ALGERICAN DEMOCRATIC AND POPULAR REPUBLIC
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
YAHIA FARES UNIVERSITE OF MEDEA



Faculty of Technology
Department of Process and Environment Engineering

Master's Thesis

Submitted in fulfillment of the requirements for the Degree of Master in Pharmaceutical Process Engineering

THEME

SYNTHESIS AND MODELING OF NEW POLYCYCLIC COMPOUNDS AND EVALUATION OF THEIR ANTIBIOLOGICAL ACTIVITY

Proposed & Supervised by:
Dr. LETTREUCH Hichem
Dr. MADANI Achwak

Prepared & presented by:
Miss. ATCHI Chahinez
Miss. MAHDI Manar

Promotion 2023/2024

Thanks

Firstly, we want to thank Allah who gave us the patience and support to achieve our goal.

Words couldn't help us expressing our deep appreciation to Dr. LETTREUCH HICHEM, Professor at YAHIA FARES UNIVERSITY OF MEDEA , for his suggesting the research project and his patience, faithful advice and unlimited efforts in this research. His guidance, recommendations, and unwavering support during our work.

we wanted to take a moment to express our heartfelt gratitude for everything you have done Dr. MADANI Achwak Your support and guidance have been invaluable, and we truly appreciate your kindness and dedication

We express our gratitude to the personnel of the University of Yahia Fares of MEDEA ,particularly the director of the technological hall and the members of the chemical laboratories (Mrs.BELLARBI and Mrs. SAFAR REMALI and Mrs.BENHAOUA and Mrs.ALIKACEM).

We also would to express our gratitude to the jury members : Professor BENKORTBI O. and Professor FERHAT S. and Dr.BELLAISSA Y. for accepting to judge our work, which is an honor for us.

In conclusion, we would like to express our deep gratitude to all the employers and administrators of our university.

DEDICATION

With all my feelings of respect, with the experience of my gratitude, I dedicate my graduation and my happiness

To my dear grandparents MANI and PAPA ALI (رحمة الله عليهما)

To my paradise, the spark of my eyes, the source of my happiness, my moon and the thread of hope that light my way, my half, Mom.

To him who made me a woman, my source of life, love and affection, my support who was always by my side to support and encourage me, to my great Dad.

To my dear brother IMED EDDINE who have never left my side and wishes me all the best

To my gorgeous sister SAFAA my soulmate who have not ceased to encourage and support me throughout my studies.

To my little princess my adorable little sister MALEK who always knows how to bring joy and happiness to the whole family.

To all the members of my big family, my aunts, uncles and cousins.

To my partner and my best friend MANAR for his moral support his patience and understanding throughout this project.

To my best friends, my companions M, N, M, M and L whose are always by my side, whose make me laugh louder and smile brighter

To all my colleagues of promotion PHARMACEUTICAL PROCESS ENGENEERING.

And finally, to all whose wish me only the best and to everyone who loves me.

CHAHINEZ

DEDICATION

to my myself, thank you for your resilience, dedication, and hard work. Your perseverance and belief in your dreams have brought you to this moment. Thank you for the late nights, and the unwavering commitment to your goals.

To my dear parents source of joy and happiness, Thank you for being my constant pillars of strength and sources of inspiration. Your unwavering support, love, and encouragement have been the driving force behind all my endeavors. words can not express my gratitude and appreciation for your dedication and sacrifices, you are always by my side to support me.

To my dear beloved, Mani Zahia (رحمة الله عليها)

To my beloved brother, I appreciate your unwavering support

To my dear sisters, I thank you for your continued support. I'm lucky to have you by my side Particularly Loubna,Chahinez,Naziha,Manel,Marwa.

To my friends H and L , thank you for every thing .
To Chahinez, dear best friend before being a partner

To all those who participated directly or indirectly in my success..

To you dear reader .

MANAR

ملخص

تم تصنيع و تعريف- 3,1-ثنائي هيدرازينو-6,4 دينيتروبنزن.
تكثيف المركب السابق مع بيروفات الميثيل اعطي ثنائي 2-نيتروفينيل هيدرازونو بيروفات المثيل والذي يمكن أن يتواجد على هيئة مماكبين هندسيين *(E/E)* و *(Z/Z)*.
تكثيف ثنائي الهيدرازين السابق مع ثنائي ميثيل أوكسالوسيتات أعطى ثنائي هيدرازون الموافق.
ارجاع ثنائي الهيدرازونات المركبة اعطى نواتج تحليلها الاولي بين إمكانية تشكل حلقات.
تم تحليل المركبات المتحصل عليها باستعمال المطيافية تحت الحمراء و نقطة الذوبان.

وأخيرا، تم التنبؤ بالنشاط البيولوجي لهذه المنتجات باستخدام نماذج QSAR. وأظهرت النتائج أن النموذج الذي تم بناؤه قد تم تقييمه ومقارنته باستخدام معايير التحقق الإحصائية المختلفة. أظهر النموذج القائم قدرة تنبؤية عالية بمعايير RMSEP(0.4147, 0.93, 0.54) R^2(0.9715,0.9003,0.9509). أظهرت دراسة نمذجة النشاط البيولوجي (MIC) أن المنتجات الاصطناعية تحتوي على MIC يبلغ 0.3 و 0.8 ملغم / م.

الكلمات المفتاحية: التكثيف، الارجاع، مكرر هيدرازون ,النمذجة

.

Abstract

The 1, 3-dihydrazino-4,6-dinitrobenzene has been synthesized and characterized.

The condensation of the latter with the methyl pyruvate to give the bis2-nitrophénylhydrazono-methyl pyruvate that can exist in two forms diastereoisomer (E/E) and (Z/Z).

Condensation of previous bishydrazine with dimethyl oxaloacetate also leads to corresponding bishydrazones.

The reduction of bishydrazones synthesized leads to these products whose preliminary characteristics show that they have a reduction, but remains to confirm the final structure.

The synthesized products were characterized by melting point and TIRF.

Finally, the prediction of the biological activity of these products was carried out using QSAR models. The results showed that the constructed model was evaluated and compared using different statistical criteria for validation. The established model demonstrated great predictive capacity with criteria of R^2(0.9715 ,0.9003 ,0.9509) RMSEP(0.4147, 0.93 , 0.54). The Biological Activity Modeling Study (MIC) showed that synthetic products have an MIC of 0.3 and 0.8 mg/m.

Keywords: condensation, reduction, bishydrazone. modeling.

Résumé

Le 1, 3-dihydrazino-4,6-dinitrobenzène a été synthétiser et caractériser.

La condensation de ce dernier avec le pyruvate de méthyle à donner le bis 2-nitrophénylhydrazono pyruvate de méthyle qui peut exister sous deux formes diastéréoisomère (E/E) et (Z/Z).

La condensation de bishydrazine précédent avec l'oxaloacétate de diméthyle conduit aussi aux bishydrazones correspondantes.

La réduction des bishydrazones synthétiser à mener à ces produits dont la caractérisation préliminaire à montrer qu'il à eux une réduction, mais reste à confirmer la structure finale.

Les produits synthétisés ont été caractérisés par point de fusion et FTIR.

Enfin, la prédiction de l'activité biologique de ces produits a été réalisée à l'aide de modèles QSAR. Les résultats ont montré que le modèle construit a été évalué et comparé en utilisant différents critères statistiques de validation. Le modèle établi a démontré une grande capacité prédictive avec les critères de R2(0,9715,0,9003,0,9509) RMSEP(0,4147, 0,93, 0,54). L'étude de modélisation de l'activité biologique (CMI) a montré que les produits synthétiques ont une CMI de 0,3 et 0,8 mg/m.

Mots clés: condensation, réduction, bishydrazone. modelisation

Table of Content

General Introduction……...…01

Chapter I: Bibliography Part

Part one: Synthesis

I.1. Introduction……..04

I.2. Methods for reducing nitro heterocyclics……04

I.2.1. Non-catalytic reduction of nitro aromatics……04

I.2.1.1. Metal hydride reductions……04

I.2.1.2. The Béchamp reduction……..05

I.2.2. Catalytic reduction of nitro aromatics……...05

I.2.2.1. Catalytic hydrogenation……..05

I.2.2.2 The sodium dithionite reduction……..06

I.3. Synthesis of nitrogenous heterocycles by reductive cyclization……..07

I.3.1. Synthesis of 5-membered nitro heterocycles ……...07

I.3.1.a. Biological activity……...10

I.3.2. Synthesis of 6-membered nitro heterocycles ……...10

I.3.2.a. Biological activity……...13

I.3.3. Synthesis of 7-membered nitro heterocycles ……...13

I.3.3.a. Biological activity……...16

I.3.4. Bis–Heterocyclic……...16

I.3.4.a. Biological activity ……..17

Part two: Modeling

I.1. Definition ……18

I.2. The history ……...18

I.3. Molecular descriptors……...19

I.4. Selection of descriptors ……...20

I.4.a. Objective selection……20

I.4.b. Subjective selection ……...21

I.5. AI-based QSAR methods…….. ……..21

I.5.1. Artificial intelligence (AI)……21

I.5.2. Machine learning……22
I.5.2.1. Support vector machine(SVM)……...23
I.6. Support Vector Regression (SVR)……...25
I.6.1. Linear problem……...26
I.6.2. Nonlinear problem……..26
I.7. validation criteria ……27
I.7.a. Correlation coefficient (R)……27
I.7.b. Coefcient of determination (R^2)……28
I.7.c. Root mean square error (RMSE)……28
I.7.d .Coefficient of prediction (Q^2) ……...29
I.7.e. Mean Absolute Error (MAE)……..30
I.7.f. The Concordance Correlation Coefficient (CCC)……30
I.8. Importance of QSAR ……...31

Chapter II: Experimental Part

Analytical equipment and techniques……34
The products used……...34
II.1. Introduction……35
II.2. Synthesis of starting compound……35
II.2.1. Synthesis of 1, 3-dihydrazino-4,6-dinitrobenzene……..35
II.3. Synthesis of Bishydrazones……35
II.3.1. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate……...35
II.3.1.1. Synthesis of (*E/E*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate……...36
II.3.1.2. Synthesis of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate……...36
II.3.2. Synthesis of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl…….37
II.4. Reductions……...37
II.4.1. Reduction of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate by sodium dithionite……37
II.4.2. Reduction of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl by sodium dithionite……38
II.5. Evaluation of the antimicrobial activity of synthesized Heterocycles using QSAR model……..38

II.5.1. Database collection……...39
II.5.2. Descriptor selection……..46
II.5.3. Database division and model development……...48
II.5.3.1. SVM model……48
II.5.3.2. Validation……...48
II.5.3.2.a. Internal validation……...48
II.5.3.2.b. External validation……..48

Chapter III: Results and discussions

III.1. Introduction……...50
III.2. Synthesis of starting compound…….50
III.2.1. Synthesis of 1, 3-dihydrazino-4,6-dinitrobenzene……..50
III.3. Synthesis of Bishydrazones ……..51
III.3.1. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate…….51
III.3.2. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl……54
III.4. Reductions…….57
III.4.1. Reduction of (Z)/(Z)-1,3-dihydrazono-4,6-dinitrophenylen-methyl pyruvate by sodium dithionite…….57
III.4.2. Reduction of (Z)/(Z)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl by sodium dithionite……60
III.5. Evaluation of the antimicrobial activity of synthesized Heterocycles using QSAR model……..63
III.5.1. Selection of representative descriptors……...63
III.5.2. QSAR model development (SVM)…….70
III.5.2.a. SVR model……..70
III.5.2.b. Internal validation……..72
III.5.2.c. External validation…….73
III.5.2. External dataset……..75
General conclusion…….78
References bibliography

List of Abbreviations

Ac : acetyl

AI: artificial intelligence

C: coal

CCC : The concordance correlation coefficient

Hcl :hydrochloric acid

IR: Infrared.

LMO : Leave-Many-Out

LOO: Leave-One-Out

M: Molar mass (in g/mol).

MAE : Mean absolute error

MIC: Minimum Inhibitory Concentration.

ML : machine learning

mmHg : Millimeter of mercury.

mp: Melting point.

MW : molecular weight

Ni : nickel

Pd : Palladium.

Pt: platinum

Q^2 : Coefficient of prediction

QSAR: Quantitative structure-activity relationships

QSPR: Quantitative structure-property relationships

R : Correlation coefficient

R^2 : Coefficient of determination

RMSE : Root mean square error

SV : Support Vector

SVM: Support Vector Machines

SVR: Support Vector Regression

3D : three dimensional

2D : two dimensional

List of schemes

Chapter I: Bibliography part

Scheme I.1: The reduction of nitrobenzene using lithium hydrides($LiAlH_4$) ……..04

Scheme I.2: The reduction of nitrophenol using Boron hydrides ($NaBH_4$) ……05

Scheme I.3: The reduction of nitrobenzene using bechamp reduction…….05

Scheme I.4: The catalytic hydrogenation of 2-nitro-1-chlorobenzene…….06

Scheme I.5: The synthesis of 3-aminopthalhydrazine from 3-nitropthalhydrazine…….06

Scheme I.6: The synthesis of new derivatives of benzimidazole…….07

Scheme I.7: The synthesis of 2 methyl-1H-pyrrolo[2,3-c] pyridine…….....07

Scheme I.8: The synthesis of 2-amino-1H-pyrrolo [3,2-b] Ehyle pyridine -3-carboxylate…….08

Scheme I.9: The synthesis of 1,3-dimethyl-1H-pyrrolo [3,2-d]pyrimidine -2,4(3H,5H)-dione…...08

Scheme I.10: The synthesis of 1,3-dimethyl-1H-pyrrolo [3,2-d]pyrimidine -2,4(3H,5H)-dione….09

Scheme I.11: The synthesis of 7-nitropyride[1,2-a]benzimidazole……..09

Scheme I.12: The synthesis of benzimidazole……..09

Scheme I.13: The synthesis of oxygenated 2-methylindolines……....10

Scheme I.14: The synthesis of 3,4-dihydroquinolin-2(1*H*)-one derivatives….....11

Scheme I.15: The synthesis of 1,2,4-benzotriazine……..11

Scheme I.16: The synthesis of pyridazino [1,6-a] benzimidazole derivative……...11

Scheme I.17: The synthesis of 3-methyl-4-methylidene-1,4-dihydroquinoline……...12

Scheme I.18: The synthesis of 1, 10-substituted benzo[*c*]cinnoline……12

Scheme I.19: The synthesis of 7-methyl-benzimidazo[1,2-c]quinazoline……...12

Scheme I.20: The synthesis of (2*E*)-3-(4-methoxyphenyl)-1-(2-nitrophenyl)prop-2-en-1-one……13

Scheme I.21: The synthesis of Tetrahydroquinoline acetamide……...13

Scheme I.22: The synthesis of 2-oxo-1,2,3,5-tetrahydro-4*H*-[1,4]diazepino [5,6-b]quinolone-4 carboxamide……..14

Scheme I.23: The synthesis of The pyrrolo[2,1-c] [1,4] benzodiazepine……....14

Scheme I.24: The synthesis of 7-chloro-1-methyl-5-phenyl-1H-1,5-benzodiazepine-2,4(3H,5H)-dione……..15

Scheme I.25: The synthesis of derivative 2R-hydroxy-1, 2, 3, 10, 11, 11as-hexahydro-5H-pyrrolo [2,1-c] [1,4] benzodiazepine- 5,11-dione……..15
Scheme I.26: The synthesis of 3-[(2-nitrophenyl)amino]propanehydrazide……16

Chapter II: Experimental part

Scheme II.1: Structure of 1, 3-dihydrazino-4,6-dinitrobenzen ……35
Scheme II.2: The structure of (*E/E*)-1,3-dihydrazono-4,6-dinitrophenyl-methyl pyruvate ……36
Scheme II.3: the structure of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenyl-methyl pyruvate …….....36
Scheme II.4: The structure of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl…….37

Chapter II: Results and discussion

Scheme III.1: Reaction mechanism of the synthesis of 1,3-dihydrazino-4,6-dinitrobenzen……...50
Scheme III.2: Reaction mechanism of synthesis of isomers (*E/E*)- and (*Z/Z*)- 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate……52
Scheme III.3: Reaction mechanism of synthesis of isomers (*E/E*) and (*Z/Z*)- 1,3-dihydrazono -4,6-dinitrophenylenoxaloacetate dimethyl……...55
Scheme III.4: Possibility of reductive cycling of the compound (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate ……57
Scheme III.5: Reaction mechanism of the possible reductive cycling of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenpyruvate of methyl leading to the triazinio benzotriazine derivative…….58
Scheme III.6: Reaction mechanism of possible reductive cycling of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenpyruvate of methyl leading to the triazépino benzotriazepine derivative……59
Scheme III.7: Possibility of reductive cycling of the compound (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl……...61
Scheme III.8: Reaction mechanism of the possible reductive cycling of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl leading to the triazinio benzotriazine derivative……..62
Scheme III.9: Reaction mechanism of the possible reductive cycling of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl leading to the triazépino benzotriazepine derivative…..…..63

List of figures

Chapter I: Bibliography part

Figure I.1: Bis-pyrazolo [1,5-a] pyrimidine derivative ……17
Figure I.2: Representation of the relationship between AI, ML and DL ……22
Figure I.3: Flow chart of machine learning and the algorithms used for training the machine ……..23
Figure I.4: Defining the margin between classes ……24
Figure I.5: ε radius insensitivity tube……...24
Figure I.6: Linear and nonlinear SVMs ……...…25
Figure I.7: Root mean square error ……..29

Chapter II: Experimental part

Figure II.1: Modeling methodology…….....39
Figure II.2: Screenshot PubChem ……40
Figure II.3: Screenshot ChemSketch ……...40
Figure II.4: Screenshot PaDEL …….....47
Figure II.5: Screenshot Stepwise MLR……47

Chapter III: Results and discussions

Figure III.1: IR spectrum of 1,3-dihydrazino-4,6-dinitrobenzene……51
Figure III.2: IR spectrum of (*E/E*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate……..53
Figure III.3: IR spectrum of (*Z/Z*)- 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate…….54
Figure III.4: IR spectrum of (Z/Z)-1,3-hydrazono-4,6- dinitrophenylenoxaloacetate of Dimethyl…….56
Figure III.5:IR spectrum of resulting product from reduction 01……60
Figure III.6: Correlation diagram ……67
Figure III.7: Bubble Cloud of the Descriptors's importance ……...68
Figure III.8: Word Cloud of the Descriptors's importance ……68
Figure III.9: Bar plot of the relative importance of descriptors ……...69
Figure III.10: The Training data……...73
Figure III.11: Test data……74
Figure III.12: SVM regression curves……..75

List of tables

Chapter II: Experimental part

Table II.1: The products used……34

Table II.2: Database collection……..41

Chapter III: Results and discussions

Table III.1: Relevant descriptors……...65

Table III.2: Selected descriptors……...66

Table III.3: Statistical criteria for the SVR model……70

Table III.4: : Representative descriptors of synthesized molecules……..76

List of Symbols

b : biais

C : penalty

F(x): output

k : Constant

n : the number of observations in the dataset.

***nx* :** Molecular descriptors

w : weight

x : input

x^- , y^- : the means of the x and y variables

xi , yi : the individual data points.

y_i : the true value.

$\hat{y}$_i : the predicted value.

ε : Epselon

***α* , *αi**:** Lagrange multipliers

σ : Radial basis function width

List of equations

Equation (I.1): Simple QSAR model……18
Equation (I.2): Linear kernel…….. 26
Equation (I.3): Nonlinear kernel……...27
Equation (I.4): The RBF Kernel ……..……..27
Equation (I.5): Correlation coefficient ……27
Equation (I.6): Coefficient of determination……28
Equation (I.7): Root mean square error……29
Equation (I.8): Coefficient of prediction……..29
Equation (I.9): Mean absolute error……..30
Equation (I.10): The concordance correlation coefficient……30

General introduction

General Introduction

Heterocycle chemistry, an important area of organic chemistry, studies chemical compounds whose cyclic carbon chain contains one or more atoms other than carbon (heteroatoms). The elements nitrogen (N), oxygen (O), and sulfur (S) are the most common heteroatoms. The presence of a heteroatom specifically impacts the properties of these compounds [1] [2].

Since heterocycles comprise more than half of all known organic molecules, their biological activity and diversity make them extremely significant. They have applicability in various industries, particularly in the pharmaceutical sector, as heterocyclic structures are found in many medications and physiologically active substances. For instance, morphine, a potent analgesic, contains the heterocyclic ring morphinan [3].

Heterocyclic compounds can be synthesized by various techniques, such as rearrangement processes, cyclization, and condensation. These synthesis techniques may vary depending on the particular structure of the target heterocycle. In recent years, the biological significance and structural variety of nitrogen-containing heterocycles have made them desirable targets for synthesis. A thorough understanding of the properties and activities of these compounds is crucial for their classification and utilization [4].

However, critical experimental characterization is challenging due to time constraints, expense, the riskiness of some tests, or ethical concerns (such as limitations on animal experiments). Integrating Quantitative Structure-Activity Relationship (QSAR) with heterocycle synthesis involves using computational methods to predict the activity of synthesized heterocycles before actual synthesis . This approach can aid in the design and optimization of new compounds with desired biological activities. This integration allows medicinal chemists to obtain valuable information for the development of new therapeutic agents and to understand the underlying mechanisms. Thus, it facilitates the design and prioritization of the synthesis of novel heterocyclic compounds with potentially desirable biological activities, thereby saving time and resources in the drug discovery process [5].

So the objective of our work is:

- Synthesis and characterization of 1,3-dihydrazino-4,6-dinitrobenzen.
- synthesis and characterization of (*E/E*) and (*Z/Z*)-2-ntrophénylhydrazone methyl pyruvate.

- Synthesis and characterization of dimethyl oxaloacetate (*Z*/*Z*)-2-ntrophenylhydrazone.
- Trial to reduce bishydrazones by excess sodium dithionte.
- Predict unknown biological activity.

Our research is divided into three main parts:

After a general introduction, the first part of this work is devoted to the theoretical study, which includes methods of synthesis of nitrogen heterocycles, as well as examples of the preparation of 5, 6, and 7 membered heterocycles and their biological properties as described in the literature. The second part comprises an experimental section that details the equipment and products used, and describes the detailed experimental protocols employed to prepare the compounds. It also includes their physico-chemical characteristics and predictions of their biological activity.

The third part covers the interpretation and discussion of the results achieved, through the study of the reactive mechanisms involved, their analysis, and the results of biological activities obtained using the QSAR model based on AI. Finally, a general conclusion is provided.

Chapter I :
Bibliography part

I.1. Introduction

Nitrogen-containing heterocycles are a valuable class of compounds that are widely used in drug development and medicinal chemistry, which exhibit a large range of pharmacological activities . Numerous methodologies used to synthesize novel heterocyclic compounds that consist of significant biological activities.

The reduction of nitrogen-containing heterocyclic compounds is an important transformation that enables the synthesis of diverse nitrogen-based scaffolds with valuable medicinal properties.

I.2. Methods for reducing nitro heterocyclics

Different catalytic and non-catalytic approaches are explored for reducing the nitro group present on an aromatic ring using various processes.

I.2.1. Non-catalytic reduction of nitro aromatics

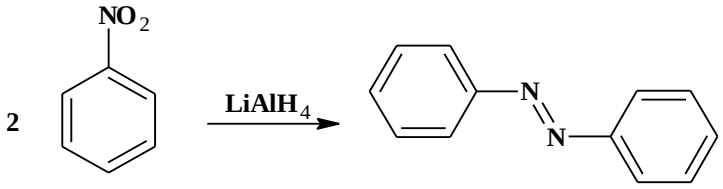
To reduce the nitro aromatics to aromatic amines . There are several non-catalytic processes. These processes often involve the use of metal hydrides such as $LiAlH_4$ or boron-based reagents or by using other processes. They are used to selectively reduce the nitro groups present in various heterocyclic compounds [6][7].

I.2.1.1. Metal hydride reductions

Certain metal hydride are synthetic equivalents of the hydride ion (H^-) synthon and as such are powerful reducing agents which react preferentially at electron-deficient centers.

- **The lithium process**

Is widely used in various synthetic pathways, $LiAlH_4$ is more strong reducing agent that can donate hydride ions to the nitro group, leading to the formation of the amine. This reaction is one of the common methods for reducing nitro groups to amines [8].

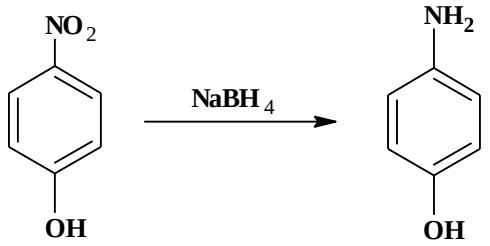


Schema I.1: The reduction of nitrobenzene using lithium hydrides ($LiAlH_4$) [9].

- **The Boron process**

Boron hydrides like sodium borohydride $NaBH_4$, or potassium borohydride $NaBH_4$, work as reducing agents. They are capable of donating hydride ions (H^-) to reduce nitro group.

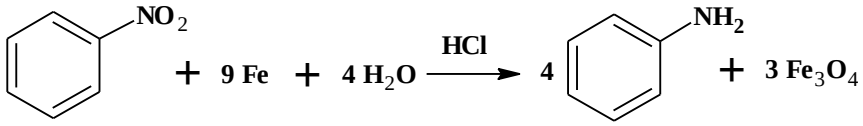
The selectivity and efficiency of boron hydrides make them valuable tools in synthetic chemistry for achieving specific reductions in complex molecules [10].



Schema I.2: The reduction of nitrophenol using Boron hydrides ($NaBH_4$).[4]

I.2.1.2. The Béchamp reduction

Also known as the Béchamp process, is a chemical reaction that converts aromatic nitro compounds to their corresponding anilines using iron as the reductant [11].



Schema I.3: The reduction of nitrobenzene using bechamp reduction.

This reaction was historically a major route to produce aniline, an important compound used in the synthesis of dyes, pharmaceuticals and other chemicals.

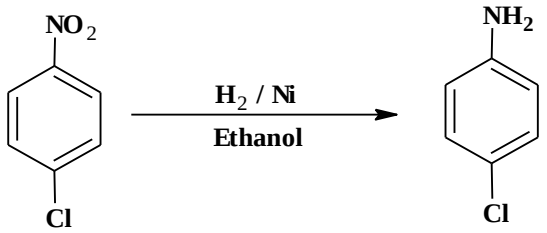
I.2.2. Catalytic reduction of nitro aromatics

I.2.2.1. Catalytic hydrogenation

Catalytic hydrogenation is a widely applicable technique for the reduction of organic compounds. Hydrogenation is a commonly used method for reducing nitrogen heterocyclic compounds. Usually, the reaction is carried out by stirring or shaking a solution of the compound, in presence of a heterogeneous catalyst, under an atmosphere of hydrogen.

Hydrogen gas is typically used in the presence of a catalyst to facilitate the reduction reaction. The catalyst, often a metal such as palladium (Pd) platinum (Pt) or nickel (Ni), helps in breaking the nitrogen-hydrogen bonds in the heterocyclic compound, leading to the formation of saturated nitrogen-containing [4].

Various catalysts can be employed to facilitate the reaction such as the metal derivatives like Pt , Pd , Ni, gold-based catalysts, sodium sulfide, sodium dithionite…ect

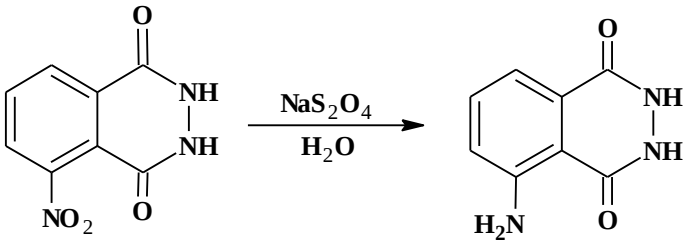


Schema I.4: Tha catalytic hydrogenation of 2-nitro-1-chlorobenzene[4].

I.2.2.2. The sodium dithionite reduction

Sodium dithionite also known as Sodium Hydrosulfite and Sodium Sulphoxylate, widely used as reducing agent in various chemical reactions because of its reducing ability. It can be employed to reduce various functional groups like Nitro groups, Carbonyl groups and Quinons.

Sodium dithionite is known for its ability to donate electrons and facilitate reduction processes. It participate in catalytic reduction reactions by transferring electrons to the nitro groups, leading to the formation of Amines [12].



Schema I.5: The synthesis of 3-aminopthalhydrazine from 3-nitropthalhydrazine.

sodium dithionite is an inexpensive and safe reducing agent, non-toxic, non-mutagenic and non-carcinogenic. Hence, it is used as an important reducing agent for its selectivity and rapidity in reaction times.

I.3. Synthesis of nitrogenous heterocycles by reductive cyclization

The reductive cyclization is commonly used in organic synthesis to construct various nitrogen-containing heterocycles efficiently.

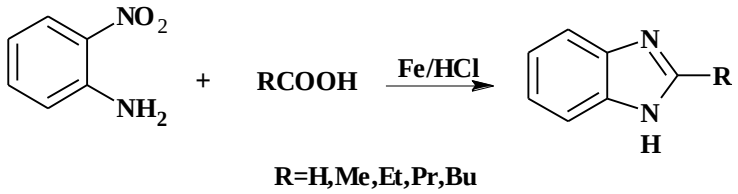
The reductive cyclization of nitrogen compounds offers an opportunity to synthesize nitrogen heterocycles of 5, 6, 7 members.

The formation of heterocycles of 5 and 6 members usually occurs spontaneously because of their stability and more favorable steric situation.

However, the formation of 7-members heterocycles does not occur spontaneously, they are more demanding .

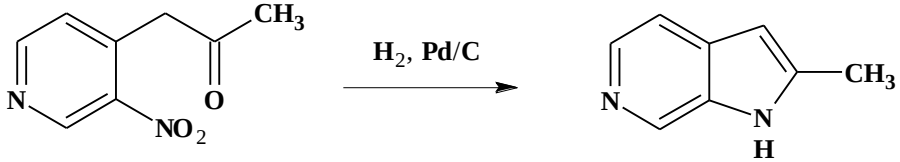
I.3.1. Synthesis of 5 membered nitro heterocycles :

Recently, new derivatives of benzimidazole with high biological activity have been obtained with variable yields, starting from orto-aniline and aliphate acids by reductive cycling with iron powder in hydrochloric acid. [13].



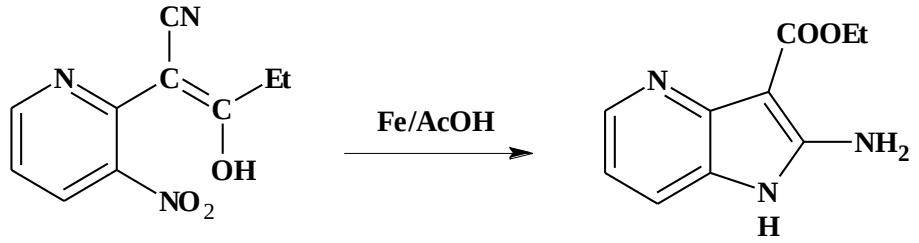
Scheme I.6: The synthesis of new derivatives of benzimidazole

2 methyl-1*H*-pyrrolo[2,3-c] pyridine is obtained by catalytic hydrogenation from (3-nitro-4-pyridyl) acetone. [14].



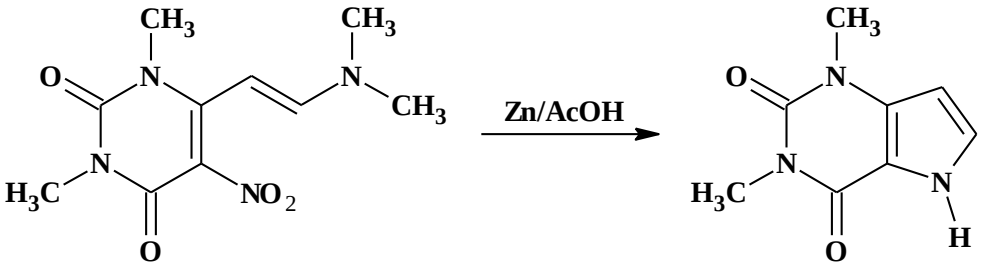
Scheme I.7: The synthesis of 2 methyl-1*H*-pyrrolo[2,3-c] pyridine

2-amino-1*H*-pyrrolo [3,2-b] Ehyle pyridine-3-carboxylate is obtained by reductive cyclization using iron powder in acid. [15]



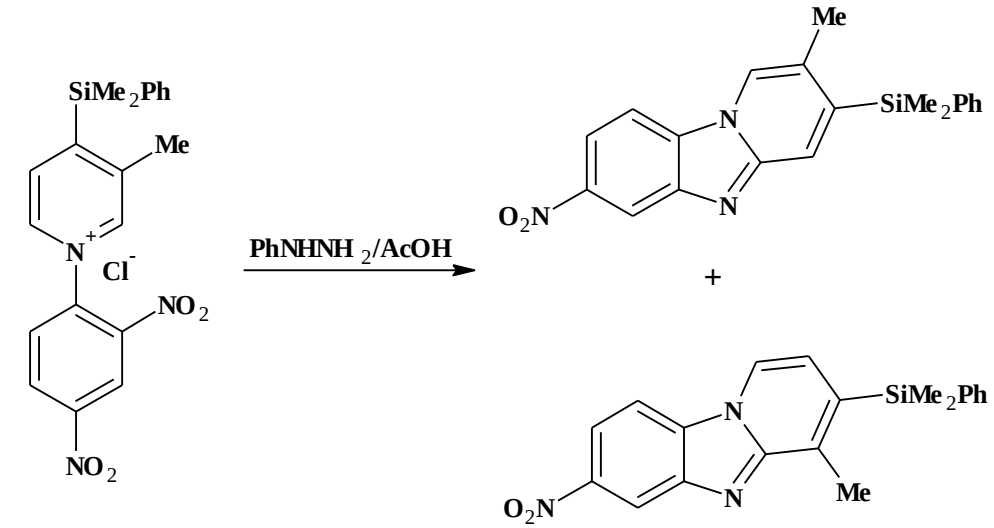
Scheme I.8: The synthesis of 2-amino-1*H*-pyrrolo [3,2-b] Ehyle pyridine -3-carboxylate

1,3-dimethyl-1*H*-pyrrolo [3,2-d]pyrimidine-2,4(3*H*,5*H*)-dione is obtained by reductive cycling with zinc powder in acetic acid
from 6-(2-dimethylamino)ethanyl-1,3-dimethyl-5-nitropyrimidine-2,4(1*H*,3*H*)-dione. [16].



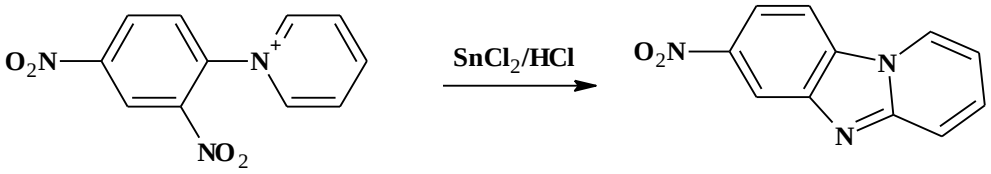
Scheme I.9: The synthesis of 1,3-dimethyl-1*H*-pyrrolo [3,2-d]pyrimidine -2,4(3*H*,5*H*)-dione

The reductive cyclication of N-(2,4-dinitrophenyl)-3-methyl-4-dimethylsilpyridium chloride with phenylhydrazine in acetic acid gives by selective reduction the 7-nitro-2- methyl and 7-nitro-4-methodyl-3-dimethylphenylsylpyrido[1,2-a]benzimidazoles. [17].



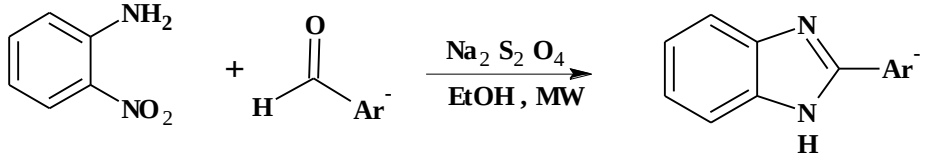
Scheme I.10: The synthesis of 1,3-dimethyl-1*H*-pyrrolo [3,2-d]pyrimidine -2,4(3*H*,5*H*)-dione

7-nitropyride[1,2-a]benzimidazole is obtained by reductive cycling from N-(2,4-dinitrophenyl)pyridinium chloride by selective reduction with stain chlorine in hydrochloric acid. [18].



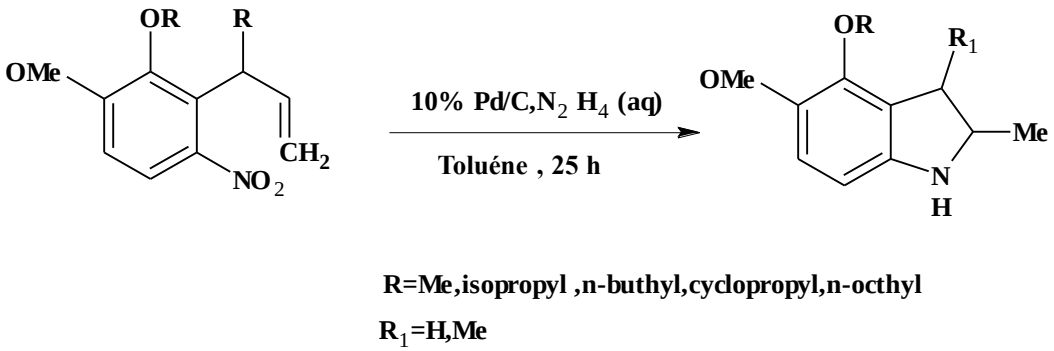
Scheme I.11: The synthesis of 7-nitropyride[1,2-a]benzimidazole

The synthesis of the benzimidazole substitute from O-nitroaniline and arylic aldehyde in the presence of sodium dithionite as a reducing agent under microwave irradiation is described.The products were obtained with excellent yields, short reaction times and soft reaction conditions. [19].



Scheme I.12: The synthesis of benzimidazole

By reducing 2-allylnitrobenzen and boiling an aqueous solution of hydrazine (80%) and toluene, a high-efficiency synthesis of oxygenated 2-methylindolines (84%) is achieved using Pd/C (10%) as a catalyst. [13].



R=Me,isopropyl ,n-buthyl,cyclopropyl,n-octhyl

R_1=H,Me

Scheme I.13: The synthesis of oxygenated 2-methylindolines

1.3.1.a. Biological activity

Imidazole derivatives have a wide range of pharmacological activities. It should be noted that imidazole and its derivatives have analgesic and anti-inflammatory activity, cardiovascular activity, anticancer activity, antineoplastic activity, antifungal activity, enzymatic inhibition activity, anti-hilmintic, antiviral and ulcerative activity.

Imidazole simple nitrate derivatives are effective as antibacterial agents. They are also useful in the treatment of infections caused by protozoa, such as Trichomonus [20].

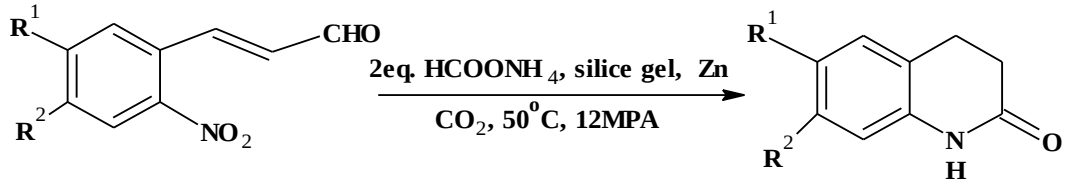
pGlu-Trp, having shown interesting anti-angiogenic activity.There are mainly two classical tests of the inhibitory activity of angiogenesis: that of the chorioallantoic membrane of the egg, and that of pseudo-tube structural formation by endothelial cells grown on a thick gel.[21]

benzyl ethers benzimidazole which have shown good antibacterial activity against S. aureus and antifungal activity against Candida albicans and Candida krusei as well as antiviral properties.[22]

I.3.2. Synthesis of 6 membered nitro heterocycles

Six-membered nitro heterocycles are a crucial and significant category among organic chemical compounds, They commonly exist as the essential backbone in diverse active medicinal natural products .

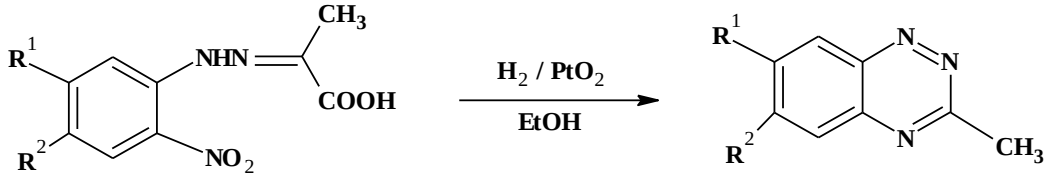
The 3,4-dihydroquinolin-2(1*H*)-one derivatives are synthesized from the derivatives of (2*E*)-3-(2-nitrophenyl)prop-2-enal in the presence of zinc powder [23].



R^1**: H, F** **R**^{2}**: H, OH, OMe**

Scheme I.14: The synthesis of 3,4-dihydroquinolin-2(1*H*)-one derivatives

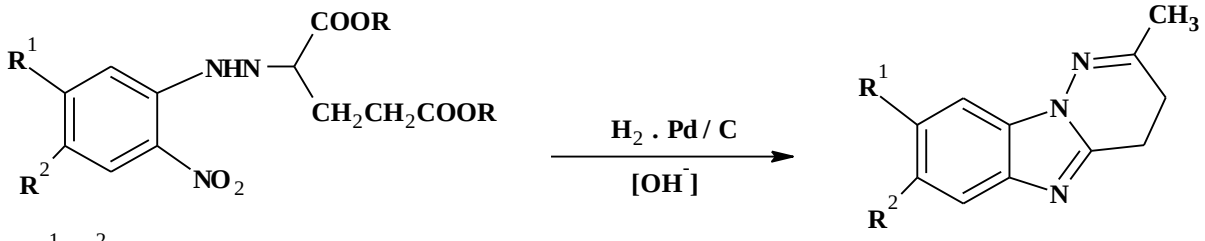
Synthesis of 1,2,4-benzotriazine by reductive cyclization [24].



R^1**: H, Cl, Br** **R**^{2}**: H, Cl, Br, NH**$_2$

Scheme I.15: The synthesis of 1,2,4-benzotriazine

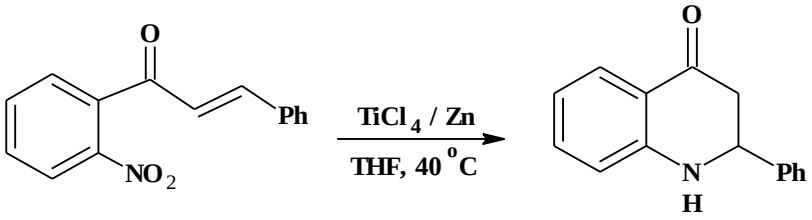
The catalytic hydrogenation of (*E*)-2-nitrophenylhydrazono dialkyl glutarate with further treatment of the reaction mixture leads to reductive cyclization, the obtained product is the pyridazino[1,6-a]benzimidazole derivative [25].



R^1**=R**^{2}**= H, Cl** **R: Me, Et**

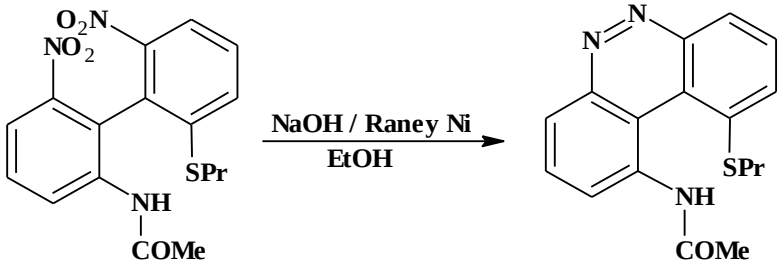
Scheme I.16: The synthesis of pyridazino[1,6-a]benzimidazole derivative

The synthesis of 2-phenylquinoline-4(1*H*)-one was achieved by reductive cyclisation of (2*E*)-1-(2-nitrophenyl)-3-phenylprop-2-to-1-one via TiCl4/Zn. This method has the advantage of the availability of raw materials [26].



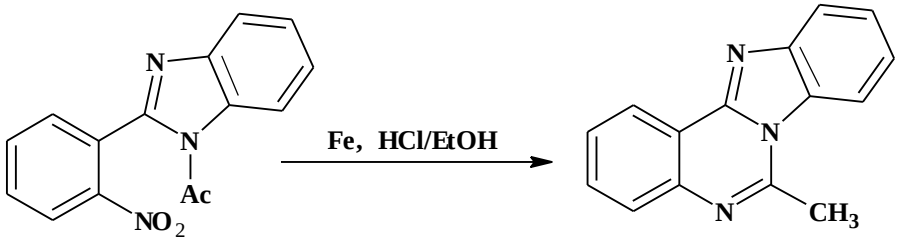
Schema I.17: The synthesis of 2-phenylquinoline-4(1*H*)-one

The 1,10-substituted benzo[*c*]cinnoline obtained by reductive cyclization using Nickel in boiling ethanol from substituted tetra diphenyls [27].



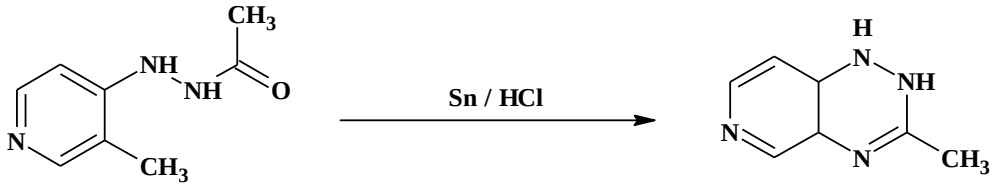
Schema I.18: The synthesis of 1,10-substituted benzo[*c*]cinnoline

The 7-methyl-benzimidazo[1,2-c]quinazoline obtained from 1-acetyl-2-(2-nitrophenyl)benzimidazole by reductive cyclization with iron in hydrochloric acid and ethanol [28].



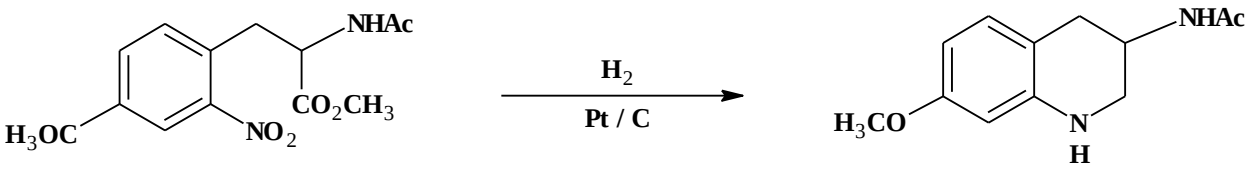
Schema I.19: The synthesis of 7-methyl-benzimidazo[1,2-c]quinazoline

The reduction of *N*-(3-nitropyridine-4-yl)acetohydrazide with tin in hydrochloric acid gives the 3-methyl-1,2-dihydropyride[3,4-e][1,2,4]triazine [29].



Schema I.20: The reduction of *N*-(3-nitropyridine-4-yl)acetohydrazide

Tetrahydroquinoline acetamide was obtained by catalytic hydrogenation with platinum from methyl nitrophenyl propionate [30].



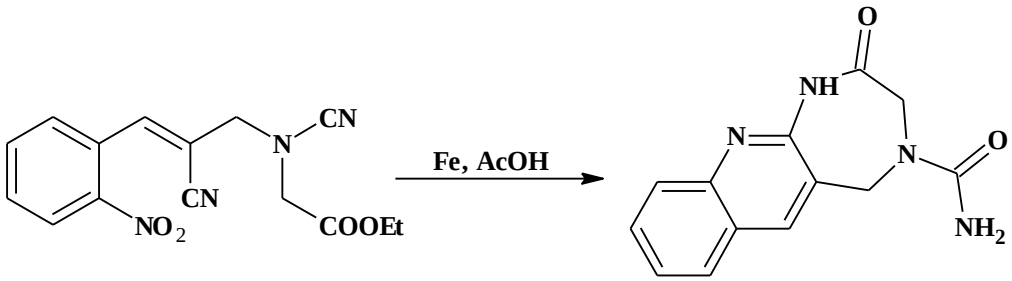
Schema I.21: The synthesis of Tetrahydroquinoline acetamide

I.3.2.a. Biological activity

Six-membered heterocyclic systems, however, do not only exist as the main structure of compounds with various biological properties utilized as traditional remedies but also, they currently are prescribed and market purchasable medicines. They have shown effective biological properties and are even used as medicine for diseases. They are also exploited as novel agents with biological activities such as anti-canser [31], anti-bacterial[32], anti-fungal[33], anti-inflammatory[34], anti-diabetic [35]; analgesic [36], antihypertensive [37] and anti-malaria [38] .

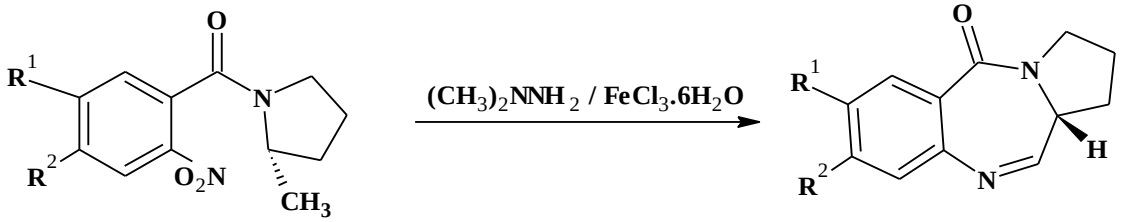
I.3.3. Synthesis of 7 members heterocycles

The 2-oxo-1,2,3,5-tetrahydro-4*H*-[1,4]diazepino[5,6-b]quinolone-4-carboxamide obtained from Ethyl acetate, 2-(cyano[(*Z*)-2-cyano-3-(2-nitrophenyl)prop-2-enyl]amino) by double reductive cyclization using iron powder in acetic acid. (in one step) [39].



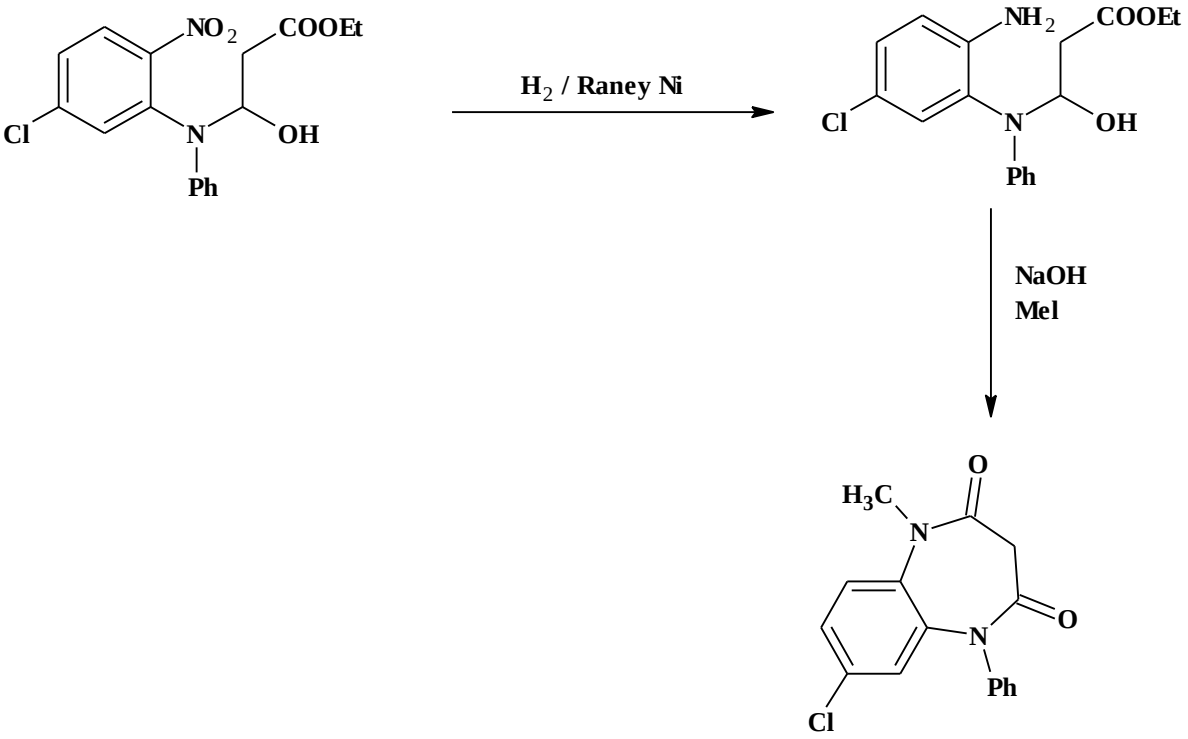
Schema I.22: The synthesis of 2-oxo-1,2,3,5-tetrahydro-4*H*-[1,4]diazepino [5,6-b]quinolone-4-carboxamide

The pyrrolo[2,1-c] [1,4] benzodiazepine is accessible in a single step by reductive cyclization using N,N-dimethylhydrazine from (2*R*)-1-(2-nitrobenzoyl)pyrrolidine-2-carbaldéhyde [40].



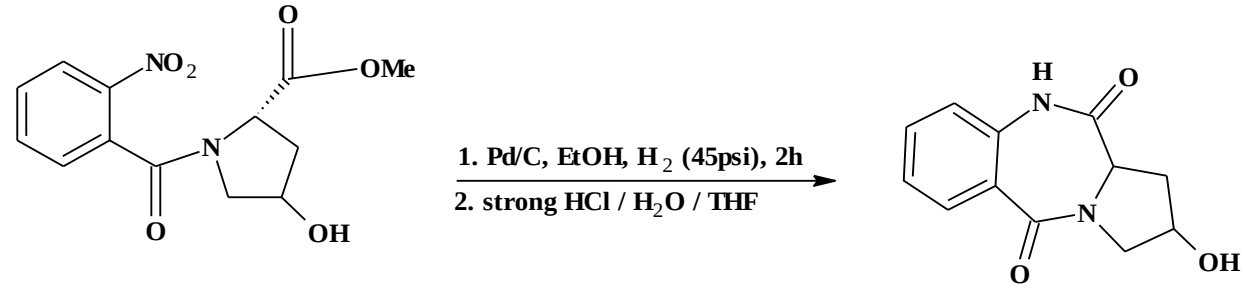
Schema I.23: The synthesis of The pyrrolo[2,1-c] [1,4] benzodiazepine

The catalytic reduction of Ethyl 3-[(5-chloro-2-nitrophenyl)(pheenyl)amino]-3oxopropanoate leads to 7-chloro-1-methyl-5-phenyl-1*H*-1,5-benzodiazepine-2,4(3*H*,5*H*)-dione, in two steps [41].



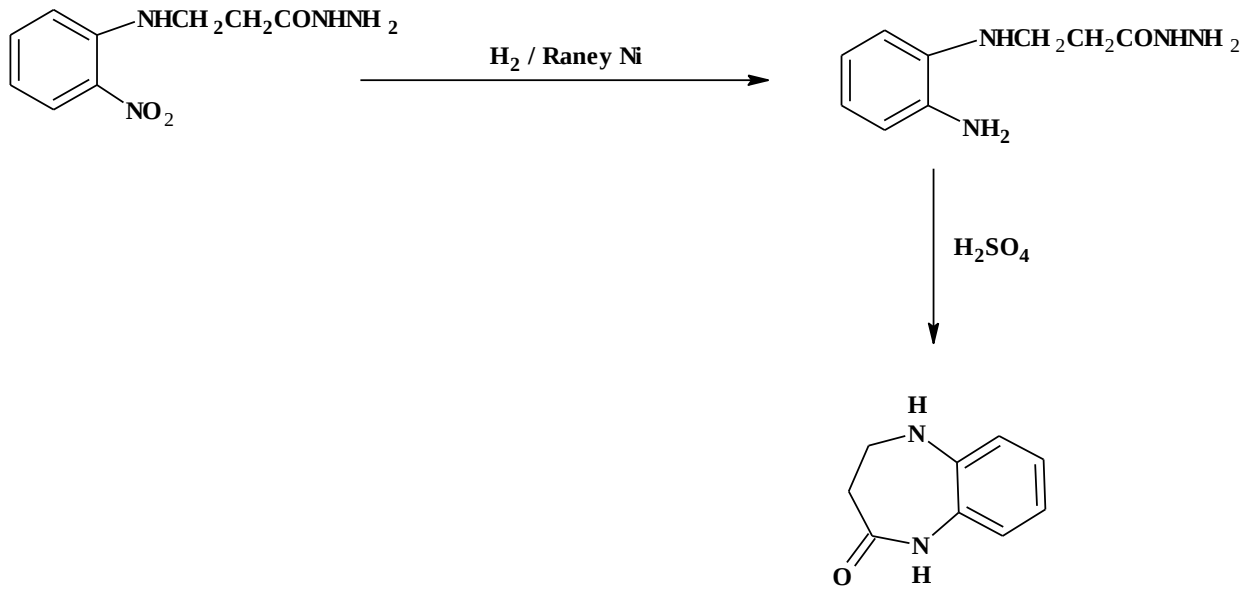
Schema I.24: The synthesis of 7-chloro-1-methyl-5-phenyl-1*H*-1,5-benzodiazepine-2,4(3*H*,5*H*)-dione

The reduction 4-hydroxy-1-(2-nitrobenzoyl)pyrrolidine-2-carboxylic Methyl ester acid in two steps give the derivative 2R-hydroxy-1,2,3,10,11aS-hexahydro-5*H*-pyrrolo[2,1-c][1,4] benzodiazepine-5,11-dione [40].



Schema I.25: The synthesis of derivative 2R-hydroxy-1,2,3,10,11a-S-hexahydro-5*H*-pyrrolo[2,1-c][1,4] benzodiazepine-5,11-dione

The reductive cyclization of 1,3,4,5-tetrahydro-2*H*-1,5-benzodiazepin-2-one guides to 3-[(2-nitrophenyl)amino]propanehydrazide in two steps [42].



Schema I.26: The synthesis of 3-[(2-nitrophenyl)amino]propanehydrazide

I.3.3.a. Biological activity

seven-membred nitro heterocycles demonstrated a variety of biological activities in the field of pharmacology, in particular benzodiazepine derivatives which are commonly used as anti-tumor, anti-cancer, anticonvulsants and hypnotics. In addition, these compounds are well known for their therapeutic uses as anxiolytics and sedatives [43].

I.3.4. Bis–Heterocyclic

Bis–Heterocyclic moieties are significant class of organic compounds that undergo a diversity of chemical reactions and acquire a great attention because of their pharmaceutical biological fields.[44] These compounds were divided based on similarities in their chemical structures and made up of two connected, bonded, or fused heterocyclic skeletons. Therefore, bis-heterocycles are primarily formed by linking heterocyclic nuclei together utilizing a variety of linkers and sites of linkage .

The Bis-heterocyclic compounds are the biggest families of biological substances and agrochemicals product in organic chemistry. They have a broad range of application in medicinal chemistry which focuses on the research and development of drugs and pharmaceuticals due to their biological and pharmacological characteristics such as (Sofosbuvir) Antiviral,

(Aripiprazole) Antipsychotic, (Esomeprazole) Antiulcerant, and (Rosuvastatin) Cholesterol regulator which have heterocycle fragments in their structural system For example [45]:

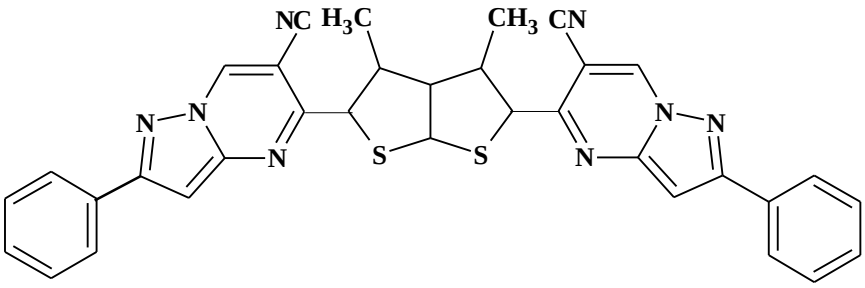


Figure I.1: Bis-pyrazolo [1,5-a] pyrimidine derivative [46].

I.3.4.a. biological activity

Two heterocyclic cycles are present in bis-heterocycles, which are organic molecules. Their unique chemical structure can cause a wide range of biological functions. These are a few instances of biological processes connected to specific bis-heterocycles :

1- Antibacterial/Antimicrobial Activity: Some bis-heterocycles have demonstrated antibacterial or anti-microbial activity, which means that they can be effective in combating bacterial infections or other microorganisms [47].
2- Anti-tumor activity: Some bis-heterocycles have shown antitumor properties, making them potential candidates for the development of anti-cancer drugs [48].
3- Antiviral activity: Some bis-heterocycles may have antiviral activities, which means that they can inhibit the replication of viruses or their entry into host cells[49].
4- Anti-inflammatory activity: Some bis-heterocycles may have anti-inflamatory properties, which means they can reduce inflammation in the body [50].
5- Antioxidant activity: Some bis-heterocycles can act as antioxidants, helping to neutralize free radicals and prevent oxidative damage in the body[51].

It is important to note that the biological activity of bisheterocycles depends heavily on their chemical structure and how they interact with specific biological processes in the organism. In addition, many bi-heterocycles can have multiple or combined activities, making them even more interesting for pharmaceutical and chemical research .

Part two: Modeling

For many years scientists have asked why some substances have a beneficial effect on the body, whilst others are toxic ,and why some are more beneficial,or more toxic, than are others.that led to QSAR .

QSAR models, known as quantitative structure-activity relationships studies, unquestionably are of great importance in modern chemistry and biochemistry. QSAR models are mathematical models which relate pharmacological or biological activity with the physicochemical characteristics termed molecular descriptors of a set of molecules based on AI, and among the most usable approaches, the SVM .

I.1. Definition

Quantitative structure–activity relationship models (QSAR models) are regression or classification models used in the chemical and biological sciences and engineering. Like other regression models, QSAR regression models relate a set of "predictor" variables (X) to the potency of the response variable (Y), while classification QSAR models relate the predictor variables to a categorical value of the response variable .

In QSAR modeling, the predictors consist of physico-chemical properties or theoretical molecular descriptors of chemicals [N] the QSAR response-variable could be a biological activity of the chemicals. QSAR models first summarize a supposed relationship between chemical structures and biological activity in a data-set of chemicals. Second, QSAR models predict the activities of new chemicals [52].

A general formula for a simple QSAR model is shown below:

$$Activity = n_1x_1 + n_2x_2 + n_3x_3 \dots\dots\dots\dots + Constant \quad \text{(I.1)[53]}$$

Where

nx = Molecular descriptors

I.2. The history

It is widely accepted that modern QSAR began in the early 1960s. However, as long ago as 1816 scientist were making predictions about physical and chemical properties .the first investigations into the correlation of biological activities with physicochemical properties such us molecular

weight and aquatic solubility began in 1841.almost 60 years before the important work of Overton and Meyer linking aquatic toxicity to lipid-water partitioning .throughout the 20th century QSAR progressed , though there were many lean years .In 1962 came the seminal work of Corwin Hansch and co-workers, which stimulated a huge interest in the prediction of biological activities . Initially that interest lay largely within medicinal chemistry and drug design, but in the 1970s and 1980s, with increasing Ecotoxicological concerns. QSAR modelling of environmental toxicities began to grow especially once regulatory authorities became involved. Since then QSAR has continued to expand ,with over 1400 publications annually from 2011 onwards [54].

I.3. Molecular descriptors

Despite of, the great advances in the field of drug design, the use of descriptors to define the molecular structure of biologically active compounds is the main method utilized to discover new lead molecules.

Descriptors are the chemical characteristic of a molecule in numerical form, used for QSAR/QSPR studies. Mathematical representation of these descriptors has to be invariant to the size of the molecule and the number of atoms it contains to enable model building with statistical methods.

Molecular descriptors have become the most significant features used in QSAR/QSPR modeling. The information encoded by descriptors generally depends on the kind of molecular representation and the defined algorithm for its calculation. Some of these include: topological indices and geometrical, constitutional and physicochemical descriptors.

- **Thermodynamic descriptors**

Are used to relate chemical structure to observed chemical behavior. e.g HF (head of formation), molRef(Molar refractivity). etc.

- **Constitutional descriptors**

Simple and commonly used descriptors reflecting the chemical information of a molecule without any information of atom connectivity. e.g Atom and bond counts, MW. etc.

- **Topological descriptors**

Based on molecular graphs Represents the connectivity of atoms in molecules. Used for modeling biological physicochemical and pharmacokinetic properties. e.g., Wiener, Zagreb connectivity indices, etc.

- **Electronic descriptors**

Used to describe electronic aspects of the molecule or atom bonds and molecular fragments. e.g Dipole moment, HOMO, LUMO energy, etc.

- **Geometrical descriptors**

Calculated from the 3D coordinates of the atoms. Capture the 3D information regarding the molecular size, shape, and atoms distribution. e.g., WHIM, MoRSE and GETAWAY, etc.[55].

Molecular descriptors typically classified as 1D, 2D, 3D, or 4D.

- **1D descriptors:** represent simple molecular properties along a single dimension, such as molecular weight, melting point, or the number of specific chemical groups present in a molecular.
- **2D descriptors:** capture more detailed information about the molecular structure in two dimensions, including topological indices, atom-centered fragments and connectivity-based descriptors.
- **3D descriptors:** provide information about the three-dimensional structure of molecules, such as shape, volume, surface area and spatial arrangements of atoms within the molecule.
- **4D descriptors:** incorporate dynamic properties or changes over time, adding a temporal dimension to the analysis of molecular structures [56].

I.4. Selection of descriptors

When a significant number of descriptors are introduced, some of them may include redundant information, causing a problem with collinearity. Furthermore, the derived descriptors do not always have an impact on the action to be modeled. The descriptors utilized should be relevant and easily interpretable from a chemical standpoint, resulting in QSAR/QSPR models that are straightforward, transparent, and phenomenologically understood. As a result, it is required to delete descriptors with less influence than the error and keep only the most relevant ones.

In general, for a descriptor to be allowed, its removal must cause a considerable decline in the model's performance. We must consequently take care not to lose critical information. Finally, the chemical relevance of the descriptors utilized must be considered, because the more chemically tied the describers are to the phenomenon, the less likely they are to be random describers. There are two steps to selecting and minimizing descriptors objective and Subjective Selection [57].

I.4.a. Objective selection

It entails selecting variables by minimizing the number of descriptors while excluding the dependent variable (biological activity). The first stage in this process is to exclude those descriptors with a large percentage of similar values across all compounds (variance is not significant). This ensures that such descriptors are not randomly inserted into the final model. Similarly, if two descriptors are highly linked and their combination has a determination coefficient that exceeds the needed threshold (R^2>0.95), only the one with the largest variance is chosen.

These approaches not only keep unsuitable descriptors out of the model, but they also cut down on the amount of descriptors that need to be processed, which lowers the cost of the follow-up analysis in terms of computation time [58].

I.4.b. Subjective selection

A/Progressive Introduction

This method is to incorporate, one by one, the variables into the model by selecting at each step the variable whose partial correlation with the modeled size is the highest.

B/ Progressive elimination

This method consists of establishing the model with all the descriptors and then keeping only those that allow the obtaining of a model with a good correlation.

C/ Step-by-step selection

It is the combination of the two methods mentioned earlier. The variables are incorporated one by one into the model by gradual selection. However, at each stage, it is verified that the partial correlations of the previously introduced variables are still significant.

I.5. AI-based QSAR methods

I.5.1. Artificial intelligence (AI)

Until this moment there is no generally accepted definition of the term "artificial intelligence".

B. Buchanan and E. Shordiffe (1983) "Artificial Intelligence is a branch of computer science dealing with symbolic, non-algorithmic methods of problem-solving".

Rich and Knight (1991) define AI as "The study of how to make computers do things which, at the moment, people do better".

John McCarthy (2007) has proposed the following definition "Artificial intelligence is the science and engineering of making intelligent machines" [59].

Knowledge-based or Artificial Intelligence techniques are used increasingly as alternatives to more classical techniques to model environmental systems with the recognition of its potential.

We review some of the techniques covered: Machine learning, Data mining, case-based reasoning, rule-based systems, artificial neural networks [60].

I.5.2. Machine learning

Machine Learning is a method for obtaining Artificial Intelligence. It is an area of computer science and a subset of AI that provides computers with the capability to think and learn without being programmed explicitly,by employing statistical techniques to allow computers to "learn" [61].

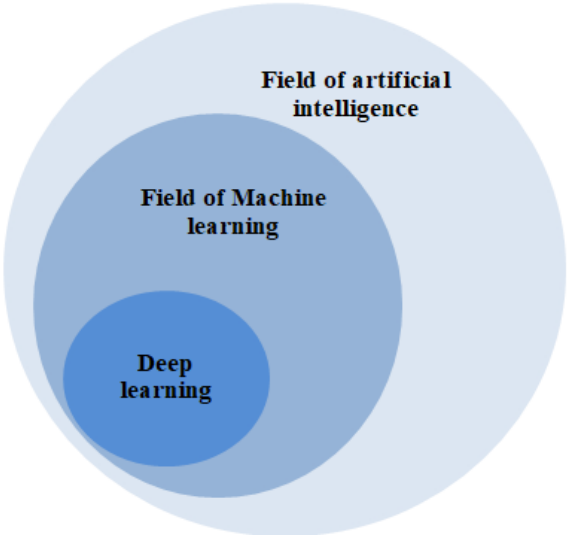


Figure I.2: Representation of the relationship between AI, ML and DL .

ML comes in a variety of forms:

Supervised learning: Systems that are exposed to vast volumes of labeled data are used.

Unsupervised learning: Refers to programs that attempt to spot trends in data to divide them into categories

Semi-supervised (Reinforcement) learning: Combines supervised and unsupervised learning to train systems with small amounts of labeled and large amounts of unlabeled data [62].

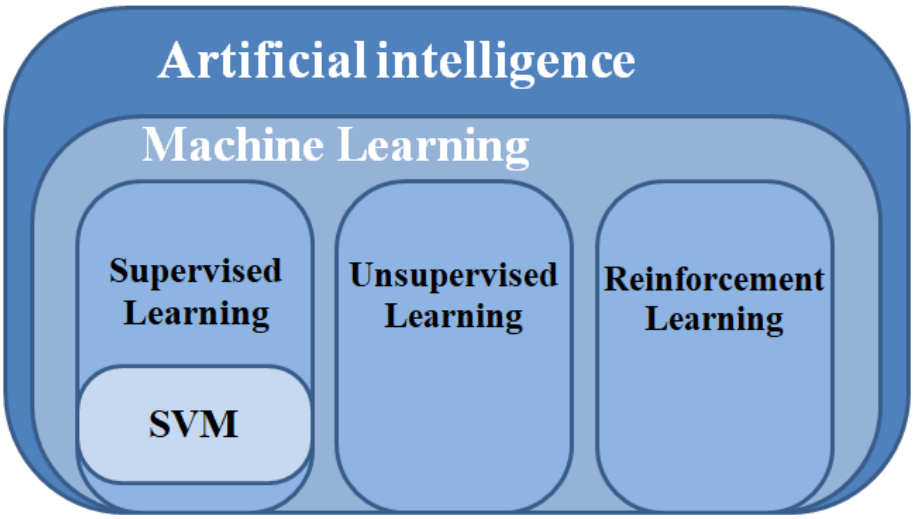


Figure I.3: Flow chart of machine learning and the algorithms used for training the machine .

I.5.2.1. Support vector machine (SVM)

A support vector machine (SVM) is a supervised learning algorithm used for many classification and regression problems, including signal processing, medical applications, natural language processing, and speech and image recognition.

The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class. “Best” is defined as the hyperplane with the largest margin between the two classes, represented by plus versus minus in the figure below. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. Only for linearly separable problems can the algorithm find such a hyperplane, for most practical problems the algorithm maximizes the soft margin allowing a small number of misclassifications [63].

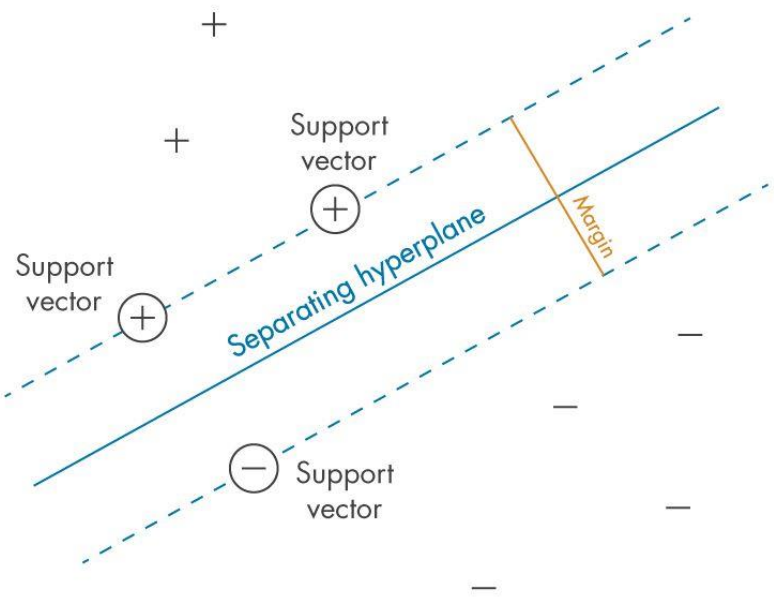


Figure I.4: Defining the margin between classes (the criterion that SVMs seek to optimize)[63].

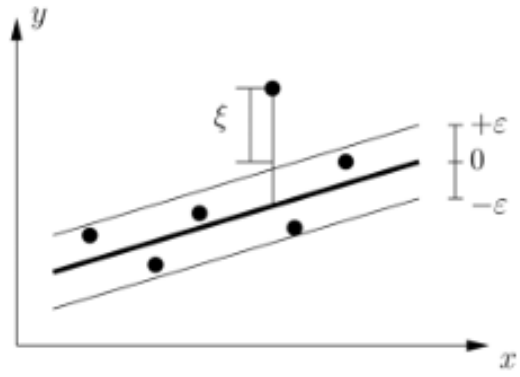


Figure I.5 : ε radius insensitivity tube[64].

- **Support vectors SV** refer to a subset of the training observations that identify the location of the separating hyperplane. The standard SVM algorithm is formulated for binary classification problems, and multiclass problems are typically reduced to a series of binary ones [63].
- **Hyperplane** is the decision boundary that is used to separate the data points of different classes in a feature space.

- **The margin** is the distance between the decision boundary and the support vectors. SVM aims to maximize this margin to achieve better separation and generalization (ε).
- **Kernel** is a mathematical function used in SVM to map the original input data points into high-dimensional feature spaces, allowing the hyperplane to be located even when the data points are not linearly separable in the original input space. SVR can use different types of kernels, Radial basis function (RBF), linear, polynomial, and sigmoid are a few of the frequently used kernel functions [65].

Types of SVM classifiers:

- **Linear SVMs**

Linear SVMs are used with linearly separable data; this means that the data do not need to undergo any transformations to separate the data into different classes [66].

- **Nonlinear SVMs**

Nonlinear SVM is necessary when the data cannot be effectively separated by a linear decision boundary in the original feature space. Nonlinear SVM addresses this limitation by utilizing kernel functions to map the data into a higher-dimensional space where linear separation becomes possible [67].

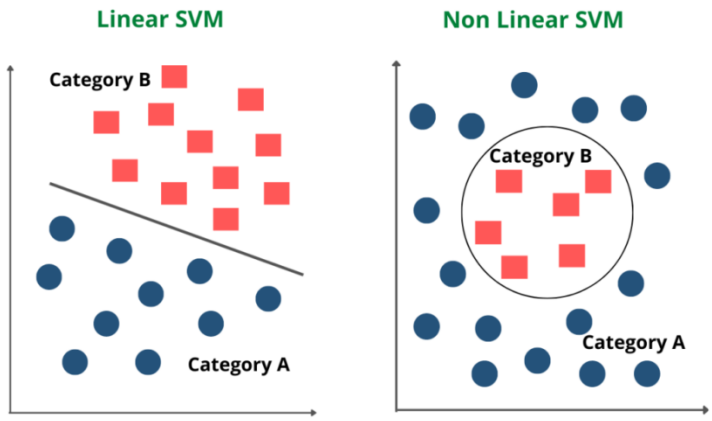


Figure I.6: Linear and nonlinear SVMs [68].

I.6. Support Vector Regression (SVR)

Support vector regression (SVR) is an extension of SVMs, which is applied to regression problems. Similar to linear SVMs, SVR finds a hyperplane with the maximum margin between data points, and it is typically used for time series prediction.
SVR differs from linear regression in that you need to specify the relationship that you're looking to understand between the independent and dependent variables [66].

SVR has several hyperparameters that you can adjust to control the behavior of the model. For example, the **'C'** parameter controls the trade-off between the insensitive loss and the sensitive loss. A larger value of **'C'** means that the model will try to minimize the insensitive loss more, while a smaller value of C means that the model will be more lenient in allowing larger errors[69].
SVR is to find the function f(x) that has at most one ε deviation from the learning examples (xi, yi), for i = 1,..., N, and which is as small as possible. This means not taking into account errors less than ε and prohibiting errors greater than ε [70]. Maximizing the flatness of the function allows to minimize the complexity of the model that affects its performance in generalization. Indeed, the theory of learning limits the error of generalization to a sum of two terms: one depending on the complexity of the model and the other dependent on the error on the learning data. SVMs methods are based on controlling the complexity of the model during learning.
The following sections present the SVR algorithm for a linear model, then for the non-linear case.

I.6.1. Linear problem

A linear kernel is a simple dot product between two input vectors. It is given by:

$$F(x) = (w, x) + b \tag{I.2}$$

Where:
w : weight
x : input
b : biais

I.11.2. Nonlinear problem

A non-linear kernel is a more complex function that can capture more intricate patterns in the data.

The extension of SVMs to the non-linear case is quite simple and relies on projecting the data into a higher dimensional space in which the problem becomes linear. The problem becomes:

$$F(x) = \sum_{i=1}^{N}(\alpha i - \alpha i *) = k(xi, x) + b \quad \text{(I.3)}$$

Where:
αi, αi*: Lagrange multipliers
k : Constant
$k(xi, x)$: kernel function

The choice of kernel depends on the data's characteristics and the task's complexity, the RBF Kernel is given by:

$$k(x, x') = \exp(-\frac{|x-x'|^2}{2\sigma^2}) \quad \text{(I.4)}$$

Where:

σ : Radial basis function width

I.7. validation criteria

I.7.a. correlation coefficient (R)

The correlation coefficient is used to study monotonous relationships whether or not they are linear. But depending on their linearity, two types of Pearson or Spearman correlation coefficients are used [71].

The correlation coefficient can be used to link a multitude of variables. For example, the relationship between size and intelligence, price and margins, population and pollution, calorie intake and disease, etc. [72].

It measures the predictive capacity of a model and is given by:

$$\mathbf{R}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad \text{(I.5)[73]}$$

Where:

- xi and yi are the individual data points.
- $\bar{x}$ and $\bar{y}$ are the means of the x and y variables, respectively.
- ∑ denotes the summation over all data points.

Depending on the result, we can interpret the relationship between the two variables. For example:

• if R is close to 0: the linear relationship is zero. Attention, it is not because there is no linear relationship between two variables that there is not a relationship at all.

• if R is close to -1: there is a strong linear relationship between the two variables, but negative. They are therefore going in opposite directions.

if R is close to 1 : there is a strong positive linear relationship between the two variables. They go in the same direction [74].

I.7.b. Coefcient of determination (R^2)

R-squared (R^2) is defined as a number that tells you how well the independent variable(s) in a statistical model explain the variation in the dependent variable. It goes from 0 to 1, where 1 indicates a perfect fit of the model to the data. These coefficient estimates and predictions are crucial for understanding the relationship between the variables.[75]

The calculation of R-squared requires several steps.This includes taking the data points (observations) of dependent and independent variables and conducting regression analysis to find the line of best fit, often from a regression model. This regression line helps to visualize the relationship between the variables. From there, you would calculate predicted values, subtract actual values, and square the results, It is given by [75]:

$$\boldsymbol{R^2 = 1 - \frac{\sum(Y_{(obs)} - Y_{(calc)})^2}{\sum(Y_{(obs)} - Y_{(obs)})^2} = \frac{ESS}{TSS}} \qquad \text{(I.6)[76]}$$

I.7.c. Root mean square error (RMSE)

The root mean square error (RMSE) measures the average difference between a statistical model's predicted values and the actual values. Mathematically, it is the standard deviation of the residuals. Residuals represent the distance between the regression line and the data points

RMSE quantifies how dispersed these residuals are, revealing how tightly the observed data clusters around the predicted values [77].

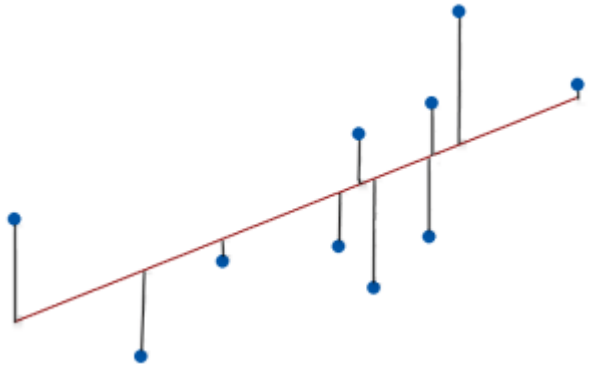


Figure II.8: Root mean square error [77].

As the data points move closer to the regression line, the model has less error, lowering the RMSE. A model with less error produces more precise predictions.

RMSE values can range from zero to positive infinity and use the same units as the dependent (outcome) variable.

Use the root mean square error to assess the amount of error in a regression or other statistical model. A value of 0 means that the predicted values perfectly match the actual values, but you'll never see that in practice. Low RMSE values indicate that the model fits the data well and has more precise predictions. Conversely, higher values suggest more error and less precise predictions, it is given by [77]:

$$\mathbf{RMSE} = \sqrt{\frac{(\boldsymbol{Y}_{(\mathbf{obs})} - \boldsymbol{Y}_{(\mathbf{calc})})^2}{\boldsymbol{n}}} \qquad \text{(I.7)[76]}$$

I.7.d. Coefficient of prediction (Q^2)

If R^2 is a measure of the correctness of the correlation, Q^2 is the measurement of the accuracy of the prediction. It measures the predictive capacity of a model and is given by [77]:

$$\boldsymbol{Q^2} = \mathbf{1} - \frac{\sum(\boldsymbol{Y\ obs}_{(\mathbf{train})} - \boldsymbol{Y\ calc}_{(\mathbf{train})})^2}{\sum(\boldsymbol{Y\ obs}_{(\mathbf{train})} - \bar{\boldsymbol{Y}}_{\mathbf{training}})^2} \qquad \text{(I.8)[76]}$$

The cross-validation by "leave-one-out" (LOO) is to recalculate the model on (n-1) observations,

and to use the model obtained to calculate the value of interest of the offset compound, noted y(i). Repeat the procedure for each of the sizes [77].

I.7.e. Mean Absolute Error (MAE)

Mean Absolute Error **(MAE)** is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model [78].

The MAE loss function formula :

$$\mathbf{MAE} = \frac{1}{n} \Sigma_{(i=1 \text{ to } n)} |y_i - \hat{y}_i| \qquad \text{(I.9)[78]}$$

where:

- **n** is the number of observations in the dataset.
- **y_i** is the true value.
- **$\hat{y}_i$** is the predicted value.

The Mean Absolute Error (MAE) is a crucial performance statistic for regression models since it is an easy-to-understand, interpretable, and reliable tool for assessing the accuracy of predictions [78].

I.7.f. The Concordance Correlation Coefficient (CCC)

The Concordance Correlation Coefficient (CCC) is a widely used statistic in various fields, including medical research, biology, psychology, and more.It quantifies the agreement and reliability between two continuous measurements or observations.[79]

It provides a single numerical value that ranges from -1 to 1, where:

- **1** indicates perfect positive agreement.
- **-1** indicates perfect negative agreement (i.e., one set of data is the exact opposite of the other).
- **0** indicates no agreement.

In practical terms, CCC combines both the correlation and bias between two sets of measurements, making it a robust measure of agreement.It is given by [79]:

$$\mathbf{CCC=1-}\ \frac{E[(Y_1-Y_2)^2]}{(u_1-u_2)^2+\sigma_1{}^2+\sigma_2{}^2} = \frac{2\sigma_{12}}{(u_1-u_2)^2+\sigma_1{}^2+\sigma_2{}^2} \quad \text{(I.10)[79]}$$

I.8. importance of QSAR

Quantitative Structure-Activity Relationship (QSAR) modeling offers several advantages over traditional methods in the fields of chemistry, pharmacology, and toxicology [80] ,Below are a few reasons :

1- Efficiency: QSAR models can predict the biological activity or properties of chemical compounds based solely on their molecular structure, without the need for time-consuming and costly experimental assays. This efficiency accelerates the drug discovery process by narrowing down the list of candidate compounds for further experimental testing [81].

2- Cost-Effectiveness: Traditional methods of testing chemical compounds often require extensive laboratory experiments, which can be expensive and resource-intensive. QSAR modeling reduces the need for such experiments by providing in silico predictions, thereby saving time and resources in drug development and toxicological studies [82].

3- Predictive Power: QSAR models can provide quantitative predictions of biological activity, toxicity, physicochemical properties, and other properties of chemical compounds. These predictions are based on mathematical relationships between molecular descriptors and observed biological activities, allowing researchers to prioritize compounds for synthesis or experimental testing with greater confidence.

4- High Throughput Screening: QSAR models enable high-throughput virtual screening of large compound libraries, allowing researchers to quickly identify potential drug candidates or lead compounds. This approach facilitates the screening of diverse chemical space and accelerates the discovery of novel bioactive molecules [83].

5- Reduced Animal Testing: By predicting the toxicity of chemical compounds in silico, QSAR modeling reduces the reliance on animal testing in toxicology studies. This aligns with ethical considerations and regulatory initiatives aimed at minimizing animal experimentation in drug development and chemical safety assessment [84].

6- Adaptability and Flexibility: QSAR models can be easily adapted and optimized for different chemical datasets, target classes, and endpoints. This flexibility allows

researchers to tailor QSAR approaches to specific research questions or application domains, enhancing their utility across diverse fields of study [85].

Overall, QSAR offers a powerful and versatile tool for drug discovery, chemical risk assessment, and understanding the molecular basis of biological activity. Its efficiency, cost-effectiveness, predictive power, and ability to provide mechanistic insights make it an indispensable tool in modern computational chemistry and pharmacology.

Chapter II :
Experimental Part

Analytical equipment and techniques

• mp melting points (°C) are determined using a FICHER-JONES plate heater (with blades).

• Fourier transformed infrared spectra were recorded on a BRUKER spectrometer. The absorption frequencies are expressed in cm^{-1} at their maximum.

• Low pressure vacuum pump for vacuum fractioning of the DREHSCHIEBER brand up to 135 mmHg.

Table II.1: The products used

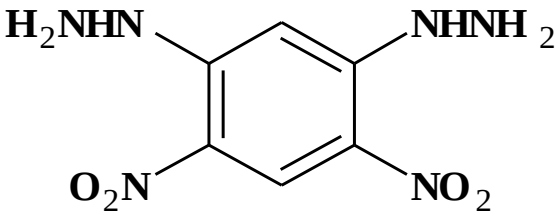
Product's Name	Formula	Purity (%)	Firm
Methanol	CH_3OH	Pure	Honeywell
Hydrochloric acid	HCl	33%	Honeywell
Hydrazine monohydrate	NH_2NH_2,H_2O	98%	Fluka
4-Chloro-2-nitroaniline	$C_6H_5ClN_2O_2$	99%	ACROS
Sodium dithionite	$Na_2S_2O_4$	85%	PANREAC
Diethyl acetylene carboxylate	$C_6H_{10}O_4$	95%	ALDRICH
1-chloro-3,4-dinitrobenzene	$C_6H_6N_3O_2Cl$	Pure	MERCK
1-chloro-2,4- dinitrobenzene	$C_6H_6N_3O_2Cl$	Pure	MERCK
3,4-dinitro toluene	$C_7H_6N_2O4$	Pure	MERCK
1,3-dihydrazino-4,6-dinitrobenzene	$C_6H_8N_6O_4$	Pure	MERCK

II.1. Introduction

A variety of techniques are used in the synthesis of nitro heterocycles, Creating effective and targeted synthesis processes for nitro heterocycles is essential to expanding their uses in various industries. Nitro heterocycles are known for their special biological qualities, which include antibacterial, anti-inflammatory, and anticancer effects. QSAR models can assist in identifying possible therapeutic candidates, forecasting their characteristics, and minimizing the requirement for rigorous experimental testing.

II.2. Synthesis of starting compound

II.2.1. Synthesis of 1, 3-dihydrazino-4,6-dinitrobenzene



Scheme II.1: Structure of 1, 3-dihydrazino-4,6-dinitrobenzen.

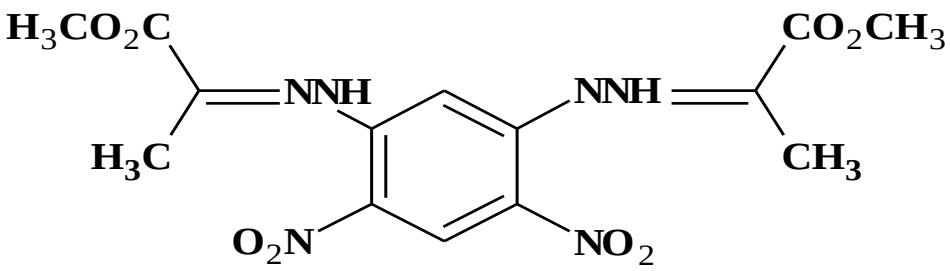
In a double-necked flat-bottom flask with a refrigerant and a dropping funnel, 2g of 1,3-dichloro-2,4-dinitrobenzene is dissolved in 40 ml of ethanol and 2 ml of hydrazine monohydrate dissolved in 20 ml of the same solvent is drained under stirring. The precipitate formed is refluxed for an hour to finish the reaction. After cooling to room temperature, the precipitate is filtered under vacuum and recrystallized into ethanol.

- $C_6H_8N_6O_4$ (228g/mol) : Red brick crystals.
- Mp =255 °C
- Yield = 98,2%
- IR (KBr): 3200-3400 cm^{-1} (NH and NH_2); 1500 cm^{-1} (NO_2). (Figure III.1).

II.3. Synthesis of Bishydrazones

II.3.1. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

II.3.1.1. Synthesis of *(E/E)*-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

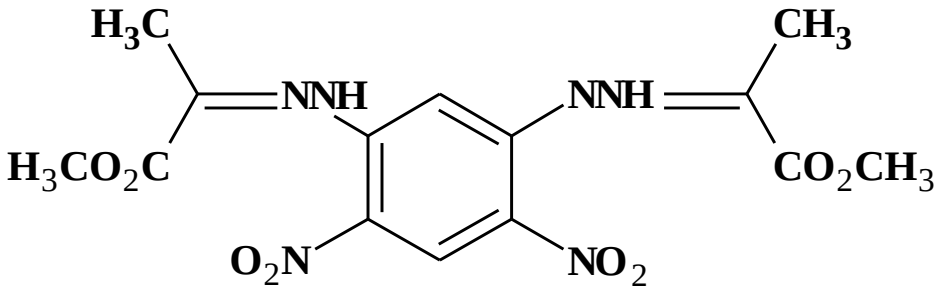


Scheme II.2: The structure of *(E/E)*-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate.

In an Erlenmeyer, 0.49g (2 mmol) of 1, 3-dihydrazino-4,6-dinitrobenzene are dissolved in 50 ml of methanol and 0.44g (2 mmol) dissolved in 20 ml methanol of methyl pyruvate are drained. The mixture is left under stirring until a precipitate is formed. The precipitate is filtered and recrystallized in the same solvent.

- $C_{14}H_{16}N_6O_8$ (396g/mole): Yellow crystals.
- mp = 98 °C
- Yield = 65%
- IR (KBr): 3299 cm^{-1} (NH), 1707 cm^{-1} (C=O), 1614cm^{-1} (C=N).(figure III.2).

II.3.1.2. Synthesis of *(Z/Z)*-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

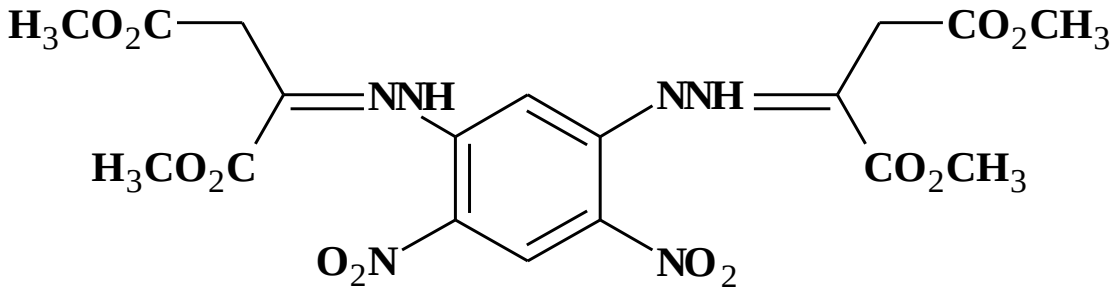


Scheme II.3: the structure of *(Z/Z)*-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

In an Erlenmeyer 0.9g of (*E/E*)-1,3-dihydrazono-4,6-dinitrophenylen-methyl pyruvate are dissolved in 50ml of methanol and 1ml of concentrated hydrochloric acid is added. Let it stir until a precipitation appears. The precipitate is filtered and recrystallized in the same solvent.

- $C_{14}H_{16}N_6O_8$ (396g/mole): Yellow crystals.
- mp = 228-230 °C
- Yield = 94.76%
- IR (KBr): 3343 cm^{-1} (NH), 1658 cm^{-1} (C=O), 1614 cm^{-1} (C=N).(figure III.3).

II.3.2. Synthesis of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl



Scheme II.4: The structure of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl.

In an Erlenmeyer 0.8g (4mmol) of 1, 3-dihydrazino-4,6-dinitrobenzene are dissolved in 50 ml of methanol and 1g (4mmol) of dimethyl acetylenedicarboxylate dissolved in 20 ml of the same solvent is drained under stirring. After the addition of 1 ml of concentrated hydrochloric acid let it stir until a precipitation appears. The precipitate is filtered and recrystallized in the same solvent.

- $C_{18}H_{20}N_6O_{12}$ (512g/mole): Yellow crystals.
- mp = 112-114 °C
- Yield = 54%
- IR (KBr): 3311 cm^{-1} (NH), 1755 cm^{-1} (C=O), 1428 cm^{-1} (NO_2).(figure III.4).

II.4. Reductions

II.4.1. Reduction of *(Z/Z)*-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate by sodium dithionite

In a double-necked flat-bottom flask with a reflux refrigerant, 2 mmoles of the *(Z/Z)*-1,3-dihydrazono-4,6-dinitrophenylen-methyl pyruvate derivative are dissolved in 40 ml of Methanol.

Bring the mixture to boil and stir through the refrigerant a solution of 17 mmoles of sodium dithionite dissolved in 20 ml of distilled water. At the end of the addition the reactive mixture is heated under reflux and agitation for five hours. After cooling to room temperature, filter the solution to remove the salt that formed and then add 2 to 3 volumes of distilled water to the solution and heat until boiling. After cooling and possibly concentrating the solution a precipitates are formed which will be filtered and recrystallized in the Methanol.

- Color: Yellow orange
- mp = °C
- IR(KBr): 3224 cm^{-1} (NH), 1730 cm^{-1} (C=O), 1604cm^{-1} (C=N). (Figure III.5)

II.4.2. Reduction of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl by sodium dithionite

As described in the previous experimental protocol, using 0.6 mmoles of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl in 40ml of Methanol and 17 mmols of sodium dithionite in 20ml of distilled water.

We have obtained a low yield value which is not uncharacterized.

Color: Brown

II.5. Evaluation of the antimicrobial activity of synthesized Heterocycles using QSAR model

The methodology adopted for the prediction of the antimicrobial activity using the SVM technique can be schematized according to the organigram above:

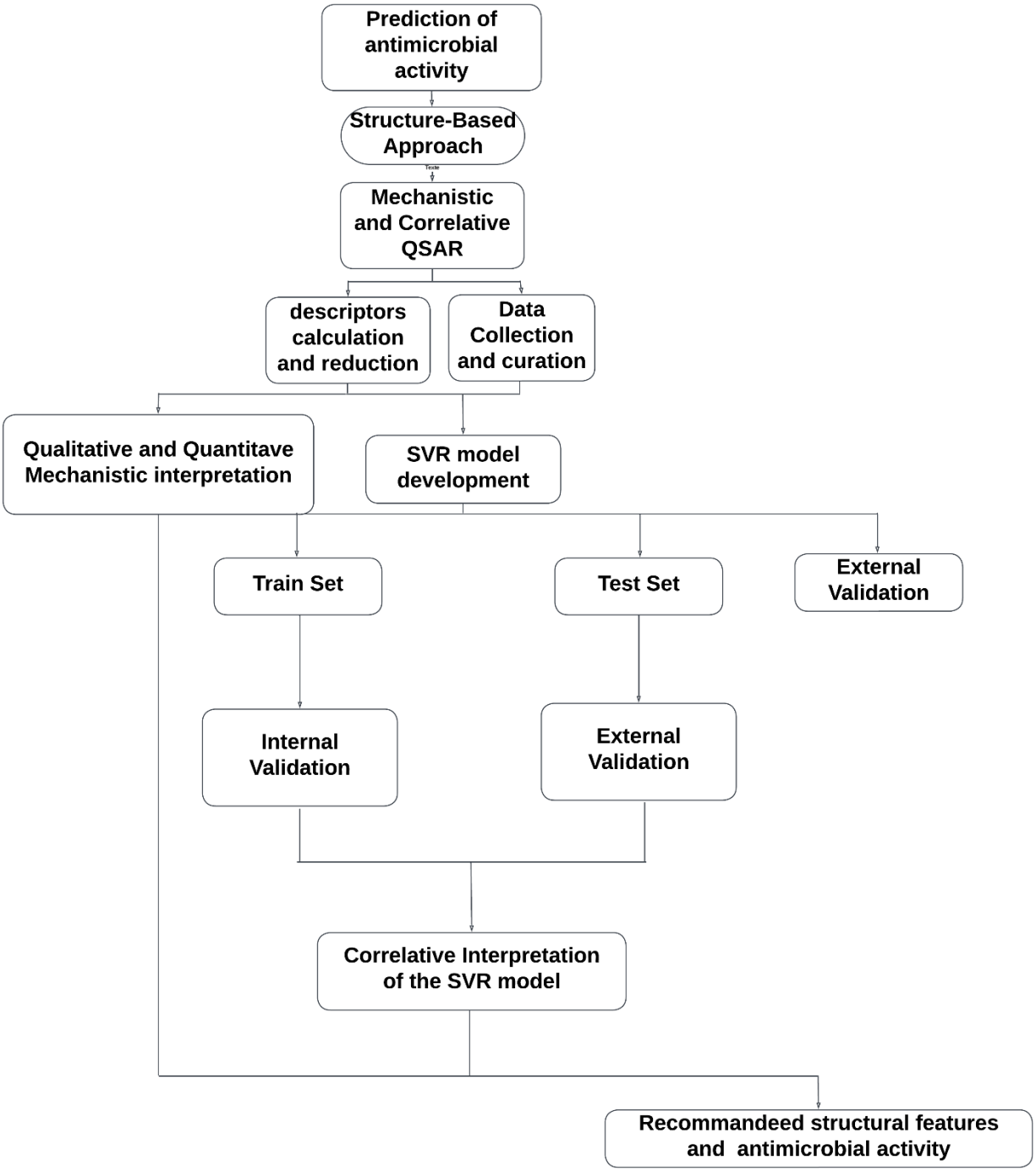


Figure II.1: Modeling methodology.

II.5.1. Database collection

The data were collected using PubChem and the literature (New Synthetic Route to Cinnoline Derivatives and Their Microbiological Activity) based on their antimicrobial activity. The data

was then visualized and preprocessed using the ChemSketsh. As well as the duplicate molecules were also removed, leaving only the molecules that had a known MIC. In order to linearize. To convert experimental MIC values (mg/ml) to a logarithmic scale:

$$(pMIC) = \ln\ (pMIC) \qquad \text{(II.1)}$$

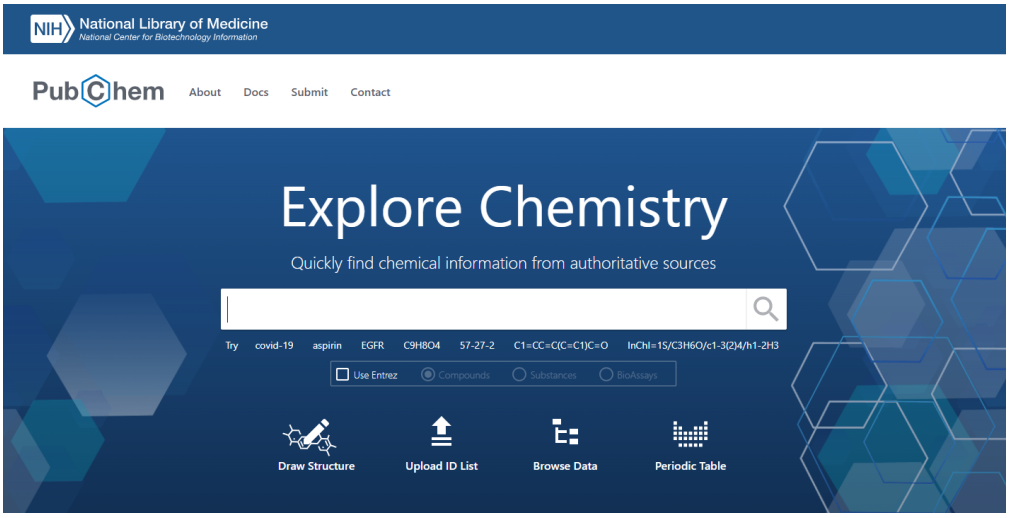


Figure II.2: Screenshot PubChem.

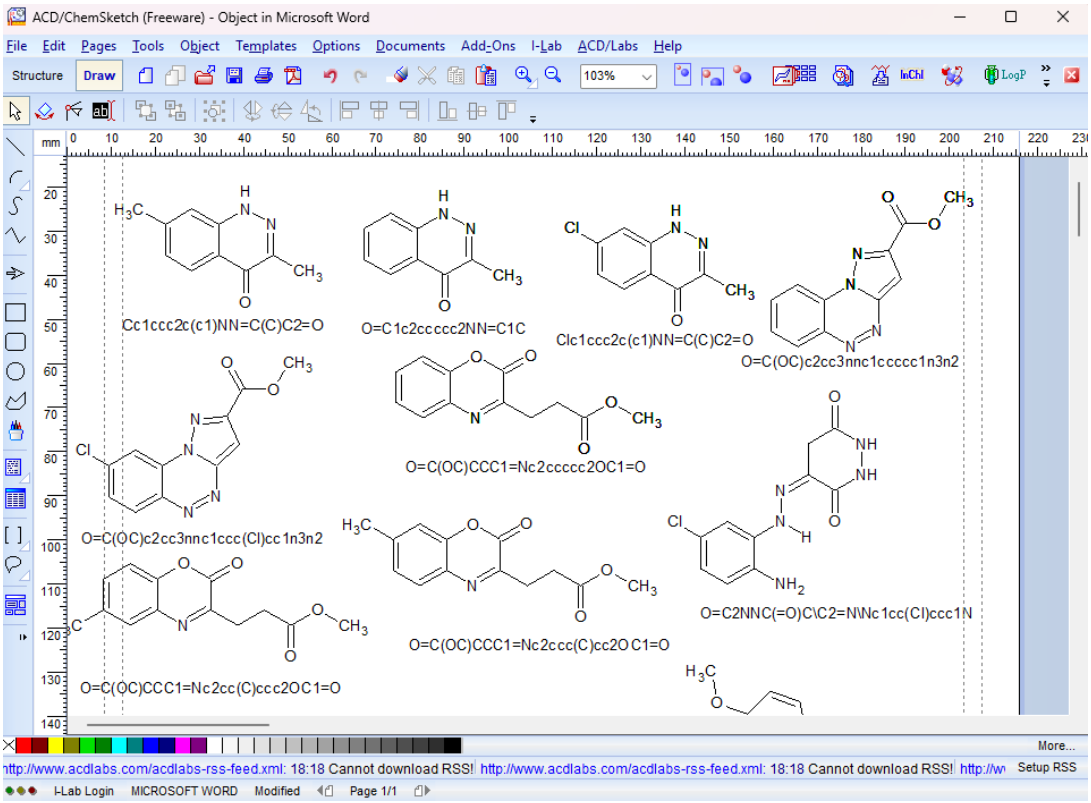
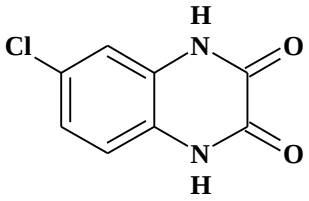
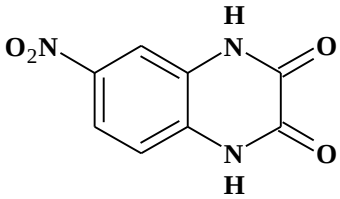
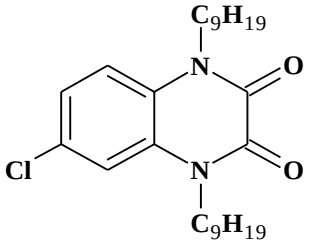
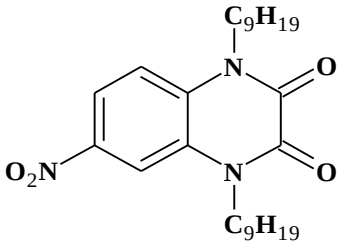
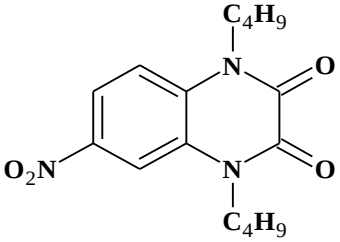
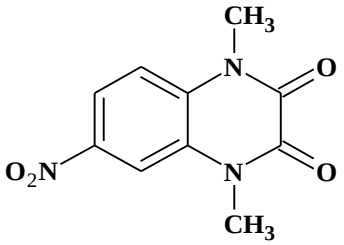
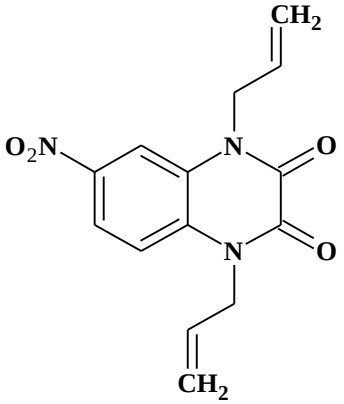
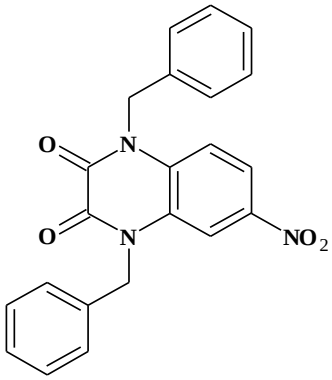
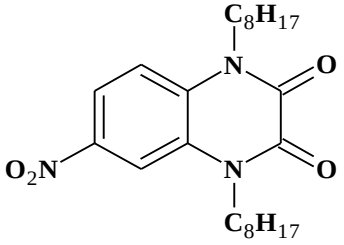
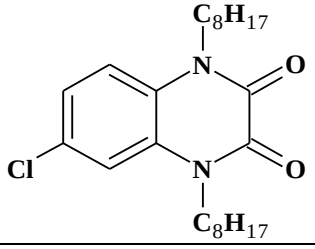
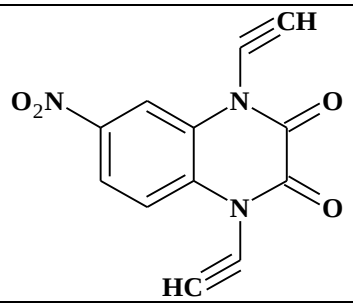
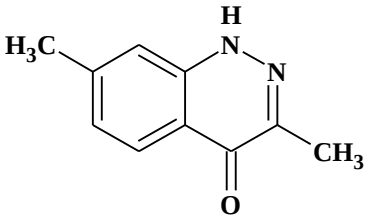
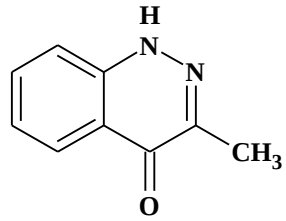
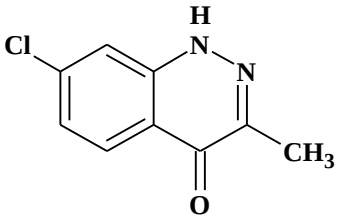
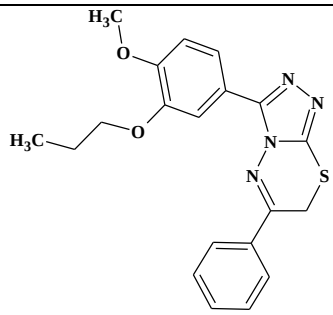
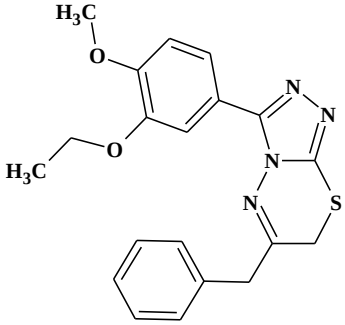


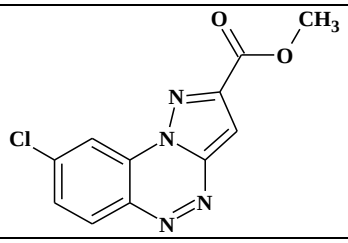
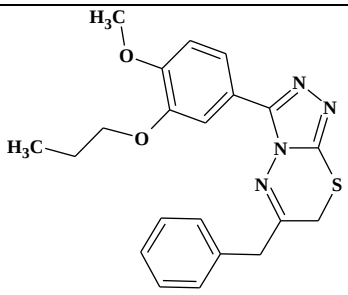
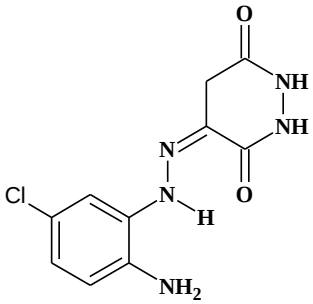
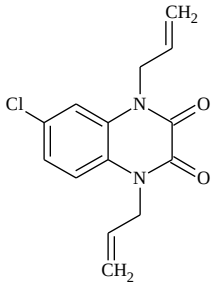
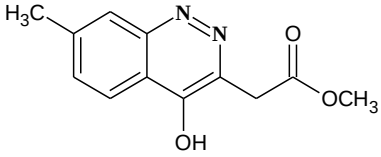
Figure II.3: Screenshot ChemSketch.

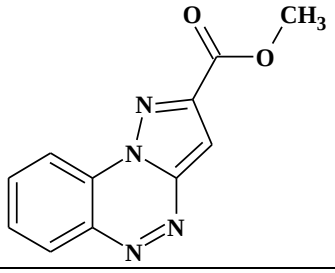
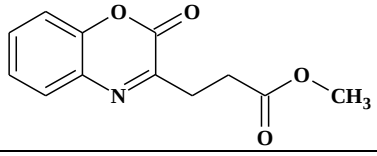
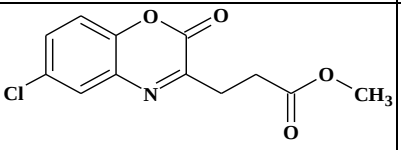
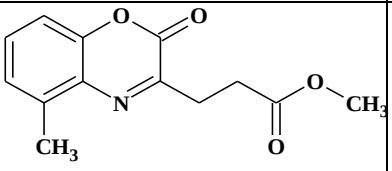
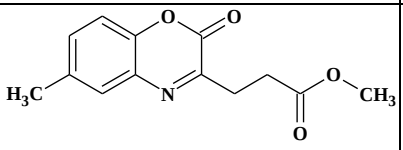
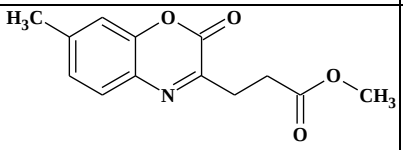
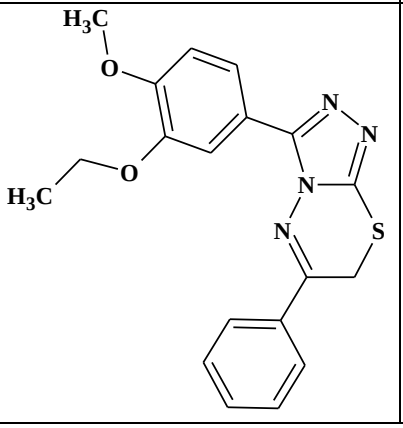
Table II.2: Database collection.

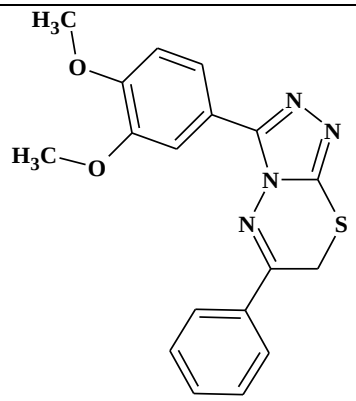
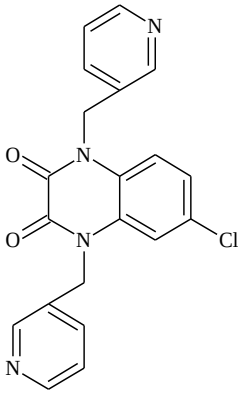
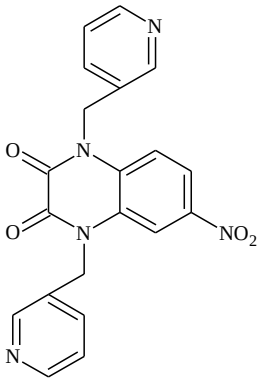
N	Structure	Smile	S.aureus MIC (mg/ml)	reference
01		Clc1ccc2NC(=O)C(=O)Nc2c1	0.14	[86]
02		[O-][N+](=O)c1ccc2NC(=O)C(=O)Nc2c1	0.07	[86]
03		Clc1ccc2c(c1)N(CCCCCCCCC)C(=O)C(=O)N2CCCCCCCCC	0.14	[86]
04		[O-][N+](=O)c1ccc2c(c1)N(CCCCCCCCCC)C(=O)C(=O)N2CCCCCCCCC	0.14	[86]
05		[O-][N+](=O)c1ccc2c(c1)N(CCCC)C(=O)C(=O)N2CCCC	0.07	[86]

06		[O-][N+](=O)c1ccc2c(c1)N(C)C(=O)C(=O)N2C	0.502	[86]
07		[O-][N+](=O)c1ccc2c(c1)N(CC=C)C(=O)C(=O)N2CC=C	0.14	[86]
08		[O-][N+](=O)c1ccc4c(c1)N(Cc2ccccc2)C(=O)C(=O)N4Cc3ccccc3	0.502	[86]
09		[O-][N+](=O)c1ccc2c(c1)N(CCCCCCCC)C(=O)C(=O)N2CCCCCCCC	0.004	[86]
10		Clc1ccc2c(c1)N(CCCCCCCC)C(=O)C(=O)N2CCCCCCCC	2.5	[86]

11		[O-][N+](=O)c1ccc2c(c1)N(C#C)C(=O)C(=O)N2C#C	0.281	[86]
12		Cc1ccc2c(c1)NN=C(C)C2=O	10	[87]
13		O=C1c2ccccc2NN=C1C	10	[87]
14		Clc1ccc2c(c1)NN=C(C)C2=O	10	[87]
15		COc1ccc(cc1OCCC)c2nnc3SCC(=Nn23)c4ccccc4	0.05	[88]
16		COc1ccc(cc1OCC)c2nnc3SCC(=Nn23)Cc4ccccc4	0.025	[88]

17		O=C(OC)c2cc3nnc1ccc(Cl)cc1n3n2	0.023	[89]
18		COc1ccc(cc1OCCC)c2nnc3SCC(=Nn23)Cc4ccccc4	0.05	[88]
19		O=C2NNC(=O)C\C2=N\Nc1ccc(N)cc1N	2.5	[89]
20		Clc1ccc2c(c1)N(CC=C)C(=O)C(=O)N2CC=C	0.017	[86]
21		C1=C(C=CC2=C1N=NC(=C2O[H])CC(=O)OC)C	10	[87]

22		O=C(OC)c2cc3nnc1ccccc1n3n2	0.035	[89]
23		O=C(OC)CCC1=Nc2ccccc2OC1=O	00	[90]
24		O=C(OC)CCC1=Nc2cc(Cl)ccc2OC1=O	00	[90]
25		O=C(OC)CCC2=Nc1c(cccc1C)OC2=O	00	[90]
26		O=C(OC)CCC1=Nc2cc(C)ccc2OC1=O	0.013	[90]
27		O=C(OC)CCC1=Nc2ccc(C)cc2OC1=O	00	[90]
28		COc1ccc(cc1OCC)c2nnc3SCC(=Nn23)c4ccccc4	0.00625	[88]

29		COc1ccc(cc1OC)c2nnc3SCC(=Nn23)c4ccccc4	0.00625	[88]
30		Clc1ccc4c(c1)N(Cc2cccnc2)C(=O)C(=O)N4Cc3cccnc3	0.281	[86]
31		[O-][N+](=O)c1ccc4c(c1)N(Cc2cccnc2)C(=O)C(=O)N4Cc3cccnc3	0.281	[86]

II.5.2. Descriptor selection

Molecular descriptors are the numerical representation of the structural characteristics of molecules, they were calculated by the **PaDEL** software to obtain the descriptor. The **Stepwise-MLR** tool has been used twice to select only representative descriptors reducing the number of descriptors to 7 and thus forming a 31x7 matrix.

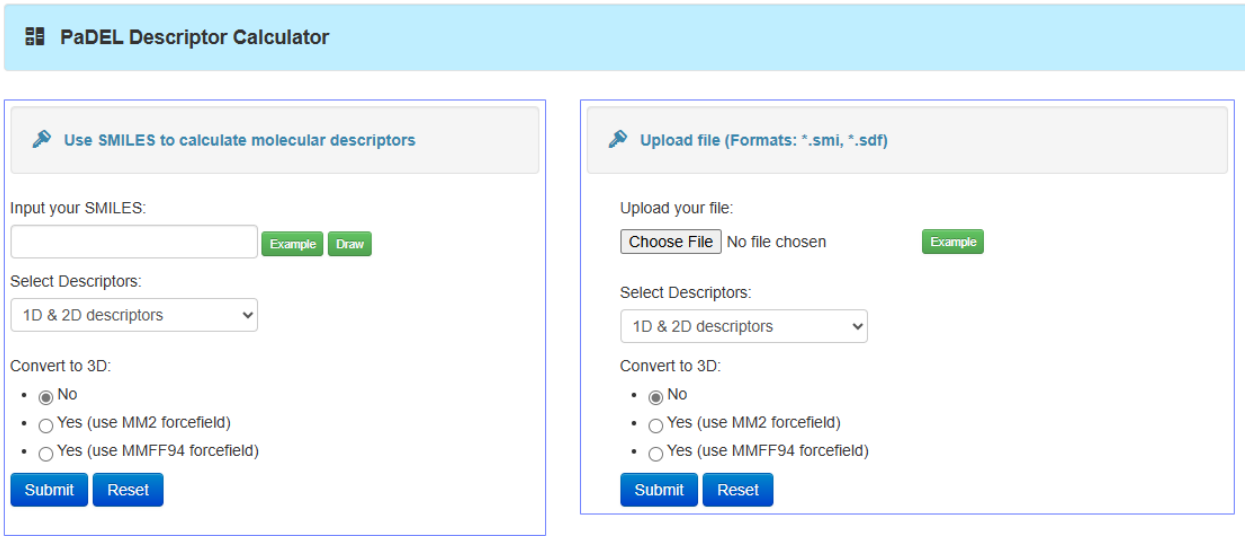


Figure II.4: Screenshot PaDEL.



Figure II.5: Screenshot Stepwise MLR.

II.5.3. Database division and model development

The data set was optimally distributed into a learning set containing 80% of the data set, and a test set Containing 20% of the data set, after two trials of division (learning: 75%, test 25%), (learning 70%, test 30%) QSAR models were developed using the approach: SVR, made using MATLAB.

II.5.3.1. SVM model

In this study, a SVR model was developed to predict the biological activity of synthesized molecules

II.5.3.2. Validation

Once developed. The model must be interpreted by analyzing all the statistical parameters of the model. Its quality must also be studied. This quality is verified by so-called validation.

A. Internal validation

The internal validation of a QSAR/QSPR model was performed using cross validation LOO (Leave-One-Out) or LMO (Leave-Many-Out), which is measured by the R2 coefficient. This process involves extracting a certain number of K molecules from the initial set to N molecule and building a new model with the remaining (N-K) molecules using the chosen descriptors (only the regression constants change). This method is then repeated to predict the values of all molecules in the learning series Depending on the number of molecules removed at each iteration.

Talking about Leave-One-Out (LOO) or Leave-Many-Out(LMO) depending on whether one or more molecules is (are) removed(s). However, internal validation is insufficient to study the predictive power of a model. For this reason, internal model validation has become a standard and a mandatory part of QSAR/QSPR modeling.

B. External validation

This method predicts the properties/activities of a series of molecules called the test series that are not part of the model's development series. This validation is defined by the parameter R^2. Several recent research have demonstrated that the R^2 parameter is insufficient to test the predictive power of QSAR/QSPR models. As a result, additional factors must be verified for this purpose. These parameters are known as "external validation criteria" or frequently called "Trophsa criteria."

Chapter III :

Results and discussion

III.1. Introduction

The synthesis strategy in this work was based on the synthesis of new bishydrazone derivatives synthesized by hydrazone condensation.

To access the corresponding bishydrazones, we first had to synthesize the starting compound which is 1,3-dihydrazino-4,6-dinitrobenzene.

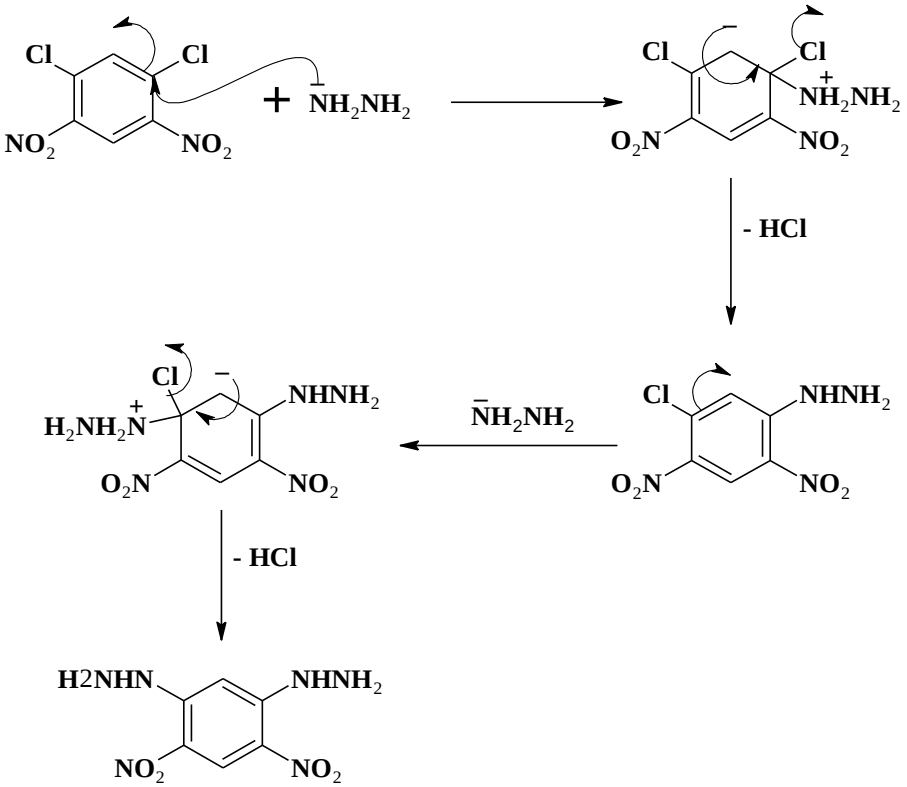
The reduction of these has led to opportunities for reductive cyclisation..

the prediction of their biological activity was made using the QSAR model.

III.2. Synthesis of starting compound

III.2.1. Synthesis of 1, 3-dihydrazino-4,6-dinitrobenzene

1,3-dihydrazino-4,6-dinitrobenzene is easily obtained from 1,3-dichloro-4, 6-dinitrobenzene. As nitro groups are electron attractors, they weaken the benzene nucleus in electron density. This makes a simple nucleophilic substitution of chlorine atoms by hydrazine in two stages mechanism (addition/elimination) (Scheme III.1).



Scheme III.1: Reaction mechanism of the synthesis of 1,3-dihydrazino-4,6-dinitrobenzen.

According to the protocol, 1,3-dichloro-4,6-dinitrobenzen is mixed in cold with excess hydrazine monohydrate in ethanol and carried under reflux for one hour. 1,3-dihydrazino-4,6-dinitrobenzene is obtained with high yield.

The IR spectrum shows the characteristic vibrations of NH and NH_2 groups between 3200 and 3400 cm^{-1}, the absorption bands of the cluster (NO_2) appear around 1500 cm^{-1} (Figure III.1).

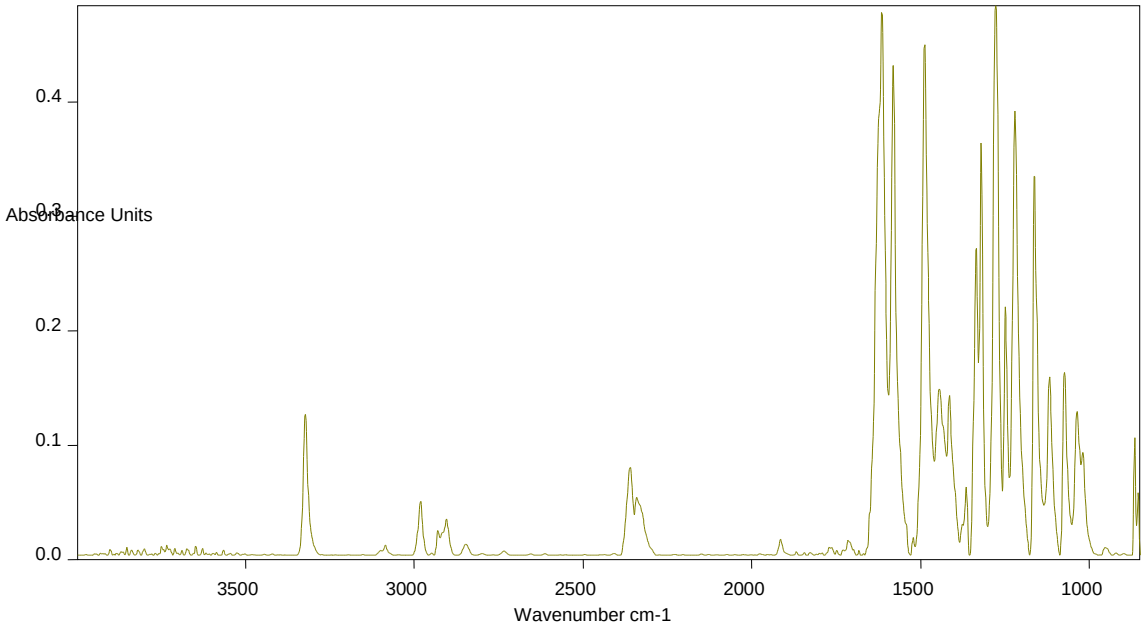
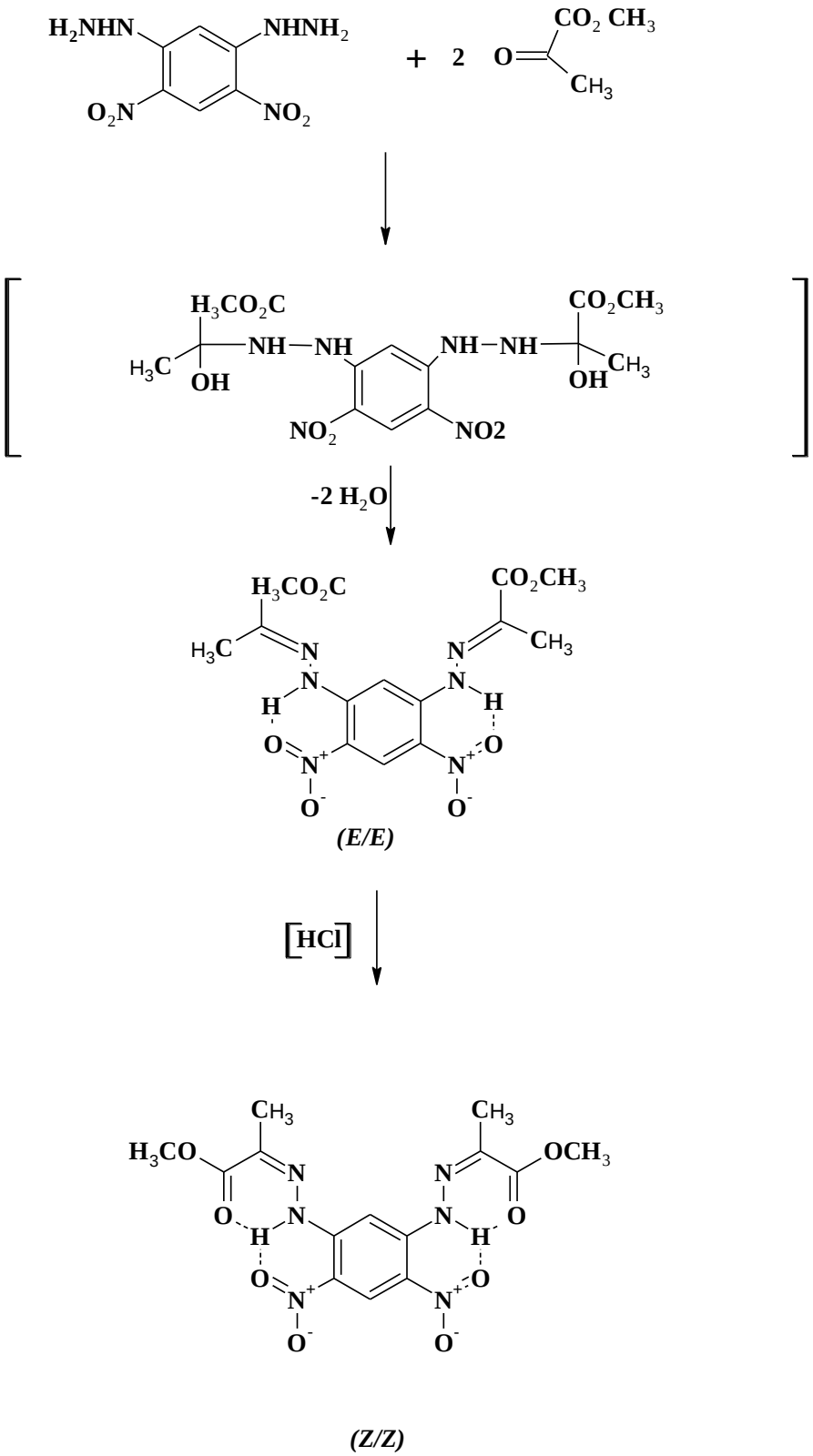


Figure III.1: IR spectrum of 1,3-dihydrazino-4,6-dinitrobenzene.

III.3. Synthesis of Bishydrazones

III.3.1. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

The synthesis of 1,3-hydrazino -4,6-dinitrophenylmethyl pyruvate occurs according to the well-known carbonyl group condensation reaction with the amino group following a two-stage mechanism (addition/ elimination) (Scheme III.2).



(E/E)

(Z/Z)

Scheme III.2: Reaction mechanism of synthesis of isomers (*E/E*)- and (*Z/Z*)- 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate

Initially, the product of the nucleophilic addition is generated, but this unstable product eventually develops into the specified hydrazone. This reaction happens at room temperature.

To an alcoholic solution of 1,3-dihydrazino-4,6-dinitrobenzen, a methanol solution of methyl pyruvate is added.

After the whole addition of the solution and stirring for two hours, the (*E/E*)- 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate precipitates as a solid product.

Through the formation of the double bond C=N, 1,3-hydrazono-4,6-dinitrophenylenmethyl pyruvate can exist in two diastereoisomeric forms: *(E/E)* and (*Z/Z*) forms. At room temperature in the methanol is formed exclusively the isomer (*E/E*).By processing the latter in methanol, or better yet in ether, in the presence of concentrated hydrochloric acid, it is easily transformed into thermodynamically more stable isomers (*Z/Z*).These two bishydrazones are distinctly distinguished by their melting points.

The IR spectrum shows the presence of a single vibrational mode of elongation of the NH link towards 3343 cm^{-1}, at 1658 cm^{-1} of strong bands of absorptions of characteristic length of carbonyl grouping (C=O), and also shows a characteristic band of elongations of the absorption (C=N) at 1614cm^{-1} as well as the absorption of the nitro group (NO_2) towards 1543 cm^{-1} (Figure III.2).

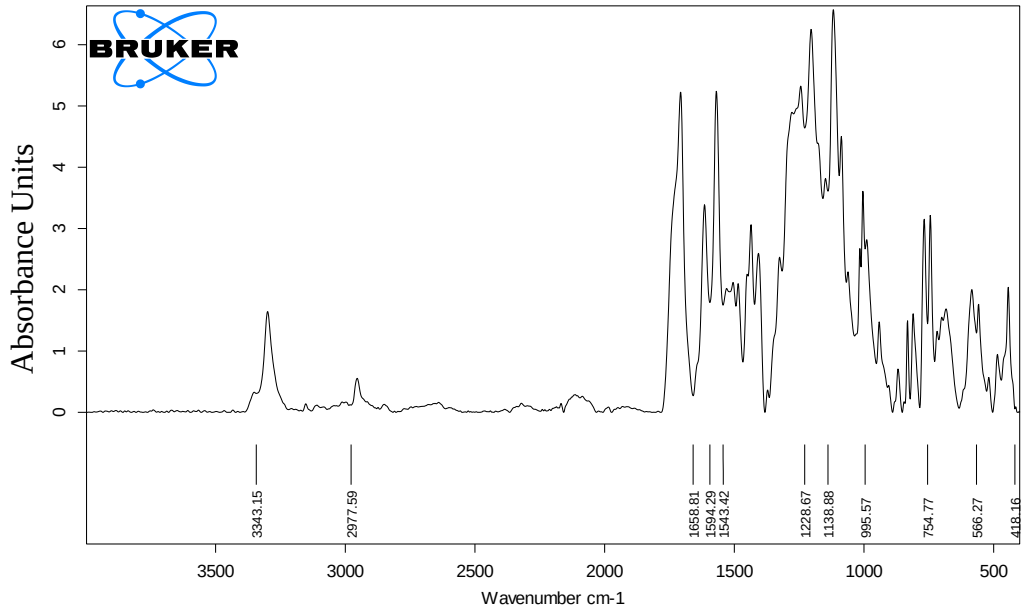


Figure III.2: IR spectrum of (*E/E*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate.

In terms of the isomer (*Z*/*Z*), the IR spectrum shows the presence of a single vibrational elongation mode of the (NH) bond towards 3347 cm^{-1}, as well as the characteristic carbonyl group bands (C=O) at 1756 cm^{-1}, and a characteristic elongation band of the absorption (C = N) at 1598 cm^{-1}, indicating the nitro group (NO_2) towards 1552 cm^{-1} (Figure III.3).

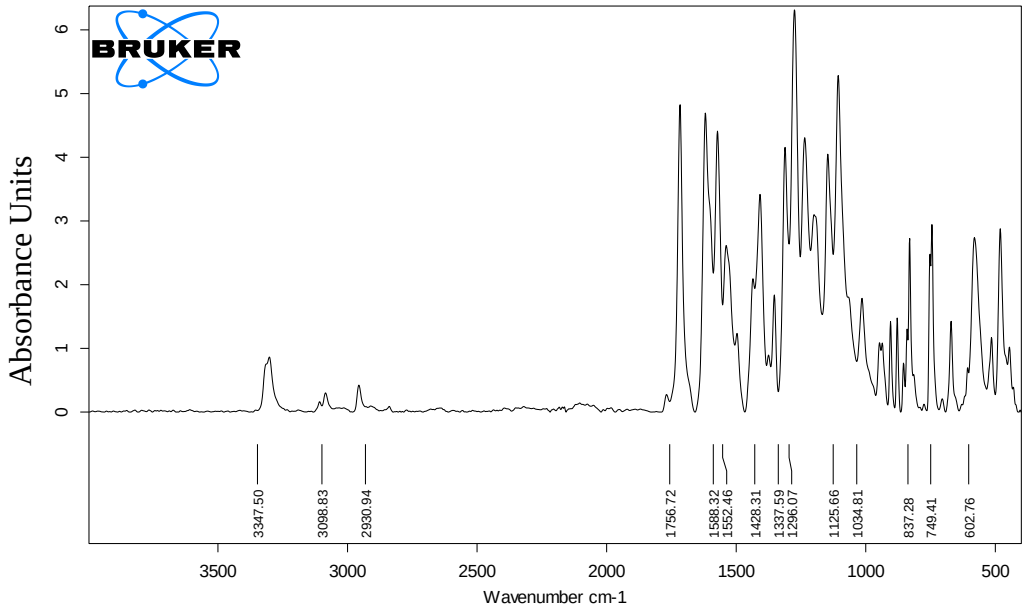
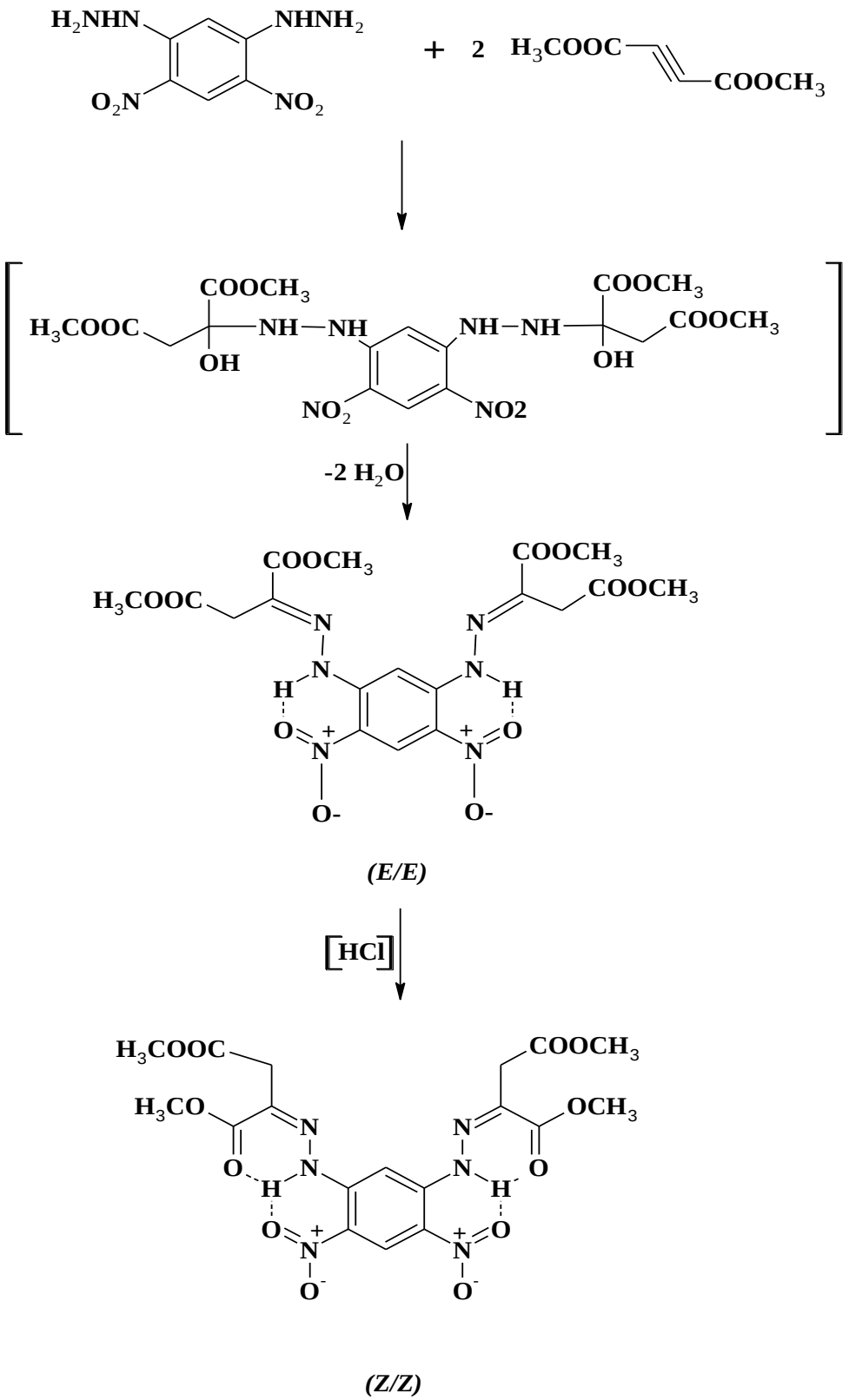


Figure III.3: IR spectrum of (*Z*/*Z*)- 1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate.

II.3.2. Synthesis of 1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl

The synthesis of 1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl is done according to the same reactionary mechanism described previously following two-stage mechanism (addition/ elimination) (scheme III.3).



Scheme III.3: Reaction mechanism of synthesis of isomers (*E/E*) and (*Z/Z*)- 1,3-dihydrazono - 4,6-dinitrophenylenoxaloacetate dimethyl.

Initially, the product of the nucleophilic addition is generated, but this unstable product eventually develops into the specified hydrazone. This reaction happens at room temperature.

To an alcoholic solution of 1,3-dihydrazino-4,6-dinitrobenzen, a methanol solution of Dimethyl acetylene dicarboxylate is added.

After the whole addition of the solution and stirring for two hours, the (*E/E*)- 1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl precipitates as a solid product.

Through the formation of the double bond C=N, 1,3-hydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl can exist in two diastereoisomeric forms: (*E/E*) and (Z/Z) forms. At room temperature in the methanol is formed exclusively the isomer (*E/E*).By processing the latter in methanol, or better yet in ether, in the presence of concentrated hydrochloric acid, it is easily transformed into thermodynamically more stable isomers (Z/Z). About this product, we were only interested in the formation of the isomer (Z/Z) by treating the starting reactive mixture with concentrated HCl.

The IR spectrum shows the presence of a single vibrational extension mode of the NH link towards 3311.45cm^{-1}, and also at 1755.11cm^{-1} strong bands of extension absorptions characteristic of carbonyl group (C=O), as well as the nitro group absorbtion (NO_2) towards 1428.95cm^{-1}. (Figure III.4).

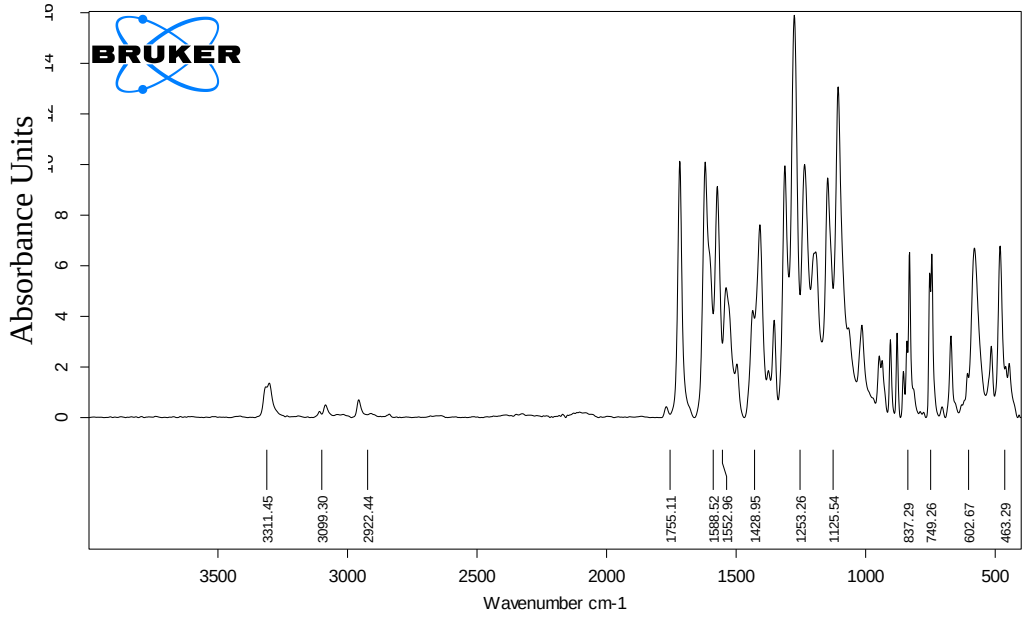


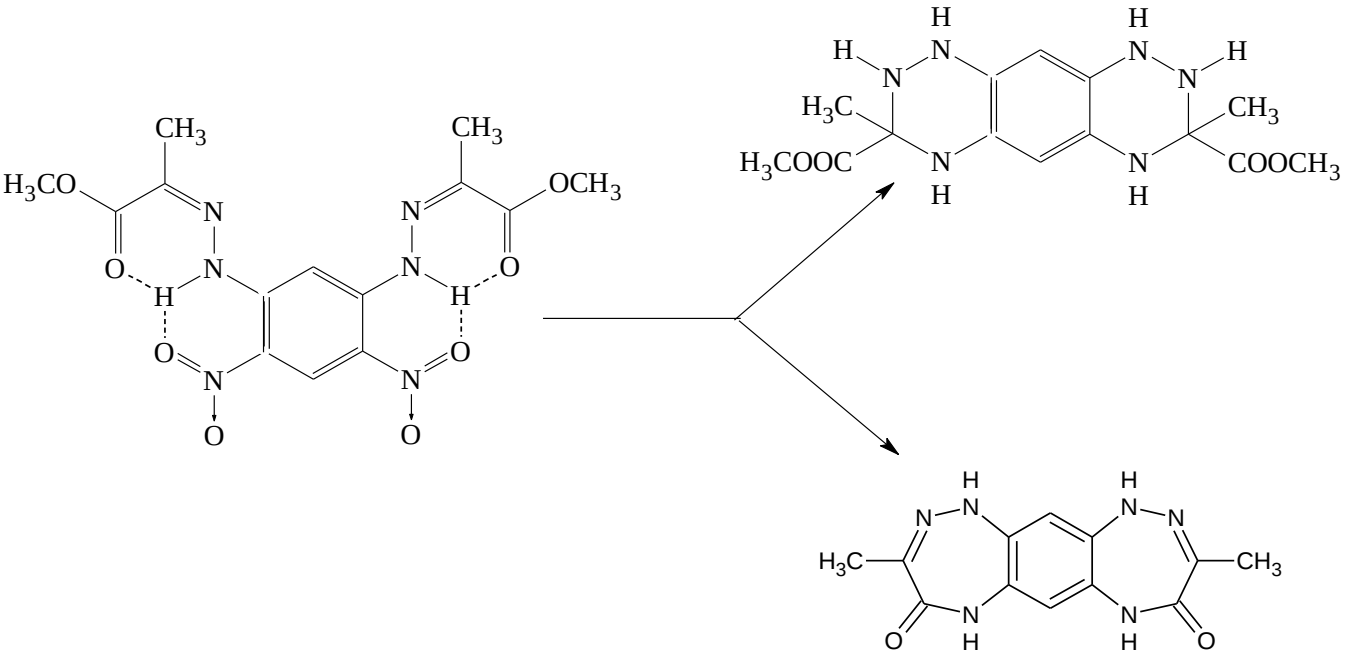
Figure III.4. IR spectrum of (Z/Z)-1,3-hydrazono-4,6- dinitrophenylenoxaloacetate of dimethyl

II.4. Reductions

After the synthesis of hydrazons, we chose to make reduction reactions by using sodium dithionite.

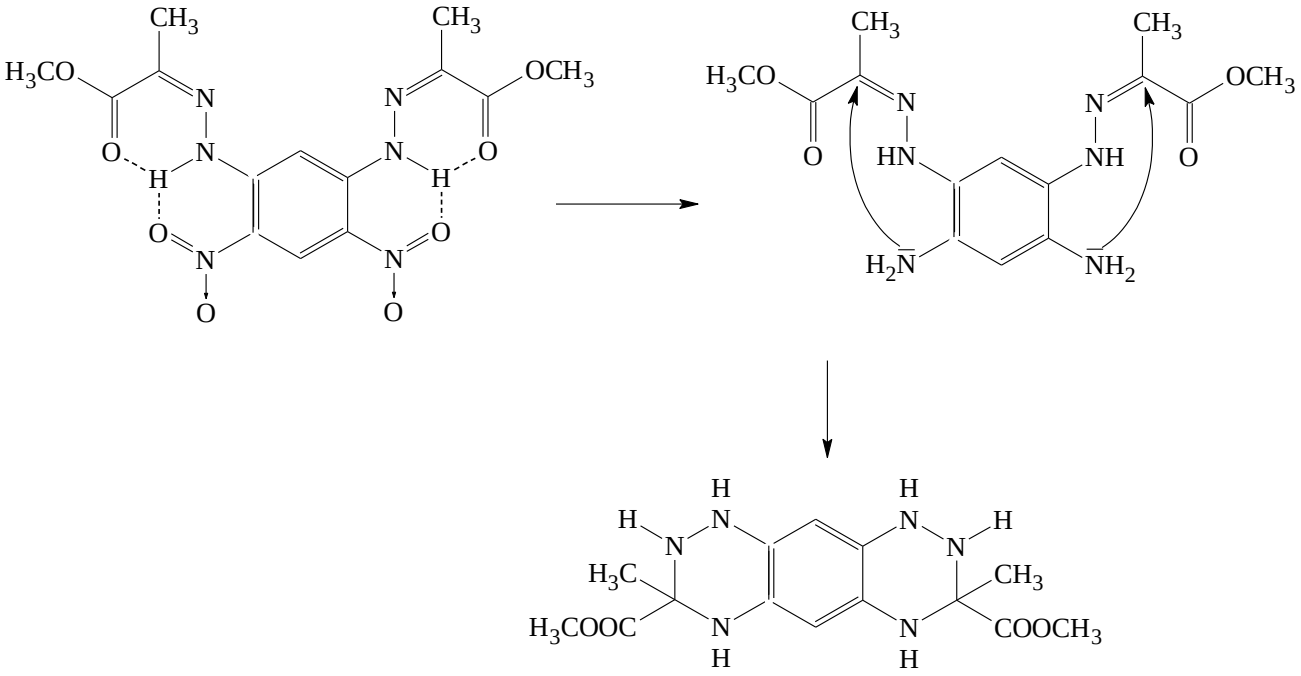
III.4.1. Reduction of (Z)/(Z)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate by sodium dithionite

In principle, two possibilities of reductive cyclisation could occur by reduction of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate



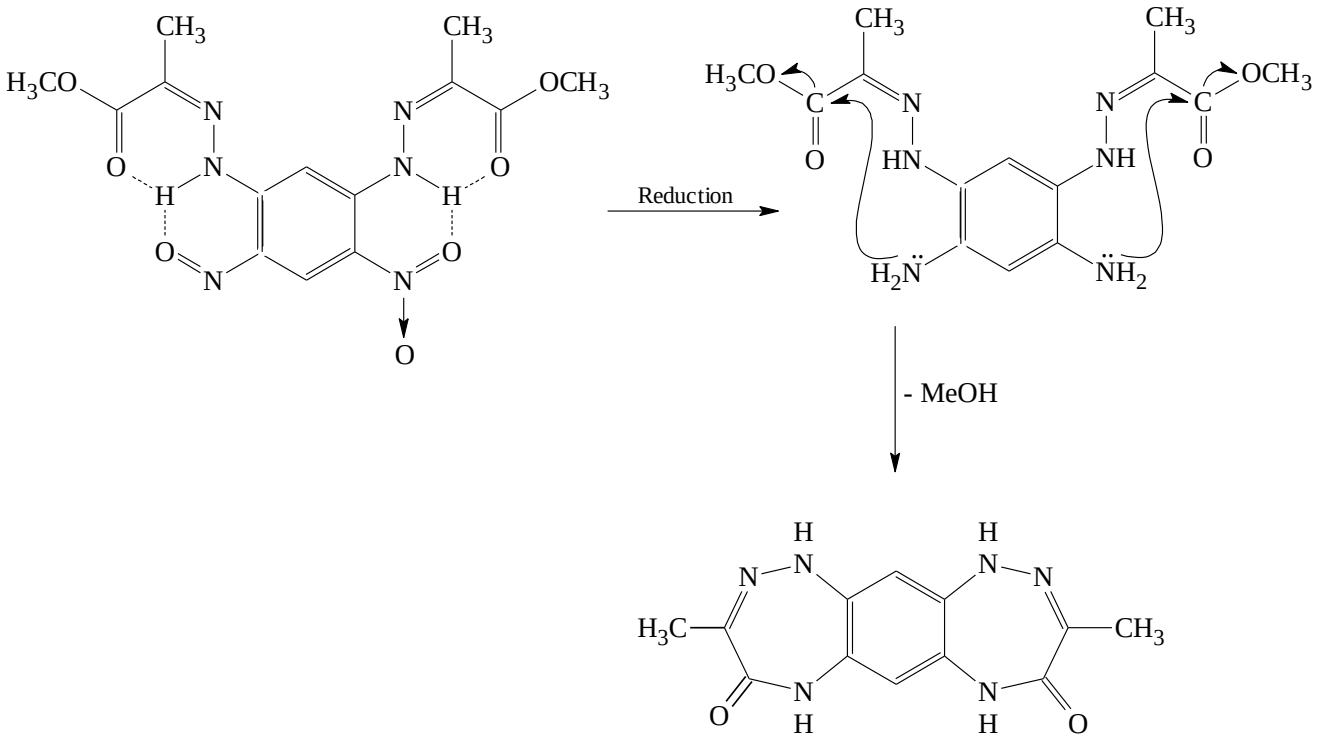
Scheme III.4: Possibility of reductive cyclisation of the compound (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate by sodium dithionite.

After reducing the two nitro groups to the amino groups, nucleophilic additions on the carbon atoms of the two azomethine N=C bonds should lead to the dimethyl derivative 3,7-dimethyl-1,2,3,4,6,7,8,9-octahydro[1,2,4]triazino[5,6-g][1, 2,4]benzotriazine-3,7-dicarboxylate.



Scheme III.5: Reaction mechanism of the possible reductive cycling of (*Z/Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate leading to the triazinio benzotriazine derivative.

Similarly, when the nitro group is reduced to the amino group, the latter may nucleophilically attack the carbonyl group in a favorable location, which would eliminate the methanol and result in a derivative. 3,9-dimethyl-7,11-dihydro-1.2.5triazepine[4,3-h][1,2,5] 4,8(1*H*,5*H*)-benzotriazepine.



Scheme III.6: Reaction mechanism of possible reductive cycling of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenmethyl pyruvate leading to the triazépino benzotriazepine derivative.

In a previous work [Lettreuch, H., Khodja, M. & Boutoumi, H. New Synthetic Route to Cinnoline Derivatives and Their Microbiological Activity. Russ J Org Chem 56, 2188–2193 (2020). https://doi.org/10.1134/S1070428020120222],we have synthesized new cinnoline derivatives by reducing methyl (*Z*)-2-nitrophénylhydrazinopyruvate derivatives using excess sodium dithionite. In our current work, we have applied the same reduction method to the obtained bishydrazone. For this purpose, methyl (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenpyruvate is suspended in methanol, and after complete dissolution of the reagent, a water solution of excess sodium dithionite.

The IR spectrum shows the presence of a single vibrational extension mode of the NH link towards 3224.07cm^{-1}, at 1730.5cm^{-1} strong bands of extension absorptions characteristic of carbonyl group (C=O), and also shows a characteristic extension band of the absorption (CN) to 1604.46cm^{-1}, as well as the nitro group absorbtion (NO_2) towards 1426.9cm^{-1}. (Figure III.).

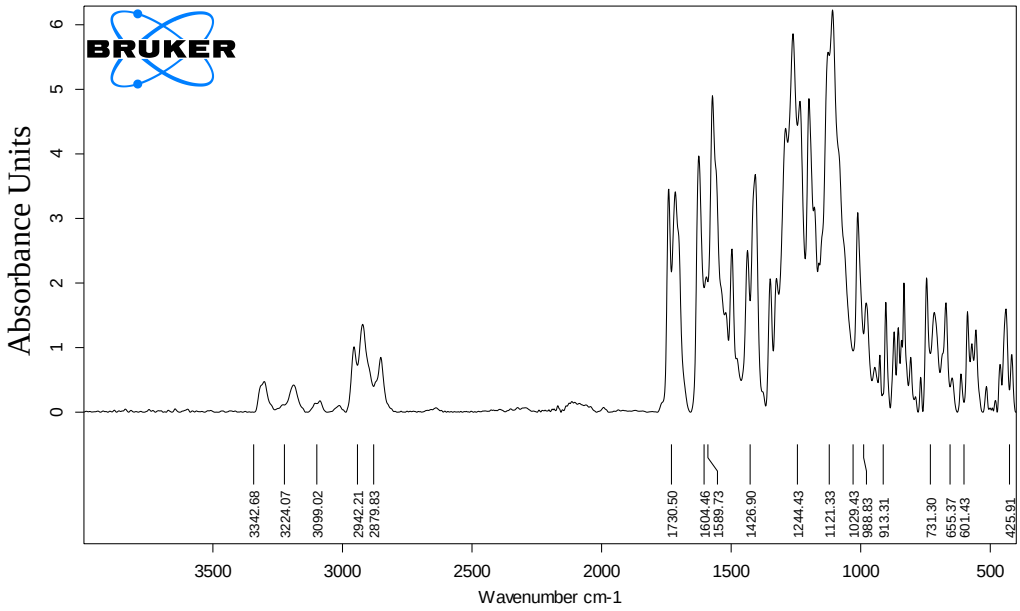
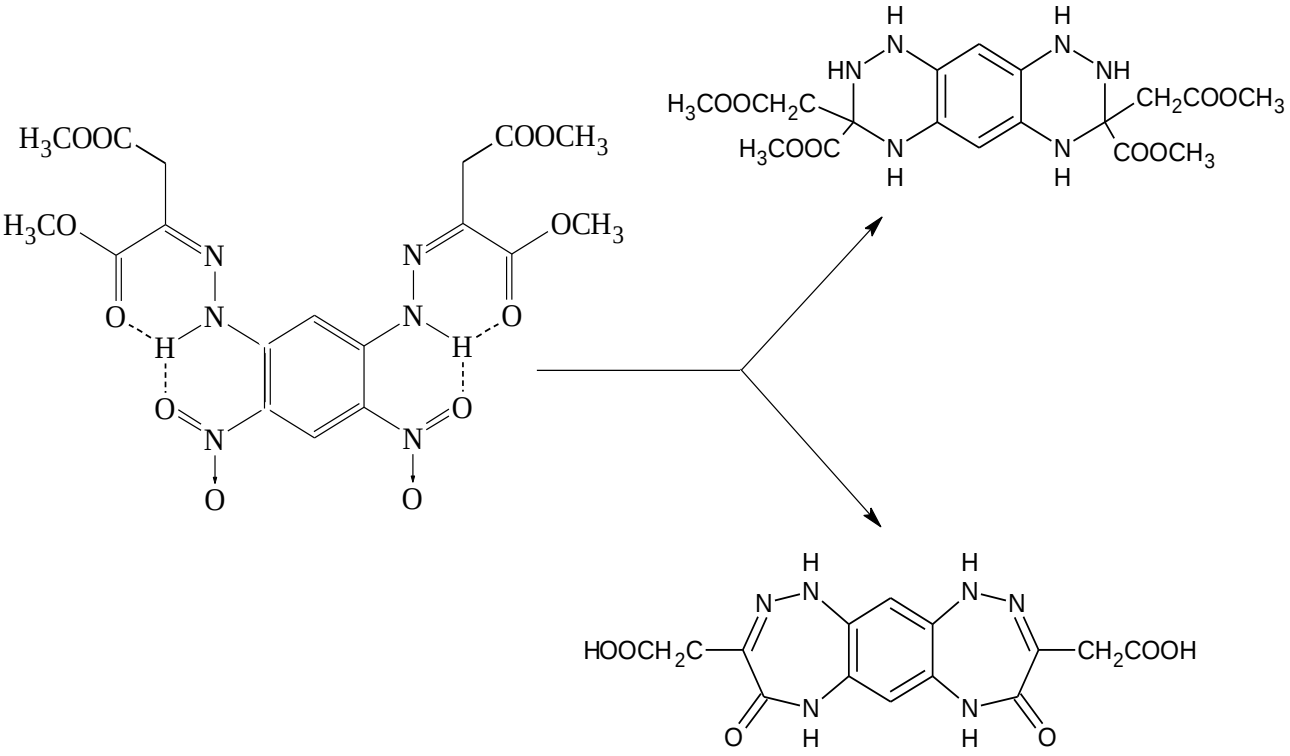


Figure III.5: IR spectrum of resulting product from reduction 1.

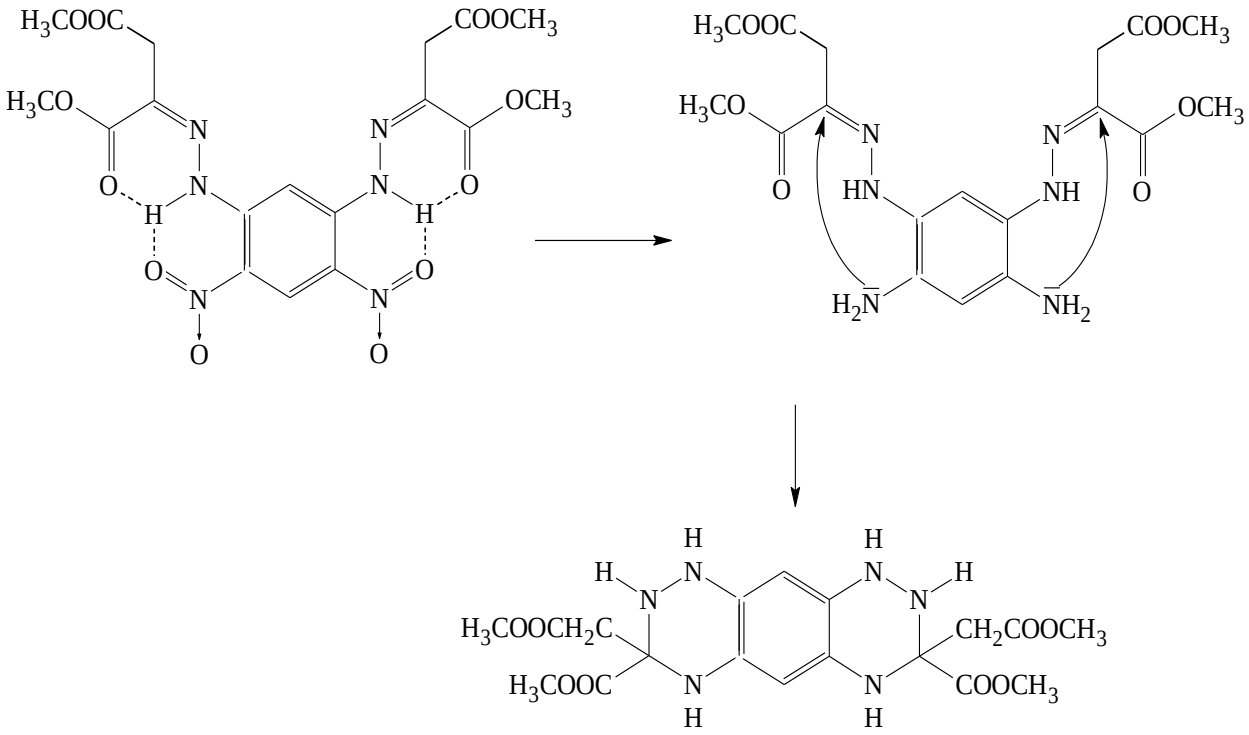
III.4.2. Reduction of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl by sodium dithionite

The isomer (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate of dimethyl and two more reductive cycling options are also considered.



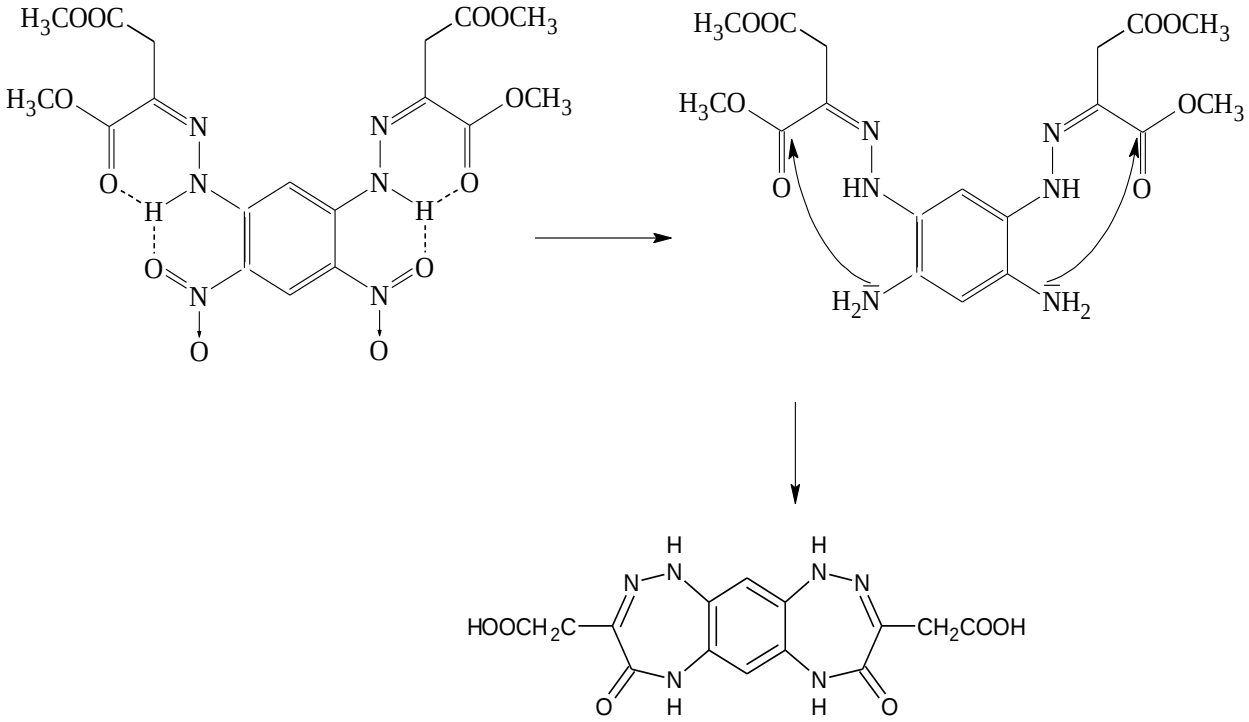
Scheme III.7: Possibility of reductive cyclisation of the compound (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl.

As previously mentioned, reduction of these (*Z*/*Z*)-isomers can reduce the nitro group to the amino group, which can also produce a derivative.



Scheme III.8: Reaction mechanism of the possible reductive cyclisation of (*Z*/*Z*)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl leading to the triazinio benzotriazine derivative.

Similarly, when the nitro group is reduced to the amino group, the latter may nucleophilically attack the carbonyl group in a favorable location, which would eliminate the methanol and result in a derivative. 2,2'-(4,8-dioxo-1,4,5,7,8,11-hexahydro[1,2,5]triazepino[4,3-*h*][1,2,5]benzotriazepine-3,9-diyl)diacetic acid



Scheme III.9: Reaction mechanism of the possible reductive cyclisation of (Z/Z)-1,3-dihydrazono-4,6-dinitrophenylenoxaloacetate dimethyl leading to the triazépino benzotriazepine derivative.

III.5. Evaluation of the antimicrobial activity of synthesized Heterocycles using QSAR model

III.5.1. Selection of representative descriptors

The selection was made using the Stepwise software revealing the seven most relevant and representative descriptors: ATSC3c, ATSC0c, nBondsT, GATS6i, GATS7v, AATS2i, minHBint4.

- **ATSC3c:**

The principle of this descriptor lies in leveraging the spatial distribution and electronic properties of atoms within a molecule, as captured by Gasteiger charges, to compute autocorrelation values that inform about the molecule's structure, behavior, and properties within the Moreau-Broto framework.

- **ATSC0c:**

Captures the electronic properties of atoms within a molecule based on their Gasteiger charges, providing valuable information for predicting molecular properties and activities in chemoinformatics and QSAR studies.

- **nBondsT:**

The "number of triple bonds in non-kekulized structure" is a descriptor used in chemistry to quantify the number of triple bonds present in a molecule when it's represented in a non-Kekulé structure.

In Kekulé structures, which are commonly used for representing aromatic compounds like benzene, each carbon-carbon bond is typically depicted as alternating single and double bonds. However, in certain molecules, especially those with delocalized electrons or resonance structures, this representation might not accurately reflect the true nature of the bonding. In such cases, a non-Kekulé structure is used, where the electron density is delocalized over multiple bonds, including fractional bond orders.

- **GATS6i :**

Geary coefficient of lag 6 weighted by the first ionization potential **;** the **First Ionization Potential (FIP)** is the energy required to remove the outermost electron from an atom in its gaseous state and **Lag 6** refers to the distance or interval between observations. So, lag 6 would imply that you're looking at the relationship between observations that are six units apart.

To incorporate the first ionization potential as a weight in the calculation of the Geary coefficient. Presumably, will use the first ionization potential as a measure of spatial relationship strength between observations.

- **GATS7V :**

Geary coefficient of lag 7 weighted by Ven der waals volume, it suggests that the spatial autocorrelation is being calculated with a specific spatial lag. A spatial lag of 7 implies considering the spatial relationships or interactions between points that are 7 units apart in some defined space.

Van der Waals (vdW) volume refers to the volume occupied by an atom or molecule when it is in close contact with other atoms or molecules, accounting for the space taken up by the electron cloud. In the context of the Geary coefficient, weighting by vdW volume will reflect not just the spatial relationships but also how the physical volume of the molecules influences these

relationships, providing a more nuanced understanding of spatial patterns in molecular structures or properties.

- **AATS2i:** Average Atom-type E-State Index. The "E-State" (Electrotopological State) is a descriptor that combines electronic, topological, and valence state information for atoms in a molecule. The number 2 here indicates the path length, which is the number of bonds separating the atoms in the molecule. this descriptor can be used to capture information about the molecular structure that is useful for predicting various properties and activities of chemical compounds. For example, in drug design, it can help in understanding how changes in molecular structure might affect biological activity.
- **minHBintu:**

The Minimum E-state Descriptors (MESDs) for potential hydrogen bonds of path length 4 refer to a set of molecular descriptors used in computational chemistry and cheminformatics. These descriptors encode the electronic structure of molecules, focusing specifically on identifying the minimum energy states within a molecular system that are indicative of potential hydrogen bonding interactions. The "path length 4" specifies that the hydrogen bond occurs between atoms separated by four covalent bonds. MESDs aim to predict the strength or likelihood of hydrogen bonding at specific sites within a molecule, aiding in the understanding of molecular properties such as solubility, reactivity, and conformational stability.

Table III.1: Relevant descriptors.

Descriptor Name	Dimension	Full Name
ATSC3c	2D	Autocorrelation of lag 3 weighted by gasteiger charge centered moreau-broto
ATSC0c	2D	Autocorrelation of lag 0 weighted by gasteiger charge
NBondsT	2D	Number of triple bonds in non-kekulized structure

GATS6i	2D	Geary coefficient of lag 6 weighted by the first ionization potencial
GATS7v	2D	Geary coefficient of lag 7 weighted by vdw volume
AATS2i	2D	Autocorrelation of lag 2 weighted by ionization potential
minHBint4	2D	Minimum E-state descriptors of strength for potential hydrogen bonds of path length 4

Table III.2: Selected descriptors.

	ATSC3c	ATSC0c	nBondsT	GATS6i	GATS7v	AATS2i	minHBint4
01	0.028642196	0.264518498	0	0.837818673	0.966360908	152.6977263	6.223948023
02	0.048938591	0.266215144	0	0.911721708	0.889165641	151.3660245	6.39292311
03	0.025203742	0.277864545	0	0.895756681	0.674962486	150.9481926	0.437042042
04	0.16324071	0.467789195	0	0.821231795	1.275670664	156.4233671	0
05	0.192808018	0.455733974	0	0.837601135	1.209921489	156.7537458	0
06	0.209795187	0.614962387	0	0.960537437	1.105664374	154.1089578	0
07	0.18072392	0.613406973	0	0.990745477	1.098061401	154.7906331	0
08	0.181537823	0.613369374	0	0.889805796	1.313128928	154.7906331	0
09	0.225739731	0.61409453	0	1.101605386	1.053219435	154.7906331	0
10	0.157612237	0.627528015	0	0.937243911	1.206252101	153.8129935	0
11	0.120283784	0.665374232	0	1.129445101	1.136545796	157.2267248	0.389513935
12	0.295423099	0.495395922	0	1.078986245	0.841082899	153.2327585	0
13	0.268676606	0.495730056	0	1.026699327	1.02864165	153.223693	0
14	0.296338974	0.500517204	0	1.079838433	1.005918845	153.2815506	0
15	0.222354563	0.500103494	0	0.989812633	1.066608791	153.2815506	0
16	0.250016931	0.504890641	0	1.040978396	1.041875735	153.3316939	0
17	0.206140116	0.485853916	0	0.595321938	0.814627629	150.7369316	0.419047302
18	0.218226292	0.517087309	0	1.010150618	1.0152889	155.6816945	0
19	0.235920466	0.514560476	0	0.733469062	0.547271475	153.0158708	0
20	0.299579298	0.480521944	0	1.036641385	0.408294731	160.8791925	0
21	0.248900198	0.545799709	0	1.053895709	0.989104186	156.1098555	0
22	0.248899083	0.537314035	0	1.060017585	0.990675242	156.3029862	0
23	0.249604535	0.503338804	0	1.203615562	1.180578992	157.6477486	0

24	0.265418671	0.554655872	0	1.030989864	1.0159078	149.8219542	0
25	0.064090941	0.488077711	2	0.962087944	1.053246548	150.750614	8.041909161
26	0.218225177	0.508601635	0	1.0124235	1.020658258	155.8386463	0
27	0.22625914	0.498800908	0	1.042346017	1.137035366	154.9149305	0
28	0.257377952	0.527513798	0	1.121415492	1.027193411	156.0144217	0
29	0.246050422	0.672276797	0	1.129686057	1.041115296	153.8781599	0
30	0.214822714	0.643563875	0	1.088424316	1.061905369	152.9892233	0
31	-0.087570216	0.542106212	0	1.038345835	1.113482634	154.0732164	0

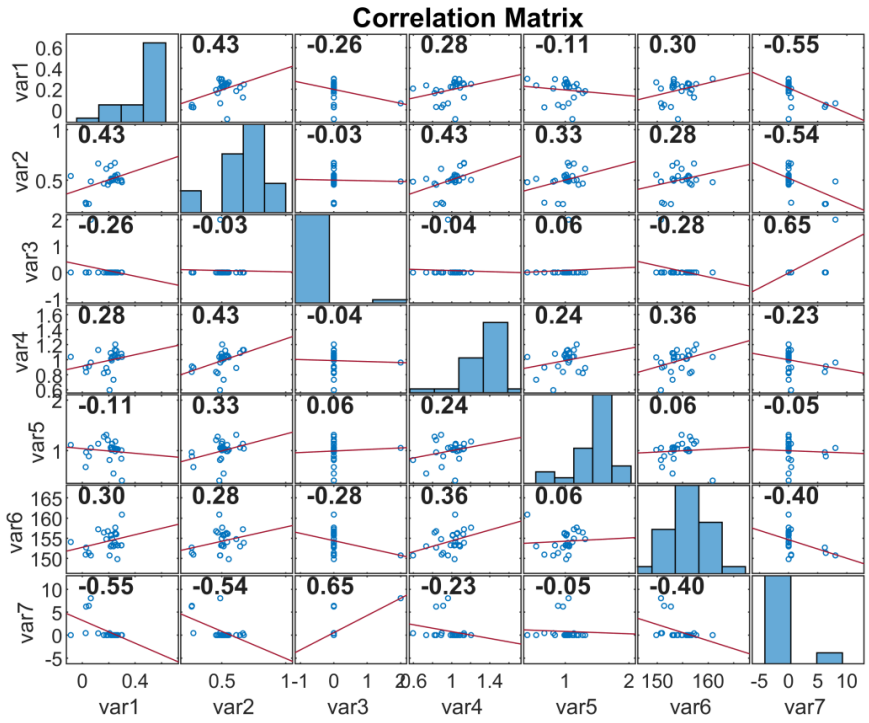


Figure III.6: correlation plot.

Positive R^2 values indicate a positive correlation, while negative R^2 values indicate a negative. This can be useful for understanding the relationships and interactions between chemical descriptors and their influences on the properties or responses of the compounds studied.

$R^2 > 0.75$ indicates a strong linear relationship, however a $R^2 < 0.75$ indicates a little to no linear relationship.

The histograms that lie on the line stretching from the top left corner to the bottom right corner of the matrix refers to the diagonal. These histograms correspond to the pairings of each variable with itsef.

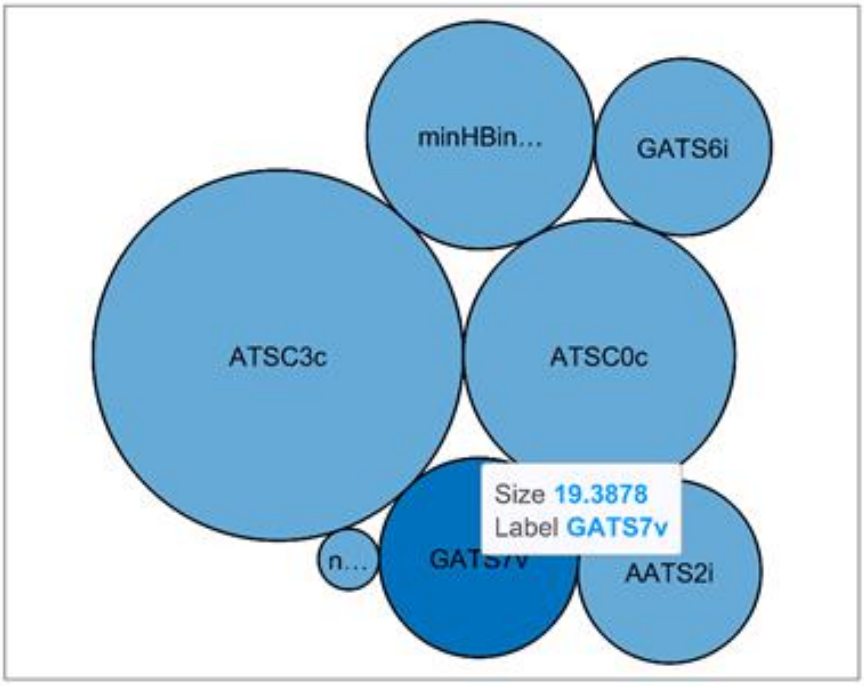


Figure III.7: Bubble Cloud of the Descriptors's importance.



Figure III.8: Word Cloud of the Descriptors's importance.

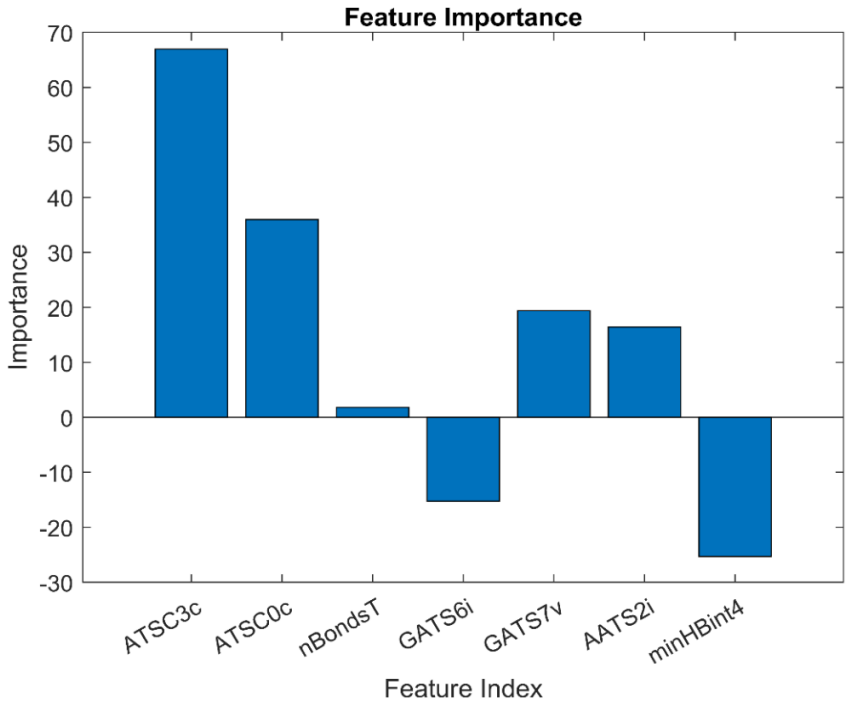


Figure III.9: Bar plot of the relative importance of descriptors.

The bar plot shows the relative importance of different descriptors in a feature importance analysis.The x-axis lists the different descriptors: ATSC3c, ATSC0c, nBondST, GATS6i, GATS7v, AATS2i, and minHBint4.The y-axis represents the importance values of these descriptors. These values indicate how significant each descriptor is in the context of the model.

- ATSC3c has the highest importance value, around 65%, making it the most influential descriptor.
- ATSC0c has a moderate importance value, around 35%.
- GATS7v and AATS2i both have positive importance values, around 15% each.
- nBondST has the lowest importance value, around 0.2%.
- GATS6i, and minHBint4 have negative importance values, with minHBint4 having the most negative value around -20, indicating these descriptors negatively influence the model.

Descriptors with positive values contribute positively to the model, while those with negative values contribute negatively.

ATSC3c and ATSC0c are key positive contributors, while minHBint4 is a significant negative contributor.

The bar plot provides insights into which features are most and least influential in the model.

- Interpritation based on organic chemistry :

The descriptors obtained have shown that there is a direct relationship between the structure (structure in space, sterile congestion) of the molecule and its biological activity. Electronic properties such as dipolar moment (electrical dipole) that is involved in the formation of the bond, as well as the relocation of electrons in a molecule that generates the formation of double and triple bonds. The energy of the first ionization allows to predict the molecule's ability to form single electrons that form covalent bonds. Van der walls bonds directly influence the structure of the molecules studied by the formation of diastérioisomers and their stability.

III.5.2. QSAR model development (SVM)

III.5.2.a. SVR model

In this study, a SVR model was developed to predict the minimum inhibition concentration of synthesized molecules.

A kernel function is required for the implicit mapping of learning models in the functionality space.

The choice of kernel and kernel parameters can be automated by optimizing the selection criteria of the SVR model and are selected by a trial and error procedure.

The choice of the hyper-parameters of the SVR model (C, ε and γ) allow us to predict the MIC.

The chosen Kernel function and the best combination of the hyper-parameters were found to be $C = 500$, $\gamma = 5$, $\varepsilon = 0.8650$, and the Radial Basis Function RBF as the kernel function.

Table III.3: Statistical criteria for the SVR model

Validation criteria	Train	Test	Global
R^2	0.9715	0.9003	0.9509
R_0^2	0.9705	0.8186	0.9455
$R_0^{2'}$	0.9690	0.7117	0.9387

Q2f1	0.9949	0.9669	0.9907
Q2f2	0.9696	0.8064	0.9454
CCC	0.9839	0.8886	0.9705
RMSEP	0.4147	0.93	0.54

- **R^2:**

The determination coefficient, is a number that ranges from 0 to 1 that indicates how much of the variation in the data is explained by the model. An evaluation of a model's capacity to forecast a molecule's biological activity from its molecular descriptors is frequently done using the determination coefficient R2. A high R2 value suggests that the model can effectively capture and predict quantitative correlations between chemical structures and activities, indicating a strong link between molecular descriptors and biological activity.

- **R_0^2:**

Is often computed using a separate test dataset that was not utilized in the model's construction. It shows the percentage of the test data's overall variance that the model can account for. A low R_0^2 test implies a poorer prediction performance, whereas a high R02 test indicates a superior predictive capacity of the model.

- **$R_0^{2'}$:**

These are squared correlation coefficients that assess how well a QSAR model's anticipated and observed values match or differ from one another. In this case, these are the square correlation coefficients between observed and predicted values obtained using the "leave-one-out" method (when a compound is excluded at once) to evaluate the performance of the QSAR model, with and without intercept (a constant in the model equation)

- **Q2F1:**

Is derived from comparing the predicted values of QSAR with the experimental values. It gauges how well the model predicts the biological activity of various chemical substances. A high Q2F1 suggests that the model is capable of reliably predicting the activities of many substances.

- **Q2F2:**

Is based on the prediction of the compounds of the test set (Q2) using a cross-validation procedure (F2) proposed by Schüürmann et al

- **CCC:**

The criterion CCC is used to evaluate the quality of a QSAR model in terms of accuracy and consistency between predictions and observations. A high CCC value indicates a good match between the model and the actual data

- **RMSEP:**

Root Mean Square Error of Prediction. It's a variation of the RMSE used specifically in the context of predictive modeling.

RMSEP is very similar to RMSE, but it's used when you're predicting new, unseen data rather than evaluating how well your model fits the data it was trained on.

III.5.2.b. Internal validation

- Training data:

The chord scheme of the training data is illustrated in (Figure III.10), which shows the output calculated by the SVR model compared with the experimental data with a chord vector. The coefficient of determination R^2 for the training data set is 0.9715, which indicates that the variation of the training data and the correlation are in good agreement.

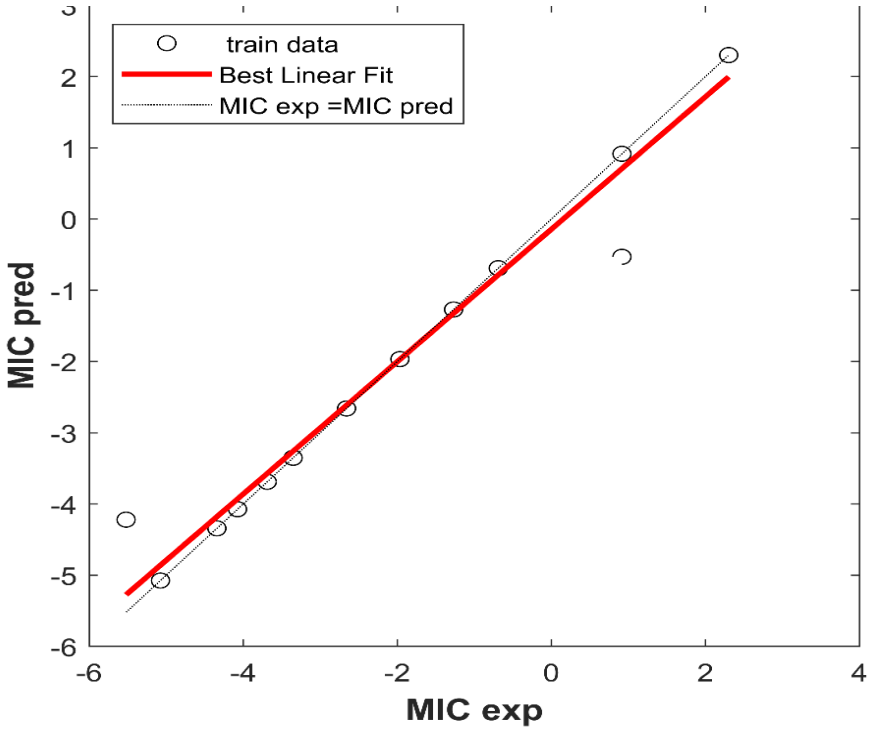


Figure III.10: The Training data.

The results show that the SVR model perform well on the training data. This is shown by the R^2 value (0.9715) and the RMSEP value which is 0.4147, and confirmed by the CCC value, the high root mean square error RMSE and the coefficient of prediction Q^2.

III.5.2.c. External validation

- **Test data:**

The chord scheme of the test data is illustrated in (Figure III.11). It shows that the output values calculated and experimental data are aligned, with a chord vector approaching the ideal.

For all the test data, the coefficient of determination R^2 is 0.9003, which shows a good agreement between the experimental data and the results predicted by SVR.

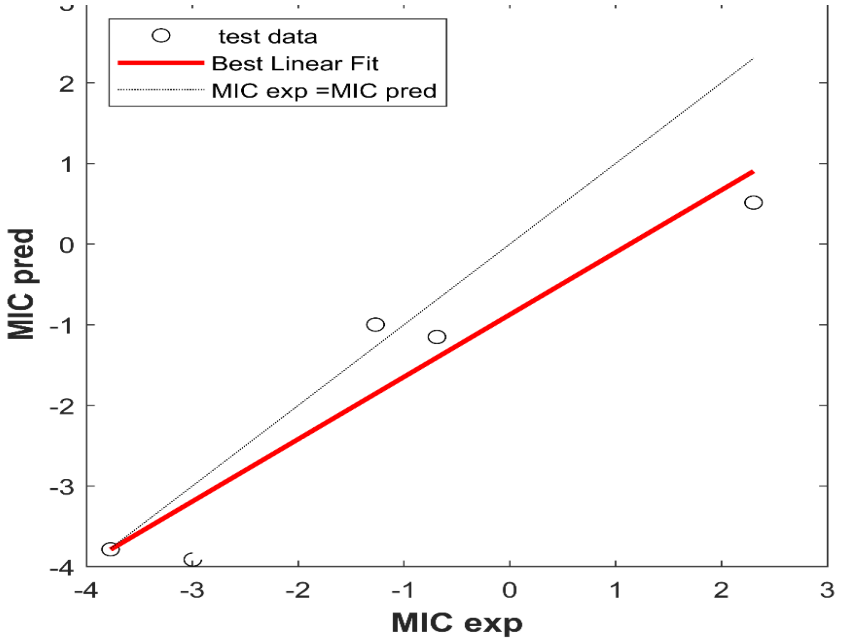


Figure III.11: Test data.

The results show that the SVR model perform well on the test data. This is shown by the R^2 value (0.9003) and the RMSEP value, and confirmed by the other validation criterias.

- Global data:

The agreement scheme of the total data is illustrated in (Figure III.12), which illustrates the output calculated by the SVR model compared with the experimental data with a vector of agreement.

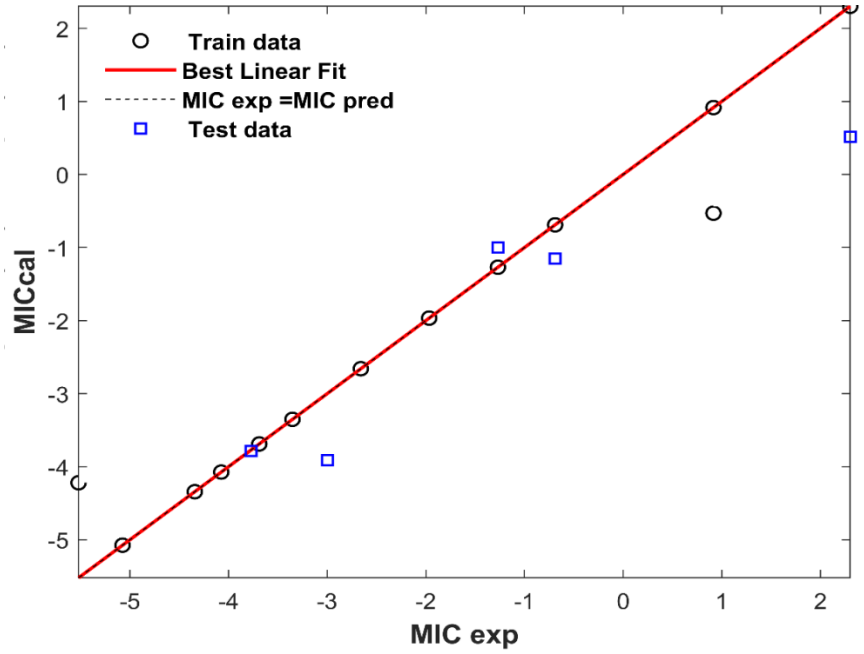


Figure III.12: SVM regression curves.

The modeled numerical results confirmed that the SVR yielded considerably smaller errors. Therefore, the proposed model (SVR) is able to find the optimal concentrations of MIC as a function of (RBF), with a minimum error. The SVR model showed good statistical significance and robustness.

III.5.3. External dataset

The developed SVR model was demonstrated to have very good performance internally, good generalization ability externally, and excellent statistical and mechanistic significance. Thus, it was utilized to predict the unknown MIC of two novelly synthesized molecules. Molecules (NAMES) were synthesized as described in the previous part. Their corresponding SMILES were used to calculate the seven descriptors mentioned above, and then fed into the developed SVR model to predict their MIC.

Table IIII.4: Representative descriptors of synthesized molecules

ATSC3c	ATSC0c	AATS2i	GATS7v	GATS6i	nBondsT	minHBint4
0.72538381	1.07011631	163.479283	1.05881303	0.80739488	0	1.15849674
0.29405333	1.5975166	163.56982	0.96408171	1.02447924	0	1.18653457

The predicted values of MIC by the SVR model are:

The first molecule = 0,33008668mg/ml

The second molecule = 0,86883678mg/ml

These predicted concentrations are relatively small, indicating that these two synthesized molecules are potential potent antimicrobials.

General conclusion

This work aimed to synthesize new bisheterocyclic systems using the following steps:

Bishydrazones were synthesized by condensation of bishydrazine with esters of pyruvic acid and oxaloacetic acid. These compounds were isolated in their diastereomeric (E/E) and (Z/Z) forms and were purified and characterized by Fusion Point and FTIR analysis.

The synthesized bishydrazones were subjected to reduction reactions using excess sodium dithionite.

Reduction reactions yielded products that showed reaction during preliminary characterization. However, the exact structure needs to be confirmed using advanced physical analysis methods such as (RMN, SM), etc…

Other compounds were obtained in low yields, hindering their characterization.

QSAR modeling has shown to be a very successful method for predicting the minimum inhibitory concentration (MIC) of compounds using (SVR), which is a strong and reliable method for estimating the biological activity of chemical compounds and characteristics of an unknown set of molecules based on equation analysis.The model's high predictive accuracy, ability to handle non-linear relationships, strong generalization capability and robustness against overfitting collectively contribute to its effectiveness with hight R^2(0.9715 ,0.9003 ,0.9509) values and low RMSEP(0.4147, 0.93 , 0.54) values ,The Biological Activity Modeling Study (MIC) showed that synthetic products have an MIC of 0.3 and 0.8 mg/m.

The obtained results allowed to validate the use of advanced machine learning techniques to enhance the understanding of structure-activity relationships and drive innovation in the development of novel antimicrobial agents.

The success of this project provides a solid foundation for future research and application of SVR in QSAR modeling, particularly for predicting critical pharmacological parameters such as MIC.

Reference Bibliography

[1] Guillaume Bélanger. (2015). *Chimie Organique Hétérocyclique*.

[2] Farhan, M. M., Al-Jumaili, M. H. A., & Bakr, E. A. (2023). New bis-heterocyclic structures: Synthesis, characterization and biological activity. *Journal of Chemical Research*, *47*(6).

[3] Salat, K., Moniczewski, A., & Librowski, T. (2013). Nitrogen, oxygen or sulfur containing heterocyclic compounds as analgesic drugs used as modulators of the nitroxidative stress. *Mini Reviews in Medicinal Chemistry*, *13*(3), 335–352.

[4] Raymond K.Machie and David M.Smith. (1983). *Organic Synthesis* .

[5] HM Patel and all, (2014). Quantitative structure–activity relationship (QSAR) studies as a strategic approach in drug discovery. *Medicinal Chemistry Research*.

[6] Formenti, D., Ferretti, F., Scharnagl, F. K., & Beller, M. (2019). Reduction of Nitro Compounds Using 3d-Non-Noble Metal Catalysts. *Chemical Reviews*, *119*(4), 2611–2680.

[7] Kadam, H. K., & Tilve, S. G. (2015). Advancement in methodologies for reduction of nitroarenes. *RSC Advances*, *5*(101), 83391–83407.

[8] Anet, F. A. L., & Muchowski, J. M. (1960). Lithium Aluminum Hydride Reduction Of Sterically Hindered Aromatic Nitro Compounds. *Canadian Journal of Chemistry*, *38*(12), 2526–2528.

[9] The structure of the compound formed, when nitrobenzene is reduced by lithium aluminum hydride($LiAlH_4$) is (collegedunia.com), Sunday, April 28, 2024

[10] Severin, T., Loske, J., & Scheel, D. (1969). Umsetzungen von Nitroaromaten mit Natriumborhydrid, V. *Chemische Berichte*, *102*(11), 3909–3914.

[11] Popat, V., & Padhiyar, N. (2013). Kinetic study of bechamp process for P-nitrotoluene reduction to P-toluidine. *Int. J. Chem. Eng. Appl*, *4*(6), 401-405.

[12] Kaplánek, R., & Krchňák, V. (2013). Fast and effective reduction of nitroarenes by sodium dithionite under PTC conditions: application in solid-phase synthesis. *Tetrahedron Letters*, *54*(21), 2600–2603.

[13] K.B. Abireimov , N.S ,.Mukhamedov, M.Zh . Aiymbetov and Kh.M.Shakhidoyatov , 2010,synthesis of nitro heterocycle by ciclisation reductive from 2-nitrophenylhydrazon of acid 2-oxo carboxyliques *Chemistry of Heterocyclic Compounds* , 46 , 941

[14] Prokopov, A.A., Yakhontov, L.N. (1977).Synthesis of 6-azaindole. , *Chemistry of heterocyclic compounds* .13, 919.
[15] .P. Diana and all , 2009,pyrazolo[3,4-b]quinolin-5-amines, *ARKIVOC*, 177.
[16].Yu.N.Tkachenko,E.B.Tsupark and A.F.Pozharkii, synthesis of nitro heterocycle by ciclisation reductive ,2000 , *Chemistry of heterocyclic compounds* .
[17].A.P.Krapivko, E.A.Savitkina, Kuanu,A.Antares, A.A.Astakhov and A.V.Varlamov, 1966 , chemistry of heterocyclic compounds , 290
[18].R.S.Begunov and G.A.Ryzvanovich , 2009, chemistry of heterocyclic compounds,1224
[19].Hossein Naeimi , Nasrin Alishahi , 25 Juillet 2014, Journalof Industrial and Engineering Chemistry , Volume 20 , Issue 4 , pages 2543-2547
[20].Meng –Yang Chang , Yi-Chia Chen , chieh-kai Chan , Tetrahedronlettersvol 55, Mars , 2014, pages 2876-2878
[21]. U.Heinz, K.Hegetschweiler, P.Acklin, B.Faller, R.Lattmann, H.Schnebli, Angew.Chem., 1999, Int. Ed. Engl , 38, 2568-2570.
[22] Güven, Ö. Ö., Erdoğan, T., Göker, H., & Yıldız, S. (2007). Synthesis and antimicrobial activity of some novel phenyl and benzimidazole substituted benzyl ethers. *Bioorganic & Medicinal Chemistry Letters*, *17*(8), 2233–2236.
[23] B. R. Dinish, A. R. Baba, K. U. Sankar, D. C. Gowda, (2008), Jornal of Chem. Researches, 5, P. 287-288.
[24] Khodja, M., Moulay, S., Boutoumi, H., & Wilde, H. (2006). Two-step syntheses of 3-methyl and 3-phenyl-1, 2, 4-benzotriazines. *Heteroatom Chemistry: An International Journal of Main Group Elements*, *17*(2), 166-172.
[25] Khodja, M., Sicker, D., & Wilde, H. (1994). Synthesis of 3, 4-dihydropyridazino [1, 6-a] benzimidazoles. *Heterocycles (Sendai)*, *37*(1), 401-411.
[26] Kentaro Okuma, Jun-ichiSeto, NoriyoshiNagahora, and Kosei Shioji,(2016) Synthèse de nouveaux dérivés hétérocycliques azotés
[27]Benin, V., Kaszynski, P., Pink, M., & Young, V. G. (2000). Formation of 1,10-disubstituted benzo[c]cinnolines. Synthesis and molecular structure of 1-amino-10-propylthiobenzo[c]cinnoline and cyclization to 4-propylcinnolino[5,4,3][c,d,e][1,2]benzothiazine. Journal of Organic Chemistry, 65(20), 6388–6397. https://doi.org/10.1021/jo0002691

[28] Valderrama, J. A., Pessoa-Mahana, H., Sarras, G., & Tapia, R. (1999). ChemInform Abstract: Access to Quinazolines from 2-Nitrobenzaldehyde and Arylamines. *ChemInform*, *30*(52).
[29] K. I. Kobrakov, A. G. Ruchkina, and I. I. Rybina, 2003, Chemistry of Heterocyclic Compounds, 3.
[30] Kundu, S. K., Mahindaratne, M. P. D., Quintero, M. v., Bao, A., & Negrete, G. R. (2007). One-pot reductive cyclization to antitumor quinazoline precursors. *Arkivoc*, *2008*(2), 33–42.
[31] Desai, N. C., Rajpara, K. M., & Joshi, V. v. (2012). Synthesis and characterization of some new quinoline based derivatives endowed with broad spectrum antimicrobial potency. *Bioorganic & Medicinal Chemistry Letters*, *22*(22), 6871–6875.
[32] Sonawane, A. E., Pawar, Y. A., Nagle, P. S., Mahulikar, P. P., & More, D. H. (2009). Synthesis of 1,4-Benzothiazine Compounds Containing Isatin Hydrazone Moiety as Antimicrobial Agent. *Chinese Journal of Chemistry*, *27*(10), 2049–2054.
[33] N. Gautam, D. H. D. C. G. (2005). Antifungal activity of some 4H-1, 4-benzothiazine compounds. *Oriental Journal of Chemistry*.
[34] Abbas, E. M. H., & Farghaly, T. A. (2010). Synthesis, reactions, and biological activity of 1,4-benzothiazine derivatives. *Monatshefte Für Chemie - Chemical Monthly*, *141*(6), 661–667.
[35] Devreux, V., Wiesner, J., Jomaa, H., van der Eycken, J., & van Calenbergh, S. (2007). Synthesis and evaluation of α,β-unsaturated α-aryl-substituted fosmidomycin analogues as DXR inhibitors. Bioorganic & Medicinal Chemistry Letters, 17(17), 4920–4923. https://doi.org/10.1016/j.bmcl.2007.06.026
[36] Pribut, N., Kaiser, T. M., Wilson, R. J., Jecs, E., Dentmon, Z. W., Pelly, S. C., Sharma, S., Bartsch, P. W., Burger, P. B., Hwang, S. S., Le, T., Sourimant, J., Yoon, J.-J., Plemper, R. K., & Liotta, D. C. (2020). Accelerated Discovery of Potent Fusion Inhibitors for Respiratory Syncytial Virus. *ACS Infectious Diseases*, *6*(5), 922–929.
[37] Kuzu, B.Gül, S.Tan, M., Menges, N., & Balci, M. (2021). Synthesis of Indolizines by Dimerization of *N*- Propargylated Pyrroles via Allene Intermediates. *ChemistrySelect*, *6*(9), 2366–2372.
[38 Kuzu, B., Gül, S., Tan, M., Menges, N., & Balci, M. (2021). Synthesis of Indolizines by Dimerization of *N*- Propargylated Pyrroles via Allene Intermediates. *ChemistrySelect*, *6*(9), 2366–2372.

[39] Singh, V. (2009). SamiranHutait and SaujayBatra. *Eur. J. Organic. Chem*, *3454*.
[40] Antonow, D., Jenkins, T. C., Howard, P. W., & Thurston, D. E. (2007). Synthesis of a novel C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine-5,11-dione library: Effect of C2-aryl substitution on cytotoxicity and non-covalent DNA binding. *Bioorganic & Medicinal Chemistry*, *15*(8), 3041–3053.
[41] Weber, K., Bauer, A., & Hauptmann, K. (1972). Benzodiazepine mit psychotroper Wirkung, III. N -Aryl- und N -Heteroaryl-1 H -1.5-benzodiazepin-2.4-[3 H .5 H]-dione. Justus Liebigs Annalen Der Chemie, 756(1), 128–138. https://doi.org/10.1002/jlac.19727560112
[42] Solomko, Z. F., Braichenko, V. T., & Malinovskii, M. S. (1972). Reductive cyclization of 3-(2-nitrophenylamino)propionic acid hydrazide. Chemistry of Heterocyclic Compounds, 8(3), 395–395. https://doi.org/10.1007/BF00475314
[43] McCall, W. V., J. Clin. Psychiatry, (2001), Vol. 62 (Suppl.),27.
[44] https://www.sigmaaldrich.com/DZ/fr/products/chemistry-and-biochemicals/building-blocks/heterocyclic-building-blocks , Saturday, April 27, 2024
[45] Martinez-Ariza, G., Ayaz, M., Medda, F., & Hulme, C. (2014). Synthesis of Diverse Nitrogen-Enriched Heterocyclic Scaffolds Using a Suite of Tunable One-Pot Multicomponent Reactions. *The Journal of Organic Chemistry*, *79*(11), 5153–5162.
[46] Martinez-Ariza, G., Ayaz, M., Medda, F., & Hulme, C. (2014). Synthesis of Diverse Nitrogen-Enriched Heterocyclic Scaffolds Using a Suite of Tunable One-Pot Multicomponent Reactions. *The Journal of Organic Chemistry*, *79*(11), 5153–5162.
[47] Babulreddy, A., Hymavathi, R. v., Hussain, Md. M., & Narayana Swamy, G. (2013). Synthesis, Characterization, and *In Vitro* Antimicrobial Activity of Methyleneamine-Linked Bis-heterocycles. *Journal of Heterocyclic Chemistry*, *50*(3), 727–733.
[48] Kassab, R. M., Gomha, S. M., Muhammad, Z. A., & El-khouly, A. S. (2021). Synthesis, Biological Profile, and Molecular Docking of Some New Bis- Imidazole Fused Templates and Investigation of their Cytotoxic Potential as Anti-tubercular and/or Anticancer Prototypes. *Medicinal Chemistry*, *17*(8), 875–886.
[49] kaur, K., Utreja, D., Dhillon, N. K., Anupam, & Buttar, H. S. (2022). Heterocyclic Moieties as Prospective Nematicides: An Overview. *Current Organic Chemistry*, *26*(18), 1703–1724. https://doi.org/10.2174/1385272827666221209094444

[50] Bari, A., Grenier, D., Azelmat, J., Syed, S. A., Al-Obaid, A. M., & Hosten, E. C. (2019). Synthesis and anti-inflammatory activity of diversified heterocyclic systems. *Chemical Biology & Drug Design*, *94*(4), 1750–1759.

[51] Kumar, H., Aggarwal, N., Marwaha, M. G., Deep, A., Chopra, H., Matin, M. M., Roy, A., Emran, T. bin, Mohanta, Y. K., Ahmed, R., Mohanta, T. K., Saravanan, M., Marwaha, R. K., & Al-Harrasi, A. (2022). Thiazolidin-2,4-Dione Scaffold: An Insight into Recent Advances as Antimicrobial, Antioxidant, and Hypoglycemic Agents. *Molecules*, *27*(19), 6763.

[52]https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Bruice)/31%3A_The_Organic_Chemistry_of_Drugs-_Discovery , Sunday, May 19, 2024

[53] https://pharmafactz.com/medicinal-chemistry-introduction-to-quantitative-structure-activity-relationships-qsar/ , Wednesday, April 17, 2024

[54] Hutter, M. C. (2010). Molecular Descriptors for Chemoinformatics (2nd ed.). By Roberto Todeschini and Viviana Consonni. *ChemMedChem*, *5*(2), 306–307.

[55] Danishuddin, & Khan, A. U. (2016). Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today*, *21*(8), 1291–1302.

[56] Recent Advances in Machine-Learning-Based Chemoinformatics: A Comprehensive Review - PMC (nih.gov) ,Wednesday, April 17, 2024

[57] Verma, J., Khedkar, V.M., Coutinho, E.C., (2010), 3D-QSAR in drug design--a review." Curr. Top. Med. Chem., vol. 10, no. 1, pp. 95–115.

[58] Golbraikh, A., & Tropsha, A. (2002). Beware of q^2, *Journal of Molecular Graphics and Modelling*, *20*(4), 269–276.

[59] Schuett, J. (2023). Defining the scope of AI regulations. *Law, Innovation and Technology*, *15*(1), 60–82.

[60] Chen, S. H., Jakeman, A. J., & Norton, J. P. (2008). Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, *78*(2–3), 379–400.

[61] Tewari, K., Vandita, S., & Jain, S. (2020). *Predictive Analysis of Absenteeism in MNCS Using Machine Learning Algorithm* (pp. 3–14).

[62] Supervised vs Unsupervised Learning - Javatpoint, Friday, April 26, 2024

[63] Support Vector Machine (SVM) - MATLAB & Simulink (mathworks.com), Saturday, April 27, 2024

[64] Ayouri, M. V., Shahbazian, M., Vafaee Ayouri, M., Moslemi, B., & Taheri, M. (2015). Robust Identification of Hydrocarbon Debutanizer Unit using Radial Basis Function Neural Networks (RBFNNs). *Journal of Automation and Control*, *3*(1), 10–17.
[65] ML | Non-Linear SVM - GeeksforGeeks, Sunday, April 28, 2024
[66] What Is Support Vector Machine? | IBM, Sunday, April 28, 2024
[67] https://www.geeksforgeeks.org/ml-non-linear-svm/?ref=ml_lbp, Sunday, April 28, 2024
[68] Day-12: Kernel SVM (Non-Linear SVM) | by Samet Girgin | PursuitOfData | Medium, Tuesday, April 30, 2024
[69] Support Vector Regression (SVR) using Linear and Non-Linear Kernels in Scikit Learn - GeeksforGeeks, Tuesday, April 30, 2024
[70] Schölkopf, B., Smola, A.J., Learning withKernels, Jun. 05- 2018, Support Vector Machines, Regularization, Optimization, and Beyond." The MIT Press.
[71] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.
[72] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.
[73] David S. Moore and George P. McCabe. (2005). *Introduction to the Practice of Statistics*.
[74] https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp - Tuesday, June 10, 2024
75-https://www.investopedia.com/terms/r/r-squared.asp, Tuesday, June 11, 2024
[76] Madani, A., Benkortbi, O., & Laidi, M. (2024). In silico prediction of the inhibition of new molecules on SARS-CoV-2 3CL protease by using QSAR: PSOSVR approach. *Brazilian Journal of Chemical Engineering*, *41*(1), 427–442..
[77] https://statisticsbyjim.com/regression/root-mean-square-error-rmse/, Wednesday, May 22, 2024
[78]https://deepchecks.com/glossary/mean-absolute-error/,Wednesday, May 22, 2024
[79] Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, *45*(1), 255.
[80] Roy, K. (Ed.). (2017). *Advances in QSAR Modeling* (Vol. 24). Springer International Publishing.

[81] Cortes-Ciriano, I. (2016). Benchmarking the Predictive Power of Ligand Efficiency Indices in QSAR. *Journal of Chemical Information and Modeling*, *56*(8), 1576–1587.
[82] Zuñiga, L., & Calvo, B. (2010). Biosimilars approval process. *Regulatory Toxicology and Pharmacology*, *56*(3), 374–377. https://doi.org/10.1016/j.yrtph.2009.11.004
[83] Kunal Roy. (2017). *Advances in QSAR Modeling*.
[84] Cronin, M. T. D. (2010). *Quantitative Structure–Activity Relationships (QSARs) – Applications and Methodology* (pp. 3–11).
[85] Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, *29*(6–7), 476–488.
[86] el JANATI Ali. (2019). Synthèse de nouveaux hétérocycles dérivant de la 6-chloro et la 6-nitroquinoxaline-2,3-diones : Caractérisation et activités antimicrobienne et anticorrosive. *Toubkal* .

[87] Lettreuch, H., Khodja, M., & Boutoumi, H. (2020). New Synthetic Route to Cinnoline Derivatives and Their Microbiological Activity. *Russian Journal of Organic Chemistry*, *56*(12), 2188–2193.
[88] Sahu JK, Ganguly S, Kaushik A. Synthesis and antimicrobial activity of some novel fused heterocyclic 1,2,4-triazolo [3,4-b][1,3,4] thiadiazine derivatives. J Adv Pharm Technol Res. 2014 Apr;5(2):90-5.
[89] Lettreuch, H., Khodja, M., Moulay, S., & Drioueche, A. (2023). Synthesis and Antimicrobial Activity of New Methyl Pyrazolo[5,1-c][1,2,4]benzotriazine-2-carboxylate Derivatives. *Russian Journal of Organic Chemistry*, *59*(4), 674–678.
[90] Hachama, K., Khodja, M., Moulay, S., Boutoumi, H., Hennig, L., & Sicker, D. (2013). Novel 3-(2-Oxo-2H-benzo[b][1,4]oxazin-3-yl)propanoates from dimethyl-2-oxoglutarate and test of their biological activity. *Journal of Heterocyclic Chemistry*, *50*(2), 413–416.