olson AJ (2010) AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comp Chem* 31:455-461.
9. BIOVIA, Dassault Systèmes (2019) Discovery studio Visualizer, San Diego: Dassault Systèmes, 2019.
10. Ouassaf M, Belaidi S, Chtita S, Lanez T, Qais FA, Amiruddin HM (2021) Combined molecular docking and dynamics simulations studies of natural compounds as potent inhibitors against SARS-CoV-2 main protease. *J Biomol Struc Dyn*. <https://doi.org/10.1080/07391102.2021.1957712>.
11. Khamouli S, Belaidi S, Ouassaf M, Lanez T, Belaouad S, Chtita S (2022) Multi-combined 3D-QSAR, docking molecular and ADMET prediction of 5-azaindazole derivatives as LRRK2 tyrosine kinase inhibitors. *J Biomol Struc Dym* 40: 1285-1298.
12. Guay AT (2001) Advances in the management of androgen deficiency in women. *Med Aspects Hum Sex* 1:32–38.
13. Mannhold R (2008) *Molecular Drug Properties. Measurement and Prediction*. Weinheim: Wiley-VHC Verlag.

14. Pires DE, Blundell TL, Ascher DB (2015) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58:4066–72.
15. Hou TJ, Zhang W, Xia K, Qiao XB, Xu XJ (2004) ADME evaluation in drug discovery. 5. Correlation of Caco-2 permeation with simple molecular properties. *J Chem Inf Comput Sci* 44:1585–600.
16. Watanabe R, Esaki T, Kawashima H, Natsume-Kitatani Y, Nagao C, Ohashi R, Mizuguchi K (2018) Predicting fraction unbound in human plasma from chemical structure: improved accuracy in the low value ranges. *Mol Pharm* 15:5302-5311.
17. Małkiewicz MA, Szarmach A, Sabisz A, Cudała WJ, Szurowska E, Winklewski PJ (2019) Blood-brain barrier permeability and physical exercise. *J Neuroinflamm* 16:1-16.
18. Biswas M (2021) Global distribution of CYP2C19 risk phenotypes affecting safety and effectiveness of medications. *Pharmacogenomics J* 21 :190–199.
19. Cerqueira NM, Gesto D, Oliveira EF, Santos-Martins D, Brás NF, Sousa SF, Fernandes PA, Ramos MJ (2015) Receptor-based virtual screening protocol for drug discovery. *Arch Biochem Biophys* 582:56–67.
20. Daina A, Michielin O, Zoete V (2019) Swiss Target Prediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res* 47:357- 364.

General Conclusion & Perspectives

The work brought together in this manuscript has made it possible to highlight the various virtual screening approaches for evaluating the biological activity and structural characteristics of 1,3,5-triazine and its derivatives for inhibiting *P. falciparum* dihydrofolate reductase (PfDHFR), with the aim of developing new bioactive molecules that are more effective against the disease Malaria. In this thesis, we discussed a number of important studies.

In the first part, we focused our attention on studying the structural and electronic properties of the basic nucleus, 1,3,5-triazine. The results demonstrate the effectiveness of the (DFT) method for performing quantum calculations and for furthering our quantum study of the 1,3,5-triazine nucleus and its derivatives.

In addition, the quantitative structure-activity relationship study (2D-QSAR) was carried out to quantitatively determine the effect of the 2D structural descriptors of the twenty-eight compounds studied on their biological activities. Mathematical models containing five descriptors were developed using the two important statistical methods MLR and ANN, to predict the specific activity of PfDHFR inhibition of new 1,3,5-triazine derivatives. The various main validation techniques (internal validation, external validation, Y-Randomization, applicability domains) demonstrated that the descriptors selected are relevant and that these QSAR models developed have strong potential, are robust and can be very useful for predicting the inhibitory activity of the PfDHFR enzyme in all the studies we carried out.

Drug likeness properties were then studied for the series of 1,3,5-triazine derivatives as well as for the new compounds. The molecules used in this study have pharmacological activities. The diversity of groups that bind to the base nucleus, i.e. the structural diversity studied, influences the physicochemical properties of the s-triazine derivatives and consequently their pharmacological properties. The computational results show that all compounds are in agreement with the druglikeness rules, suggesting that these compounds theoretically have high ADME properties and will not have problems with oral bioavailability.

In the second part, the quantitative 3D structure-activity relationship (QSAR-3D) has aroused the interest of medicinal chemists, leading to the discovery of several major diaminodihydrotriazine-derived molecules. For this reason we devoted our attention to the design of the best model obtained using the CoMSIA comparative molecular similarity index analysis method, we used the statistical method Partial Least Squares (PLS) to develop the 3D-QSAR model, displayed a good statistical parameter of internal and external validation. With the aid of obtained model we could evaluate the Fields contributions of CoMSIA model with the most active chemical (compound 8) (the electrostatic (E), steric (S), hydrophobic (H), hydrogen bond acceptor (A) and hydrogen bond donor (D) fields)where we allowed to understand the essential molecular criteria and from these fields we can design new PfDHFR inhibitors.

Subsequently, an *in silico* pharmacokinetics and ADMET study was performed for the selected more active compound (no. 8) and the reference ligand to assess their pharmacokinetic properties and toxicities. Analysis of the results obtained shows that the most active compound

is non-toxic and has pharmacokinetic parameter values within the acceptable range for human use, better than the reference ligand.

Finally, to study and visualize the predominant interactions and affinity energies of these two molecules with the PfDHFR enzyme, applying Molecular docking investigations. These two selected molecules were docked into the active site of PfDHFR (PDB ID: 1J3K) in order to assess and gain an in-depth insight into their ability to bind into the active site of the target enzyme. The docking analysis revealed that the compounds identified (no. 8) associated with the active site residues better than the reference inhibitor.

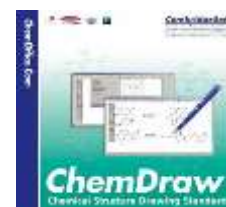
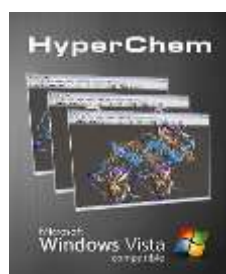
In the future, to remedy the inadequacy of our results and further our investigations, we intend to continue this work by using other VS techniques on the molecular structure of our study series in order to develop our own pharmacophore model. We will also carry out a molecular dynamics study to gain an insight into the biological environment using more effective tools, to confirm the stability of our developed compounds; then we may examine their biological activity at the other stage (in vitro and in vivo).



Appendix

Appendix A: Table of calculated descriptors of 2D-QSAR model.

obsval	PIC50	eh	log p	ref (A3)	pol (A3)	MW(aum)	HBA	log D	PSA	NRB	ENOMO	ELUMO	Egap	Pc (cm3)	n	τ	d	μ
1	4.3032	-10.42	2.18	71.77	26.59	231.3	5	-1.03	74.8	2	-0.209	-0.026	-0.183	494.1	1.681	54.6	1.33	3.6664
2	5.0048	-9.35	3.21	83.5	31.32	271.37	5	0.24	74.8	2	-0.208	-0.026	-0.182	570.1	1.724	61.8	1.39	3.2201
3	4.7812	-9.61	2.21	81.24	30.35	279.77	6	-0.11	74.8	3	-0.214	-0.034	-0.18	561.5	1.683	54.9	1.41	2.3197
4	4.9539	-17.1	-2.34	109.98	40.11	407.26	9	2.66	117.94	5	-0.229	-0.105	-0.124	743.7	1.765	70.8	1.66	6.8671
5	5.0752	-10.81	1.74	104.9	38.18	380.25	8	2.36	74.8	4	-0.22	-0.045	-0.175	690.4	1.721	58.4	1.57	5.5321
6	4.9469	-11.68	1.42	87.96	32.83	309.8	7	-0.35	84.03	5	-0.21	-0.029	-0.18	619.3	1.663	54.3	1.41	5.1093
7	5.1481	-13.67	1.77	106.77	38.82	357.84	7	1.25	84.03	6	-0.213	-0.029	-0.184	719.1	1.701	58.6	1.43	5.7522
8	5.172	-13.29	1.55	111.49	40.74	392.29	8	1.94	84.03	6	-0.203	-0.024	-0.179	748	1.711	59.8	1.51	7.3667
9	5.3645	-12.22	1.33	116.21	42.67	426.73	9	2.62	84.03	6	-0.212	-0.05	-0.163	776.8	1.72	60.8	1.58	4.5555
10	6.0044	-16.32	0.04	97.63	35.98	347.8	7	0.61	97.17	6	-0.218	-0.028	-0.189	776.8	1.72	60.8	1.58	5.4915
11	5.1314	-18.66	-2.91	111.99	40.66	402.84	9	1.095	117.94	7	-0.227	-0.103	-0.124	743.7	1.765	70.8	1.66	7.2161
12	5.1765	-13.33	1.55	111.49	40.74	392.29	8	1.94	129.85	6	-0.218	-0.031	-0.187	748	1.711	59.8	1.51	5.0794
13	5.3382	-10.45	2.56	103.82	39.39	363.89	7	1.32	84.03	6	-0.211	-0.029	-0.182	719.1	1.701	58.6	1.45	5.1005
14	5.1002	-10.39	2.45	99.69	37.56	349.86	7	0.89	84.03	5	-0.21	-0.028	-0.181	695.4	1.699	10.2	1.45	5.0695
15	5.1765	-15.29	0.78	113.15	41.29	387.87	8	1.14	93.26	7	-0.211	-0.028	-0.184	769.4	1.685	56.6	1.43	5.5876
16	4.3492	-10.17	0.82	120.39	43.84	400.91	8	1.39	87.27	7	-0.181	-0.01	-0.171	815.4	1.685	55.2	1.39	7.1367
17	5.2069	-13.3	1.17	106.9	38.73	375.83	8	1.46	84.03	6	-0.217	-0.032	-0.186	719.3	1.691	56.5	1.48	5.2232
18	5.2565	-13	2.34	111.99	40.38	425.84	10	2.26	84.03	7	-0.248	-0.047	-0.201	773.1	1.655	52.1	1.52	5.1938
19	5.8861	-12.86	2.12	113.22	41.18	373.9	10	1.92	100.1	6	-0.214	-0.032	-0.182	749.5	1.73	61.2	1.45	5.0198
20	5.1057	-10.48	1.77	94.4	35.19	325.86	6	0.33	100.1	5	-0.208	-0.04	-0.168	649.7	1.696	57.2	1.43	4.6868
21	5.3439	-13.25	1.77	106.77	38.82	357.84	7	1.24	84.03	6	-0.21	-0.027	-0.183	719.1	1.701	58.6	1.43	5.2635
22	5.3449	-10.12	1.55	111.49	40.74	392.29	8	1.94	84.03	6	-0.221	-0.033	-0.188	748	1.711	59.8	1.51	5.5135
23	4.9646	-13.55	1.39	102.19	36.8	341.39	7	0.75	84.03	6	-0.21	-0.027	-0.183	690.4	1.68	55.3	1.4	5.1168
24	5.4123	-20.13	-2.68	107.27	38.73	368.4	8	0.81	129.85	7	-0.225	-0.1	-0.126	743.7	1.722	67.1	1.49	4.852
25	4.9097	-15.88	1	108.43	38.73	353.42	7	0.44	93.26	7	-0.206	-0.025	-0.181	740.5	1.674	55.5	1.35	5.1592
26	5.1427	-11.86	1.87	92.6	34.66	323.83	7	0.12	84.03	6	-0.209	-0.031	-0.177	657.9	1.655	53.3	1.38	3.9113
27	5.567	-13.9	2.22	111.42	40.65	371.87	7	1.71	84.03	7	-0.212	-0.031	-0.181	757.7	1.691	57.4	1.4	3.7588
28	5.8665	-13.63	2.62	116.02	42.49	385.9	7	2.15	84.03	8	-0.212	-0.03	-0.181	796.3	1.682	56.4	1.38	5.272

Appendix B: Computational software using in virtual screening drug design & discovery.

Appendix C: Computational servers using to pharmacokinetic & drug-likeness properties.

- <http://biosig.unimelb.edu.au/pkcsml/prediction>.
- <https://www.swissadme.ch>.

Appendix D: Table of Y-randomization test (chapter III).

Model	R	R ²	Q ²	Model	R	R ²	Q ²
Original	0.900819	0.811475	0.578526	Original	0.900819	0.811475	0.578526
Random 1	0.617257	0.381006	-0.02843	Random 51	0.346931	0.120361	-0.57225
Random 2	0.77679	0.603402	0.323255	Random 52	0.464752	0.215994	-0.22802
Random 3	0.42713	0.18244	-0.52865	Random 53	0.554302	0.307251	-0.07466
Random 4	0.234666	0.055068	-0.65744	Random 54	0.522282	0.272778	-0.37024
Random 5	0.233215	0.054389	-0.41374	Random 55	0.524479	0.275079	-0.83486
Random 6	0.3493	0.12201	-1.65824	Random 56	0.508119	0.258185	-0.69826
Random 7	0.340565	0.115984	-0.78713	Random 57	0.515293	0.265527	-0.23084
Random 8	0.320447	0.102686	-0.74846	Random 58	0.495707	0.245725	-0.34212
Random 9	0.361597	0.130752	-0.78091	Random 59	0.321234	0.103191	-0.6664
Random 10	0.583126	0.340036	-0.24084	Random 60	0.288373	0.083159	-0.55529
Random 11	0.405991	0.164828	-0.75545	Random 61	0.427141	0.18245	-0.56825
Random 12	0.22291	0.049689	-0.63856	Random 62	0.525296	0.275935	-0.6413
Random 13	0.444058	0.197188	-0.57031	Random 63	0.460026	0.211624	-0.25021
Random 14	0.339929	0.115552	-0.63148	Random 64	0.393568	0.154896	-0.73274
Random 15	0.246054	0.060543	-0.61934	Random 65	0.616157	0.379649	-0.20779
Random 16	0.323343	0.10455	-0.50165	Random 66	0.432564	0.187111	-0.27518
Random 17	0.522765	0.273283	-0.17442	Random 67	0.602638	0.363173	-0.20981
Random 18	0.592669	0.351256	-0.36364	Random 68	0.591421	0.349779	-0.35715
Random 19	0.738147	0.54486	-0.32618	Random 69	0.59783	0.357401	-1.09344
Random 20	0.404043	0.163251	-0.95317	Random 70	0.180864	0.032712	-0.98801
Random 21	0.488713	0.238841	-0.52674	Random 71	0.707515	0.500578	0.02752
Random 22	0.419124	0.175665	-0.80826	Random 72	0.292741	0.085697	-1.86562
Random 23	0.497638	0.247644	-0.59841	Random 73	0.621268	0.385974	-0.10699
Random 24	0.385613	0.148697	-0.66182	Random 74	0.402456	0.161971	-0.55345
Random 25	0.499614	0.249614	-0.69918	Random 75	0.42794	0.183133	-1.17911
Random 26	0.411151	0.169045	-0.59274	Random 76	0.501694	0.251696	-0.16322
Random 27	0.229169	0.052518	-0.41813	Random 77	0.511167	0.261292	-0.80432
Random 28	0.540232	0.29185	-0.72733	Random 78	0.456522	0.208412	-0.92419
Random 29	0.781857	0.6113	0.209298	Random 79	0.488548	0.238679	-0.55169
Random 30	0.428501	0.183613	-0.53444	Random 80	0.669117	0.447717	-0.22213
Random 31	0.221082	0.048877	-1.37835	Random 81	0.200553	0.040221	-0.70711
Random 32	0.384212	0.147619	-0.57722	Random 82	0.48849	0.238622	-0.3919
Random 33	0.504544	0.254565	-0.21691	Random 83	0.663582	0.440342	-0.0074
Random 34	0.270896	0.073384	-0.48148	Random 84	0.382505	0.14631	-0.72194

Random 35	0.352039	0.123932	-1.16022	Random 85	0.677534	0.459052	- 0.16915
Random 36	0.559946	0.31354	-0.49041	Random 86	0.475431	0.226035	-0.59752
Random 37	0.55168	0.304351	-0.91549	Random 87	0.460155	0.211742	-0.30341
Random 38	0.643765	0.414433	-0.1615	Random 88	0.354559	0.125712	-0.26973
Random 39	0.266378	0.070957	-0.6093	Random 89	0.240598	0.057887	-0.49398
Random 40	0.301611	0.090969	-0.38155	Random 90	0.410288	0.168336	-0.58431
Random 41	0.254418	0.064728	-1.29233	Random 91	0.610211	0.372358	-0.26122
Random 42	0.504036	0.254052	-0.26302	Random 92	0.536479	0.287809	-0.59456
Random 43	0.264827	0.070133	-0.63649	Random 93	0.551948	0.304647	-0.17704
Random 44	0.327991	0.107578	-0.77359	Random 94	0.251424	0.063214	-0.767
Random 45	0.361148	0.130428	-0.87839	Random 95	0.567729	0.322316	-0.39188
Random 46	0.513164	0.263337	-0.91081	Random 96	0.534569	0.285764	-0.66548
Random 47	0.334278	0.111742	-0.69702	Random 97	0.439052	0.192766	-1.45673
Random 48	0.501247	0.251248	-0.31005	Random 98	0.240507	0.057844	-0.94552
Random 49	0.483714	0.233979	-1.22811	Random 99	0.315873	0.099776	-0.40235
Random 50	0.482727	0.233025	-0.81549	Random 100	0.379495	0.144016	-0.51855