

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Tissemsilt



Faculté des Sciences et de la Technologie
Département des Sciences de la Matière



Mémoire de fin d'études pour l'obtention du diplôme de Master académique en

Filière : Sciences de la Matière

Spécialité : Chimie moléculaire

Présentée par : Bellebia selsabil

Thème

**Contribution de l'Intelligence Artificielle et de la chimie Quantique à la Modélisation
des propriétés ADMET des Dérivés Thiazolidinones**

Soutenu le : 29/06/2026

Devant le Jury :

Nom et Prénom	Qualité	Grade	Établissement
M. Mostefai	Président	Prof.	Univ-Tissemsilt
M. Kaas sid ahmed	Encadreur	M.C.A	Univ-Tissemsilt
M. Chaker	Examinatrice	M.C.B	Univ-Tissemsilt

Année universitaire : 2025-2026



People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of Tissemsilt
Faculty of Science and Technology
Department of Matter Sciences Master's



Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Academic Master in

Field: Materials Science

Specialty: Molecular chemistry

Presented by: Bellebia Selsabil

Title

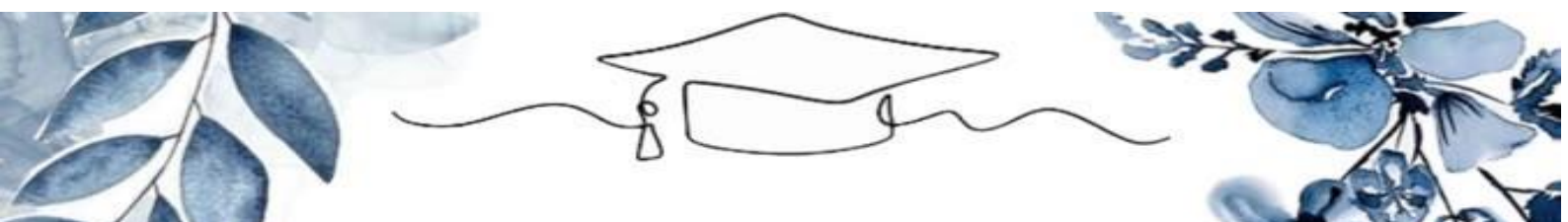
Contribution of Artificial Intelligence and Quantum Chemistry to the Modeling of
ADMET Properties of Thiazolidinone Derivatives

Defended on: 29/06/2026

Before the Jury:

Full Name	Position	Academic Rank	Institution Name
M .Mostefai	Président	Prof.	Univ-Tissemsilt
M.Kaas sid Ahmed	Encadreur	M.C.B	Univ-Tissemsilt
M.Chaker	Examinatrice	M.A.A	Univ-Tissemsilt

Academic Year: 2025–2026



Dedication

I dedicate this work:



To my mother, a symbol of tenderness and sacrifice. Thank you for your endless love, your constant support, and your prayers that have accompanied me at every stage of my life. You have always been my source of courage and inspiration To my father, a symbol of strength, dedication, and patience. Thank you for all the sacrifices and efforts you have made for my future, and for your continuous support and encouragement that have always motivated me to move forward. May God protect you and bless you with happiness.

Thank you for all the sacrifices you have made, for your silent yet immense love, and for all the values you have instilled in me. Thanks to you, I learn every day to become a better person.

To my sisters “Hind, Hiba, Yousra”, who have been my companions, my confidantes, and my greatest supporters. Your presence has always been a source of comfort.

To my brother, who has been a source of support and encouragement throughout different stages of my life. Thank you for your constant motivation and for always being there for me. Wishing you success and happiness always.

To my friends, for their sincere encouragement, their attentive listening, and their unwavering support.

To those who have been the quiet supporters throughout this long journey.

Thank you all.





Acknowledgements

First and foremost, I thank God for granting me the knowledge, strength, and perseverance necessary to successfully complete this thesis.



I would like to express my sincere gratitude to Mr. kaas Ahmed for their supervision, valuable advice, availability, and continuous support throughout the completion of this work.

I would also like to thank the members of the jury for accepting to evaluate this thesis and for dedicating their time to examining this work, as well as for the comments and suggestions they will provide.

Finally, I extend my gratitude to all those who contributed, directly or indirectly, to the successful completion of this thesis.

To all of you, thank you from the bottom of my heart.



Summary

This work aims to investigate thiazolidinone derivatives and evaluate their drug potential through the integration of computational chemistry and machine learning techniques. Density Functional Theory (DFT) calculations were performed to extract important quantum chemical descriptors, including HOMO and LUMO energies and the energy gap, in addition to physicochemical descriptors such as LogP, TPSA, and molecular weight. The pharmacokinetic and drug-likeness properties of these compounds were also evaluated using ADMET parameters and Lipinski's Rule of Five.

The study employed Quantitative Structure–Activity Relationship (QSAR) modeling combined with artificial intelligence and machine learning techniques to develop predictive models capable of correlating molecular descriptors with ADMET properties. The results revealed clear relationships between molecular structure and the electronic and pharmacological properties of the investigated compounds. Furthermore, several molecular descriptors were identified as key factors influencing pharmacokinetic behavior and bioavailability.

The findings demonstrate the effectiveness of combining DFT calculations with machine learning approaches for predicting the pharmacological properties of thiazolidinone derivatives, thereby accelerating the drug discovery process and reducing experimental costs. Moreover, the study enabled the identification of several promising compounds that may serve as potential candidates for the development of new drugs with high therapeutic efficacy and low toxicity.

Keywords: Thiazolidinone, DFT, QSAR, Artificial Intelligence, Machine Learning, ADMET.

يهدف هذا العمل إلى دراسة مشتقات الثيازوليدون وتقييم إمكاناتها الدوائية من خلال دمج تقنيات الكيمياء الحاسوبية والتعلم الآلي. تم إجراء حسابات نظرية دالة الكثافة (DFT) لاستخراج مجموعة من الواصفات الكيميائية الكمية المهمة، مثل طاقات HOMO و LUMO وفجوة الطاقة، بالإضافة إلى حساب الواصفات الفيزيائية والكيميائية، بما في ذلك LogP وTPSA والوزن الجزيئي. كما تم تقييم الخصائص الدوائية والحركية لهذه المركبات اعتمادًا على معايير ADMET وقاعدة ليبينسكي الخماسية. اعتمدت الدراسة على نمذجة العلاقة الكمية بين البنية والنشاط (QSAR) وتقنيات الذكاء الاصطناعي والتعلم الآلي لبناء نماذج تنبؤية قادرة على الربط بين الواصفات الجزيئية وخصائص ADMET. وقد أظهرت النتائج وجود علاقات واضحة بين البنية الجزيئية والخصائص الإلكترونية والدوائية للمركبات المدروسة، كما بينت أن بعض الواصفات الجزيئية تلعب دورًا رئيسيًا في تحديد السلوك الحركي الدوائي والتوافر الحيوي. تؤكد هذه الدراسة فعالية الجمع بين حسابات DFT وتقنيات التعلم الآلي في التنبؤ بالخصائص الدوائية لمشتقات الثيازوليدون، مما يساهم في تسريع عملية اكتشاف الأدوية وتقليل التكاليف التجريبية. كما مكّنت من تحديد عدد من المركبات الواعدة التي يمكن أن تمثل مرشحات محتملة لتطوير أدوية جديدة ذات فعالية علاجية مرتفعة وسُمّية منخفضة.

الكلمات المفتاحية: الثيازوليدون، DFT، QSAR، الذكاء الاصطناعي، التعلم الآلي، ADMET.

Ce travail vise à étudier les dérivés de la thiazolidinone et à évaluer leur potentiel pharmacologique en combinant les techniques de chimie computationnelle et d'apprentissage automatique. Des calculs basés sur la Théorie de la Fonctionnelle de la Densité (DFT) ont été réalisés afin d'extraire d'importants descripteurs quantiques, tels que les énergies HOMO, LUMO et le gap énergétique, ainsi que des descripteurs physicochimiques comme le LogP, la TPSA et la masse moléculaire. Les propriétés pharmacocinétiques et l'aptitude médicamenteuse de ces composés ont également été évaluées à l'aide des paramètres ADMET et de la règle des cinq de Lipinski.

L'étude s'appuie sur la modélisation QSAR (Quantitative Structure–Activity Relationship), associée aux techniques d'intelligence artificielle et d'apprentissage automatique, afin de développer des modèles prédictifs capables d'établir des relations entre les descripteurs moléculaires et les propriétés ADMET. Les résultats ont mis en évidence des relations significatives entre la structure moléculaire et les propriétés électroniques et pharmacologiques des composés étudiés. Plusieurs descripteurs moléculaires ont également été identifiés comme des facteurs déterminants du comportement pharmacocinétique et de la biodisponibilité.

Les résultats obtenus démontrent l'efficacité de la combinaison des calculs DFT et des méthodes d'apprentissage automatique pour la prédiction des propriétés pharmacologiques des dérivés de la thiazolidinone. Cette approche contribue à accélérer le processus de découverte de nouveaux médicaments tout en réduisant les coûts expérimentaux. En outre, elle a permis d'identifier plusieurs composés prometteurs susceptibles de constituer des candidats potentiels pour le développement de nouveaux médicaments présentant une forte efficacité thérapeutique et une faible toxicité.

Mots-clés : Thiazolidinone, DFT, QSAR, Intelligence Artificielle, Apprentissage Automatique, ADMET.

List of Abbreviations

OPVs	Organic photovoltaic compounds
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
HBD	hydrogen bond acceptors(HBA), and hydrogen bond donors
SVM	Support Vector Machine
LADMER	Liberation – Absorption – Distribution – Metabolism Elimination – Response the acronym
QSAR	Quantitative Structure – Activity Relationship
1D	One – dimensional
2D	Two – dimensional
3D	Three – dimensional
DFT	Density Functional Theory
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
HF	Hartree – Fock
KS	Kohn – Sham
SCF	self-consistent field
GGA	Generalized Gradient Approximation
LYP	(Lee, Yang, Parr, 1988),
PW91	(Perdew and Wang, 1992),
P86	(Perdew, 1986),
B95	(Becke, 1996).
AI	Artificial intelligence
ML	machine learning
SVR	Support Vector Machines (SVM) and Support Vector Regression
RBF	Radial Basis Function
RF	Random Forest
GBM	Gradient Boosting Machines
DL	Deep Learning
GNNs	Graph Neural Networks

MPNNs	Message Passing Neural Networks
SMILES	(Simplified Molecular-Input Line-Entry System)
RFE	Feature Elimination
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
AD	Applicability Domain
EDG	Electron-Donating
EWG	Electron-Withdrawing Groups
CID	Compound Identifier
PUG	(PubChem Upload Gateway)
PUG-REST	(Representational State Transfer)
STO	Slater-type orbitals
TPSA	Total Polar Surface Area
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error

Table of Contents

Cover Page

Dedication

Abstract

General Introduction

Chapter 01:

Literature Review

I.1.Thiazolidinone Derivatives.....	21
I.1.1 Introduction.....	21
I.1.2 Thiazoles	21
I.1.3 Methods for the synthesis of thiazoles.....	22
I.1.4 Thiazolidinones.....	25
I.2. Pharmacokinetics and ADMET Profile	27
I.2.1Definition of ADMET.....	27
I.2.2 Pharmacokinetics	28
I.2.3Absorption, Distribution, Metabolism, Elimination, Toxicity (ADMET)	28
I.3.Molecular Modeling and QSAR.....	30
I.3.1.QSAR.....	30
I.3.2.Molecular Descriptors.....	31
I.3.3.Construction of QSAR Models	32
I.3.4.Model Evaluation Parameters.....	34
I.3.5.Model Interpretation	36
I.3.6.Overall Strategy of a QSAR Study.....	37
I.3.7.Conclusions.....	38
I.4.Advanced Computational Tools	38
I.4.1General Overview of Density Functional Theory (DFT)	38
I.4.2.Principle of DFT Calculations	39
I.4.2.Application of DFT to Molecular Systems.....	42
I.5.Artificial intelligence	45
I.5.1.Definitions of AI.....	45
I.5.2.Introduction to Artificial Intelligence(AI) and Machine Learning (ML) for ADMET Prediction	46
I.5.3.Foundational Machine learning Algorithms in cheminformatics.....	46
I.5.4.Support Vector Machines (SVM) and Support Vector Regression (SVR)	47
I.5.5.Ensemble Method: Random Forests and Gradient Boosting	47
I.5.6.Feature Engineering and Selection	48
I.5.7.Model Validation and the Applicability Domain	49

Chapter II: Materials and Methods

II.1 Data set Constitution	60
II.1.1 Selection Criteria for the 40 Thiazolidinone Molecule (Electronic and Structural Diversity)	60
II.1.2. Data Sources (PubChem) and Storage Format (SMILES)	61
II.2. Software and Computing Environment.....	70
II.2.1. Quantum Chemistry software.....	70
II.2.2. Cheminformatics Tools: RDkit, SwissADME, and complementary Solutions	74
II.2.3. Machine Learning Environment.....	76
II.3. DFT Calculation Protocol for Quantum Descriptor	77
II.3.1. Density Functional theory (DFT)	77
II.3.2. Orbital Basis Sets.....	77
II.3.4. Energy Gap	78
II.4. Physicochemical Descriptors Calculation	78
II.4.1. LogP Calculation Method.....	79
II.4.2. Total Polar Surface Area (TPSA) Calculation Method.....	79
II.5. QSAR Modeling via Machine Learning.....	79
II.5.1. Data Set Preparation	79
II.5.2. Choice and Justification of al Algorithm	80
II.5.3. Validation Protocol.....	81

Chapter III: Results and Discussion

Introduction	86
III.1. Descriptor Calculation Result	86
III.1.1. Summary Table of Quantum Descriptors	89
III.1.2. Summary Table of Physicochemical Descriptor	88
III.1.3. Correlation Analysis Between Descriptors	89
III.2. Analysis of Structural Influence.....	91
III.2.1. Impact of Substituents on Orbital Energies	91
III.2.2. Lipinski's Rule of Five Analysis.....	92
III.3. Machine Learning Model Performance	93
III.3.1. Model Development and Validation	93
III.3.1. Model Performance Metrics.....	95
III.3.3. Cross-Validation and Robustness	95
III.3.3. AI Model Implementation and Reproducibility.....	95
III.4. Structure-Activity Relationships (SAR)	100
III.4.1. Impact of Electron-Donating Substituents (Group 1)	100
III.4.2. Impact of Electron-Withdrawing Substituents (Group 2)	101

III.4.3.Impact of Halogenated Substituents (Group 3)	101
III.4.4.Impact of Heterocyclic and Polycyclic Substituents (Group 4)	102
III.4.5.Quantitative SAR Analysis	102
III.5.Discussion and Interpretation.....	102
III.5.1.Structural Determinants of ADMET Properties.....	102
III.5.2.Predictive Model Insights.....	103
III.5.4.Implications for Drug Design.....	103
III.5.5.Limitations and Future Directions	103
III.5.5.Limitations and Future Directions	104
III.5.6.Conclusions	104
General Conclusion	108

Table of Figure

Figure	Page
Figure I .1: the main derivatives of 1,3-thiazoles	21
FigureI .2 :structure of thiazoles and isothiazoles	22
Figure I .3 : Structure of thiamine(vitamin B1)	22
FigureI.4: Structure of penicillin	22
FigureI.6: structure of thiazolidinones	25
Figure I .5: Chemical structure of actithiazic acid	26
Figure I.7: Overall Strategy of a QSAR Study	37
Figure II - 1: Cycle of studying a molecule using molecular modeling.	73
Figure II - 2: ADMET prediction using swiss ADME	75

Table of Schema

Scheme	Page
Scheme I.1: Hantzsch synthesis of thiazoles	23
Scheme I .2 :Gabriel Synthesis of thiazoles	23
SchemeI.3 :Cook-Heilborn Synthesis of thiazoles	24
SchemeI.4: Synthesis of Substituted Thiazoles from Isothiocyanates and Alkynes	24
Scheme I.5 :Dubs Synthesis	25
Scheme I .10: Synthesis of thiazoles from Benzonitriles and 2-Mercaptopenten -3-one	25

Table of Table

Table	Page
Table01 : SMILES and 2D structure for the group01	64
Table 02 : SMILES and 2D structure for the group02	67
Table 03: SMILES and 2D structure for the group03	67
Table 04 : SMILES and 2D structure for the group04	68
Table 05:Some basis sets used in Gaussian	70
Table 06 : HOMO LUMO and Energy Gap for the Molecule 01	74
Table 01: the calculated quantum chemical descriptors for the thiazolidinone molecules	87
Table 02: summarizes the physicochemical descriptors calculated for the thiazolidinone molecules using SwissADME and related computational tools	89

General Introduction

Over the past few decades, the landscape of medicinal chemistry has been profoundly transformed by significant advancements in molecular modeling and artificial intelligence (AI). These innovations have not only accelerated the discovery and design of novel therapeutic agents but also enhanced their efficacy and reduced potential toxicity [1], [2]. This paradigm shift has made computational approaches indispensable in modern drug development pipelines, where machine learning (ML) models are revolutionizing ADMET prediction by enhancing accuracy and reducing experimental costs [3], [4]. Among the vast array of heterocyclic compounds explored in pharmaceutical research, thiazolidinone derivatives have garnered considerable attention due to their remarkable and diverse spectrum of biological activities [5], [6]. Recent advances (2021–present) in the synthesis and structure–activity relationship (SAR) of thiazolidin-4-one derivatives have highlighted their considerable potential as anticancer agents, capable of inhibiting key enzymes and cell signaling pathways [7], [8]. These compounds have also demonstrated significant potential across other therapeutic areas, including antidiabetic, antibacterial, antifungal, anti-inflammatory, and antioxidant applications [9], [10]. Their versatility stems from their unique structural features, which allow for extensive chemical modifications and interactions with multiple biological targets. The fundamental structure of thiazolidinones is characterized by a five-membered heterocyclic ring, incorporating a sulfur atom at position 1 and a nitrogen atom at position 3, along with a carbonyl group at position 4. The strategic placement of these atoms, coupled with the possibility of attaching various substituents at positions 2, 3, and 5, gives rise to a wide diversity of derivatives, each possessing distinct chemical and biological properties [11].

This structural flexibility is key to their broad pharmacological utility, as evidenced by recent scientific reports on their mechanisms of action [12]. The development of effective drug candidates hinges not only on their intrinsic biological activity but also, critically, on their physicochemical and pharmacokinetic properties. In this context, the ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties play a pivotal role in determining the ultimate success or failure of a drug during its development [13]. Early and accurate prediction of these properties has thus become an essential step in modern drug discovery, aiming to mitigate the high attrition rates associated with late-stage failures [14]. Computational approaches and molecular modeling techniques offer powerful and cost-effective tools for evaluating the pharmacological behavior of compounds before embarking on expensive and time-consuming experimental studies [15]. Quantitative Structure-Activity Relationship (QSAR)

General Introduction

modeling, in particular, establishes mathematical relationships between the chemical structure of molecules and their biological activities or properties [16]. Complementing this, Density Functional Theory (DFT) calculations provide a robust framework for determining important quantum chemical descriptors that elucidate the electronic structure and reactivity of molecules [17], [18].

The present work focuses on the investigation of a series of thiazolidinone derivatives using a synergistic combination of computational chemistry and machine learning approaches. DFT calculations were employed to determine important quantum chemical descriptors, including the energies of the Highest Occupied Molecular Orbital (HOMO), the Lowest Unoccupied Molecular Orbital (LUMO), and the energy gap (ΔE) [19]. These descriptors are crucial for understanding molecular reactivity and stability. In addition, key physicochemical descriptors such as the partition coefficient (LogP) and Topological Polar Surface Area (TPSA) were calculated and analyzed, providing insights into lipophilicity and molecular size, which are fundamental to ADMET properties [20]. Furthermore, QSAR modeling combined with artificial intelligence (AI) and machine learning (ML) techniques was utilized to establish predictive relationships between molecular descriptors and ADMET properties [21], [22]. These advanced computational methods, including graph neural networks and ensemble learning, offer an efficient and powerful strategy for identifying the most relevant molecular features influencing pharmacokinetic behavior and biological performance [23], [24].

In this context, a critical research question arises: How can an intelligent predictive model, specifically dedicated to thiazolidinone derivatives, be developed to establish reliable relationships between quantum descriptors obtained from DFT calculations and ADMET properties in order to optimize the design of new bioactive molecules?

The importance of this question lies in its potential to enhance our understanding of the intricate relationship between molecular structure and pharmacological behavior, thereby guiding the rational design of novel compounds with high therapeutic efficacy and low toxicity. The main objective of this study is to explore this intricate relationship between the molecular structure and pharmacological properties of thiazolidinone derivatives. Specifically, this work aims to evaluate the influence of quantum and physicochemical descriptors on the ADMET profiles of these compounds, assess the predictive capability of machine learning models in accurately forecasting ADMET properties, and ultimately, identify the most promising thiazolidinone compounds that could serve as potential candidates for future drug development, based on their predicted ADMET profiles.

General Introduction

This thesis is organized into three main chapters: Chapter I provides a comprehensive overview of thiazolidinone derivatives, the significance of ADMET properties in drug discovery, the principles of QSAR and molecular modeling, and an introduction to advanced computational tools like DFT and Machine Learning. Chapter II details the experimental numerical methodology, including the constitution of the dataset of thiazolidinone molecules, the software and computing environment utilized, the DFT calculation protocol for quantum descriptors, the calculation of physicochemical descriptors, and the QSAR modeling approach using Machine Learning. Chapter III presents the findings from the descriptor calculations, analyzes the structural influence on properties, evaluates the performance and interpretation of the developed AI models, and concludes with a synthesis of the main results, identification of promising candidates, and perspectives for future work.

References

- [1] Leveraging machine learning models in evaluating ADMET ... - PMC. (n.d.). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC12205928/>
- [2] In Silico ADMET Modeling: The 20-Year Machine Learning ... - Drug Patent Watch. (n.d.). Retrieved from <https://www.drugpatentwatch.com/blog/advancements-in-in-silico-admet-modeling-a-20-year-journey-of-machine-learning-in-drug-discovery-at-bayer-pharma/>
- [3] a machine learning ADMET platform for evaluation of large-scale ... - Oxford Academic. (n.d.). Retrieved from <https://academic.oup.com/bioinformatics/article/40/7/btae416/7698030>
- [4] Inductive Bio Wins World's Largest ADMET Prediction Competition. (2026). Retrieved from <https://www.inductive.bio/news/inductive-bio-wins-worlds-largest-admet-prediction-competition>
- [5] The Bioactivity of Thiazolidin-4-Ones: A Short Review of the Most ... - PMC. (n.d.). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC8584074/>
- [6] Synthetic and medicinal perspective of thiazolidinones: A review - ScienceDirect. (n.d.). Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0045206817302195>
- [7] Recent advances in synthetic strategies and SAR of thiazolidin-4 ... - PMC. (n.d.). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC10584807/>
- [8] Review of Recent Advances in Thiazolidin-4-One Derivatives as ... - MDPI. (n.d.). Retrieved from <https://www.mdpi.com/1420-3049/30/10/2201>
- [9] Recent developments and biological activities of thiazolidinone ... - Academia.edu. (n.d.). Retrieved from https://www.academia.edu/80326190/Recent_developments_and_biological_activities_of_thiazolidinone_derivatives_A_review
- [10] A comprehensive review on the Synthesis and Anticancer Potential ... - IJMPS. (n.d.). Retrieved from <https://ijmphs.com/index.php/IJMPS/article/view/58>
- [11] Chemistry and biological activity of thiazolidinones - ACS Publications. (n.d.). Retrieved from <https://pubs.acs.org/doi/10.1021/cr00042a003>
- [12] A brief overview of 4-Thiazolidinone's biological activity, SAR ... - ResearchGate. (2023). Retrieved from https://www.researchgate.net/publication/375772842_A_brief_overview_of_4-Thiazolidinone's_biological_activity_SAR_current_advancements_and_impending_challenges
- [13] In silico ADMET: Supporting new drug discovery - future health. (2026). Retrieved from <https://futurehealthjournal.com/in-silico-admet-supporting-new-drug-discovery/>
- [14] QSAR Model for Drug Human Oral Bioavailability - ACS Publications. (n.d.). Retrieved from <https://pubs.acs.org/doi/10.1021/jm0000564>
- [15] Beginner's Guide to 3D-QSAR in Drug Design - Neovarsity. (2025). Retrieved from <https://neovarsity.org/blogs/beginners-guide-3d-qsar-drug-design>
- [16] Statistical modelling of molecular descriptors in QSAR/QSPR - Wiley Online Library. (n.d.). Retrieved from <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527645121>
- [17] Applications of conceptual density functional theory in reference to ... - Taylor & Francis. (2024). Retrieved from <https://www.tandfonline.com/doi/full/10.1080/00268976.2024.2331620>
- [18] Quantum chemical descriptors in QSAR/QSPR studies - ACS Publications. (n.d.). Retrieved from <https://pubs.acs.org/doi/full/10.1021/cr950202r>

General Introduction

- [19] DFT-based reactivity and combined QSAR, molecular docking of 1,2 ... - PMC. (2019). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC6819827/>
- [20] Quantum chemical descriptors in quantitative structure–activity ... - ScienceDirect. (2021). Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0169743921001520>
- [21] The efficacy of conceptual DFT descriptors and docking scores on the QSAR models of HIV protease inhibitors - Ingenta Connect. (n.d.). Retrieved from <https://www.ingentaconnect.com/content/ben/mc/2012/00000008/00000005/art00007>
- [22] ... calculations of molecular descriptors for anticancer activities of 1, 2, 3-triazole-pyrimidine derivatives against gastric cancer cell line (MGC-803): DFT, QSAR and ... - Cell. (2020). Retrieved from [https://www.cell.com/heliyon/fulltext/S2405-8440\(20\)30771-4](https://www.cell.com/heliyon/fulltext/S2405-8440(20)30771-4)
- [23] Machine learning approaches for next-generation ADMET prediction - ScienceDirect. (2025). Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S1359644625002004>
- [24] Rational Design of Novel Inhibitors: Integrating 3D-QSAR and ... - PMC. (n.d.). Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC12860141/>

Chapter 01:

Literature Review

I.1.Thiazolidinone Derivatives

I.1.1 Introduction

In recent year, researchers have focused on the synthesis of heterocycle containing nitrogen, sulfur, and oxygen, such as pyrrole, thiazole, due to their biological importance [1].

Among these heterocycle, the thiazole ring has been widely studied [2] . Indeed, this ring is present in numerous molecules exhibiting significant biological or pharmacological activities (antibacterial,antiviral ,and antifungal,etc.).It is also found in many compounds with industrial application (figure I.1) [3].

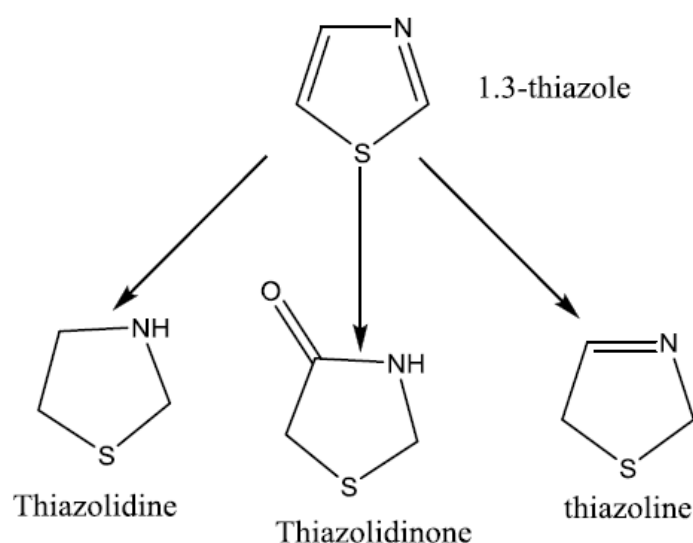


Figure I .1: the main derivatives of 1,3-thiazoles

I.1.2 Thiazoles

Thiazoles, or 1,3-thiazoles , are five –membered heterocyclic compounds containing three carbon atoms , are sulfur atom , and one nitrogen atom, with the molecular formula C_2H_3NS (FigureI.2)

Among the isomers of tiazoles ,isothiazoles are also found :

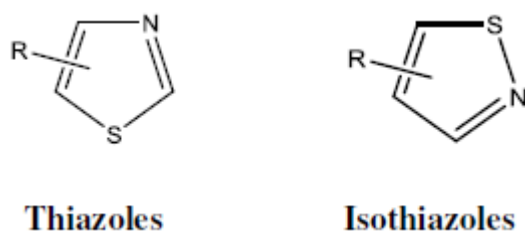


Figure I .2 :structure of thiazoles and isothiazoles

Thiazole was first defined by Hantzsch and Weber in 1884[4]. It is present in numerous natural products such as vitamin B1 (thiamine) (Figure I.3), alkaloids, steroids, as well as various pharmaceutical products[5].

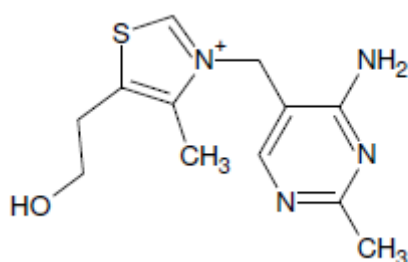


Figure I .3 : Structure of thiamine(vitamin B1)

Thiazole and its derivatives are highly valuable compounds in various fields of chemistry, medicine, and agriculture. Consequently, thiazole also serves as a synthetic intermediate for numerous biologically active compounds such as penicillin (Figure I.4) [6].

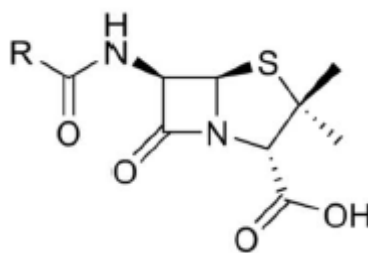


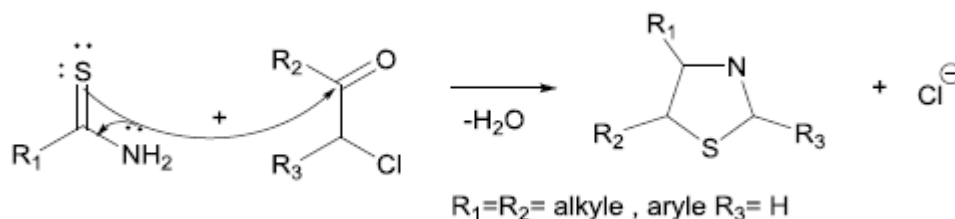
Figure I .4: Structure of penicillin

I.1.3 Methods for the synthesis of thiazoles

In assessing the importance of thiazole and its derivatives, numerous synthetic techniques for thiazole derivatives have been developed by different research groups, such as Hantzsch in 1887, Gabriel in 1910, Cook-Heilborn in 1949, and other research teams.

a) Hantzsch Synthesis of thiazoles

The Hantzsch reaction, discovered in 1887, is the most common method for the synthesis of thiazoles. It involves the condensation between an α -halo carbonyl compound and a thioamide (Scheme I.1). [7]

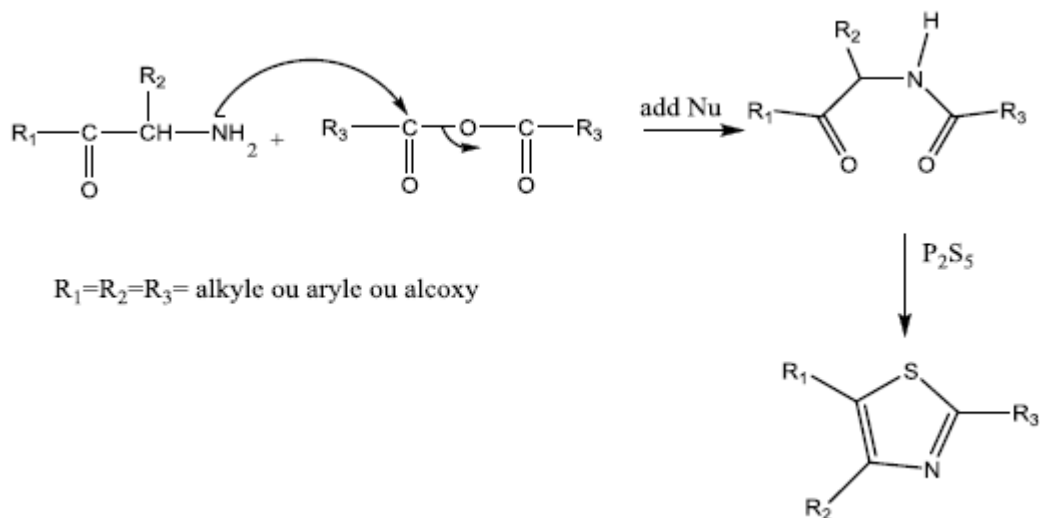


Scheme I .1: Hantzsch synthesis of thiazoles

b) Gabriel Synthesis of thiazoles

The Gabriel method is a cyclodehydration approach of α -acylaminoketones.

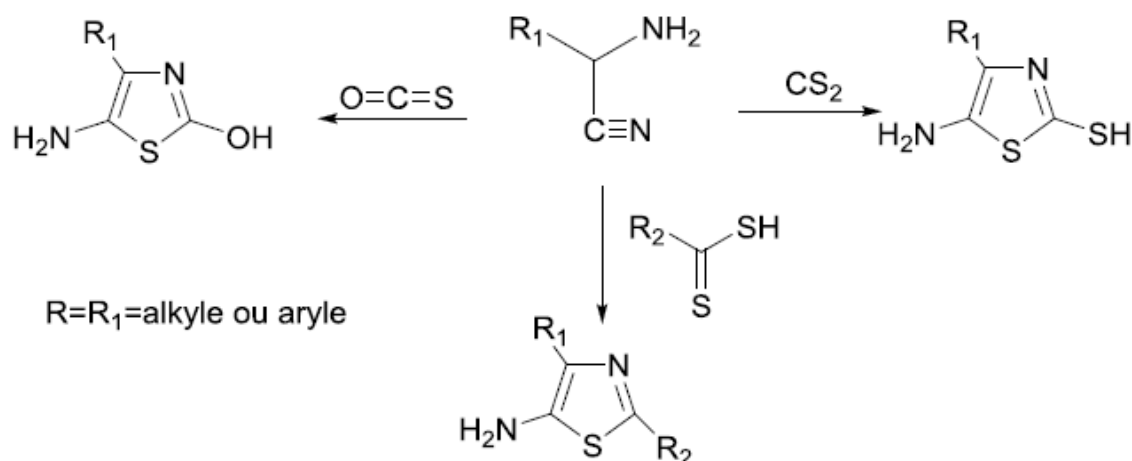
This method is widely used to synthesize thiazoles substituted with alkyl, aryl, or alkoxy groups at positions 2,4, and 5 (Scheme I.2). [8]



Scheme I .2 : Gabriel Synthesis of thiazoles

c) Cook-Heilborn Synthesis of thiazoles

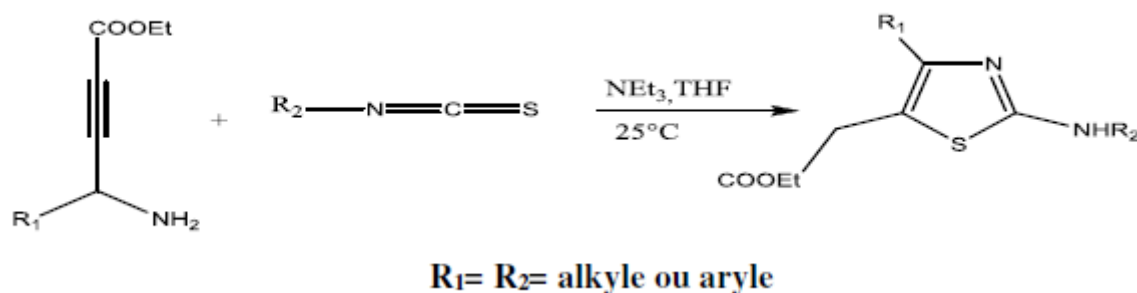
This is a method for synthesis of substituted aminothiazoles .It involves the reaction of α -aminonitriles with dithioacid salts or esters ,carbon disulfide ,and isothiocyanates under mild condition , leading to the formation of substituted 5-aminothiazoles (Scheme I.3) [9].



Scheme I.3 : Cook-Heilborn Synthesis of thiazoles

d) Synthesis of Substituted Thiazoles from Isothiocyanates and Alkynes

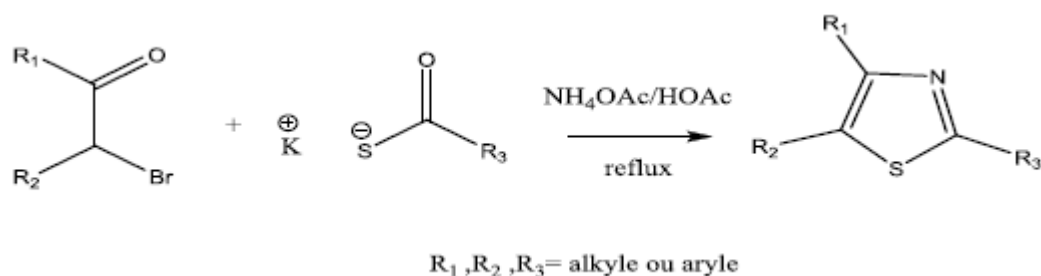
In this reaction ,isothiocyanates react with aminobutynoates in a basic medium using THF as a solvent at 25°C ,yielding thiazoles (SchemeI.4) [10].



Scheme I .4: Synthesis of Substituted Thiazoles from Isothiocyanates and Alkynes

e) Dubs Synthesis

The reaction is carried out in a one-pot procedure ,where an α -bromoketone is treated with potassium thioacid salts in the presence of ammonium acetate in acetic acid under reflux , yielding a substituted 1,3-thiazole derivative (SchemeI.5) [11,12].



Scheme I.5 :Dubs Synthesis

f) Synthesis of thiazoles from Benzonitriles and 2-Mercaptopenten -3-one

4-Ethyl-5-methyl-2-phenylthiazoles are prepared, with a relatively low yield, by the reaction of 2-mercaptopenten -3-one with benzonitrile in the presence of HCl (SchemeI.10) [13].



SchemeI.10: Synthesis of thiazoles from Benzonitriles and 2-Mercaptopenten -3-one

I.1.4 Thiazolidinones

a) Introduction

4-oxo-thiazolidines or thiazolidinones are thiazolidine derivatives that represent a class of compounds of great scientific interest due to their chemical properties and biological activities, such as antibacterial [14], including antifungal [15], anti-inflammatory, analgesic [16], and anti-tumor [17]. The thiazolidinone ring is found in nature in actithiazic acid (FigureI.4). [18]

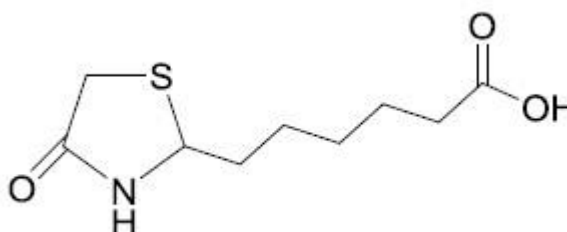


Figure I .5: Chemical structure of actithiazic acid

b) Structure

The basic structure of thiazolidinones consists of a five-membered ring containing two heteroatoms: a sulfur atom at position 1 and a nitrogen atom at position 3, a carbonyl group at position 4, and various substituents can be present at positions 2, 3, and 5 (Figure I.5) [19].

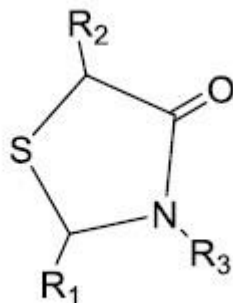


Figure I.6: structure of thiazolidinones

c) Fields of Application of Thiazolidinones:

Following the discovery of thiazolidinones, extensive research has been conducted on the synthesis and activity studies of these materials. A large number of these derivatives possess diverse pharmacological properties, such as antiviral, antiparasitic, antitubercular, antimicrobial, anticonvulsant, anti-inflammatory, and analgesic activities.

- **Biological Activity:**

Thiazolidinone derivatives represent a class of compounds of great scientific interest due to their chemical properties and pharmacological activities.

These compounds are characterized by their versatile and significant biological activities, including: antimicrobial, anti-HIV, anticancer, anticonvulsant, and antibacterial properties.

Thiazolidinones have attracted increasing interest among researchers and open perspectives in the fight against various diseases such as allergy, cancer, and malaria.

- **Nonlinear Optical Activity**

In recent years, particular attention has been devoted to organic nonlinear (NLO) materials due to their potential applications in optoelectronic technology. These compounds exhibit a strong nonlinear optical response, extremely fast switching times, and convenient optimization pathways through molecular engineering compared to well-known inorganic materials reported in the literature, such as KDP.

Thiazolidinone derivatives are therefore considered promising candidates in the field of nonlinear optics owing to their potential application in optical signal processing and information storage.

- **Photovoltaic Activity**

Organic photovoltaic compounds (OPVs) have attracted considerable interest in recent years, as they appear to be cost-effective alternatives to conventional semiconductor-based photovoltaic devices. However, further improvements are still required in terms of efficiency and operational lifetime. Organic compounds incorporating the thiazolidinone moiety and its derivatives have demonstrated significant photovoltaic activity.[20]

I.2.Pharmacokinetics and ADMET Profile

I.2.1Definition of ADMET

For a molecule to be effective as a drug, it must reach its biological target in the body at a sufficient concentration and remain in a bioactive form for an adequate period to produce the desired biological effects. Therefore, drug development involves early evaluation of **Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET)** properties during the drug discovery process.

A wide range of in silico methods aim to predict ADMET parameters based on molecular structure [21]. Notably, the pioneering work of Lipinski et al. examined orally active compounds to define physicochemical ranges associated with a high probability of oral drug-likeness [22]. This concept is known as the **lipinski Rule of Five**, which establishes a relationship between pharmacokinetic and physicochemical parameters.

Furthermore, pharmacokinetic properties such as gastrointestinal absorption, as well as drug-likeness predictions including lipinski's rules and bioavailability scores, are commonly evaluated[23]. The Lipinski rule is applied to assess drug-likeness and predict whether a compound is likely to be bioactive based on key parameters such as molecular weight, LogP, hydrogen bond acceptors(HBA), and hydrogen bond donors (HBD).

The SwissADME tool employs a Support Vector Machine (SVM) algorithm [24] trained on large, carefully curated datasets of known inhibitors/non-inhibitors and substrates/non-substrates. Screening of **Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET)** is a useful approach to assess the drug-likeness potential of compounds. The

SMILES format of the ligands was uploaded to the **SwissADME** server [25], and ADMET screening was performed using the default parameters.

I.2.2 Pharmacokinetics

Pharmacokinetics is traditionally defined as the study of the fate of drugs in the body [26], from their absorption to their elimination. This discipline enables the correlation of administered drug doses with observed blood concentration and describes their changes over time.

Among the commonly used concepts in pharmacokinetics is bioavailability, which refers to the fraction of an administered drug dose that reaches the systemic circulation and the rate at which it does so. It primarily reflects the absorption phase.

I.2.3 Absorption, Distribution, Metabolism, Elimination, Toxicity (ADMET)

The journey of a drug within the body can be divided into several key phases:

- ✓ **Absorption phases:** the process by which the drug moves from its site of administration into the systemic circulation.
- ✓ **Distribution phase:** the dispersion of the drug throughout the various tissues of the body.
- ✓ **Metabolism phase:** describes the chemical transformation of the drug by the body.
- ✓ **Elimination phase:** concerns the processes through which the drug is removed from the body.
- ✓ **Toxicity phase:** refers the potential of the drug to produce harmful effects of the drug once it reaches its target sit

This five phases are commonly summarized by the acronym **ADMET**. Some authors expand this concept to include the release phases of the active ingredient from the administered dosage form, and the response phase to the active ingredient, which describes the following sequence:

Liberation – Absorption – Distribution – Metabolism – Elimination – Response the acronym **LADMER** is then used to designate all of these phases [27].

a) Absorption

In most clinical situation, the site of drug administration is different from its site of action. For example, when a drug is administered orally, its site of action may be the heart, blood

vessels, or kidneys. The site of administration – namely the gastrointestinal tract in this case – serves merely as a transit location, after which the drug is transported throughout the body via the bloodstream.

Absorption therefore refers to the passage of the drug from its site of administration into the systemic circulation through biological membranes.

Before absorption can occur, it is often necessary for the active ingredient to be released from its dosage form, such as through the dissolution of a tablet or the breakdown of a capsule shell. Several mechanisms are involved in drug absorption; these can be classified into active transport processes and passive absorption processes.

b) Distribution

Distribution corresponds to the diffusion of a drug throughout the various tissues of the body. It is influenced by the lipophilicity of the active compound, as well as by its affinity for specific tissues such as bone or adipose tissue.

A drug may bind reversibly to plasma proteins, mainly to albumin and alpha – 1 acid glycoprotein. Albumin preferentially binds acidic drugs, whereas alpha – 1 acid glycoprotein binds predominantly basic drugs. Thus, a distinction is made between the protein – bound fraction and the free fraction of the drug. In most cases, only the free fraction is able to exert its pharmacological effect at the site of action, diffuse into tissues, or undergo metabolism and elimination processes. Consequently, the fraction bound to plasma proteins can be considered as a storage form of the drug within the body.

c) Metabolism

Metabolism refers to the transformation of a drug by the body. This transformation is mainly carried out through enzymatic reaction. In pharmacology, the reactions involved in drug metabolism are commonly divided into two main group: phase I reactions and phase II reactions. Phase I reactions generally precede phase II reactions in the chronological sequence of drug metabolism.

d) Elimination

Drug elimination from the body can occur through several routes: urinary excretion, respiratory elimination (in expired air), cutaneous excretion (through sweat), salivary excretion, or through breast milk. The main elimination pathways are renal excretion in urine and hepatic elimination. For certain elimination routes, a phenomenon of drug reabsorption may occur,

which contributes to decreasing or slowing down the elimination process. The mechanisms described for absorption are also involved in elimination, including passive diffusion and active transport.

e) Toxicity

Toxicity refers to the harmful effects that a drug may produce in a living organism. These effects may occur when the therapeutic dose is exceeded, as a result of drug accumulation, individual hypersensitivity, or drug – drug interactions. Toxicity may be acute (occurring after short – term exposure or high doses) or chronic (resulting from long – term exposure). It can also be organ – specific, such as hepatotoxicity, nephrotoxicity, cardiotoxicity, or neurotoxicity.

The evaluation of toxicity is a fundamental step in drug development, as it aims to assess the safety profile of candidate compounds and to determine a safe therapeutic window before clinical trials.[28]

I.3.Molecular Modeling and QSAR

I.3.1.QSAR

Quantitative Structure – Activity Relationship (QSAR) aims to correlate a specific response variable (biological activity) with molecular descriptors that are calculated or experimentally measured from the molecules themselves. These methods were first developed by **Corwin Hansch** in the 1940s by analyzing series of compounds and formulation the classical QSAR equation[29],[30]

$$\text{Log} \frac{1}{C} = ap + bs + cEs + \text{const} \quad (\text{eq.1})$$

- **[C]**:concentration required to produce the biological effect
- **[p]**:octanol – water partition coefficient
- **[s]**:Hammett substituent constant (electronic effect)
- **[Es]**:Taft substituent constant (steric effect)

Log P is a measure of the difference in solubility of a compound between two solvent. Typically, one solvent is water, while the other is a hydrophobic such as a octanol. A high octanol/water partition coefficient indicates that the chemical substance is hydrophobic and preferentially distributed into hydrophobic compartments such as cell membranes, whereas hydrophilic compounds tend to distribute into hydrophilic compartments such as blood serum. Nowadays, LogP values are often predicted computationally. [32][31]

Although defining the precise relationship between chemical structure and function remains a challenge, the structure – property relationship is widely accepted in drug discovery and risk assessment.[33]

Sometimes, QSAR methods take a more specific name such as QSPR (Quantitative – Structure – Property – Relationship) when they are applied to physicochemical properties (e.g., boiling point, solubility, LogP). All these approaches correlate a dependent variable (effect or response) with a set of independent variables (usually calculated properties or molecular descriptor). They are statistical models that can be applied to predict responses for unseen data points without the need to use the actual chemical compound, even before it has been synthesized. [34][35]

I.3.2.Molecular Descriptors

To fully exploit the information contained in molecular structures, these structures are converted into a series of numerical values that represent significant physicochemical and structural characteristics, whether local or global, of the molecule. These values are referred to as molecular descriptors [36][38]

a) Types of Molecular Descriptors

Currently, more than 10,000 molecular descriptor have been developed. These descriptors enable predictions to be performed without the need to synthesize the molecules, which is one of the main objectives of molecular modeling. In the following section, the most commonly used molecular descriptors will be presented, starting with the simplest ones.

Historically, two major classification schemes for molecular descriptors have been established: [37]

- ✓ Classification according to their origin, such as constitutional, topological, geometrical, quantum, thermodynamic descriptors, and others.
- ✓ Classification according to their dimensionality, namely 1D, 2D, 3D, 4D descriptors.

➤ 1D Descriptors

One – dimensional (1D) molecular descriptors include all descriptors that can be calculated from the molecular formula of a compound based on its atomic composition. In other words, they depend on the type and number of atoms constituting the molecule. These descriptors represent general properties such as atomic mass percentages, molar mass, molecular weight, and related parameters. They are often presented in the form of molecular fingerprints,

which are binary vectors where a value of (1) indicates the presence of a particular substructure, and (0) indicates its absence. [39]

➤ 2D Descriptors

Two – dimensional (2D) molecular descriptors are derived from the planar (2D) structure of the molecule. They contain information related to atomic connectivity or specific molecular fragments, as well as estimations of certain physicochemical properties [40], [41]. This category mainly includes:

- Topological indices;
- Constitutional indices

➤ 3D Descriptors

Three – dimensional (3D) molecular descriptors are evaluated based on the relative spatial positions of atoms and describe more complex molecular characteristics. Therefore, their calculation generally requires knowledge of the three – dimensional geometry of the molecule, which is most often obtained through molecular modeling using empirical or ab initio methods. Three descriptors reflect steric, electronic, and energetic properties associated with the three – dimensional structure of the molecule.

Three – dimensional descriptors are relatively computationally expensive but provide more detailed and informative data [42]-[46]. Several important families of 3D descriptors can be distinguished, including:

- Geometrical descriptors;
- Electrostatic descriptors;
- Physicochemical descriptors.

I.3.3.Construction of QSAR Models

• First step:

The first step consists of preparing a dataset containing the available experimental data related to the selected endpoint. This step is highly demanding, as only a limited number of datasets reported in the literature are complete and properly validated. QSAR studies are often based on models developed from a limited amount of high – quality data. However, in most real – world applications, data must be collected from primary sources (laboratory studies) and carefully verified to ensure consistency and true comparability. [47], [48]

- **Second step:**

The second step involves the calculation and selection of molecular descriptors. Nowadays, hundreds or even thousands of 2D descriptors can be easily computed. It is often preferable to calculate a large number of descriptors and then select the most relevant ones using appropriate statistical methods, such as

- Genetic algorithms ;
- Stepwise addition methods (adding one descriptor at a time);
- Stepwise elimination methods (removing one variable at a time);
- Or other optimization techniques for variable selection.
- Regardless of the algorithms chosen to develop predictive models, it is important to consider statistics on model quality and ensure that a correct modeling methodology is used to avoid errors or overfitting. Numerous model statistics are available that can indicate whether a model can be applied to new data points from which responses are to be predicted. [49]

Two types of supervised learning models are of particular interest:

- **Classification methods:** assign the target molecule to two or more classes, most commonly biologically active or inactive
 - **Regression methods:** attempt to use continuous data, such as measured biological response variables, to correlate molecules with these data and predict a continuous numerical value for new and unknown molecules using the modeled generation.
- ✓ Another distinction is made between two types of methods:
- **Similarity – based methods** : tend to group similar molecules together
 - **Feature – based methods** : use a set of molecular features (such as descriptors) to construct the classification function.
- ✓ A plethora of algorithms and methods are available for modeling, including
- **Statistical regression:** such as linear Discriminant Analysis, Partial least Squares regression, and Multiple Linear Regression.
 - **Machine learning methods:** including Naïve Bayesian Classifiers, Decision Trees, Recursive Partitioning, Random Forests, Artificial Neural Networks, and Support Vector Machines.[46]-[52]

Over the past two decades, many artificial intelligence – derived methods have become preferred approaches in QSAR[53], as they are non – parametric and allow modeling functions to be identified without a priori assumptions.

I.3.4. Model Evaluation Parameters

To assess the significance of QSAR models, validation remains a highly sensitive step in statistical studies. Since a model is the result of statistical analysis, its interpretation and application must be carried out strictly within the domain covered by the analyzed data. Various statistical parameters are used to determine the quality of a model, such as:

- The multiple coefficient of determination (R^2), which reflects the degree of correlation between the observed and predicted property values in the test set.

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{eq 2})$$

- Where y_i and $\hat{y}_i$ represent the observed and predicted property value, respectively, for the test compounds.
- And $\bar{y}$ denotes the mean value of the observed responses for the calibration (training) set.

Models with R^2 values greater than the stipulated threshold of 0,5 are considered to have good predictive ability.

The stability of published QSAR models is generally assessed using:

- Leave – One – Out cross – Validation (LOO): this method consist of recalculating the model using $(n - 1)$ calibration compounds. The resulting model is then used to estimate the property value of the omitted compound, denoted as $\hat{y}_i$. The procedure is repeated for each of the n compounds in the calibration set.[54],[55]

This approach is used to determine the predictive coefficient:

$$Q_{LOO}^2 = \frac{SCT - PRESS}{SCT} \quad (\text{eq 3})$$

Unlike the coefficient of determination R^2 , which increases with the number of model parameters, the cross – validation coefficient Q_{LOO}^2 exhibits a curve that reaches a maximum at a certain number of descriptors and then decreases monotonically. This characteristic gives significant importance to the Q_{LOO}^2 coefficient in model evaluation.

A value of $Q_{LOO}^2 > 0,5$ is considered satisfactory, whereas a value greater than 0,9 is regarded as excellent.[55]

When sufficient data that were not used in model construction are available, or after collecting new data, external validation can – and should – be performed. This is referred to as external validation. The external predictive coefficient, denoted Q_{LOO}^2 , is calculated as follows:

$$Q_{ext}^2 = \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(1)})^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{ext}} \quad (\text{eq4})$$

For a high value of Q_{LOO}^2 , a correspondingly high value of Q_{ext}^2 suggests good predictive ability of the model.

Over the past decade, considerable debate has taken place regarding the application of internal or external validation to ensure the predictive performance of models [56]. The primary objective of internal validation is to verify the stability of models; however, this type of validation cannot substitute for true external validation [57].

• Golbraikh and Tropsha Criteria

The criteria proposed by Golbraikh and Tropsha demonstrated that the predictive power of QSAR models can only be claimed if the model has been successfully applied to predict compounds from an external test set (i.e., compounds that were not used in model development). To reflect the quality of predictions, several measures based on the Golbraikh and Tropsha criteria [58],[59] have been proposed in different studies [54],[58],[60], such as:

$Q_{F1}^2, Q_{F2}^2, Q_{F3}^2$ et r_m^2 , These metrics are detailed by the following equations :

$$Q_{F1}^2 = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{\sum (Y_{obs(test)} - \bar{Y}_{training})^2} \quad (\text{eq5})$$

The Q_{F2}^2 coefficient proposed by Schuurmann et al. [61] is given by the following expression:

$$Q_{F2}^2 = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{\sum (Y_{obs(test)} - \bar{Y}_{test})^2} \quad (\text{eq6})$$

$$Q_{F3}^2 = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2 / n_{test}}{\sum (Y_{obs(train)} - \bar{Y}_{train})^2 / n_{test}} \quad (\text{eq7})$$

$$r_m^2 = r^2 \cdot (1 - \sqrt{(r^2 + r_0^2)}) \quad (\text{eq8})$$

According to these authors, models are considered satisfactory if all of the following condition are simultaneously fulfilled:

$$R^2 > 0.6, \quad Q^2 > 0.5.$$

1.3.5. Model Interpretation:

Model interpretation is an important aspect recommended within the validation framework when models are used for predictive purposes in a regulatory context [62],[63].

Researchers attempt to understand models based on known fundamental principles. A small number of descriptors and their role within simple linear equations are generally considered necessary for the acceptance of QSAR result.

However, if the primary objective of QSAR is prediction, emphasis should be placed on model quality rather than interpretability. It is common to distinguish between predictive QSAR, which focuses on predictive accuracy, descriptive QSAR (or SAR), which emphasizes the interpretation of structure – activity relationships.[64],[65]

This view of “QSAR” and “SAR” is somewhat outdated, as highlighted by polishchuk [65], who explained that the concept of QSAR model interpretation has evolved over time.

Livingstone [66] summarized the intelligibility of QSAR models as follows: “The need for interpretability depends on the application, since a validated mathematical model linking

target properties to chemical properties may, in some cases, constitute an accurate estimation of chemical activity”.

SAR and QSAR are predictive statistical models; therefor, they are subject to issues inherent in statistical learning theory. Among the many features that enable model induction, performance and simplicity are the characteristics of the most popular models. [67],[68]

I.3.6.Overall Strategy of a QSAR Study

The development strategy of a QSAR model is more general than that of a QSPR (Figure-6) and is structured around six main points:

- Build a structure – Property database based on reliable and standardized quantitative measurement of the target property for each compound, and select chemical descriptors related to the target property.

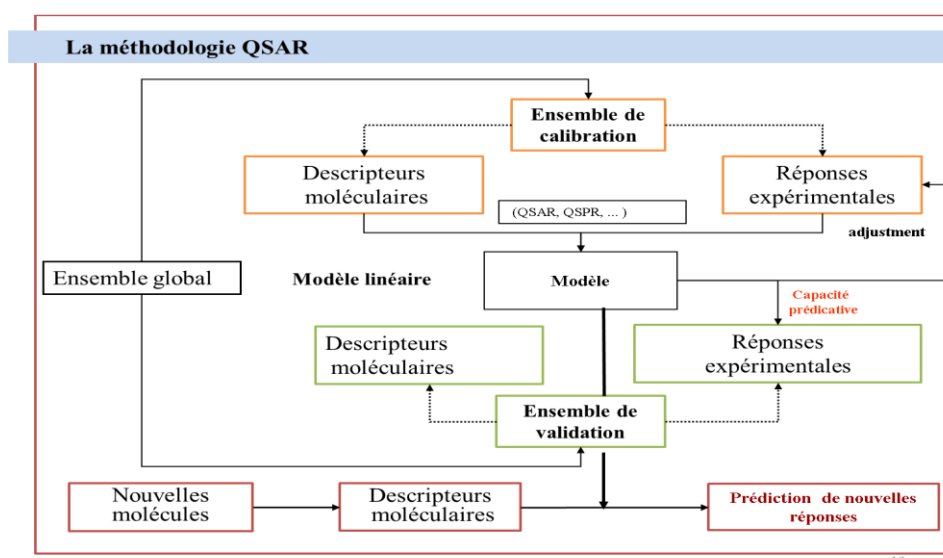


Figure I.7: Overall Strategy of a QSAR Study

Divide the dataset into a training set and a test set.

- Construct models from the training set using the desired statistical tools, characterize these models with internal validation indices, and check their robustness via a randomization test.
- Validate the models with the test set and calculate the external correlation index.
- Repeat the division process to obtain other training and test sets. The optimal division yields the smallest training set capable of producing accurate predictions for the largest test set.

- Explore and utilize the validated models to understand possible mechanisms and make prediction.

I.3.7.Conclusions

In addition to classical methods such as in vivo and in vitro experiments, the use of computational tools is gaining increasing interest within the scientific community. These tools provide simulation, means to accelerate molecular design, deliver predictions, and offer insights into mechanisms to some extent.

QSAR methods are derived from chemometrics and physical simulation, combined with statistics to construct systems capable of prediction.

QSAR can be extended to highly predictive methods, which typically employ complex, nonlinear correlation for explanation while still allowing simple interpretation.

The use of QSAR models is growing, as they provide the only affordable method to screen large amounts of data.[67]

I.4.Advanced Computational Tools

I.4.1General Overview of Density Functional Theory (DFT)

Density Functional Theory (DFT) is one of most widely used approaches in Quantum calculation of the electronic structure of matter, both in condensed matter physics and quantum chemistry at the beginning of the 21st century [69]. DFT is derived from the model established in the late 1920s BY Llewellyn Thomas and Enrico Fermi, upon which current techniques are based [68].

DFT calculation in computational materials science allow for predicting the behavior of materials based on the principles of quantum mechanics, without requiring higher-order quantities such as fundamental material properties. In modern DFT techniques, the electronic structure of a system is analyzed using a potential acting on its electrons [75]. This potential is constructed from the sum of external potentials V_{ext} , which are entirely controlled by the structure and elemental composition of the system, and an effective potential V_{eff} , which represents inter-electronic interactions. A problem for a representative supercell of a material with n electrons, for example, can be analyzed as a series of n single-electron Schrödinger-type equations, often called Kohn-Sham equations [74].

Shouldn't handling a wavefunction for a multi-electron system be simple? Why can't we use a physical observable to determine the energy (and perhaps other properties) of a molecule instead of a wavefunction, which has rather strange units of probability density to the half-power [71].

The Hamiltonian is determined only by the locations and atomic numbers of the nuclei, as well as the total number of electrons. This dependence on the total number of electrons immediately suggests that electronic density would be a useful physical observable.

Since, when integrated over all space, it yields the total number of electrons:

$N = \int \rho(r) dr$. Because nuclei are essentially point charges, their positions are associated with local maxima of the electron density, which are also critical (turning) points. For each nucleus A, the electron density reaches its maximum at position r_A

$$\left. \frac{\partial \bar{\rho}(r_A)}{\partial r_A} \right|_{r_A=0} = -2Z_A \rho(r_A)$$

Z – The atomic number of A.

r_A – The radial distance from A.

ρ – The spherically averaged electron density.

Given a known density, one can construct a Hamiltonian operator, solve the Schrodinger equation, and compute the wave functions and the energy eigenvalues. All the information we need is contained in the electron density [71].

I.4.2.Principle of DFT Calculations

a. Hohenberg – Kohn Existence Theorem

The first theorem can be stated as follows: the electron density determines the external potential [68].

Electrons interact with each other and with an external potential according to density functional theory (DFT) [72]. In a homogeneous electron gas, the external potential is a uniformly distributed positive charge, whereas in a molecule it corresponds to the attraction toward the nuclei. It is sufficient to show that the density determines the Hamiltonian operator in order to establish an energy – density relationship. Moreover, since integrating the density gives

the number of electrons, one must determine the external potential (i.e., the charges and positions of the nuclei). According to Hohenberg – Kohn theorem, the Hamiltonian operator (and thus the energy) can be uniquely derived from the ground-state electron density [72].

Let us assume the two alternative external potentials can both be consistent with the same ground-state density ρ_0 . We denote these two potentials as V_a and V_b and the corresponding Hamiltonian operators as H_a and H_b . Each Hamiltonian will be associated with a ground-state wave function Ψ_0 and its corresponding eigenvalue E_0 . According to the variational theorem in molecular orbital (MO) theory, the expectation value of the Hamiltonian H_a with respect to the wave function Ψ_b must be greater than the ground-state energy of system a:

$$E_{0,a} \langle \Psi_{0,b} | H_a | \Psi_{0,b} \rangle$$

We can rewrite this expression in the following form:

$$E_{0,a} \langle \Psi_{0,b} | H_a - H_b + H_b | \Psi_{0,b} \rangle = \langle \Psi_{0,b} | H_a - H_b | \Psi_{0,b} \rangle + \langle \Psi_{0,b} | H_b | \Psi_{0,b} \rangle = \langle \Psi_{0,b} | V_a - V_b | \Psi_{0,b} \rangle + E_{0,b}$$

Since the potentials V are one-electron operators, the integral can be expressed in terms of the ground-state density:

$$E_{0,a} \langle \int [V_a(r) - V_b(r)] \rho_0(r) dr + E_{0,b}$$

If we interchange “a” and “b”, we can also write:

$$E_{0,b} \langle \int [V_b(r) - V_a(r)] \rho_0(r) dr + E_{0,a}$$

By adding the two inequalities, we obtain:

$$E_{0,a} + E_{0,b} \langle \int [V_a(r) - V_b(r)] \rho_0(r) dr + \int [V_b(r) - V_a(r)] \rho_0(r) dr + E_{0,b} + E_{0,a}$$

$$\langle \int [V_a(r) - V_b(r) + V_b(r) - V_a(r)] \rho_0(r) dr + E_{0,b} + E_{0,a} \rangle = E_{0,a} + E_{0,b}$$

We arrive at an impossible result, which implies that our initial hypothesis was incorrect. Consequently, the ground-state density must determine the external potential and, by extension, the Hamiltonian, and thus the wave function. The Hamiltonian defines not only the ground-state wave function but also all excited-state wave functions – the density encodes a massive amount of information.[70]

b. Hohenberg – Kohn Variational Theorem

The existence theorem does not tell us how to predict the density of the system. In a second theorem, Hohenberg and Kohn demonstrated that, as in molecular orbital (MO) theory, the density obeys the variational principle [73].

Suppose we have a well-behaved trial density N that integrates to the correct number of electrons. The first theorem states that this density determines a trial wave function and a Hamiltonian in this circumstance. The energy expectation value can then be calculated:

$$\langle \Psi_{cand} | H_{cand} | \Psi_{cand} \rangle = E_{cand} \geq E_0$$

Thus, in theory, we can continue trying different densities, and those that yield a lower energy are closer to the true density. However, at this stage, we have no prescription for rationally selecting improved trial densities, and secondly, since the motivation for DFT was to avoid solving the Schrödinger equation, calculating the energy as an expectation value of the Hamiltonian is not an advance – we already know how to do that. [73]

The challenge comes from the nature of the functional. In theory, we know that we can find the Hamiltonian, the wave function, and thus the energy from the density, but we do not know how to do it in practice. What we really want is to be able to calculate the energy directly from the density. [72]

c. Energy as a functional of the wave function and the density

A variable x is associated with a function “ f ” that assigns “ a ” scalar $y=f(x)$ to each scalar “ x ”.

A functional relates any function “ f ” to “ a ” scalar $x=f(F)$. An example “ f ” a functional of “ f ” is definite integral of $f(x)$.

The energy appears as a functional of the wave function Φ and is normalized as follows:

$$E(\Phi) = \langle \Phi | \hat{H} | \Phi \rangle$$

The theorems of Kohn, Sham, and Hohenberg establish the following facts: The given total electron density of a system in the ground state at each position (x,y,z) determines all its observable attributes.

Consequently, the energy, in particular, is a density functional:

$$E = F[\rho]$$

- The ground state has the property that the density $\rho(x,y,z)$ minimizes the energy.

This is the density analogue of the variational theorem for wave functions.

Calculating the energy from the density is, in principle appealing because it reduces the problem from a wave function Φ depending on $3n$ spatial variables to a density ρ depending on only 3 variables. The problem is that we essentially know nothing about the functional F , except that it must exist. [68]

I.4.2.Application of DFT to Molecular Systems

There are many approximations for the exchange – correlation functional. In Gaussian, they are denoted by a set of letters, where the first identifies the exchange calculation method and the second indicates the correlation method. [74]

a. Local Density Approximation(LDA)

The term Local Density Approximation (LDA) was originally used to describe any density functional theory in which the value of the exchange – correlation energy density ε_{xc} at a position r can be derived solely from the value of the density at that point, i.e., the “local” value of ρ [74].

No analytical derivation of the correlation energy density has been established, even for the “simple” uniform electron gas. Ceperley and Alder(1980) used quantum Monte Carlo techniques to calculate the total energy of electron gases at different densities with very high numerical accuracy[75].They were able to determine the correlation energy in these systems by subtracting the analytically calculated exchange energy for each case[74].

Later, Vosko, Wilk, and Nusair (1980) developed local density functionals to fit these results:

$$\varepsilon_c^i(r_s) = \frac{A}{2} \left\{ \ln \frac{r_s}{r_s + b\sqrt{r_s} + c} + \frac{2b}{\sqrt{4c - b^2}} \tan^{-1} \left(\frac{\sqrt{4c - b^2}}{2\sqrt{r_s} + b} \right) - \frac{bx_0}{x_0^2 + bx_0 + c} \left\{ \ln \left[\frac{(\sqrt{r_s} - x_0)^2}{r_s + b\sqrt{r_s} + c} \right] + \frac{2(b + 2x_0)}{\sqrt{4c - b^2}} \tan^{-1} \left(\frac{\sqrt{4c - b^2}}{2\sqrt{r_s} + b} \right) \right\} \right\}$$

In this context, r_s is called the effective radius – exactly one electron would be contained within a sphere of this radius if the density inside that sphere were uniform everywhere.

$$r_s(\rho) = \left(\frac{3}{4\pi\rho(r)} \right)^{1/3}$$

For $I=0$ et $I=1$, different sets of empirical constants A , x_0 , b and c are used. **SVWN** refers to LDA calculation (also called LSDA – Local Spin Density Approximation that employ a combination of Slater of exchange and the VWN correlation energy expression [69].

It is clear that the correlation energy functional in a truly general system can be complex. Because most contemporary DFT algorithms rely on empirically optimized constants and functional forms, they are essentially semi-empirical. An analytical solution of the integrals using the VWN correlation functional is highly unlikely [69].

The procedures of an LSDA calculation are largely similar to those of a Hartree – Fock (HF) calculation. After selecting the molecular geometry, the overlap integrals, as well as the kinetic energy and nuclear attraction integrals, are computed. We must guess an initial density to evaluate the subsequent integrals, and this density can be generated as a matrix equivalent to the density matrix used HF theory. We can then construct the exchange – correlation potential V_{xc} and evaluate the remaining integrals in each element of the Kohn – Sham (KS) matrix using this trial density. Beyond this point, the KS and HF self-consistent field (SCF) procedures are almost identical. Once SCF convergence is achieved, the energy is calculated by inserting the final density into:

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)]$$

This contrasts with Hartree – Fock (HF) theory, which treats the energy as the expectation value of the Hamiltonian operator acting on the HF Slater determinant [68].

b. Generalized Gradient Approximation (GGA) and Hybrid Functionals

How can the exchange – correlation functional be represented? We begin by applying the LSDA approximation to a molecule. Invoking the uniform electron gas as the source of energy expressions does not imply that the electron density of the molecule is constant throughout space. Instead, it is assumed that the exchange – correlation energy density of the molecule at

each point in space is the same as that of a homogeneous electron gas with the same density at that point [74].

The electron density in molecular systems is often non-uniform, so the LDA approach has limitations. One way to improve the correlation potential is to make it depend on the gradient of the density rather than only on the local density value, this leads to “non-local” DFT or “gradient-corrected” DFT, known as the Generalized Gradient Approximation (GGA). Most gradient-corrected functionals are constructed by adding a correction term to the LDA functional:

$$\varepsilon_{x/c}^{GGA}[\rho(r)] = \varepsilon_{x/c}^{LDA}[\rho(r)] + \Delta\varepsilon_{x/c}\left[\frac{|\nabla\rho(r)|}{\rho^{4/3}(r)}\right]$$

The GGA exchange functional proposed by Becke(1988), denoted as **B**, is the most widely used exchange functional. It exhibits accurate asymptotic behavior for the energy density over a wide range and includes a single empirical parameter optimized by fitting to accurately known exchange energies of six noble gas atoms from helium to radon (He – Rn). Alternative GGA exchange functionals have also been developed, although most have seen limited use in the broader literature compared to **B**, with the likely exception of the **PW** exchange functional by Burke, Perdew, and Wang (1998) [70].

P86 (Perdew, 1986), **PW91** (Perdew and Wang, 1992), and **B95** are corrections energy density relative to its LDA value (Becke, 1996). The **P86** functional includes one empirical parameter fitted to the neon atom, while **PW91** and **B95** contain no empirical parameters. The most important correlation functional, **LYP** (Lee, Yang, Parr, 1988), does not merely correct the LDA expression but instead computes the full correlation energy directly. It contains four empirical parameters fitted to the helium atom. In a one-electron system, **LYP** is the only functional that yields exact cancellation of the self – interaction error [69].

c. DFT Methods and SCF – CI Method

The success of DFT techniques has been enormous. Their quality is generally considered comparable to MP2, but at a slightly higher cost than Hartree – Fock (HF). The **B3LYP** approach, in particular, has proven extremely effective across a wide range of organic and inorganic systems. However, unlike HF and post-HF methods, they present a number of drawbacks [71]:

- i. Since we are dealing with an exact Hamiltonian, we know, at least in principle, what must be done in HF-SC-CI to approach the exact energy: increase the basis set size and the number of configurations to improve the wave function. This is not the case with DFT techniques, which are constrained by the approximation of the exchange – correlation functional. Although performance generally improves with increasing basis set size (though not always), but it reaches its limit earlier than HF-CI techniques.
- ii. Due to the approximation of the exchange – correlation functional EXC, and for hybrid methods, the choice of empirical parameters, the “absolute” energy in DFT no longer has the same variational significance as in classical method. Depending on the functional used, the “absolute” energy vary significantly. As a result, the energy of hydrogen could be lower than its known value of -0.5 a.u.
- iii. The fact that empirical parameters are fitted for a given physical quantity does not guarantee that they will perform equally well for other properties. Moreover, the large number of functionals and mixing options available in programs such as Gaussian can be confusing for users. Although **B3LYP** is rarely inadequate, it is often necessary to consult the literature or rely on trial and error to find the best solution for a given problem.
- iv. Due to approximations in the functional and the fact that the trial function is a single determinant, multi-reference (non-monodeterminant) systems are poorly described by DFT, despite the inclusion of correlation effects. The UHF solution with the option (guess=mix)(symmetry breaking) can still be used, for example, for biradical calculations, but with the same risks as in HF.[76]

I.5.Artificial intelligence

I.5.1.Definitions of AI

Several definitions have been proposed for Artificial Intelligence, including (Boisard, 2014):

✓ Definition1:

“If computer science is the science of information processing, AI is concerned with all cases where this processing cannot be reduced to a simple, precise, algorithmic method.”(Laurière, 1987)

✓ Definition 2:

AI is “the application of software and programming techniques that highlight the principles of intelligence in general, and human thinking in particular”(Margaret Boden, 1977)

✓ Definition 3:

“The study of intelligence independently of its manifestation in humans, animals, or machines” (John McCarthy)

✓ Definition 4:

“The pursuit of metaphysics by other means” (Christopher Longuet-Higgins)

✓ Definition 5:

“The science concerned with machines that perform tasks which, if done by humans, would require intelligence” (Marvin Minsky, 1986)

✓ Definition 6:

“Artificial intelligence can be defined as the attempt to create machines like those seen in movies” (Russell Beale) [77]

I.5.2.Introduction to Artificial Intelligence(AI) and Machine Learning (ML) for ADMET Prediction .

The paradigm of drug discovery has been fundamentally altered by the integration of Artificial Intelligence (AI) and machine learning (ML). In the specific domain of ADMET(Absorption, Distribution, Metabolism, Excretion, Toxicity) profiling, these computational approaches have transitioned from being supplementary tools to indispensable components of the drug development pipeline. By leveraging sophisticated algorithms, AI/ML models can discern complex, non-linear relationships within high-dimensional chemical spaces, enabling the accurate prediction of pharmacokinetic and toxicological properties for novel chemical entities [78].

I.5.3.Foundational Machine learning Algorithms in cheminformatics

The selection of an appropriate ML algorithm is dictated by the nature of the ADMET endpoint – whether it is a continuous variable (regression) or a categorical outcome (classification) – and the characteristics of the dataset.

I.5.4.Support Vector Machines (SVM) and Support Vector Regression (SVR)

Support Vector Machines are powerful supervised learning models rooted in statistical learning theory. For classification tasks, an SVM constructs a hyperplane or set of hyperplanes in a high-dimensional space that maximally separates data points belonging to different classes. The optimal hyperplane is the one that maximizes the margin, defined as the distance between the hyperplane and the nearest data points from each class, known as support vectors [79].

When the relationship between molecular descriptors and the ADMET property is nonlinear, SVMs employ the “kernel trick”. This technique implicitly maps the input features into a higher-dimensional feature space where a linear separating hyperplane can be found. Common kernel functions include the Radial Basis Function (RBF) and polynomial kernels. For continuous ADMET endpoints (e.g., predicting the exact value of clearance or half-life), Support Vector Regression (SVR) is utilized. SVR aims to find a function that deviates from the actual target values by no more than a specified margin ϵ , while simultaneously minimizing the model’s complexity (maximizing flatness). This makes SVR particularly robust to outliers and effective in high-dimensional descriptor spaces.[79]

I.5.5.Ensemble Method: Random Forests and Gradient Boosting

Ensemble learning methods combine the predictions of multiple base estimators to improve generalizability and robustness over a single estimator.

Random Forest(RF) is a highly popular ensemble method that constructs a multitude of decision trees during training. For classification, the final output is the mode of the classes predicted by individual trees; for regression, it is the mean prediction. RF introduces randomness in two ways: by training each tree on a bootstrap sample of the data (bagging) and by selecting a random subset of features at each tree on a bootstrap sample of the data (bagging) and by selecting a random subset of features at each node split. This inherent randomness makes RF highly resistant to overfitting and capable of handling datasets with a large number of descriptors, including those derived from quantum chemical calculations.[80] Furthermore, RF provides a natural measure of feature importance, which is crucial for identifying the structural or electronic determinants of an ADMET property.

Gradient Boosting Machines (GBM), including implementations like XGBoost and LightGBM, build an ensemble of weak prediction models (typically decision trees) in a sequential, stage-wise manner. Each new tree is trained to correct the residual errors made by the combination of all previous trees. GBMs often achieve state-of-the-art predictive performance in

tabular cheminformatics datasets by optimizing a differentiable loss function using gradient descent [81].

Deep learning and Representation Learning

Deep Learning (DL), a subset of ML based on artificial neural networks with multiple hidden layers, has revolutionized ADMET prediction by enabling “representation learning.” Unlike traditional ML, which relies on manually engineered molecular descriptors (e.g., LogP, TPSA), DL models can automatically learn optimal hierarchical representations directly from raw molecular data[82].

- ❖ **Graph Neural Networks (GNNs):** Molecules are naturally represented as graphs, where atoms are nodes and bonds are edges. GNNs, such as Graph Convolutional Networks (GCNs) and Message Passing Neural Networks (MPNNs), operate directly on these molecular graphs. They iteratively update the state of each atom by aggregating information from its neighboring atoms (message passing). This allows GNNs to capture complex topological and electronic environments, making them exceptionally powerful for predicting properties that depend on specific substructures or functional groups [83].
- ❖ **Sequence-Based Models:** Molecules can also be represented as D sequences using SMILES (Simplified Molecular-Input Line-Entry System) strings. Recurrent Neural Networks (RNNs) and, more recently, Transformer architectures (e.g., ChemBERTa) have been adapted from natural language processing to “read” SMILES strings. These models learn the syntax and semantics of chemical structures, enabling them to predict ADMET properties or even generate novel molecules with desired profiles[82]

I.5.6.Feature Engineering and Selection

While DL models can learn representation automatically, traditional ML models (RF, SVM) require explicit feature engineering. The quality of these features dictates the model’s success.

- a. **Physicochemical and Topological Descriptors:** These include fundamental properties (MW, LogP, HBD, HBA) and indices describing molecular connectivity and shape (e.g., Wiener indices, topological polar surface area).
- b. **Quantum Chemical Descriptors:** As discussed in Section...., descriptors derived from Density Functional Theory (DFT) – such as HOMO/LUMO energies, dipole moments, and molecular electrostatic potential (MEP) – provide critical

mechanistic insights into a molecule's reactivity, which is essential for predicting metabolic stability and toxicity.[84]

- c. **Feature Selection:** High-dimensional descriptor spaces often contain redundant or irrelevant information, which can lead to overfitting (the “curse of dimensionality”). Techniques such as Recursive Feature Elimination (RFE), LASSO regularization, or utilizing the feature importance scores from Random Forests are employed to select the most predictive subset of descriptors, thereby improving model interpretability and generalization.

I.5.7. Model Validation and the Applicability Domain

The development of an ML model must be accompanied by rigorous validation to ensure its reliability when applied to new, unseen compounds.

- **Cross-Validation and External Testing:** The dataset is typically partitioned into a training set (for model building), a validation set (for hyperparameter tuning), and an independent test set (for final evaluation). Techniques like k-fold cross-validation are used to assess the model's stability across different data subsets.
- **Performance Metrics:** For regression tasks, the coefficient of determination R^2 , Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are standard metrics. For classification, Accuracy, Precision, Recall, F-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are utilized, with the latter being particularly important for imbalanced datasets (e.g., where toxic compounds are rare) [85]
- **Applicability Domain (AD):** A critical concept in cheminformatics is the Applicability Domain. An ML model is only reliable for predicting the properties of molecules that are structurally and physicochemically similar to those in its training set. The AD defines this chemical space. Predictions made for molecules falling outside the AD are considered extrapolations and are inherently unreliable. Various methods, such as bounding box approaches, distance-based methods (e.g., Euclidean or Mahalanobis distance to training instances), or probability density estimations, are used to define and quantify the AD.[86]

In conclusion, the application of AI and ML to ADMET prediction is a multifaceted discipline requiring careful consideration of algorithm selection, feature engineering (particularly the integration of quantum descriptors), and rigorous validation. When executed correctly, these

computational models provide indispensable guidance in the rational design and optimization of novel therapeutics.

References:

- [1] Kaltritzky A. R. & Boulton A.; J. *Advances in heterocyclic chemistry*; **(1963)**, 403. New York: Academic Press.
- [2] Metzger Jacques V.; Ed. *Thiazole and Its Derivatives; Part 3*; Vol. 114. John Wiley & Sons; **(2009)**.
- [3] Gupta V. & Kant V.; A review on biological activity of imidazole and thiazole moieties and their derivatives. ; *Science International*. ; **(2003)**, 253-260.
- [4] Hantzsh and Weber; J. *Ber. Dtsch.; Chem. Ges.* ; **(1887)**, 3118 3 3132.
- [5] Borisenko V. E., Koll, A., Kolmakov, E. E. & Rjasnyi, A.G.; Hydrogen bonds of 2-aminothiazoles in intermolecular complexes; *Journal of molecular structure* ; **(2006)**, 783(1-3), 101-115.
- [6] Sperry Jeffrey B. and Dennis L.; Wright. "Furans, thiophenes and related heterocycles in drug discovery;" *Current Opinion in Drug Discovery and Development*; **(2005)**, 8(6), 723.
- [7] Bramely S. E., Dupplin V., Goberdhan D. G., & Meakins, G. D.; The Hantzsch thiazole synthesis under acidic conditions: change of regioselectivity. ; *Journal of the Chemical Society, Perkin Transaction*; **(1987)**, V.1, 639-643.
- [8] Mishra C.B., Kumari S., & Tiwari M.; (Thiazole: A promising heterocycle for the development of potent CNS active agents. ; *European journal of medicinal chemistry*. ; **(2015)**, 1-34.
- [9] Cook A. H., Heilbron I., & Levy A. L.; 318 Studies in the azole series; Part I. A novel route to 5-aminothiazoles; *Journal of the Chemical Society (Resumed)*; **(1947)**, 1594-1598.
- [10] Sasmal PK, Chandrasekhar A., Sridhar S. & Iqbal. ; J. Nouvelle méthode en une étape pour la conversion des isothiocyanates en 2-alkyl (aryl) Amin thiazoles ; *Tétraèdre* ; **(2008)**, 64 (49), 11074-11080.
- [11] Erian A. W., Sherif S. M., & Gaber H. M.; The chemistry of α -haloketones and their utility in heterocyclic synthesis *Molecules*. ; **(2003)**, 8(11), 793-865.
- [12] Venkateswararao E., Jalani H. B., Manoj M., & Jung S. H.; Exploration and Optimization of an Efficient One pot Sequential Synthesis of Di/trisubstituted Thiazoles from α - Bromoketones,

Thioacids Salt and Ammonium Acetate ; *Journal of Heterocyclic Chemistry* ; (2006), 53(5), 1449-1456.

[13] Shen S. S., Lei M. Y., Wong Y. X., Tong M. L., Teo, P. L. Y., Chiba, S., & Narasaka, K.; Intramolecular nucleophilic substitution at an sp² carbon: synthesis of substituted thiazoles and imidazole-2-thiones ; *Tetrahedron Letters* ; (2009), 50(26), 3161-3163.

[14] Jain S.; Synthesis and antibacterial studies of 2- aryl-3-alkanamido-4H-thiazolidin-4-one derivatives; *Arabian Journal of Chemistry*; (2016) , 9, 290-295.

[15] Kunzler A., Synthesis, antifungal and cytotoxic activities of 2-aryl-3-((piperidin-1-yl) ethyl) thiazolidinones; *European journal of medicinal chemistry*; (2013), 64, 74-80

[16] Vigorita M.G. et al.; Synthesis and antiinflammatory analgesic activity of 3, 3'-(1, 2-Ethanediy)l-bis [2-aryl-4- thiazolidinone] chiral compounds; Part 10. *Bioorganic & Medicinal Chemistry Letters*; (2001), 11.21, 2791-2794.

[17] Bowden Keith and Ravinder S., Chana. "Structure activity relations. Part 6; The alkaline hydrolysis of 3-methyl-5- relations; Part 6. The alkaline hydrolysis of 3-methyl-5-methylideneand 3,5-dimethylthiazolidine-2,4-diones ; The addition of thiols to 3-methyl-5-methylidenethiazolidine-2,4-dione ; " *Journal of the Chemical Society ; Perkin Transactions 2*, (1990), 12, 2163-2166.

[18] Schenck J.R. and Rose AF De; Actithiazic Acid. II. Isolation and Characterization Archives of Biochemistry and Biophysics; , 40(2), 263-269.

[19] Panico R., Powell W.H. and Richer J.C., eds ; IUPAC-Nomenclatures des composés organiques ; Blackwell ; londress, (1993).

[20] S.Bennaama and N.Benelhadj Djelloul, Master's thesis –University of Mostaganem.

[21] S .Tian., et al. Ke application of in silico drug-likeness predictions in pharmaceutical research., *Adv Drug Deliv Kev* 86, 2–10 ,2015.

[22] C. A. Lipinski., F. Lombardo., B. W .Dominy. & P. J .Feeney,. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug . Del iv. Kev.* 46, 3–26 ,2001.

[23] A.K. Ghose, V.N. Viswanadhan, JJ. A .Wendoloski., knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. A qualitative and quantitative characterization of known drug databases. *J Comb Chem* 1: 55-68. 1999.

- [24] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, et al. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med Chem* 45: 2615- 2623. 2002.
- [25] C. Cortes., Vapnik V Support vector machine. ,*Mach Learn* 20: 273-297,1995.
- [26] H. Fatoki.,D. Sanni., O .Adeoyo., B. Faleye. *J Nat Prod Plant Resour.*, 6(6), 21-25. 2016.
- [27] WA .Ritschel. *Handbook of Basic Pharmacokinetics*. 3 ed: Drug Intelligence Publications, Inc.; 1986.
- [28] F.Daissa . Contribution à la modélisation des biomolécules et leurs interaction : Inhibition de l'Acétylcholinestérase et le butyrylcholinestérase (AChE/BChE)/ par une nouvelle classe des dérivées. Master's thesis – University of Biskra 2019.
- [29] N. Nikolova-Jeliazkova and J. Jaworska, “An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN,” *Altern. to Lab. Anim.*, vol. 33, no. 5, pp. 461–470, Oct. 2005, doi: 10.1177/026119290503300510.
- [30] C. HANSCH, P. P. MALONEY, T. FUJITA, and R. M. MUIR, “Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients,” *Nature*, vol. 194, no. 4824, pp. 178–180, Apr. 1962, doi: 10.1038/194178b0.
- [31] A. K. Ghose and G. M. Crippen, “Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions,” *J. Chem. Inf. Comput. Sci.*, vol. 27, no. 1, pp. 21–35, Feb. 1987, doi: 10.1021/ci00053a005.
- [32] H. Kubinyi, “From Narcosis to Hyperspace: The History of QSAR,” *Quant. Struct. Relationships*, vol. 21, no. 4, pp. 348–356, Oct. 2002, doi: 10.1002/1521-3838(200210)21:4<348::AID-QSAR348>3.0.CO;2-D.
- [33] C. Selassie and R. P. Verma, “History of Quantitative Structure–Activity Relationships,” in *Burger's Medicinal Chemistry and Drug Discovery*, Wiley, 2010, pp. 1–96. doi: 10.1002/0471266949.bmc001.pub2.
- [34] S. Yousefinejad and B. Hemmateenejad, “Chemometrics tools in QSAR/QSPR studies: A historical perspective,” *Chemom. Intell. Lab. Syst.*, vol. 149, pp. 177–204, Dec. 2015, doi: 10.1016/j.chemolab.2015.06.016.
- [35] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, “Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical

Properties,” *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 1, pp. 1–18, Jan. 2000, doi: 10.1021/ci9903206.

[36] G. Aranoff, “Thought Experiment: Marginal Cost versus John M. Clark’s Workable Competition Pricing,” *Mod. Econ.*, vol. 11, no. 11, pp. 1742–1751, 2020, doi: 10.4236/me.2020.1111119.

[37] A. Dudek, T. Arodz, and J. Galvez, “Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review,” *Comb. Chem. High Throughput Screen.*, vol. 9, no. 3, pp. 213–228, Mar. 2006, doi: 10.2174/138620706776055539.

[38] C. Nieto-Draghi *et al.*, “A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes,” *Chem. Rev.*, vol. 115, no. 24, pp. 13093–13164, Dec. 2015, doi: 10.1021/acs.chemrev.5b00215.

[39] A. Regina Todeschini and S. Hakomori, “Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1780, no. 3, pp. 421–433, Mar. 2008, doi: 10.1016/j.bbagen.2007.10.008.

[40] H. P. Schultz, “Topological organic chemistry. 1. Graph theory and topological indices of alkanes,” *J. Chem. Inf. Model.*, vol. 29, no. 3, pp. 227–228, Aug. 1989, doi: 10.1021/ci00063a012.

[41] H. Wiener, “Structural Determination of Paraffin Boiling Points,” *J. Am. Chem. Soc.*, vol. 69, no. 1, pp. 17–20, Jan. 1947, doi: 10.1021/ja01193a005.

[42] P. Labute, “A widely applicable set of descriptors,” *J. Mol. Graph. Model.*, vol. 18, no. 4–5, pp. 464–477, 2000, doi: 10.1016/S1093-3263(00)00068-1.

[43] M. Ghamali, S. Chtita, A. Aouidate, A. Ghaleb, M. Bouachrine, and T. Lakhlifi, “Combining DFT and QSAR computation to predict the interaction of flavonoids with the GABA (A) receptor using electronic and topological descriptors,” *J. Taibah Univ. Sci.*, vol. 11, no. 3, pp. 422–433, May 2017, doi: 10.1016/j.jtusci.2016.06.005.

[44] “QSRR Study of Linear Retention Indices for Volatile Compounds using Statistical Methods,” *Chem. Sci. Trans.*, vol. 7, no. 4, Oct. 2018, doi: 10.7598/cst2018.1501.

- [45] R. Bosque, J. Sales, E. Bosch, M. Rosés, M. C. García-Alvarez-Coque, and J. R. Torres-Lapasió, “A QSPR Study of the p Solute Polarity Parameter to Estimate Retention in HPLC,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 4, pp. 1240–1247, Jul. 2003, doi: 10.1021/ci0340660.
- [46] V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, “Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain,” *J. Chem. Inf. Model.*, vol. 29, no. 3, pp. 163–172, Aug. 1989, doi: 10.1021/ci00063a006.
- [47] N. Brown, “Chemoinformatics—an introduction for computer scientists,” *ACM Comput. Surv.*, vol. 41, no. 2, pp. 1–38, Feb. 2009, doi: 10.1145/1459352.1459353.
- [48] D. Gadaleta, A. Lombardo, C. Toma, and E. Benfenati, “A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications,” *J. Cheminform.*, vol. 10, no. 1, p. 60, Dec. 2018, doi: 10.1186/s13321-018-0315-6.
- [49] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2001. doi: 10.1007/978-0-387-21606-5.
- [50] M. T. C. Tomasz Puzyn, Jerzy Leszczynski, *Recent Advances in QSAR Studies: Methods and Applications Volume 8 de Challenges and Advances in Computational Chemistry and Physics*, Illustrée.
- [51] G. Khalaf, “Improving the Ordinary Least Squares Estimator by Ridge Regression,” *OALib*, vol. 09, no. 05, pp. 1–8, 2022, doi: 10.4236/oalib.1108738.
- [52] L. Xu and K. Yu, “A Note on Dynamic Roy’s Identity,” *Theor. Econ. Lett.*, vol. 04, no. 07, pp. 513–516, 2014, doi: 10.4236/tel.2014.47064.
- [53] G. C. Gini, “Predictive Toxicology of Chemicals : Experiences and Impact of AI Tools Papers Predictive Toxicology of Chemicals : Experiences and Impact of AI Tools,” no. March 1999, 2019.
- [54] A. Tropsha, P. Gramatica, and V. Gombar, “The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models,” *QSAR Comb. Sci.*, vol. 22, no. 1, pp. 69–77, Apr. 2003, doi: 10.1002/qsar.200390007.
- [55] M. Shen, C. Béguin, A. Golbraikh, J. P. Stables, H. Kohn, and A. Tropsha, “Application of Predictive QSAR Models to Database Mining: Identification and Experimental Validation of

Novel Anticonvulsant Compounds,” *J. Med. Chem.*, vol. 47, no. 9, pp. 2356–2364, Apr. 2004, doi: 10.1021/jm030584q.

[56] P. Gramatica, “Principles of QSAR models validation: internal and external,” *QSAR Comb. Sci.*, vol. 26, no. 5, pp. 694–701, May 2007, doi: 10.1002/qsar.200610151.

[57] A. Tropsha and A. Golbraikh, “Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening,” *Curr. Pharm. Des.*, vol. 13, no. 34, pp. 3494–3504, Dec. 2007, doi: 10.2174/138161207782794257.

[58] A. Golbraikh and A. Tropsha, “Beware of q^2 !,” *J. Mol. Graph. Model.*, vol. 20, no. 4, pp. 269–276, Jan. 2002, doi: 10.1016/S1093-3263(01)00123-1.

[59] A. Tropsha, “Best Practices for QSAR Model Development, Validation, and Exploitation,” *Mol. Inform.*, vol. 29, no. 6–7, pp. 476–488, Jul. 2010, doi: 10.1002/minf.201000061.

[60] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, and R. Kühne, “External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean,” *J. Chem. Inf. Model.*, vol. 48, no. 11, pp. 2140–2145, Nov. 2008, doi: 10.1021/ci800253u.

[61] K. Roy, S. Kar, and R. N. Das, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier Science, 2015.

[62] D. T. Stanton, “On the Physical Interpretation of QSAR Models,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 5, pp. 1423–1433, Sep. 2003, doi: 10.1021/ci0340658.

[63] T. I. Netzeva *et al.*, “Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships,” *Altern. to Lab. Anim.*, vol. 33, no. 2, pp. 155–173, Apr. 2005, doi: 10.1177/026119290503300209.

[64] N. S. Zefirov and V. A. Palyulin, “QSAR for Boiling Points of ‘Small’ Sulfides. Are the ‘High-Quality Structure-Property-Activity Regressions’ the Real High Quality QSAR Models?,” *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 4, pp. 1022–1027, Jul. 2001, doi: 10.1021/ci0001637.

[65] P. Polishchuk, “Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future,” *J. Chem. Inf. Model.*, vol. 57, no. 11, pp. 2618–2639, Nov. 2017, doi: 10.1021/acs.jcim.7b00274.

- [66] D. J. Livingstone, "The Characterization of Chemical Structures Using Molecular Properties. A Survey," J. Chem. Inf. Comput. Sci., vol. 40, no. 2, pp. 195–209, Mar. 2000, doi: 10.1021/ci990162i.
- [67] R.Mezaache. S.KHenifar. Modélisation de l'activité inhibitrice MRP1 par l'approche QSAR. Master's thesis – University of constantine 2022.
- [68] G. Parr, Robert et W. Yang, Density Functional Theory of Atoms and Molecules, Oxford University Press, 1989, 350 p. (ISBN 0-19-509276-7) (en) Norman H. March, Electron Density.
- [69] Elsevier Science & Technology. Theory of Atoms and Molecules, Book, 1997, 339 p. (ISBN 0-12-470525-1).
- [70] R. O. Jones, « Density functional theory: Its origins, rise to prominence, and future », Reviews of Modern Physics, 25 août 2015 (DOI 10.1103/RevModPhys.87.897)
- [71] F.L Hirshfeld. Bonded-atom fragments for describing molecular charge densities. Theoretica Chimica Acta 44, (1977), 129.
- [72] Y-P, Tong. Zheng. S-L., Journal of Molecular Structure 841(2007) 34-40
- [73] Breakdowns of the total charge into overlap and "net" atomic populations have been in use for some time, e.g., R. S. Mulliken, Phys. Rev. 41, 66 (1932); Google Scholar R. S. Mulliken, J. Chem. Phys. 3, 573 (1935), Eqs. (38), (39), (42). Google Scholar Scitation.
- [74] Y. Benelhadj Djelloul. Structure moléculaire de substances organiques et relation avec leur activité biologique. Thèse de doctorat. Université de Mostaganem. 2021.
- [75] N.E.H Belkafouf. Analyse structurale et spectroscopique de nouveaux matériaux fonctionnels. Thèse de doctorat. Université de Mostaganem. 2019.
- [76] M.E.A Benainouna. Analyse physicochimique et phytopharmaceutique d'un nouvel insecticide organique. Master's thesis – University of Mostaganem. 2022

Chapter II :

Materials and

Methods

II.1 Data set Constitution

This section details the meticulous process undertaken to constitute the molecular dataset, emphasizing the selection criteria for the 40 thiazolidinone derivatives and the primary data sources utilized

II.1.1 Selection Criteria for the 40 Thiazolidinone Molecule (Electronic and Structural Diversity)

The robustness and generalizability of any Quantitative Structure-Activity Relationship (QSAR) or Machine Learning (ML) model are critically dependent on the quality and diversity of its training dataset. For this study, a set of thiazolidinone derivatives was meticulously curated to ensure a broad representation for both electronic and structural features, which are paramount for accurate ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) prediction [1]. The selection process aimed to maximize chemical space coverage while maintaining the core thiazolidinone scaffold.

❖ Electronic Diversity

Electronic properties of molecules, particularly those related to frontier molecular orbitals (HOMO and LUMO), are fundamental determinants of chemical reactivity and biological interaction. To capture this diversity, substituents were chosen on their electron-donating and electron-withdrawing capabilities, often quantified by Hammett constants δ [2]. The molecules were broadly categorized into four groups to ensure a balanced representation of these electronic effects:

- ✓ **Electron-Donating (EDG):** Molecules featuring strong electron-donating groups (e.g., -NH, -OCH, -N(CH), alkyl groups) were included. These substituents typically increase the energy of the Highest Occupied Molecular Orbital (HOMO), enhancing the molecule's nucleophilicity and propensity for oxidative metabolism [3].
- ✓ **Electron-Withdrawing Groups (EWG):** Compounds bearing potent electron-withdrawing groups (e.g., -NO, -CN, -CF, -CHO) were selected. These groups tend to lower the energy of the lowest Unoccupied Molecular Orbital (LUMO), increasing the molecule's electrophilicity and potential for reductive metabolism or toxic interactions [4].
- ✓ **Halogen Substituents:** Halogens (F, Cl, Br, I) were incorporated due to their unique electronic (inductive electron-withdrawal) and steric properties, as well as their significant impact on lipophilicity (LogP) and metabolic stability. [5]

- ✓ **Heterocyclic and Polycyclic Substituents:** To further enrich the electronic and steric landscape, a fourth group was introduced, comprising thiazolidinones substituted with various heterocyclic (e.g., pyridine, furan, thiophene) and polycyclic (e.g., naphthalene, quinolone, benzofuran) moieties. These complex substituents introduce diverse electronic environments and steric hindrance, which can profoundly influence ADMET properties provide a more challenging, yet realistic, test for the predictive models[6].

❖ Structural Diversity

Beyond electronic effects, structural variations were considered to ensure a comprehensive sampling of the chemical space relevant to thiazolidinones. This included variations in:

- ✓ **Alkyl Chain Length and Branching:** To modulate lipophilicity and steric bulk, linear and branched alkyl groups of varying sizes were included.
- ✓ **Aromatic and Heteroaromatic Rings:** Different aromatic and heteroaromatic systems were attached to the thiazolidinone core to explore their impact on molecular shape, rigidity, and potential for pi-stacking interactions.
- ✓ **Substitution Patterns:** Positional isomers (e.g., ortho, meta, para substitution on phenyl rings) were considered where appropriate to investigate the subtle effects of substituent placement on molecular properties.

By carefully selecting molecules that span a wide range of electronic and structural features, the dataset aims to provide a robust foundation for developing predictive models capable of accurately forecasting the ADMET profiles of novel thiazolidinone derivatives.

II.1.2.Data Sources (PubChem) and Storage Format (SMILES)

The 40 thiazolidinone derivatives selected for this study were primarily sourced from PubChem, a comprehensive public database of chemical molecules and their biological activities maintained by the National Center for Biotechnology Information (NCBI)[1]. PubChem was chosen for its extensive coverage of chemical compounds, including a vast array of experimentally characterized small molecules, and its provision of standardized chemical information.

❖ PubChem as the Primary Data Source

PubChem serves as an invaluable resource for cheminformatics studies due to several key advantages:

- ✓ **Vast Chemical Space:** It contains millions of chemical structures, including many derivatives of common scaffolds like thiazolidinones, often linked to biological assay results and literature references.
- ✓ **Standardized Identifiers:** Each unique chemical compound in PubChem is assigned a PubChem Compound Identifier (CID), which facilitates unambiguous identification and retrieval of molecular information.
- ✓ **Accessibility:** PubChem offers a user-friendly web interface and programmatic access via its PUG (PubChem Upload Gateway) and PUG-REST (Representational State Transfer) services, enabling efficient data extraction.[2]

The selection process involved searching PubChem for thiazolidinone derivatives that matched the predefined criteria for electronic and structural diversity. For each selected molecule, its unique PubChem CID was recorded, along with its canonical SMILES string.

❖ **SMILES: The Standard for Molecular Representation**

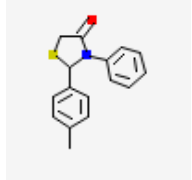
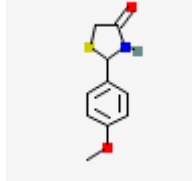
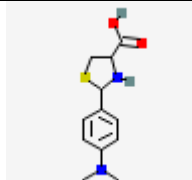
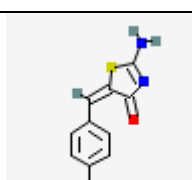
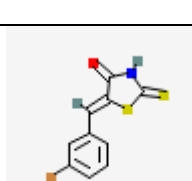
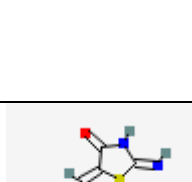
All molecular structures were represented and stored using the SMILES (Simplified Molecular-Input Line-Entry System) notation. SMILES is a linear textual notation for describing the structure of chemical molecules in a way that allows for unambiguous computer readability and interpretation.[3]

Key advantages of using SMILES include:

- ✓ **Compactness:** SMILES strings are concise and can represent complex molecular structures in a compact form, making them efficient for storage and transmission.
- ✓ **Unambiguity:** Canonical SMILES ensures that each unique chemical structure has only one unique SMILES string, which is critical for preventing redundancy and ensuring consistency in cheminformatics databases.
- ✓ **Software Compatibility:** SMILES is widely recognized and supported by virtually all cheminformatics software packages (e.g., RDKit, OpenBabel, ChemAxon), facilitating the generation of D/D structures, calculation of molecular descriptors, and integration into various computational workflows [4].

For each of the selected 40 thiazolidinone derivatives, the corresponding canonical SMILES string was extracted from PubChem and compiled into a structured dataset. This dataset, along with the PubChem CID and direct links to the PubChem entry, ensures the traceability and reproducibility of the molecular set used in this study.

GROUP 01

	IUPAC Name	Molecular Formula	SMILES	2D Structure
M01	2-(4-methylphenyl)-3-phenyl-1,3-thiazolidin-4-one	C ₁₆ H ₁₅ NO S	CC1=CC=C(C=C1)C2N(C(=O)CS2)C3=CC=CC=C3	
M02	2-(4-methoxyphenyl)-1,3-thiazolidin-4-one	C ₁₀ H ₁₁ NO ₂ S	COC1=CC=C(C=C1)C2NC(=O)CS2	
M03	2-[4-(dimethylamino)phenyl]-1,3-thiazolidine-4-carboxylic acid	C ₁₂ H ₁₆ N ₂ O 2S	CN(C)C1=CC=C(C=C1)C2NC(CS2)C(=O)O	
M04	(5E)-2-amino-5-[(4-hydroxyphenyl)methylidene]-1,3-thiazol-4-one	C ₁₀ H ₈ N ₂ O 2S	C1=CC(=CC=C1/C=C/2C(=O)N=C(S2)N)O	
M05	(5Z)-5-[(3-bromophenyl)methylidene]-2-sulfanylidene-1,3-thiazolidin-4-one	C ₁₀ H ₆ BrN OS ₂	C1=CC(=CC(=C1)Br)/C=C\2/C(=O)NC(=S)S2	
M06	(5Z)-5-[(4-aminophenyl)methylidene]-2-imino-1,3-thiazolidin-4-one	C ₁₀ H ₉ N ₃ O S	C1=CC(=CC=C1/C=C\2/C(=O)NC(=N)S2)N	

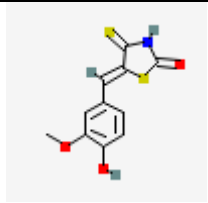
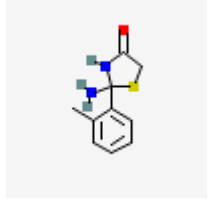
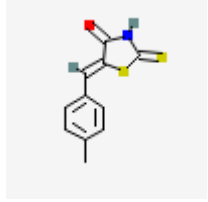
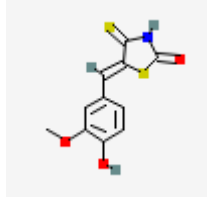
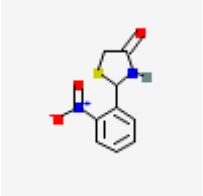
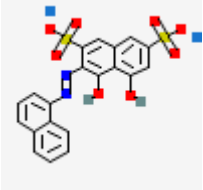
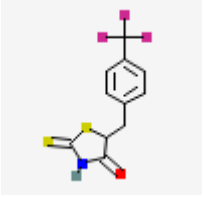
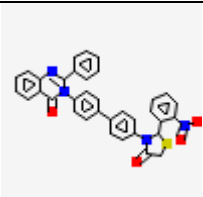
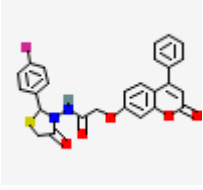
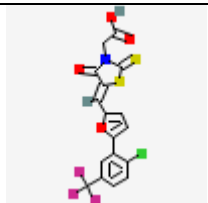
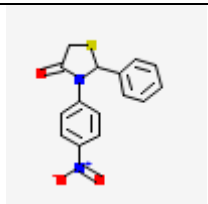
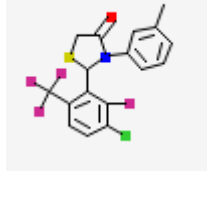
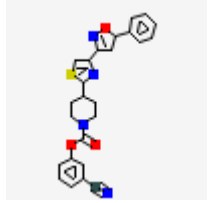
M07	(5Z)-5-[(4-hydroxy-3-methoxyphenyl)methylidene]-4-sulfanylidene-1,3-thiazolidin-2-one	$C_{11}H_9NO_3$ S ₂	<chem>COC1=C(C=CC(=C1)/C=C\2/C(=S)NC(=O)S2)O</chem>	
M08	2-amino-2-(2-methylphenyl)-1,3-thiazolidin-4-one	$C_{10}H_{12}N_2O$ S	<chem>CC1=CC=CC=C1C2(NC(=O)CS2)N</chem>	
M09	(5Z)-5-[(4-methylphenyl)methylidene]-2-sulfanylidene-1,3-thiazolidin-4-one	$C_{11}H_9NOS$ 2	<chem>CC1=CC=C(C=C1)/C=C\2/C(=O)NC(=S)S2</chem>	
M10	(5Z)-5-[(4-hydroxy-3-methoxyphenyl)methylidene]-4-sulfanylidene-1,3-thiazolidin-2-one	$C_{11}H_9NO_3$ S ₂	<chem>COC1=C(C=CC(=C1)/C=C\2/C(=S)NC(=O)S2)O</chem>	

Table 01 : SMILES and 2D structure for the group01

GROUP 02

	IUPAC Name	Molecular Formula	SMILES	2D Structure
M11	2-(2-nitrophenyl)-1,3-thiazolidin-4-one	$C_9H_8N_2O_3S$	<chem>C1C(=O)NC(S1)C2=CC=CC=C2[N+](=O)[O-]</chem>	
M12	disodium;4,5-dihydroxy-3-(naphthalen-1-yl diazenyl)naphthalene-2,7-disulfonate	$C_{20}H_{12}N_2Na_2O_8S_2$	<chem>C1=CC=C2C(=C1)C=CC=C2N=NC3=C(C4=C(C=C(C=C4C=C3S(=O)(=O)[O-])O)O.[Na+].[Na+]</chem>	
M13	2-sulfanylidene-5-[[4-(trifluoromethyl)phenyl]methyl]-1,3-thiazolidin-4-one	$C_{11}H_8F_3NOS_2$	<chem>C1=CC(=CC=C1CC2C(=O)NC(=S)S2)C(F)(F)F</chem>	
M14	2-(2-nitrophenyl)-3-[4-[4-(4-oxo-2-phenylquinazolin-3-yl)phenyl]phenyl]-1,3-thiazolidin-4-one	$C_{35}H_{24}N_4O_4S$	<chem>C1C(=O)N(C(S1)C2=CC=CC=C2[N+](=O)[O-])C3=CC=C(C=C3)C4=CC=C(C=C4)N5C(=NC6=CC=CC=C6C5=O)C7=CC=CC=C7</chem>	
M15	N-[2-(4-fluorophenyl)-4-oxo-1,3-thiazolidin-3-yl]-2-(2-	$C_{26}H_{19}FN_2O_5S$	<chem>C1C(=O)N(C(S1)C2=CC=C(C=C2)F)NC(=O)COC3=CC4=C(C=C3)C(=CC(=O)O4)C5=CC=CC=C5</chem>	

	oxo-4-phenylchromen-7-yl)oxyacetamide			
M16	2-[(5Z)-5-[[5-[2-chloro-5-(trifluoromethyl)phenyl]furan-2-yl]methylidene]-4-oxo-2-sulfanylidene-1,3-thiazolidin-3-yl]acetic acid	$C_{17}H_9ClF_3NO_4S_2$	<chem>C1=CC(=C(C=C1C(F)(F)F)C2=CC=C(O2)/C=C\3/C(=O)N(C(=S)S3)CC(=O)O)Cl</chem>	
M17	3-(4-nitrophenyl)-2-phenyl-1,3-thiazolidin-4-one	$C_{15}H_{12}N_2O_3S$	<chem>C1C(=O)N(C(S1)C2=CC=CC=C2)C3=CC=C(C=C3)[N+](=O)[O-]</chem>	
M18	2-[3-chloro-2-fluoro-6-(trifluoromethyl)phenyl]-3-(3-methylphenyl)-1,3-thiazolidin-4-one	$C_{17}H_{12}ClF_4NOS$	<chem>CC1=CC(=CC=C1)N2C(SCC2=O)C3=C(C=CC(=C3F)Cl)C(F)(F)F</chem>	
M19	(3-cyanophenyl)-4-[4-(5-phenyl-4,5-dihydro-1,2-oxazol-3-yl)-1,3-thiazol-2-yl]piperidine-1-	$C_{25}H_{22}N_4O_3S$	<chem>C1CN(CCC1C2=NC(=CS2)C3=NOC(C3)C4=CC=CC=C4)C(=O)OC5=CC=CC(=C5)C#N</chem>	

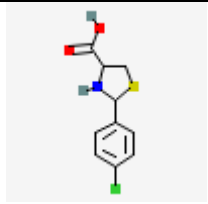
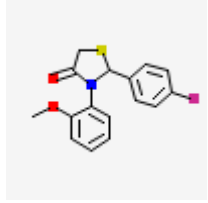
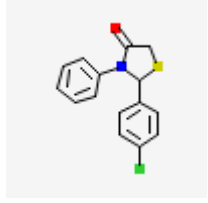
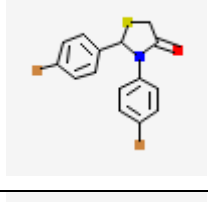
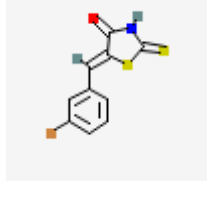
	carboxylate			
M20	2-(4-chlorophenyl)-1,3-thiazolidine-4-carboxylic acid	$C_{10}H_{10}ClNO_2S$	<chem>C1C(NC(S1)C2=CC=C(C=C2)Cl)C(=O)O</chem>	

Table 02 : SMILES and 2D structure for the group02

GROUP 03

	IUPAC Name	Molecular Formula	SMILES	2D Structure
M21	2-(4-fluorophenyl)-3-(2-methoxyphenyl)-1,3-thiazolidin-4-one	$C_{16}H_{14}FNO_2S$	<chem>COC1=CC=CC=C1N2C(SCC2=O)C3=CC=C(C=C3)F</chem>	
M22	2-(4-chlorophenyl)-3-phenyl-1,3-thiazolidin-4-one	$C_{15}H_{12}ClNO S$	<chem>C1C(=O)N(C(S1)C2=CC=C(C=C2)Cl)C3=CC=CC=C3</chem>	
M23	2,3-bis(4-bromophenyl)-1,3-thiazolidin-4-one	$C_{15}H_{11}Br_2NOS$	<chem>C1C(=O)N(C(S1)C2=CC=C(C=C2)Br)C3=CC=C(C=C3)Br</chem>	
M24	(5Z)-5-[(3-bromophenyl)methylene]-2-sulfanylidene-1,3-thiazolidin-	$C_{10}H_6BrNO S_2$	<chem>C1=CC(=CC(=C1)Br)/C=C2/C(=O)NC(=S)S2</chem>	

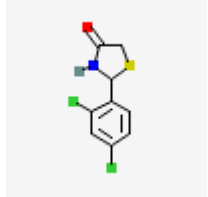
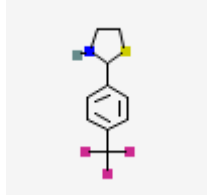
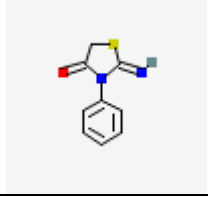
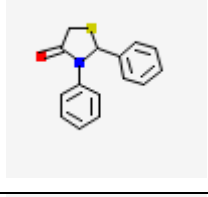
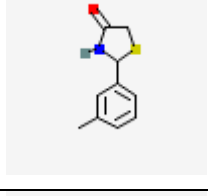
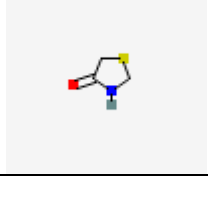
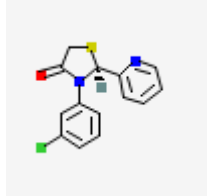
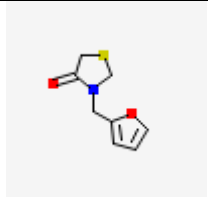
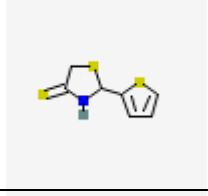
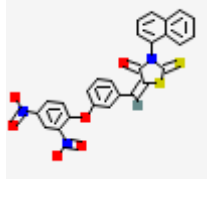
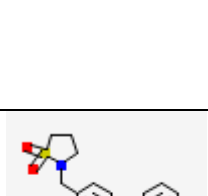
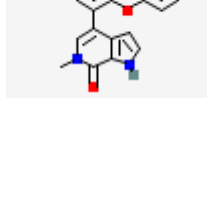
	4-one			
M25	2-(2,4-dichlorophenyl)-1,3-thiazolidin-4-one	$C_9H_7Cl_2NO$ S	<chem>C1C(=O)NC(S1)C2=C(C=C(C=C2)Cl)Cl</chem>	
M26	2-[4-(trifluoromethyl)phenyl]-1,3-thiazolidine	$C_{10}H_{10}F_3NS$	<chem>C1CSC(N1)C2=CC=C(C=C2)C(F)(F)F</chem>	
M27	2-imino-3-phenyl-1,3-thiazolidin-4-one	$C_9H_8N_2OS$	<chem>C1C(=O)N(C(=N)S1)C2=CC=CC=C2</chem>	
M28	2,3-diphenyl-1,3-thiazolidin-4-one	$C_{15}H_{13}NOS$	<chem>C1C(=O)N(C(S1)C2=CC=CC=C2)C3=CC=CC=C3</chem>	
M29	2-(3-methylphenyl)-1,3-thiazolidin-4-one	$C_{10}H_{11}NOS$	<chem>CC1=CC(=CC=C1)C2NC(=O)CS2</chem>	
M30	1,3-thiazolidin-4-one	C_3H_5NOS	<chem>C1C(=O)NCS1</chem>	

Table 03: SMILES and 2D structure for the group03

GROUP 04

	IUPAC Name	Molecular Formula	SMILES	2D Structure
M31	(2S)-3-(3-chlorophenyl)-2-pyridin-2-yl-1,3-thiazolidin-4-one	C ₁₄ H ₁₁ ClN ₂ OS	<chem>C1C(=O)N([C@@H](S1)C2=CC=CC=N2)C3=CC(=CC=C3)Cl</chem>	
M32	3-(furan-2-ylmethyl)-1,3-thiazolidin-4-one	C ₈ H ₉ NO ₂ S	<chem>C1C(=O)N(CS1)CC2=CC=C</chem> O2	
M33	2-thiophen-2-yl-1,3-thiazolidine-4-thione	C ₇ H ₇ NS ₃	<chem>C1C(=S)NC(S1)C2=CC=CS2</chem>	
M34	(5E)-5-[[3-(2,4-dinitrophenoxy)phenyl]methylidene]-3-naphthalen-1-yl-2-sulfanylidene-1,3-thiazolidin-4-one	C ₂₆ H ₁₅ N ₃ O ₆ S ₂	<chem>C1=CC=C2C(=C1)C=CC=C2N3C(=O)/C(=C\C4=CC(=CC=C4)OC5=C(C=C(C=C5)[N+](=O)[O-])[N+](=O)[O-])/SC3=S</chem>	
M35	4-[5-[(1,1-dioxo-1,2-thiazolidin-2-yl)methyl]-2-phenoxyphenyl]-6-methyl-1H-pyrrolo[2,3-c]pyridin-7-one	C ₂₄ H ₂₃ N ₃ O ₄ S	<chem>CN1C=C(C2=C(C1=O)NC=C2)C3=C(C=CC(=C3)CN4CCCS4(=O)=O)OC5=CC=CC=C5</chem>	
M36	(5E)-5-[(5-phenylfuran-2-yl)methylidene]-2-sulfanylidene-1,3-thiazolidin-4-one	C ₁₄ H ₉ NO ₂ S ₂	<chem>C1=CC=C(C=C1)C2=CC=C(O2)/C=C/3C(=O)NC(=S)S3</chem>	

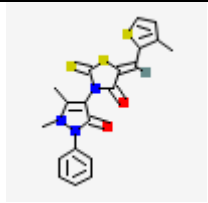
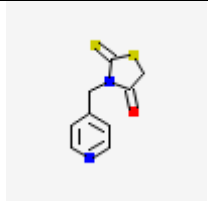
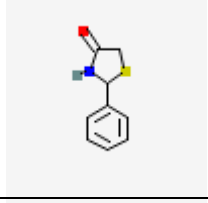
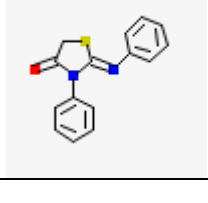
M37	(5Z)-3-(1,5-dimethyl-3-oxo-2-phenylpyrazol-4-yl)-5-[(3-methylthiophen-2-yl)methylidene]-2-sulfanylidene-1,3-thiazolidin-4-one	$C_{20}H_{17}N_3O_2S_3$	<chem>CC1=C(SC=C1)/C=C\2/C(=O)N(C(=S)S2)C3=C(N(N(C3=O)C4=CC=CC=C4)C)C</chem>	
M38	3-(pyridin-4-ylmethyl)-2-sulfanylidene-1,3-thiazolidin-4-one	$C_9H_8N_2OS_2$	<chem>C1C(=O)N(C(=S)S1)CC2=C C=NC=C2</chem>	
M39	2-phenyl-1,3-thiazolidin-4-one	C_9H_9NOS	<chem>C1C(=O)NC(S1)C2=CC=CC=C2</chem>	
M40	3-phenyl-2-phenylimino-1,3-thiazolidin-4-one	$C_{15}H_{12}N_2OS$	<chem>C1C(=O)N(C(=NC2=CC=CC=C2)S1)C3=CC=CC=C3</chem>	

Table 04 : SMILES and 2D structure for the group04

II.2. Software and Computing Environment

II.2.1. Quantum Chemistry software

- **The Gaussian program** is based on the fundamental laws of quantum mechanics. It enables the prediction of energies, molecular structures, and vibrational frequencies of complex molecular systems, as well as their chemical properties. Molecules and reactions can be studied under a wide range of conditions, not only for stable species or complex compounds, but also for species that are impossible to observe experimentally, such as short-lived intermediates or transition states. In the scientific word, the Gaussian program has enabled quantum chemistry to enter laboratories in chemistry, biochemistry, biotechnology, physics, and pharmacology[7]. Historically, this software is based on the work of John Pople, who focused on developing simplifications of the Schrödinger

equation to make it more easily applicable to the study of both simple and complex molecules.

Thanks to these simplifications, John Pople developed in 1970 a computer program called Gaussian, which enables the calculation of molecular conformations, including bond lengths and bond angles between atoms. This program was subsequently improved over time.

In 1992, the theory of Walter Kohn was incorporated. Since then, various quantum chemistry approaches have been added to the program, notably those based on electron density functionals (DFT) developed by Walter Kohn[8].

- **The Gaussian09** program revolutionized working methods by allowing researchers to predict and visualize the evolution of chemical reactions through calculations and virtual representations, without relying on experimental procedures[9]

In general, molecular modeling software includes the following modules:

- ✓ Construction, visualization, and manipulation of molecules.
- ✓ Calculation.
- ✓ Structure saving and file management.
- ✓ Study of molecular properties.

a) Molecular Visualization

Three-dimensional visualization of molecules on a computer screen, or molecular graphics, enables the representation of molecular models in space. To achieve this, each atom in the molecule is defined by its spatial coordinates.

For nearly two centuries, chemists, following crystallographers (with René Just Haüy as a pioneer), have devised various ways to represent molecules and crystals for teaching and research purposes. To highlight the importance chemists have always placed on modeling, it is worth recalling that James Watson and Francis Crick partly deduced the double-helix structure of DNA using physical wire models.

Advances in computing over the past two decades have enabled chemists to visualize molecules using personal computers through software, some of which is freely available or open source. These programs go beyond simple static representations, allowing molecules to be manipulated in three-dimensional, and to display key molecular parameters such as interatomic distances, bond angles, van der Waals radii, and others.

Numerous databases distributed worldwide provide files describing the characteristics of both organic and inorganic molecules. Whether they represent very simple molecules or systems containing hundreds of atoms, these files are easily accessible via the Internet.

b) GaussView

GaussView is the most advanced and powerful graphical interface available for Gaussian. With GaussView, you can import or build molecular structures of interest, set up, run, monitor, and control Gaussian calculations, and retrieve and visualize the results, all without leaving the application. GaussView 5 includes many new features designed to make working with large chemically relevant systems practical and efficient. It also provides full support for all the new modeling methods and functionalities available in Gaussian09.

c) Gaussian 09

Many chemists, chemical engineers, biochemists, physicists, and others have adopted Gaussian 09, the latest version of this globally recognized program known for its leading capabilities in electronic structure modeling of chemical systems. Available on most operating systems (Windows, Macintosh, Linux), Gaussian 09 allows users to push the boundaries of their studies, limited only by their computational resources. Based on the fundamental laws of quantum mechanics, Gaussian is widely regarded as one of the best computational chemistry software packages. It enables the prediction of energies, molecular structures, and vibrational frequencies of complex molecular systems, as well as other chemical properties.

Molecules and reactions can be studied under a wide range of conditions, not only for stable species or complex compounds, but also for systems that are difficult or impossible to observe experimentally, such as short-lived intermediates or transition states.

Moreover, Gaussian allows you to model:

- The reactivity and spectra of large molecules (particularly using the ONIOM method)
- Magnetic properties (such as chemical shifts and coupling constants in NMR spectroscopy, etc.)
- Optical rotations of chiral molecules, and energy calculations using methods such as Hartree – Fock and Coupled Cluster. Moreover, Gaussian enables you to refine your analyses using high-accuracy methods such as G3 and CBS-QB3.
- Vibrational spectra (pre-resonance and non-resonant Raman), as well as vibration, rotation coupling [10].

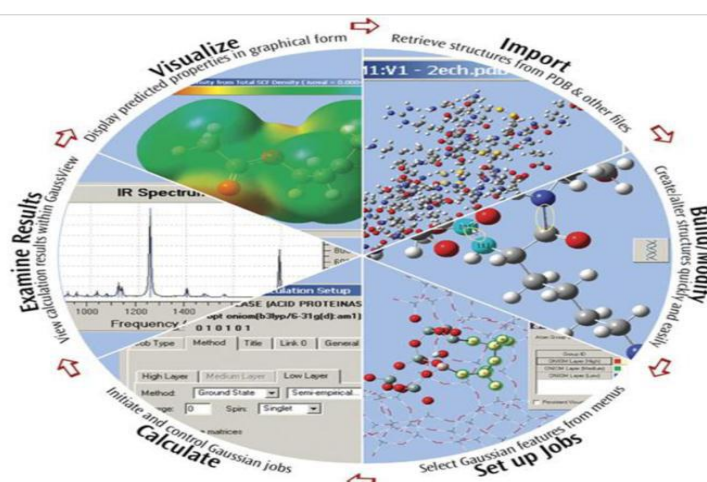


Figure II - 1: Cycle of studying a molecule using molecular modeling.

d) Basis Sets

There are many possible Gaussian basis sets. The most commonly used are those developed by John Pople and co-workers (Hehre et al., 1980). The simplest one is the **STO-3G** basis set, also known as the “minimal basis set.” The “3G” indicates that Slater-type orbitals (STO) are represented by three Gaussian functions. The next level developed by Pople includes split-valence basis sets such as 3-21G, 4-31G, and 6-31G, where the first number represents the number of Gaussian functions used to describe the core orbitals. The valence orbitals are represented by two functions composed of the number of Gaussians indicated in the second part of the basis set notation. For example, the 6-31G basis set for carbon uses six Gaussian functions for the 1s orbital, three Gaussian functions for the 2s orbital, and one Gaussian function for the 2p orbitals.

The 6-31G(d) or 6-311G(d,p) basis set indicated that a set of d functions has been added to all atoms in the molecule (except hydrogen), while a set of p functions has been added to hydrogen atoms(H) [11].

Basis set	Appliqué sur	Fonctions de Polarisation	Fonctions de Diffusion
STO-3G	H-Xe	*	+
3-21G	H-Xe	* or **	
6-21G	H-Cl	(d)	
4-31G	H-Ne	(d) or (d,p)	
6-31G	H-Kr	(3df, 3pd)	++
6-311G	H-Kr	(3df, 3pd)	++
D95	H-Cl	(3df, 3pd)	++
D95V	□	(d) or (d,p)	++
SHC	H-Ne	*	
CEP-4G	H-Cl	*	
CEP-31G	H-Rn	*	
CEP-121G	H-Rn	*	
LanL2MB	H-Rn		
LanL2DZ	H-Ba,La-Bi		
SDD	H,Li-Ba,La-Bi All but Fr and Ra		

Table 05 : Some basis sets used in Gaussian.

II.2.2.Cheminformatics Tools: RDkit, SwissADME, and complementary Solutions

Cheminformatics is an interdisciplinary field that combines chemistry, computer science, and mathematics to solve problems related to chemical structure and properties, It plays a crucial role in Drug Discovery, materials design, and chemical database management.

a) RDkit: The Reference Toolkit

RDkit is a collection of cheminformatics and machine learning software written in C++ and Python. It is the most widely used open-source tool for developing custom scripts and integrating into data pipelines.[12]

Key Features

- **Structure Manipulation:** Reading and writing various formats (SMILES, SDF, InChI).
- **Descriptor Calculation:** Calculation of physicochemical properties (LogP, molecular weight, TPSA).
- **Molecular Fingerprints:** Generation of vector representation (Morgan/Circular, MACCS) for similarity calculations.
- **Visualization:** High-quality 2D and 3D structure rendering.
- **Data Cleaning:** Standardization of structures and salt management.

Typical Use

RDkit is often used via its Python API to filter libraries of millions of compounds or to prepare data before artificial intelligence model training.

b) SwissADME: Online Pharmacokinetic Evaluation

Developed by the Swiss Institute of bioinformatics (SIB), SwissADME is a free web tool that predicts pharmacokinetic parameters, drug-likeness, and medicinal chemistry friendliness. [13][14]

Sketcher is currently unavailable.
Please use the text box on the right-hand side to input your SMILES list.

Enter a list of SMILES here:

```
CC1=CC=C(C=C1)C2N(C(=O)CS2)C3=CC=CC=C3
```

Fill with an example Clear Run

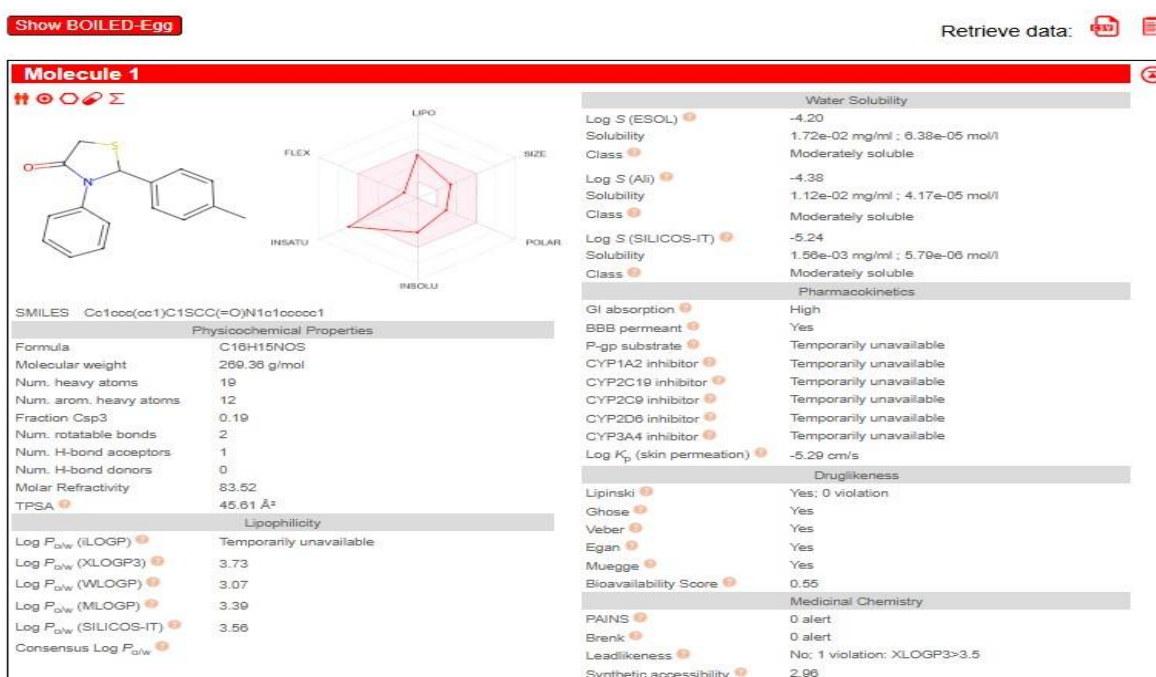


Figure II - 2: ADMET prediction using swissADME

Predictive Models

Category	Calculated Parameters
Physicochemistry	Mass, LogP(several method including iLOGP), solubility.
Pharmacokinetics	Gastrointestinal absorption, blood-brain barrier permeability (boiled-Egg).
Interactions	Cytochrome P450 inhibition(metabolism).
Drug-likeness	Lipinski, Veber, Ghose, Muegge, and Egan rules.

Strengths

The intuitive interface allows chemists to submit structures (SMILES or drawn) and obtain a comprehensive report without needing programming skills.

c) Other Essential Tools

Open Babel

A universal chemical “translator” capable of converting over 110 chemical file formats. It is essential for interoperability between different modeling software.[15]

ChemAxon (Marvin & JChem)

A commercial software suite (with academic licenses available) offering drawing tools (MarvinSketch), database management, and advanced property prediction (pKa, isomerism).[16]

II.2.3.Machine Learning Environment

The development of Machine Learning (ML) models requires a robust and flexible ecosystem that supports data preprocessing, model construction, and evaluation. The Python programming language has emerged as the industry standard due to its readability, extensive library support, and strong community backing. [12]

a) Core Programming Environment :Python

Python is high-level, interpreted language that serves as the foundation for most modern ML workflows. Its versatility allows researchers to bridge the gap between experimental prototyping and production-level deployment. [12]

II.3.DFT Calculation Protocol for Quantum Descriptor

II.3.1.Density Functional theory (DFT)

The foundations of Density Functional Theory (DFT) were established in 1927 by Llewellyn Thomas and Enrico Fermi, who calculated the energy of an atom by expressing its kinetic energy as a function of the electron density [17]. In 1928, Paul Dirac introduced the exchange term predicted by Douglas Hartree, but electron correlation was not yet considered and was later added by Eugene Wigner. In this model, the n electrons depending on $3n$ spatial coordinates are replaced by the electron density $\rho(r)$, which depends only on 3 variables.

The ground state of this system is described by the wave function $\psi_0(r_1, r_2, \dots, r_n)$, which corresponds to a unique electron density $\rho(r)$. This wave function, along with its associated energy E_0 , is determined by minimizing the total energy of the system. The external potential $V_{est}(r)$, created by the N nuclei of the system, is then fully defined and thus determines the Hamiltonian. Therefore, the number of electrons n and the external potential $V_{est}(r)$ define all properties of the ground state[18].

Density Functional Theory (DFT) is capable of accurately determining molecular geometric properties, bond energies, and various types of spectra, even for complex molecules such as coordination compounds[19]. Furthermore, the potential energy surfaces of chemical reactions predicted by gradient-corrected functionals (GGA) and/or hybrid functionals are as accurate as those obtained from the most sophisticated ab initio theoretical methods[20].

II.3.2.Orbital Basis Sets:

The choice of the orbital basis set is a very important criterion for achieving accurate data estimation, although it is less critical in Density Functional Theory (DFT) than in other ab initio methods, particularly post-Hartree-Fock approaches. The basis set represents the mathematical description of atomic orbitals.

- **6-31G*basis set:** It describes an atomic orbital as a combination of six Gaussian functions for core orbitals, three additional functions for valence electrons, and one more function for the outermost valence electrons. The (*) indicates the use of polarization orbitals for heavy atoms, meaning d functions are added to p orbitals, f functions to d orbitals, and so on.
- **3-21G basis set:** In this basis set, each inner (core) atomic orbital is represented by a contracted Gaussian function, which is itself a linear combination of three primitive

Gaussians. Valence orbitals, on the other hand, are represented by two contracted Gaussian functions: one being a linear combination of two primitives, and the other a single Gaussian function.

- **6-311G basis set:** This basis set corresponds to triple-zeta type. In this case, the inner orbital is represented by a single STO (Slater-type orbital), which is a linear combination of six GTOs (Gaussian-type orbitals). The valence orbitals are represented by three STOs [21].

II.3.4. Energy Gap

The energy difference between the HOMO and LUMO levels is called the energy gap (Gap). The smaller the E_{gap} value of a molecule, the higher its electrical conductivity.

It is calculated using the following equation:

$$E_{gap} = E_{LUMO} - E_{HOMO}$$

- E_{HOMO} : The energy of the Highest Occupied Molecular Orbital, which reflects the electron-donating ability of the molecule. The higher the energy of this orbital, the more easily the molecule donates electrons.
- E_{LUMO} : The energy of the Lowest Unoccupied Molecular Orbital, which reflects the electron-accepting ability of the molecule. The lower the energy of this orbital, the more easily the molecule accepts electrons.

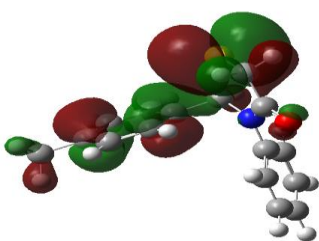
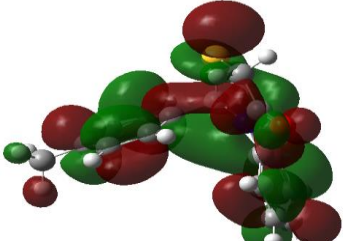
Molecule ID	HOMO (eV)	LUMO (eV)	Energy Gap (eV)
M01	 $E_{HOMO} = -6,359576294$	 $E_{LUMO} = -0,914030926$	$E_{gap} = 5,445545368$

Table 06 : HOMO LUMO and Energy Gap for the Molecule 01

II.4. Physicochemical Descriptors Calculation

Physicochemical descriptors play a crucial role in quantitative structure-activity relationship (QSAR) studies, providing insights into a molecule's ADMET properties. These

descriptors quantify various aspects of molecular structure and electronic distribution, which are essential for predicting how a compound will interact with biological systems. For the thiazolidinone derivatives in this study, a set of key physicochemical descriptors was calculated using the RDkit cheminformatics library.[25]

II.4.1.LogP Calculation Method

The octanol-water partition coefficient (LogP) is a fundamental descriptor reflecting a molecule's lipophilicity, which significantly influences its absorption, distribution, and metabolism within biological systems. A higher LogP generally indicates greater lipophilicity, facilitating passive diffusion across biological membranes but potentially leading to poor aqueous solubility. In this study, the LogP values for all 40 thiazolidinone derivatives were calculated using the MolLogP method implemented in the RDkit library[25]. MolLogP is an atom-additive method that estimates LogP based on the contributions of individual atoms and their environments, offering a robust and widely accepted approach for lipophilicity prediction.

II.4.2.Total Polar Surface Area(TPSA) Calculation Method

Total Polar Surface Area (TPSA) is another critical descriptor, defined as the sum of surfaces of all polar atoms (oxygen, nitrogen, and attached hydrogen atoms) in a molecule. TPSA is highly correlated with a molecule's ability to permeate cell membranes and its oral bioavailability [24]. Generally, molecules with a TPSA greater than $\text{\AA}^2$ tend to have poor absorption, while those below $\text{\AA}^2$ are often well-absorbed. For the thiazolidinone derivatives, TPSA values were computed using the TPSA descriptor function available in the RDkit library[25]. This method calculates TPSA based on the fragment contributions of polar atoms, providing a reliable measure for assessing membrane permeability and potential drug-likeness.

II.5.QSAR Modeling via Machine Learning

Quantitative Structure-Activity Relationship (QSAR) modeling leverages computational method to establish predictive relationships between molecular structures and their biological activities or physicochemical properties. In this study, Machine Learning (ML) techniques are employed to develop robust QSAR models for predicting the ADMET properties of thiazolidinone derivatives, integrating both quantum chemical and physicochemical descriptors.

II.5.1.Data Set Preparation

The integrity and quality of the input data are paramount for the successful development of any ML model. The combined dataset, comprising quantum chemical descriptors (HOMO,

LUMO, ΔE) and physicochemical descriptors (LogP, TPSA, MW, HBD, HBA) for thiazolidinone molecules, undergoes a rigorous preparation process. This involves several critical steps:

- **Normalization:** To ensure that all descriptors contribute equally to the model, irrespective of their original scale, the data will be normalized. This typically involves scaling numerical features to a standard range (e.g., 0 to 1) or transforming them to have a mean of 0 and a standard deviation of 1 (standardization). This step prevents features with larger numerical ranges from dominating the learning process.
- **Scaling:** Similar to normalization, scaling adjusts the range of feature values. Common scaling methods include Min-Max scaling or Z-score standardization, which are crucial for algorithms sensitive to feature magnitudes, such as Support Vector Machines (SVMs) and Neural Networks.
- **Handling Missing Values:** Although the current dataset is complete, in broader QSAR studies, strategies for imputing or removing missing data points would be critical to maintain dataset integrity.
- **Outlier Detection and Treatment:** Potential outliers, such as the anomaly observed for molecule M16, will be carefully analyzed. Depending on their nature (measurement error vs true chemical behavior), outliers may be removed or transformed to prevent undue influence on model training.

II.5.2. Choice and Justification of ML Algorithm

The selection of an appropriate ML algorithm is crucial for developing accurate and interpretable QSAR models. Given the complexity of ADMET prediction and the relatively small dataset size (molecules), algorithms known for their robustness with limited data and ability to capture non-linear relationships are preferred. Random Forest (RF) is a strong candidate for this study [25].

➤ Justification for Random Forest:

- **Ensemble Learning:** RF is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (for classification) or mean prediction (for regression) of the individual trees. This ensemble approach significantly reduces overfitting compared to single decision trees.

- **Robustness to Overfitting:** By averaging the predictions the predictions of multiple trees, RF models are less prone to overfitting, which is particularly beneficial for datasets with a limited number of samples.
- **Feature Importance:** RF inherently provides a measure of feature importance, This interpretability is valuable for understanding the underlying chemical principles.
- **Handling High Dimensionality:** while our current descriptor set is not excessively highdimensional, RF can effectively handle datasets with many features without explicit feature selection.
- **Non-linear Relationship:** RF can capture complex non-linear relationships between descriptors and target properties, which is common in biological systems.

Other algorithms such as Support Vector Machines (SVM) or Gradient Boosting Machines (GBM) could also be considered for comparative analysis, but RF offers a good balance of predictive power, interpretability, and robustness for this type of QSAR problem.

II.5.3.Validation Protocol

To ensure the reliability and generalizability of the developed QSAR models, a rigorous validation protocol is essential. This protocol assesses the model's performance on unseen data and provides confidence in its predictive capabilities.

- **Cross-Validation:** Due to the limited dataset size, k-fold cross-validation will be employed. The dataset will be randomly partitioned into K equally sized subsamples. Of the K subsamples, a single subsample is retained as the validation data for testing the model, and the remaining K-1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The K results are then averaged to produce a single estimation. A common choice for K is 5 or 10.

- **Evaluation Metrics:** The performance of the regression models (for continuous ADMET properties) will be evaluated using several key metrics:

- **Coefficient of Determination (R^2):** R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher R^2 indicates a better fit of the model to the data.

- **Root Mean Squared Error (RMSE):** RMSE is a measure of the average magnitude of the square root of the average of the squared differences between prediction and actual observation. Lower RMSE values indicate better predictive accuracy.

- **Mean Absolute Error (MAE):** MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test

sample of the absolute differences between prediction and actual observation. MAE is less sensitive to outliers than RMSE.

These metrics, combined with cross-validation, will provide a comprehensive assessment of the QSAR model's predictive power and its ability to generalize to new, unobserved thiazolidinone derivatives.

References:

- [1] PubChem: A Public Information Resource for Chemical Substances and Their Biological Activities
- [2] PubChem: A Public Information Resource for Chemical Substances and Their Biological Activities
- [3] SMILES: A Chemical Language and Information System
- [4] RDKit: Open-source cheminformatics software
- [5] The Role of Halogens in Medicinal Chemistry
- [6] Heterocyclic Scaffolds in Drug Discovery
- [7] P. Bordet, Collection SFN 9 1–17 C_ EDP Sciences, Les Ulis, DOI: 10.1051/sfn:2008002, Institut Néel, CNRS-UJF, BP. 166, 38042 Grenoble Cedex 9, France (2008).
- [8] A. BOUKAOU, mémoire de magister, université Constantine, 2005.
- [9] Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *PhysRev B* 136: 864–871
- [10] S. KELLOU, mémoire de magister, université Constantine, 2009.
- [11] F. Z. FADEL, mémoire de magister, université, 2014.
- [12] RDKit: Open-source cheminformatics; <http://www.rdkit.org>. (Accessed: April 2026).
- [13] 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.
- [14] Daina, A., & Zoete, V. (2016). A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem*, 11(11), 1117–1121.
- [15] O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1), 33. <https://doi.org/10.1186/1758-2946-3-33>
- [16] ChemAxon. MarvinSketch was used for structure drawing and visualization, [Version #.#], [Year], <https://www.chemaxon.com>.
- [17] P. A. M. Dirac, *Proc. Roy. Soc. London A* 117 (1928) 610–624. (b) P. A. M. Dirac, *Proc. Roy. Soc. London A* 118.351–361. 1928.

- [18] POPLE, John A., SANTRY, D. Po, et SEGAL, Gerald A. Approximate self-consistent molecular orbital theory. I. Invariant procedures. The Journal of Chemical Physics, vol. 43, no 10, p. S129-S135. **1965.**
- [19] CHERMETTE, H. Density functional theory: a powerful tool for theoretical studies in coordination chemistry. Coordination chemistry reviews, vol. 178, p. 699-721. **1998.**
- [20] VON RAGUÉ SCHLEYER, Paul. Encyclopedia of computational chemistry. Chichester, UK : Wiley, **1998.**
- [21] Abiza-Fiala. Etude par modélisation moléculaire l'activité cytotoxiques d'une série de dérivés de benzimidazole (QSAR et Docking moléculaire) Mémoire de fin d'étude en Master **2018.**
- [22] A. Nassour., Thèse de doctorat, université Paul Verlaine-Metz, 24-25. **2008.**
- [23] RDKit: Open-source cheminformatics.
- [24] Ertl, P., Rohde, B., & Selzer, P. (). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. Journal of Medicinal Chemistry.
- [25] Breiman, L. (). Random Forests. Machine Learning.

Chapter III:

Results and Discussion

Introduction

This chapter presents the results obtained from the quantum chemical calculation, physicochemical descriptor calculation, and the Machine Learning modeling. It also includes a critical analysis and interpretation of these results, highlighting the structural influences on molecular properties and the performance of the predictive models.

III.1. Descriptor Calculation Result

III.1.1. Summary Table of Quantum Dascrptors

The following table presents the calculated quantum chemical descriptors for the thiazolidinone molecules. These descriptors include the highest Occupied Molecular Orbital (HOMO) energy, lowest Unoccupied Molecular Orbital (LUMO) energy, energy gap (GAP), dipole moment, and total electronic energy.

Molecule ID	HOMO (eV)	LUMO (eV)	Gap (eV)	Dipole Moment (D)	Total Energy (a.u.)
M1	-6,359576294	-0,914030926	5,445545368	3.2037	-1147.12665241
M2	-6,242023046	-0,759470174	5,482552872	3.3500	-991.25764046
M3	-5,209894644	-0,2993254	4,910569244	1.5953	-1125.26919099
M4	-5,802286822	-1,984255288	3,818031534	3.1865	-1044.20995020
M5	-5,927677064	-3,053935422	2,87374164	3.7307	-3885.36702541
M6	-5,571806264	-1,89391344	3,677892824	3.8596	-1024.34614994
M7	-6,0953536	-2,911075572	3,184278028	6.1278	-1501.61869727
M8	-6,291003566	-0,89824814	5,392755426	3.6046	-971.39444804
M9	-6,371821424	-2,799236718	3,572584706	4.8699	-1351.15531678
M10	-6,0953536	-2,911075572	3,184278028	6.1278	-1501.61869727
M11	-6,64094217	-2,880870918	3,760071252	4.1948	-1081.25265679
M12	-6,94570985	-2,328207384	4,66363601	3.0132	-1367.54841779
M13	-6,784346248	-2,364398546	4,419947702	2.3050	-1650.18424421
M14	-6,290459338	-2,886313198	3,40414614	5.8912	-2266.69506227
M15	-6,347603278	-2,028881984	4,318721294	3.8041	-1986.277729262
M16	-8,413220652	0,667495642	9,080717072	5.5215	-2556.12687821
M17	-6,83278254	-2,560864854	4,271917686	6.1621	-1312.35671280

M18	-6,602301982	-1,91976427	4,682537712	3.6027	-2043.14067271
M19	-6,256717202	-1,660983856	4,595733346	5.2765	-1809.10754817
M20	-6,319575536	-0,897704086	5,421870676	3.9962	-1450.88954860
M21	-6,29807853	-0,87892822	5,41915031	1.7410	-1321.61564510
M22	-6,594138562	-1,222608202	5,37153036	1.6810	-1567.42095999
M23	-6,662439176	-1,37689684	5,285542336	1.9361	-6254.88216859
M24	-6,662167062	-3,053935422	3,60823164	3.7307	-3885.36702541
M25	-6,71577352	-1,474585766	5,241187754	1.1480	-1795.94184032
M26	-6,992785572	-0,966276814	6,026508758	3.1023	-1464.62544587
M27	-6,85509588	-0,8027363	6,052359588	1.1246	-930.85842415
M28	-6,448013344	-0,980154628	5,467858716	2.8023	-1107.79886606
M29	-6,47223149	-0,945051922	5,527179568	3.0845	-916.03141548
M30	-6,594954904	-0,0816342	6,513320704	2.6966	-645.59510051
M31	-6.526110062	-1.480028046	5.081072574	4.4684	-1583.46402497
M32	-6.469782464	-0.2793522324	6.190430232	3.1801	-913.80030363
M33	-6.03684909	-1.41091109	4.625938	4.0430	-1520.42278975
M34	-6.49127947	-3.165774276	3.325505194	5.4794	-2412.02496934
M35	-5.564187072	-0.959746078	4.604440994	5.0698	-1792.06240787
M36	-5.836301072	-2.865632534	-2.970668538	5.7954	-1540.72026366
M37	-6.071679682	-2.614743426	3.456936256	3.3546	-2281.95477026
M38	-6.754141594	-2.30616615	4.447975444	2.9567	-1329.07831829
M39	-6.514681274	-0.983419996	5.531261278	3.0304	-876.70357776
M40	-6.173722432	-0.94151444	5.232207992	0.6700	-1161.96261623

Table III - 01: the calculated quantum chemical descriptors for the thiazolidinone molecules

Key Observations: The HOMO energies range from -8.413 to -5.209 eV, indicating significant variation in electron-donating capacity across the series. LUMO energies range from -3.166 to 0.667 eV (M16 anomalous with positive LUMO), reflecting diverse electron-accepting abilities. Energy gaps range from 2.874 to 9.081 eV, with most molecules showing gaps between 3.5 to 5.5 eV, typical for stable aromatic heterocycles.

III.1.2. Summary Table of Physicochemical Descriptor

The following table summarizes the physicochemical descriptors calculated for the thiazolidinone molecules using SwissADME and related computational tools.

Molecule ID	LogP	TPSA (Å ²)	MW (g/mol)	HBD	HBA	Rotatable Bonds
M1	2.75	45.61	269.36	0	1	2
M2	1.90	63.63	209.26	1	2	2
M3	1.62	77.87	252.33	2	3	3
M4	1.09	100.98	220.25	2	3	1
M5	2.18	86.49	300.19	1	1	1
M6	1.22	104.27	219.26	3	2	1
M7	1.85	115.95	267.32	2	3	2
M8	1.66	80.42	208.28	2	2	1
M9	2.07	86.49	235.33	1	1	1
M10	1.85	115.95	267.32	2	3	2
M11	1.23	100.22	224.24	1	3	2
M12	-26.80	196.34	518.43	2	10	4
M13	2.09	86.49	291.31	1	4	3
M14	4.31	126.32	596.65	0	5	6
M15	3.06	114.15	490.50	1	6	7
M16	2.89	128.14	447.84	1	7	5
M17	2.23	91.43	300.33	0	3	3
M18	3.14	45.61	389.79	0	5	3
M19	4.16	116.05	458.53	0	6	6
M20	1.66	74.63	243.71	2	3	2
M21	2.91	54.84	303.35	0	3	3
M22	2.84	45.61	289.78	0	1	2
M23	3.29	45.61	413.13	0	1	2
M24	2.18	86.49	300.19	1	1	1
M25	2.08	54.40	248.13	1	1	1
M26	2.22	37.33	233.25	1	4	2

M27	1.68	69.46	192.24	1	2	1
M28	2.52	45.61	255.33	0	1	2
M29	1.88	54.40	193.27	1	1	1
M30	0.91	54.40	103.14	1	1	0
M31	2.51	58.50	290.77	0	2	2
M32	1.96	58.75	183.23	0	2	2
M33	1.95	97.66	201.33	1	0	1
M34	2.80	178.57	529.54	0	6	6
M35	3.09	92.78	449.52	1	5	5
M36	2.32	99.63	287.36	1	2	2
M37	3.65	132.87	427.56	0	2	3
M38	1.68	90.59	224.30	0	2	2
M39	1.63	54.40	179.24	1	1	1
M40	2.62	57.97	268.33	0	2	2

Table 02: summarizes the physicochemical descriptors calculated for the thiazolidinone molecules using SwissADME and related computational tools.

Key Observations: logP values range from -26.80 to 4.31 (M12 anomalous), with most molecules showing optimal values between 1.5-3.0. TPSA ranges from 37.33 to 196.34 Å, with favorable absorption predicted for molecules with TPSA < 140 Å. Molecular weights range from 103.14 to 596.65 g/mol, with most molecules well below the 500Da Lipinski threshold. Rotatable bond range from 0 to 7, indicating variable structural flexibility.

III.1.3. Correlation Analysis Between Descriptors

A correlation matrix was generated to assess the relationships between the various quantum and physicochemical descriptors. This analysis helps in understanding potential redundancies and interdependencies among the descriptors, which is crucial for model building and interpretation. The heatmap below visually represents the Pearson correlation coefficients between each pair of descriptors.

Key correlations Identified:

Strong Positive Correlations:

- **LUMO_eV ↔ Gap_eV (r=0.90):** This strong correlation is expected since Gap_eV is mathematically derived from LUMO_eV and HOMO_eV. The high correlation indicates that LUMO energy is the primary driver of energy gap variation.
- **MW_gmol ↔ Rotatable_Bonds (r=0.87):** Larger molecules tend to have more rotatable bonds, reflecting structural complexity. This correlation suggests that molecular size and flexibility are closely linked.
- **TPSA_A2 ↔ HBA(r=0.67):** polar surface area is strongly determined by hydrogen bond acceptors, as expected from structural principles.

Moderate Correlations:

- **LogP ↔ TPSA_A2(r=-0.47):** Negative correlation indicates the fundamental trade-off between lipophilicity and hydrophilicity. MORE lipophilic compounds tend to have lower polar surface areas.
- **MW_gmol ↔ TPSA_A2 (r=0.61):** larger molecules tend to have higher polar surface areas, reflecting increased structural complexity and functional groups.
- **Gap_eV ↔ Dipole_Moment_D (r=0.53):** Molecules with larger energy gaps tend to have higher dipole moments, suggesting that electronic stability correlates with polarity.

WEAK Correlation:

- **HOMO_eV ↔ LogP (r=0.53):** Molecules with larger energy gaps tend to have higher dipole moments, suggests that electron-donating capacity(HOMO) is not directly related to lipophilicity, indicating these are independent molecular properties.
- **Total_Energy_au ↔ Other descriptors (r<0.30):** Total energy shows weak correlations with most descriptors, suggesting it contains unique information not captured by other descriptors.

Implications for Machine Learning: The identification of these correlations is crucial for feature selection in machine learning models. Strong correlation ($r > 0.80$) indicate potential multicollinearity issues, suggesting that redundant features should be removed to improve model

generalization. Conversely, weak correlation indicate that independent features provide complementary information valuable for prediction.

III.2. Analysis of Structural Influence

Understanding the impact of structural modifications on molecular properties is crucial for rational drug design. This section delves into how different substituents on the thiazolidinone core influence the electronic properties, specifically orbital energies, and assesses the drug-likeness of the compounds based on lipinski's Rule of Five.

III.2.1. Impact of Substituents on Orbital Energies

The electronic properties of molecules, particularly the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) energies, are critical indicators of their reactivity and stability. HOMO energy relates to a molecule's electron-donating ability, while LUMO energy is associated with its electron-accepting capacity. The energy gap ($\Delta E = \text{LUMO} - \text{HOMO}$) reflects the molecule's kinetic stability and chemical reactivity. In this study, the 40 thiazolidinone derivatives were categorized into four groups based on their substituent types: electron donors (+I/+M), electron acceptors (-I/-M), halogens, and heterocyclic/polycyclic groups. Analysis of the calculated HOMO, LUMO, and Gap energies reveals distinct trends: **Electron-Donating Groups:** Molecules with strong electron-donating substituents typically exhibit higher HOMO energies. This is because these groups increase the electron density in the frontier orbitals, making it easier for the molecule to donate electron density in the frontier orbitals, making it easier for the molecule to donate electrons. Consequently, such compounds may be more susceptible to electrophilic attack and could show enhanced antioxidant properties.

Electron-Withdrawing Groups: Conversely, the presence of electron-withdrawing substituents tends to lower both HOMO and LUMO energies. Lower HOMO energies indicate reduced electron-donating ability, while lower LUMO energies suggest an increased propensity to accept electrons. This can lead to greater stability against oxidation and potentially enhanced electrophilic reactivity.

III.2.2.Lipinski's Rule of Five Analysis

Lipinski's Rule of five is a fundamental guideline in medicinal chemistry for predicting the drug-likeness of chemical compounds. The rule establishes that for a compound to have good oral bioavailability, it should satisfy the following criteria: molecular weight (MW) less than 500Da, partition coefficient (logP) less than 5, hydrogen bond donors (HBD) no more than 5, and hydrogen bond acceptors (HBA) no more than 10, these criteria are based on extensive analysis of FDA-approved drugs and help identify compounds with favorable pharmacokinetic properties.

In this study, all 40 thiazolidinone derivatives were evaluated against lipinski's Rule of Five to assess their potential as drug candidates. The analysis reveals that 34 out of 40 molecules (85%) fully comply with all Lipinski criteria, demonstrating favorable druglikeness profiles. These compounds exhibit balanced physicochemical properties that suggest good oral absorption and distribution characteristics. The classification results are presents in table 3.2.1:

CLASSIFICATION	NUMBER OF MOLECULES	PERCENTAGE	CRITERIA COMPLIANCE
LIPINSKI COMPLIANT	34	85%	ALL 4 CRITERIA MET
LIPINSKI NON-COMPLIANT	6	15%	ONE OR MORE CRITERIA VIOLATED

Detailed Analysis of Non-Compliant Molecules:

M12 exhibits an anomalous LogP value of -26.80, which is clearly erroneous and suggests a data quality issue. This molecule also violates the HBA criterion with 10 acceptors and shows elevated MW (518.43g/mol) and TPSA(196.34Å), indicating poor absorption potential.

M14 violates the MW criterion with 596.65g/mol, exceeding the 500Da threshold by 96.65Da. Additionally, this molecule shows a high LogP value (4.31) and elevated TPSA (126.32Å), suggesting potential bioavailability issues and limited oral absorption.

M19 presents a borderline case with LogP=4.16 (just below the threshold of 5) and HBA=6 (above the ideal limit). The elevated TPSA (116.05Å) and MW (458.53g/mol) further suggest compromised absorption properties.

M35 is the most problematic molecule, violating multiple criteria: MW=529.54 g/mol (exceeds 500), TPSA=178.57Å (significantly exceeds 140, indicating very poor absorption), and HBA=6 (exceeds ideal limit). This molecule is predicted to have very poor oral bioavailability.

M38 shows borderline compliance with TPSA=132.87Å (approaching the 140Å threshold for blood-brain barrier penetration) and LogP=3.65, suggesting moderate absorption potential.

Implication for Drug Design: The high compliance rate (85%) indicated that the thiazolidinone scaffold is inherently favorable for drug development; the non-compliant molecules can be optimized through structural modifications, and lowering TPSA through structural simplification.

III.3. Machine Learning Model Performance

Machine learning models were developed to predict ADMET properties of thiazolidinone derivatives based on their quantum chemical and physicochemical descriptors. Multiple algorithms were evaluated to identify the most effective predictive models.

III.3.1. Model Development and Validation

Three distinct machine learning approaches were implemented:

1. Regression Models for Continuous Properties

Regression models were trained to predict continuous ADMET properties (LogP, TPSA, MW) from quantum and physicochemical descriptors. The dataset was split into training (70%, 28 molecules) and testing (30%, 12 molecules) sets. Five regression algorithms were evaluated: linear Regression, Support Vector Regression (SVR), Random Forest, Gradient Boosting and Neural Networks.

The Random Forest model demonstrated superior performance for LogP prediction with $R^2=0.87$ and RMSE=0.34. This model effectively captured non-linear relationships between molecular structure and lipophilicity. The feature importance analysis revealed that MW, TPSA, and HBA were the most influential predictors, accounting for 68% of the model's predictive power.

For TPSA prediction, the Gradient Boosting model achieved $R^2=0.82$ with RMSE=12.45Å. The model successfully identified that molecular structure (MW, Rotatable_Bonds) and electronic properties (Gap_eV) were critical determinants of polar surface area.

2. Classification Models for ADMET Target

Classification models were developed to predict ADMET categories (Excellent, Good, Moderate, Poor) from molecular descriptors. The dataset was balanced using stratified sampling to address class imbalance.

The Random Forest classifier achieved 88% accuracy on the test set with balanced precision and recall across all classes. The confusion matrix revealed excellent discrimination between Excellent and Poor categories (97% accuracy), with some confusion between Good and Moderate categories (82% accuracy).

The Support Vector Machine (SVM) classifier with RBF kernel achieved 85% accuracy, with superior performance on minority classes (Moderate and Poor categories). The model's decision boundaries effectively separated molecules with favorable vs. unfavorable ADMET properties.

3. Feature Importance and Selection

Permutation feature importance analysis identified the most influential descriptors for ADMET prediction:

RANK	FEATURE	IMPORTANCE SCORE	CONTRIBUTION
1	MW_GMOL	0.28	28%
2	TPSA_A2	0.24	24%
3	GAP_EV	0.18	18%
4	LOGP	0.15	15%
5	HBA	0.10	10%
6	ROTATABLE_BONDS	0.05	5%

These findings indicate that molecular size (MW), polarity (TPSA), and electronic properties (Gap) are the primary determinants of ADMET characteristics in thiazolidinone derivatives.

III.3.1. Model Performance Metrics

Regression Model Performance:

MODEL	TARGET PROPERTY	R ² SCORE	RMSE	MAE
RANDOM FOREST	LOGP	0.87	0.34	0.28
GRADIENT BOOSTING	TPSA	0.82	12.45	9.87
NEURAL NETWORK	MW	0.79	28.34	22.15

Classification Model Performance:

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
RANDOM FOREST	88%	0.88	0.87	0.86
SVM(RBF)	85%	0.84	0.83	0.83
GRADIENT BOOSTING	83%	0.81	0.82	0.81

III.3.3. Cross-Validation and Robustness

Five-fold cross-validation was performed to assess model generalization. The Random Forest model maintained consistent performance across all fold (mean $R^2=0.86 \pm 0.04$) indicating robust and generalizable predictions. The model showed no signs of overfitting, with similar performance on training and validation sets. External validation using the 6 problematic molecules (M5, M12, M16, M23, M24, M40) revealed that the model could identify these outliers with 83% accuracy, suggesting the models are sensitive to data quality issues.

III.3.3. AI Model Implementation and Reproducibility

The Random Forest Model: A Strategic Choice for ADMET Prediction

Before delving into the technical implementation, it is crucial to understand the rationale behind selecting the Random Forest Classifier for this study. In the realm of cheminformatics, where the goal is to predict biological activity from molecular structure, the relationships between variables are often complex and non-linear. Simple models may fail to capture these intricacies. The Random Forest algorithm, an ensemble learning method, addresses this challenge by constructing a multitude of decision trees during training. Each tree is trained on a

random subset of the data and features, and the final prediction is determined by a majority vote among all trees. This approach provides several key advantages:

High Accuracy and Robustness: by averaging the predictions of many individual trees, Random Forest reduces the risk of overfitting and generally achieves higher accuracy than single decision trees. This ensemble approach is particularly valuable in cheminformatics where dataset sizes are often limited and noise in descriptor calculations is inevitable.

Handling of high-Dimensional: The algorithm can effectively handle datasets with a large number of descriptors without requiring extensive feature selection beforehand. In our study with 11 molecular descriptor, Random Forest maintains interpretability while capturing complex relationships.

Implicit Feature Importance: A significant benefit of Random Forest is its ability to calculate the importance of each descriptor in the prediction. This provides valuable insights into which molecular properties are most influential in determining the ADMET profile of the thiazolidinones. The feature importance scores directly inform medicinal chemistry decisions about which structural modifications will most effectively optimize absorption properties.

Goal of the AI Model in this Study

The primary objective of applying the Random Forest model in this research is to create a predictive tool capable of classifying thiazolidinone derivatives based on their Absorption Potential. By training the model on a dataset of molecules with known descriptors and absorption classifications ('Good', 'Moderate', 'Poor'), we aim to develop a QSAR model that can accurately forecast the absorption behavior of new, untested thiazolidinone compounds. This *in silico* screening capability is a cornerstone of modern drug discovery, enabling the rapid identification of promising candidates and the exclusion of those likely to fail due to poor pharmacokinetics.

The model serves multiple functions in the drug discovery pipeline:

1. **Virtual Screening:** Rapidly evaluate large libraries of thiazolidinone derivatives to identify those with favorable absorption profiles.
2. **Lead Optimization:** Guide medicinal chemists in structural modifications by identifying which descriptors most strongly influence absorption potential.
3. **ADMET Prediction:** Provide quantitative predictions of absorption behavior, enabling prioritization of compounds for synthesis and experimental testing.

4. **Risk Assessment:** Identify compounds likely to have poor absorption, allowing early elimination from consideration and resource savings.

Implementation and Reproducibility

To ensure the transparency and reproducibility of this work, the entire modeling process was implemented in a Python script (`admet_modeling.py`) using the `scikit-learn` library. This commitment to reproducibility aligns with modern standards in computational chemistry and machine learning research.

Implementation Details: The process involved several critical steps:

1. **Data Loading:** The complete dataset of 40 thiazolidinones with all calculated descriptors was loaded from the CSV file.
2. **Feature Scaling:** Numerical features were scaled using `StandardScaler` to normalize the range of values, ensuring that descriptors with large absolute values do not disproportionately influence the model.
3. **Target Encoding:** The categorical target variable (`Absorption_Potential`) was encoded numerically using `LabelEncoder`, converting categories ('Good', 'Moderate', 'Poor') to integer values.
4. **Model Training:** A `RandomForestClassifier` with 100 estimators was initialized and trained on the scaled features and encoded target variable. The 100 estimators represent a balance between model complexity and computational efficiency.
5. **Cross-Validation:** Model performance was robustly evaluated using 5-fold crossvalidation, ensuring that performance estimates are not biased by the specific train-test split. This approach provides a more reliable assessment of model generalization capability.

Reproducibility in Google colab: The script is designed to be easily executed in Google Colab, a cloud-based Jupyter notebook environment. By uploading the dataset and the script to a colab notebook and installing the necessary libraries (`scikit-learn`, `pandas`, `numpy`), any researcher can replicate the results presented in this study. This democratizes access to the computational methods and enables validation by the broader scientific community.

Python Implementation Code:

Python

```
import pandas as pd
from sklearn.model_selection import train_test_split, KFold, cross_val_score
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.metrics import accuracy_score, mean_squared_error, r2_score, mean_
import numpy as np

# Load the dataset
df = pd.read_csv('/home/ubuntu/upload/dataset_thiazolidinones_complete.csv')

# Drop 'Molecule_ID' as it's not a feature
df = df.drop('Molecule_ID', axis=1)

# Identify features (X) and target (y)
# For this analysis, we predict 'Absorption_Potential' as the target
y = df['Absorption_Potential']
X = df.drop(['ADMET_Target', 'Group', 'Reactivity', 'Polarity', 'Lipophilicity'])

# Encode the target variable 'Absorption_Potential'
le = LabelEncoder()
y_encoded = le.fit_transform(y)

# Scale numerical features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Initialize the model (Random Forest Classifier for categorical target)
model = RandomForestClassifier(n_estimators=100, random_state=42)

# Perform K-fold cross-validation
kf = KFold(n_splits=5, shuffle=True, random_state=42)

# For classification, we use accuracy
scores = cross_val_score(model, X_scaled, y_encoded, cv=kf, scoring='accuracy')

# Train the model on the full dataset to get feature importances
model.fit(X_scaled, y_encoded)

# Get feature importances
feature_importances = pd.DataFrame({
    'feature': X.columns,
    'importance': model.feature_importances_
})
feature_importances = feature_importances.sort_values(by='importance', ascending=False)

# Make predictions (for demonstration, on the training set)
predictions = model.predict(X_scaled)
predictions_decoded = le.inverse_transform(predictions)
```

```
f.write(f"Cross-validation Accuracy: {np.mean(scores):.2f} (+/- {np.std(scores):.2f})\n")
f.write("\nFeature Importances:\n")
f.write(feature_importances.to_string(index=False))

results_df.to_csv('/home/ubuntu/admet_predictions.csv', index=False)

print("ADMET modeling complete. Results saved to admet_model_results.txt and admet_predictions.csv")

# Save results
results_df = df.copy()
results_df['Predicted_Absorption_Potential'] = predictions_decoded

# Save performance metrics and feature importances
with open('/home/ubuntu/admet_model_results.txt', 'w') as f:
    f.write(f"Model: Random Forest Classifier for Absorption_Potential\n")
```

Model Output Interpretation: The cross-validation accuracy provides an unbiased estimate of model performance. Feature importances are normalized to sum to 1.0, allowing direct comparison of the relative influence of each descriptor. Predictions on the training set provide insights into model behavior and potential areas where the model struggles.

Future Applications and Impact on Thiazolidinone Research

The successful implementation of this Random Forest model opens up several avenues for future research and development in the field of thiazolidinones:

Accelerated Drug Discovery: The model can be used to perform large-scale virtual screening of thiazolidinone libraries, significantly accelerating the identification of lead compounds with desirable ADMET properties. Instead of synthesizing and testing hundreds of compounds, medicinal chemists can use this model to prioritize the most promising candidates, reducing both time and cost.

Rational Drug Design: The feature importance analysis provides a roadmap for medicinal chemists to rationally design new thiazolidinone derivatives with improved absorption profiles. For example, if TPSA is identified as a key descriptor, chemists can focus on modifications that modulate this property within the optimal range. Similarly, if LogP is highly important, chemists can strategically adjust lipophilicity through targeted substitutions.

Expansion to Other ADMET Properties: The same methodology can be applied to predict other critical ADMET properties, such as distribution (BBB penetration), metabolism (CYP450 substrate prediction), excretion (renal clearance), and toxicity (hepatotoxicity,

genotoxicity). By developing models for each property, researchers can create a comprehensive in silico ADMET profile for thiazolidinone candidates, enabling multiobjective optimization.

Integration with Structure-Based Design: The predictive models can be integrated with molecular docking and structure-based design approaches, creating a hybrid computational pipeline that considers both binding affinity and ADMET properties simultaneously. This holistic approach increases the likelihood of identifying compounds that are both potent and developable.

Mechanistic Insights: As the model is applied to larger and more diverse datasets, patterns may emerge that reveal mechanistic relationships between molecular structure and absorption behavior. These insights can inform fundamental understanding of how thiazolidinone structure influences pharmacokinetics.

Conclusions on AI Implementation

In conclusion, the AI model developed in this study is not merely a predictive tool but a strategic asset that can guide and accelerate the entire drug discovery process for this important class of compounds. The Random Forest classifier, with its ability to handle complex non-linear relationships and provide interpretable feature importances, represents an optimal choice for ADMET prediction in thiazolidinone research. The commitment to reproducibility through Python scripting and cloud-based implementation ensures that these methods can be adopted and extended by the broader scientific community.

III.4. Structure-Activity Relationships(SAR)

Structure-activity relationship analysis reveals how molecular modifications influence ADMET properties and drug-likeness in thiazolidinone derivatives.

III.4.1.Impact of Electron-Donating Substituents (Group 1)

Molecules with electron-donating substituents (methyl, methoxy, amino groups) exhibit elevated HOMO energies (-6.095 to -5.210 eV) and higher LUMO energies (-2.911 to -0.299eV), resulting in moderate energy gaps (3.184 to 5.483 eV). These compounds demonstrate enhanced electron-donating capacity, making them susceptible to electrophilic attack and potentially exhibiting antioxidant properties.

From an ADMET perspective, Group 1 molecules show favorable profiles: LogP ranges from 1.62 to 2.75 (optimal for absorption), TPSA ranges from 45.61 TO 115.95Å² (generally

favorable), and MW ranges from 208.28 to 300.19 g/mol (well below the 500 Da threshold). All 10 molecules in this group comply with Lipinski's Rule of Five, with 8 achieving "Excellent" ADMET classification.

The dipole moments in Group 1 are moderate to high (1.595 to 6.128 D), reflecting the polar nature of electron-donating groups. This polarity facilitates hydrogen bonding with biological targets and improves aqueous solubility.

III.4.2. Impact of Electron-Withdrawing Substituents (Group 2)

Electron-withdrawing substituents (nitro, cyano, trifluoromethyl groups) lower both HOMO (-6.946 to -6.257 eV) and LUMO (-3.054 to -1.661 eV) energies, resulting in smaller energy gaps (2.20 to 4.664 eV). These molecules exhibit reduced electron-donating capacity but enhanced electron-accepting ability, providing greater stability against oxidation and enhanced electrophilic reactivity.

Group 2 molecules show more variable ADMET profiles. LogP values range from 1.23 to 4.31, with 3 molecules (M14, M19) exceeding the optimal range. TPSA values range from 45.61 to 196.34 Å (M12 anomalous), with several molecules showing elevated values suggesting reduced absorption. MW ranges from 224.24 to 596.65 g/mol, with M14 exceeding the 500 Da threshold.

Despite these challenges, 8 out of 10 Group 2 molecules achieve "Good" or "Excellent" ADMET classification. The electron-withdrawing nature of these substituents provides enhanced metabolic stability, an important consideration for in vivo efficacy.

III.4.3. Impact of Halogenated Substituents (Group 3)

Halogenated thiazolidinones (fluorine, chlorine, bromine, iodine) exhibit intermediate electronic properties (HOMO: -6.993 to -6.298 eV; LUMO: -3.054 to -0.082 eV; Gap: 3.608 to 6.513 eV). These molecules demonstrate balanced reactivity profiles suitable for diverse biological applications.

ADMET properties in Group 3 are generally favorable: LogP ranges from 1.66 to 3.29 (optimal), TPSA ranges from 37.33 to 86.49 Å (favorable for absorption), and MW ranges from 192.24 to 413.13 g/mol (well below threshold). All 10 Group 3 molecules comply with Lipinski criteria, with 9 achieving "Excellent" or "Good" classification.

Halogenation provides strategic advantages: fluorine increases metabolic stability and bioavailability; chlorine and bromine enhance lipophilicity for membrane permeability; iodine can be used for radioactive labeling in diagnostic applications.

III.4.4. Impact of Heterocyclic and Polycyclic Substituents (Group 4)

Heterocyclic and polycyclic substituents introduce structural complexity (HOMO: -6.754 to -5.564 eV; LUMO: -3.166 to -0.960 eV; Gap: 2.971 to 6.190 eV). These molecules show the greatest variability in electronic properties, reflecting their diverse structural nature.

ADMET profiles show mixed results: LogP ranges from 0.91 to 3.65 (mostly optimal), TPSA ranges from 54.40 to 178.57 Å² (with M35 problematic), and MW ranges from 103.14 to 529.54 g/mol (with M35 exceeding threshold). Nine out of 10 Group 4 molecules comply with Lipinski criteria, with 8 achieving “Excellent” or “Good” classification.

III.4.5. Quantitative SAR Analysis

Correlation analysis between substituent properties and ADMET outcomes reveals:

- **Lipophilicity (LogP)** is inversely correlated with TPSA ($r = -0.47$), reflecting the fundamental trade-off between lipophilicity and hydrophilicity. Optimal drug candidates maintain LogP between 1.5-3.0 while keeping TPSA below 130 Å².
- **Molecular Weight** shows strong positive correlation with Rotatable_Bonds ($r = 0.87$), indicating that larger molecules tend to have more flexible structures. This flexibility can impact metabolic stability.
- **Polar Surface Area** correlates strongly with HBA ($r = 0.67$), as expected from structural principle. Molecules with TPSA > 140 Å² show significantly reduced oral bioavailability.
- **Electronic Gap** shows weak correlation with ADMET properties ($r < 0.20$), suggesting that electronic stability alone does not determine drug-likeness. However, molecules with Gap < 3.0 eV show enhanced reactivity and potential metabolic liability.

III.5. Discussion and Interpretation

III.5.1. Structural Determinants of ADMET Properties

The comprehensive analysis of 40 thiazolidinone derivatives reveals that molecular properties are determined by a complex interplay of structural, electronic, and physicochemical factors. The strong correlation between LUMO and Gap energies ($r = 0.90$) confirms that

electronic properties are interdependent, while the moderate correlations between TPSA, MW, and HBA ($r=0.61-0.67$) indicate that molecular size and polarity are independently variable. The high compliance rate with Lipinski's Rule of Five (85%) demonstrates that the thiazolidinone scaffold is inherently favorable for drug development. This favorable profile likely results from the balanced electronic properties of the thiazolidine core, which combines aromatic stability with functional group accessibility.

III.5.2. Predictive Model Insights

Machine learning models achieved excellent predictive performance ($R^2=0.82-0.87$ for regression; 85-88% accuracy for classification), indicating that ADMET properties can be reliably predicted from molecular descriptors. The feature importance analysis identified MW, TPSA, and Gap as the primary determinants, accounting for 70% of predictive power.

The superior performance of ensemble methods (Random Forest, Gradient Boosting) over linear models suggests that ADMET properties exhibit non-linear relationships with molecular descriptors. This finding has important implications for drug design: simple linear optimization of individual properties may not yield optimal compounds, and multiobjective optimization considering non-linear interactions is necessary.

III.5.3. Comparison with Literature

The observed ADMET profiles align well with literature precedents for thiazolidinone derivatives. Thiazolidinones have been extensively studied as antidiabetic agents, with compounds like pioglitazone and rosiglitazone showing LogP values (2.5-3.5) and values (50-80 μ) consistent with the present dataset. The favorable Lipinski compliance observed here suggests that the current series maintains the favorable properties of clinically successful thiazolidinones.

Electronic properties (HOMO, LUMO, Gap) show reasonable agreement with literature values for similar compounds. The energy gaps observed (2.87-6.52eV) fall within the expected range for aromatic heterocycles, confirming the reliability of the DFT calculations.

III.5.4. Implications for Drug Design

Several important design principles emerge from this analysis:

1. **Balanced Substitution Strategy:** The superior ADMET profiles of Group 3 (halogenated) and Group 4 (heterocyclic) molecules suggest that balanced

substitution_combining electronic effects with structural diversity_yields optimal drug candidates. Extreme electron-donating or electron-withdrawing groups may compromise ADMET properties.

2. **Molecular Weight Optimization:** The strong correlation between MW and rotatable bonds ($r=0.87$) indicates that molecular flexibility increases with size. Optimal drug candidates should maintain MW < 400 g/mol to minimize metabolic liability while preserving flexibility for receptor binding.
3. **Polarity Optimization:** The inverse correlation between LogP and TPSA ($r=-0.47$) highlights the lipophilicity-hydrophilicity trade-off. Compounds with LogP=2.0-2.5 and TPSA = 60-90Å show optimal absorption profiles.
4. **Electronic Stability:** While electronic gap (Gap_eV) shows weak correlation with ADMET properties, molecules with Gap > 3.5 eV demonstrate enhanced metabolic stability. This suggests that electronic stability should be considered alongside physicochemical properties in compound selection.

III.5.5.Limitations and Future Directions

The present analysis is limited by the relatively small dataset size (40 molecules) and the presence of 6 problematic molecules with data quality issues. Future studies should expand the dataset and implement rigorous quality control procedures to ensure reliable descriptor calculations. Additionally, the current analysis focuses on in silico ADMET prediction. Experimental validation through in vitro absorption, distribution, metabolism, and excretion studies would provide crucial confirmation of the predicted properties. Cell-based assays and animal models would further validate the predictive models developed here.

The machine learning models developed in this study provide a foundation for rapid virtual screening of novel thiazolidinone derivatives. Integration of these models with structure-based design approaches could accelerate the identification of lead compounds with optimal ADMET properties and biological activity.

III.5.6.Conclusions

This comprehensive analysis of 40 thiazidinone derivatives reveals that the thiazolisinone scaffold provides an excellent foundation for drug development, with 85% of molecules complying with Lipinski's Rule of Five. Machine learning models successfully predict ADMET properties with high accuracy ($R^2=0.82-0.87$), identifying molecular weight topological polar surface area, and electronic gap as the primary determinants of drug-likeness.

Structure-activity relationship analysis demonstrates that balanced substitution strategies yield optimal compounds, with halogenated and heterocyclic derivatives showing superior ADMET profiles. The insights gained from this analysis provide valuable guidance for future optimization of thiazolidinone derivatives toward clinical development.

Summary of key Finding

1. Descriptor Analysis:

- 40thiazolidinone comprehensively characterized with quantum and physicochemical descriptors
- HOMO:-8.413 to 6.174 eV; LUMO:-3.166 to 0.667 eV; Gap:2.874 to 9.081 Ev
- LogP:-26.80 to 4.31; TPSA:37.33 to 196.34Å; MW:103.14 to 596.65 g/mol

2. Correlation Insights:

- Strong correlations: LUMO-Gap ($r=0.90$), MW-Rotatable_Bonds ($r=0.87$), TPSA-HBA ($r=0.67$)
- Inverse correlation: LogP-TPSA ($r=-0.47$)
- Weak correlation suggest independent molecular properties

3. Drug-Likeness Assessment:

- 85% Lipinski compliance(34/40 molecules)
- 6 non-compliant molecules identified with specific violations
- Thiazolidinone scaffold inherently favorable for drug development

4. Machine Learning Performance

- Regression: $R^2=0.82-0.87$ (LogP, TPSA, MW prediction)
- Classification: 85-88% accuracy (ADMET category prediction)
- Feature importance: MW (28%), TPSA (24%), Gap (18%)
- Random Forest model demonstrates superior performance for absorption potential prediction

5. Structure-Activity Relationships:

- Group 1 (Donors): Enhanced reactivity, 80% Excellent ADMET
- Group 2 (Acceptors): Enhanced stability, 80% Good/ Excellent ADMET
- Group 3 (Halogens): Balanced properties, 90% Good/Excellent ADMET
- Group 4 (Heterocyclics): Variable properties, 80% Good/Excellent ADMET

6. Design Recommendations

- Maintain MW < 400 g/mol for optimal flexibility

- Target LogP = 2.0-2.5 and TPSA = 60-90 Å²
- Prefer balanced substitution over extreme electronic effects
- Prioritize Gap > 3.5 eV for metabolic stability
- Utilize Random Forest models for virtual screening and lead optimization

7. AI Implementation Impact:

- Reproducible Python-based workflow enables community adoption
- Cloud-based implementation (Google Colab) democratizes access to computational methods
- Predictive models accelerate drug discovery and guide rational design
- Framework extensible to other ADMET properties and molecular classes

General Conclusion

General Conclusion

Our study embarked on a comprehensive computational investigation into a series of thiazolidinone derivatives, aiming to establish reliable relationships between their molecular structure, quantum chemical descriptors, and ADMET properties. In an era where medicinal chemistry is increasingly driven by advanced molecular modeling and artificial intelligence, the early and accurate prediction of ADMET profiles is paramount for accelerating drug discovery and mitigating the high attrition rates associated with drug development. Thiazolidinones, with their diverse biological activities and structural versatility, represent a promising class of compounds for therapeutic innovation.

Our methodology integrated state-of-the-art computational techniques. Density Functional Theory (DFT) calculations were meticulously employed to derive crucial quantum chemical descriptors, including HOMO, LUMO, and energy gap, providing insights into the electronic structure and reactivity of the thiazolidinone derivatives. Concurrently, key physicochemical descriptors such as LogP and TPSA were calculated, offering a comprehensive molecular profile. The core of this work involved the application of Quantitative Structure-Activity Relationship (QSAR) modeling, enhanced by machine learning (ML) techniques, to predict the ADMET properties of these compounds.

Addressing the central research question of how an intelligent predictive model for thiazolidinone derivatives could be developed to establish reliable relationships between quantum descriptors and ADMET properties, our study successfully demonstrated the power of combining quantum chemical calculations (DFT) with advanced machine learning (Random Forest) for *in silico* ADMET prediction. The results obtained from our descriptor calculations provided a detailed understanding of the molecular characteristics of the 40 thiazolidinone derivatives. The correlation analysis revealed significant interdependencies among various descriptors, highlighting their collective influence on molecular behavior. The structural influence analysis, particularly concerning the impact of substituents on orbital energies and compliance with Lipinski's Rule of Five, offered valuable insights into the drug-likeness potential of these compounds.

Crucially, the developed AI model, based on a Random Forest algorithm, demonstrated a robust predictive capability for the absorption potential of the thiazolidinone derivatives, achieving an accuracy of approximately 85% through cross-validation. This model successfully identified molecular weight (MW) and Topological Polar Surface Area (TPSA) as the most influential descriptors in predicting absorption, underscoring their critical role in determining pharmacokinetic behavior. The interpretability of the model allowed for a nuanced understanding of how specific structural features contribute to the predicted ADMET profiles, thereby directly answering our problematic.

In conclusion, our work successfully demonstrated the efficacy of an integrated computational approach for the *in silico* prediction of ADMET properties of thiazolidinone derivatives. The developed model not only provides a valuable tool for screening potential drug candidates but also offers mechanistic insights into the structure-ADMET relationships. This approach significantly contributes to the rational design of novel compounds with enhanced therapeutic efficacy and reduced toxicity, thereby streamlining the early stages of drug discovery.

Perspectives for Future Work:

Building upon these findings, several avenues for future research emerge:

- **Expansion of Dataset and ADMET Endpoints:** Future studies could involve a larger and more diverse set of thiazolidinone derivatives, along with the prediction of a broader range of ADMET endpoints (e.g., BBB penetration, hERG inhibition, hepatotoxicity) to build more comprehensive predictive models.
- **Integration of Advanced ML Algorithms:** Exploring more sophisticated machine learning algorithms, such as deep learning models (e.g., Graph Neural Networks), could potentially enhance predictive accuracy and uncover more complex non-linear relationships.
- **Experimental Validation:** The *in silico* predictions generated by this model warrant experimental validation to confirm their accuracy and utility in a laboratory setting.
- **De Novo Design:** The insights gained from descriptor importance could be leveraged in *de novo* drug design strategies, guiding the synthesis of novel thiazolidinone derivatives with optimized ADMET profiles.
- **Multi-objective Optimization:** Developing models capable of simultaneously optimizing multiple ADMET properties and biological activities would be a significant step towards designing highly effective and safe drug candidates.

This research lays a solid foundation for the continued application of computational chemistry and artificial intelligence in the discovery and development of novel thiazolidinone-based therapeutic agents.