



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



Faculty of Exact Sciences and Science of Nature and Life

Department of Matter Sciences

Ref : .....

Thesis Presented to obtain the degree of

**Doctorate in Chemistry**

**Option: Pharmaceutical chemistry**

Entitled:

**Multi-combined Drug-likeness, 3D-QSAR Modeling, Molecular docking  
and Molecular Dynamics Analysis of several series of pharmaceutical  
interest**

**Presented by:**

**Fattouche Maroua**

Publicly defended on: 02 / 07 / 2024

**In front of the Jury committee composed of:**

<b>Mr. LANEZ Touhami</b>	Professor	University of El-Oued	President
<b>Mr. BELAIDI Salah</b>	Professor	University of Biskra	Supervisor
<b>Mme. OUASSAF Mebarka</b>	MCA	University of Biskra	Examiner
<b>Mme. MAZRI Radhia</b>	MCA	University of Biskra	Examiner
<b>Mr. HOCHLAF Majdi</b>	Professor	University of Gustave Eiffel of Paris	Guest

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*“None of us, acting alone, can achieve success”*

*Nelson Mandela, May 10, 1994*

*This thesis is profoundly dedicated to all.....*

*To my beloved family (My daughter Aicha,*

*My mother, my father and my siblings).*

*To all those who are dear to me.*

## Acknowledgments

First and foremost, I offer my sincerest prayers to Allah, the almighty, for bestowing upon me this incredible opportunity and granting me the necessary capability to navigate through this endeavor with great success. It is with immense gratitude that I express my utmost appreciation to my supervisor, Prof. **BELAIDI Salah**, whose vast knowledge and logical way of thinking have proven to be invaluable to me throughout this journey. His exceptional understanding, unwavering support, and personal guidance have laid a solid foundation for the development of this thesis, for which I am truly grateful.

I am filled with an immense sense of appreciation as I express my profound gratitude to my esteemed professors, Prof. **OUASSAF Mebarka**, Prof. **MAZRI Radhia** an LMCE Laboratory at the University of Biskra, and Professor **HOUHLAF Majdi** at the University Gustave Eiffel, COSYS/IMSE Laboratory, Champs sur Marne, France, and words cannot express how proud I am of my ability to cooperate with you and gain a wealth of knowledge through your extensive experience. Under your guidance, I have imbibed the values of perseverance, scientific integrity, and the importance of an interdisciplinary approach in broadening my intellectual horizons. I am sincerely grateful for your benevolence, your unwavering support, your unwavering dedication to your work, and your exceptional patience in responding to my inquiries. Your invaluable advice and professional conduct played a pivotal role in the successful completion of this dissertation, and for this I am eternally grateful. Furthermore, I would like to express my gratitude, with their kind agreement to study and evaluate the fruits of my work as members of the committee. Their commitment to scientific excellence and willingness to devote their time and expertise to this endeavor is truly admirable. Allow me to extend my heartfelt thanks to Professor **LANEZ Touhami**, a distinguished professor at the University of El-Oued, for graciously accepting the role of chairing the dissertation of my thesis. His willingness to contribute his expertise and oversee this significant aspect of my academic pursuit is truly commendable and deeply appreciated.

I would like to extend my gratitude to Professor **CHTITA Samir** and every member of his team from the Laboratory of Analytical and Molecular Chemistry at Hassan II University in Casablanca, Morocco. It is important for me to acknowledge the support and guidance provided by him throughout this project. His expertise, commitment to scientific rigor, and willingness to assist have been instrumental in driving my motivation to successfully complete this research endeavor.

Moreover, I would like to take this opportunity to express my profound gratitude to all the members of the group of computational and pharmaceutical chemistry at the LMCE Laboratory, Biskra University. It has been a privilege to work alongside such exceptional individuals who have not only created a friendly and cooperative atmosphere but have also provided invaluable feedback and insightful comments on my work. Their contributions have undoubtedly enriched the quality and depth of my research, for which I am sincerely thankful.

## Abstract

Neuropathic pain syndrome has a profoundly negative and agonizing impact on the lives of the individuals it afflicts. In order to find an effective treatment for this condition, extensive and thorough scientific studies have demonstrated that the  $\sigma_1$  receptor serves as an exceptional target for therapeutic compounds. The 3D-QSAR studies were constructed using the technique of comparative molecular similarity indice analysis (CoMSIA). The outcomes of these studies demonstrated the reliability of CoMSIA model (with  $R^2$  train value of 0.96 and  $Q^2$  value of 0.54) in accurately predicting the activity of various compounds. By assimilating the valuable insights gathered from the field contributors of the 3D-QSAR models and conducting molecular docking studies on the highly potent compound C48, a total of sixteen new compounds were successfully designed to exhibit enhanced efficacy against neuropathic pain. In addition to the comprehensive 3D-QSAR analysis, the newly synthesized compounds were subjected to an absorption, distribution, metabolism, excretion, and toxicity evaluation. This evaluation aimed to assess the pharmacokinetic and toxicological properties of the compounds, providing valuable insight for future in vitro investigations. Calculations DFT of the new compounds Mol2, Mol3, and Mol4, including the analysis of their molecular properties, geometric optimization, frontier molecular orbital (FMOs), density states (DOS), and energy evaluation, In order to compare these compounds with the reference ligand 61w, their respective properties were thoroughly investigated. The most stable orientations for the compounds Mol2, Mol4, and Mol3 were determined to be the ones that yielded the highest stability and efficiency. The significant advancements made in this study, should serve as a strong motivation for future in vitro investigations on these compounds.

**Keywords:**  $\sigma_1$  receptor antagonists; Pyrimidine derivatives; 3D-QSAR; Molecular Docking, ADMET, DFT, FMOs, DOS.

## الملخص

متلازمة آلام الاعتلال العصبي لها تأثير سلبي ومؤلم للغاية على حياة الأفراد الذين تعاني منهم. من أجل إيجاد علاج فعال لهذه الحالة، أظهرت الدراسات العلمية المكثفة والشاملة أن مستقبل  $\sigma 1$  يعمل كهدف استثنائي للمركبات العلاجية. تم إنشاء الدراسات **QSAR-3D** باستخدام تقنية تحليل مؤشر التشابه الجزيئي المقارن (**CoMSIA**).

أظهرت نتائج هذه الدراسات موثوقة نموذج **CoMSIA** (بقيمة  $R^2$  0,96 وتدريب وقيمة  $Q^2$  0,54) في التنبؤ الدقيق بنشاط المركبات المختلفة. من خلال مقارنة القيم المحصل عليها من نموذج **3D-QSAR** وإجراء دراسات الالتحام الجزيئي على **C48** المركب عالي الفعالية، تم تصميم ما مجموعه ستة عشر مركبًا جديدًا بنجاح لإظهار فعالية معززة ضد آلام الأعصاب. بالإضافة إلى تحليل **QSAR-3D** الشامل، خضعت المركبات المصنعة حديثًا لتقييم الامتصاص والتوزيع والإستقلاب والإفراز والسمية.

يهدف هذا التقييم إلى تقييم الخصائص الدوائية والسمية للمركبات، مما يوفر نظرة ثاقبة للتحقيقات المستقبلية في المختبر. حسابات **DFT** للمركبات الجديدة **Mol2** و **Mol3** و **Mol4**، بما في ذلك تحليل خصائصها الجزيئية، والتحسين الهندسي، والمدار الجزيئي الحدودي (**FMOs**)، وحالات الكثافة (**DOS**)، وتقييم الطاقة، من أجل مقارنة هذه المركبات مع الرابطة المرجعي **61w**. تم فحص خصائصها بدقة. تم تحديد التوجهات الأكثر استقرارًا للمركبات **Mol2** **Mol3** **Mol4** على أنها تلك التي أسفرت عن أعلى درجات الاستقرار والكفاءة. يجب أن تكون التطورات المهمة التي تم إحرازها في هذه الدراسة بمثابة دافع قوي للتحقيقات المختبرية *in vivo* المستقبلية على هذه المركبات.

### الكلمات المفتاحية:

سيكما 1، مضادات المستقبل، مشتقات بيريميدين، التحليل الكمي للعلاقة بنية-نشاط ثلاثي الأبعاد، الأرساء الجزيئي، امتصاص-توزيع-معالجة-أطراح-سمية، نظرية الكثافة الوظيفية، المدار الجزيئي الحدودي، حالات الكثافة.



## Contents

<b>Abstract</b> .....	I
<b>Contents</b> .....	III
<b>List of works</b> .....	VIII
<b>List of abbreviations</b> .....	X
<b>List of figures</b> .....	XIV
<b>List of tables</b> .....	XVII
<b>General Introduction</b> .....	2

### CHAPTER I

#### **The pyrimidine and inhibition of human $\sigma$ 1 receptor for neuropathic pain syndrome treatment**

1. The pyrimidine.....	8
1.1 Introduction.....	8
1.2 Pyrimidine ring structure.....	9
1.3 Medicinal Properties of Pyrimidines.....	10
2. Pain: A simplified model.....	11
2.1 Nociceptive input.....	12
2.2 Modulatory output.....	13
2.3 Local control.....	14
2.4 Neuropathic pain.....	16
3. The sigma-1 receptor.....	16
3.1 History and discovery of Sigma-1 Receptor.....	16
3.2 The pharmacology of Sigma-1 Receptor.....	18

References.....	20
-----------------	----

## CHAPTER II

### Overview on Computational methods in drug discovery

1. Introduction.....	25
2. Computer-aided drug design.....	25
2.1 CADD applications in drug discovery and development.....	26
2.2 Classification of CADD.....	28
2.2.1 Ligand-based drug design (LBDD).....	28
2.2.2 Structure-based drug discovery (SBDD).....	28
2.3 Molecular mechanics.....	29
3. Quantitative structure–activity relationship.....	30
3.1 3D-QSAR.....	31
3.1.1 The CoMFA method.....	32
3.1.2 The CoMSIA method.....	32
3.1.3 Alignment of the structures.....	32
3.1.4 Calculation of molecular interaction fields.....	33
3.1.5 Interpretation and validation of a QSAR model.....	34
3.1.5.1 Internal validation.....	34
3.1.5.2 Standard statistical tests and coefficients.....	35
3.1.5.3 External validation.....	36

3.1.6 Applications of the QSAR study .....	37
3.2 Structure-based virtual screening.....	39
4. Molecular docking.....	39
4.1 General protocol.....	40
4.1.1 Ligand preparation.....	41
4.1.2 Target preparation.....	41
4.1.3 Binding site detection.....	42
4.1.4 Docking validation.....	42
4.2 Types of molecular docking.....	43
4.2.1 Rigid docking.....	43
4.2.2 Semi-flexible docking.....	43
4.2.3 Flexible-flexible docking.....	43
4.3 Scoring Functions (SFs).....	45
4.3.1 Physics-based SFs.....	45
4.3.2 Empirical SFs.....	46
4.3.3 Knowledge-based SFs.....	47
4.3.4 Consensus scoring.....	48
4.3.5 Machine-learning-based SFs.....	49
5. Pharmacokinetics Properties and Computational tools employed in ADMET.....	50
6. Quantum mechanics.....	51

7. Density Functional Theory (DFT).....	51
7.1 Conceptual Density Functional Theory (DFT).....	51
7.1.1 Fundamental and Computational Aspects of DFT.....	52
References.....	54

## CHAPTER III

### Materials and methods

1. Introduction.....	62
2. Computational methods.....	62
3. Data sets.....	65
4. Equilibrium structure optimizations.....	66
5. Generation of 3D-QSAR studies and CoMSIA analysis.....	66
6. Molecular docking.....	67
6.1. Preparation of the membrane protein 5HK1.....	67
6.2. Docking of the proposed molecules into the 5HK1 binding site.....	67
7. Prediction of synthetic accessibility and ADMET characteristics.....	67
8. DFT calculations.....	68
References.....	70

## CHAPTER IV

### Results and discussion

1. Introduction.....	75
2. Optimized geometries of the compounds of the series.....	75
3. CoMSIA results.....	75
3.1. CoMSIA model.....	75
3.2. CoMSIA graphical interpretation and contour analysis.....	76
4. Design and selection of new human $\sigma_1$ receptor inhibitors.....	78
5. Docking results and analysis.....	79
6. ADMET prediction and synthetic accessibility.....	84
7. DFT calculations.....	88

7.1. The trend of reactivity and FMO study.....	88
7.2. Density of states (DOS).....	90
References.....	92
<b>General conclusion.....</b>	<b>95</b>
<b>Appendices.....</b>	<b>97</b>

## List of works

### Publications

- [1] **Fattouche, M.**, Belaidi, S., Ouassaf, M., Chtita, S., Al-Mogren, M. M., & Hochlaf, M. (2024). Computational studies of pyrimidine derivatives as inhibitors of human  $\sigma_1$  receptor using 3D-QSAR analysis, molecular docking, ADMET properties and DFT investigation. *Chemical Physics Impact*, 8, 100463.
- [2] Belaidi, S., **Fattouche, M.**, & Lanez, T. (2024). Structural investigations and structure-activity relationships in a serie of new antibiotic 16-membered macrolides. *Journal of Fundamental and Applied Sciences*, 16(1), 36-49.

### International conferences

- [1] **M. Fattouche**, In Silico Analysis and Study by Molecular Modeling in Biomolecules Applied in Drug Design; 4. International Anatolian Congress on Multidisciplinary Scientific Research, 17-20 February 2023, Kars, Turkiye. ORAL.
- [2] **M. Fattouche**, Study By QSAR Modeling Of Several Series Of New Heterocyclic Molecules Of Pharmaceutical Interest; Ahi Evran 3<sup>rd</sup> International Conference on Scientific Research, 3-4 May 2023, Odlar Yurdu University, Baku-Azerbaijan, Turkiye. ORAL.
- [3] **M. Fattouche**, Pharmacophore Modeling and Molecular Docking Studies On New Series of Heterocyclic Molecules; 5. International scientific research and innovation congress, 20-21 May 2023, Ankara, Turkey. ORAL.
- [4] **M. Fattouche**, Studying A New And Effective Multi-Activity Series Using Computational Modeling And Atom Based 2D-QSAR; 5. International scientific research and innovation congress, 20-21 May 2023 Ankara, Turkey, ORAL.
- [5] **M. Fattouche**, Molecular Docking And In Silico Evaluation of The Inhibitory Activity of Certain Isothiazole Derivatives of Pharmaceutical Interest; 10th international Mardin Artuklu scientific researches conference, 19-21 May 2023, Mardin, Türkiye, ORAL.

### National conferences

[1] **M. Fattouche**, Study By QSAR Modeling Of Several Series Of Heterocyclic Molecules Of Pharmaceutical Interest; First National Virtual Conference on chemical Process and Environmental Engineering NVCCPEE 2021, 15-16 December 2021, Biskra, Algeria. POSTER.

[2] **M. Fattouche**, Etude Par Docking Moléculaires Des Interactions Entre L'enzyme NS5B Et Des Dérivés D'isothiazole; 1er Séminaire National sur l'évaluation des Activités Biologiques des Plantes Médicinales et Docking Moléculaire SNABPMDM-1, 22 Mars 2022, University of Mostefa Ben Boulaid (Batna2). Batna, Algeria. POSTER.

[3] **M. Fattouche**, Contribution A La Découverte De Médicaments Par Une Etude Computationnelle Des Séries De Molécules Hétérocycliques; Cinquième Colloque Maghrébin sur la Chimie Hétérocyclique CMCH 5-2022, in webinar by the Faculty of Sciences, Moulay Ismail University, Meknes in Morocco and the Tunisian Association of Heterocyclic Chemistry and its Applications (STCHA). 13-15 Juillet 2022, Tunisia. ORAL.

## List of Abbreviations

### A

**ACD:** Available Chemical Directory

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

**Ala:** Alanine

**Asp:** Aspartic acid

**ATP:** Adenosine Triphosphate

### B

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

**BBB:** Blood-Brain Barrier

### C

**Caco-2:** human colorectal adenocarcinoma cells

**CADD:** Computer Aided Drug Design

**CCK:** cholecystokinin

**CHEMBL:** Chemical European Molecular Biology Laboratory

**CGRP:** Calcitonin Gene-Related Peptide

**CNS:** central nervous system

**COMFA:** Comparative Molecular Field Analysis

**COMSIA:** Comparative Molecular Similarity Indices Analysis

**CSD:** Cambridge Structural Database

**CV:** cross-validation method

**CYP:** cytochrome enzyme

### D

**DFT:** density functional theory



**DHEA:** *Dehydroepiandrosterone*

**DNA:** Deoxyribonucleic Acid

**DOS:** Density of states

**DYLOMMS:** Dynamic Lattice-Oriented Molecular Modeling System

**E**

**EA:** electron affinity

**F**

**FMOs:** The frontier molecular orbitals

**FWHM:** full width at half maximum

**G**

**Glu:** Glutamic a

**H**

**HIV:** Human Immunodeficiency Virus

**HOMO:** Highest Occupied Molecular Orbital

**I**

**IP:** ionization potential

**L**

**LBDD:** Ligand-based drug design

**LEU:** Leucine

**LGO:** leave-group-out

**LMO:** Leave-many-Out

**LOO:** leave-one-out

**LUMO:** Lowest Unoccupied Molecular Orbital

**M**

**MD:** Molecular dynamics

**MDDR:** MDL Drug Data Report

**Met:** Methion

**MM:** molecular mechanics

**MOE:** Molecular Operating Environment

**MRSA:** methicillin-resistant Staphylococcus aureus

**MW:** molecular weight

## **N**

**NCEs:** new chemical entities

**NE:** norepinephrine

**NMDA:** N-methyl-D-aspartate

**NMR:** nuclear magnetic resonance

**NPY:** Neuropeptide Y

**NRM:** nucleus raphe magnus

**NSAIDs:** Non-steroidal anti-inflammatory drugs

## **P**

**PCA:** principal component analysis

**PDB:** Protein Data Bank

**PHE:** Phenylalanine

**PGP:** permeability of glycoprotein

**PLS:** Partial Least Squares

## **Q**

**QM:** quantum mechanics

**QSAR:** Quantitative Structure-Activity Relationship

## **R**

**RMS:** root mean square

**RMSD:** Root-Mean-Square Deviation

**RNA:** Ribonucleic Acid

## **S**

**S1R:** Sigma-1 Receptor

**SAR:** Structure-activity relatio

**SBDD:** Structure-based drug discovery

**SF:** scoring function

**SFs:** Scoring Functions

**SMILES:** Simplified Molecular Input

## **T**

**TRP:** Transient Receptor Potential

**TST:** spino-thalamic tract

**TTC:** thalamo-cortical tract

**Tyr:** Tyrosin

## **V**

**Val:** valine

**VHTS:** Virtual high-throughput screening

**VNUT:** vesicular nucleotide transporter

## List of Figures

### Chapter I

- Figure 1** Pyridine and different isomeric forms of diazine family.....9
- Figure 2** (a) Pyrimidine. (b) Pyrimidine containing natural and synthetic products.....10
- Figure 3** First neuron of the ascending pain pathway, Second ascending neuron leading to supra-spinal centers, Descending modulatory output from supra-spinal centers.....11
- Figure 4** TRPV1, TRPA1, DRG, CLR-RAMP, TST, TTC.....12
- Figure 5** Receptors reducing nociception, Receptors enhancing nociception, Endogenous Opioids.....14
- Figure 6** Extracellular adenosine generated by metabolic break-down of nucleotides activates adenosine receptors that decrease the sodium and calcium permeability of TRP receptors. Adenosine is metabolized by deaminases.....15

### Chapter II

- Figure 1** CADD in drug discovery.....28
- Figure 2** Flow chart of CADD processes.....29
- Figure 3** General shape of the classic potentials used by CoMFA (thin lines) and the potential used by CoMSIA (bold line).....34
- Figure 4** The classical analogy of ligand/target complex as "lock and key" and the current analogy as "hand and glove".....40
- Figure 5** General workflow of molecular docking simulation.....41
- Figure 6** The described strategies for including receptor flexibility in docking simulation...44

<b>Figure 7</b> The SF role as pose selector.....	45
<b>Figure 8</b> The description for physics-based SF.....	46
<b>Figure 9</b> The general pathway of knowledge-based SFs.....	48
<b>Figure 10</b> Workflow of training machine-learning-based SFs.....	49

### Chapter III

<b>Figure 1</b> Structure of the studied compounds. See Table 1 for the designation of R <sub>1</sub> , R <sub>2</sub> , R <sub>3</sub> , R <sub>4</sub> , R <sub>5</sub> and R <sub>6</sub> substituents.....	62
--	----

### Chapter IV

<b>Figure 1</b> The plot showcases the correlation between the experimental activities and the predicted activities as determined by the CoMSIA model.....	76
<b>Figure 2</b> CoMSIA contour maps of C48 compound. ....	77
<b>Figure 3</b> Chemical structures of the proposed new human $\sigma_1$ receptor inhibitors. ....	78
<b>Figure 4</b> Docking interactions of the 5HK1 enzyme with the three best inhibitors Mol2, Mol15 and Mol16 as well with the reference ligand 61W and C48 compound, as well as examining the intricate interactions for the antagonists BD-1047 and S1RA, as well as the agonists PRE-084 and Blarcamesine. ....	82
<b>Figure 5</b> HOMO-LUMO structures of the compound Mol2, Mol3 and Mol4 with ligand of reference 61w.....	89
<b>Figure 6</b> DOS plot of the compound Mol2.....	91

### Appendices

<b>Figure S1</b> Boiled-egg drawing of the proposed new compounds.....	102
<b>Figure S2</b> Docking interactions of 61W.....	104

<b>Figure S3</b> Docking interactions of C48.....	104
<b>Figure S4</b> Docking interactions of Mol1.....	105

## List of Tables

### Chapter II

<b>Table 1</b> Standard tests for model validation.....	36
---	----

### Chapter III

<b>Table 1</b> Series of 54 compounds (Cp.) containing pyrimidine of S1R antagonists investigated presently is for the test set.....	63
--	----

<b>Table 2</b> Series of 54 compounds (Cp.) containing pyrimidine as S1R antagonists with their experimental, predicted and residual values of pKi (CoMSIA model). t is for the test set. The experimental values.....	65
--	----

### Chapter IV

<b>Table 1</b> PLS statistics of CoMSIA model. R <sup>2</sup> : Non-cross- validated determination coefficient. Q <sup>2</sup> : Cross-validated determination coefficient. SEE: Standard error of the estimate. F: F-test value.Str: Steric field, Ele: Electrostatic field, Hyd: Hydrophobic field, HBA and HBD correspond to H-bonds Acceptor and Donor.....	75
---	----

<b>Table 2</b> Designation of the R <sub>1</sub> <sup>'</sup> ,R <sub>2</sub> <sup>'</sup> and R <sub>3</sub> <sup>'</sup> substituents, pKi (CoMSIA model) and total scoring values of the proposed new human $\sigma_1$ receptor inhibitors and of the active compound C48.....	78
---	----

<b>Table 3</b> Important computed ADMET properties and computed safety end points of the screened best sixteen compounds of and of the most active compound C48.....	86
--	----

<b>Table 4</b> Metabolism and toxicity predictions for the best sixteen compounds and for the active compound C48.....	87
--	----

<b>Table 5</b> The E <sub>HOMO</sub> , E <sub>LUMO</sub> , energy gap ( $\Delta E$ ), ionization potential (IP), electron affinity (EA), electronegativity ( $\chi$ ), chemical potentials ( $\mu$ ), Chemical hardness ( $\eta$ ), Softness ( $\sigma$ ), and electrophilicity index ( $\omega$ ) of the subject ligands have been accurately calculated in electron volt (eV) units.....	90
--	----

### Appendices

<b>Table S1</b> Structures of the best newly compounds, Chemical nomenclature and their masses (Mol2, Mol3 and Mol4).....	97
---	----

<b>Table S2</b> Optimized 3D structures (Cp.) of the series of 54 optimized compounds containing pyrimidine of S1R antagonists currently studied.....	98
---	----

<b>Table S3</b> SMILES notation for the proposed new molecules.....	99
<b>Table S4</b> Interactions of the enzyme 5HK1 with the three best inhibitors and reference ligand 61W and C48.....	100
<b>Table S5</b> Structures of the antagonists BD-1047 and S1RA, as well as the agonists PRE-084 and Blarcamesine, Chemical nomenclature and their Scores.....	103



# *General Introduction*

Peripheral or central nervous system injury produces a syndrome called neuropathy, which greatly affects about 6-10% of the world's population. Recent studies expect that this syndrome will increase in prevalence due to aging in the coming years, and this leads to an increase in the survival of patients with chronic diseases related to neuropathic pains, which are associated also with the emergence of specific symptoms in patients presenting this syndrome. For the treatment of these diseases, some medications are prescribed, such as opiates, anticonvulsants and antidepressants, which have a certain ability to block ion channels responsible for these neuropathic pains. However, the effectiveness of these drugs is weak in addition to their confirmed significant side effects. Therefore, it is necessary to reduce the symptoms of this syndrome by discovering new effective drugs<sup>1</sup>.

Due to its ease of transport to many parts of the cell, various studies have shown that human  $\sigma_1$  receptor, S1R, uniquely treats neuropathic pain. S1R is a transmembrane protein, which is predominantly located in the endoplasmic reticulum and the mitochondrial reticulum<sup>2</sup>. Such receptors were widely investigated because of their important role in correcting the cell function and stability<sup>3</sup>. Intracellular S1R regulates all signaling and modulation between the endoplasmic and mitochondria, the membrane and the cell nucleus. Also, it acts on the second messenger systems such as protein kinase C, phospholipase C and Inositol 1,4,5-trisphosphate system. Besides, preclinical studies conducted on several models of memory impairment revealed the appropriateness of S1R links for the treatment of memory, concentration, and cognitive impairments<sup>4</sup>. Indeed, S1R has many biological functions, making it suitable for the future treatments of cancers, epilepsy, Parkinson's disease, Alzheimer's disease, sclerosis, schizophrenia, etc. Also, it has been classified as a promising target for the treatment of many types of mental disorders, mainly neurological and neurodegenerative ones. It acts via correcting the oxidative stress responses as well as the depressive states<sup>2</sup>. Such studies have targeted S1R for drug discovery. It is suggested as an effective alternative to the available treatment methods at least via its action on regulating the  $\text{Ca}^{2+}$  ion<sup>1</sup>.

Nowadays, the pharmaceutical industry is increasingly utilizing in silico technologies for the purpose of advancing upstream research, with the aim of expediting the process of developing and uncovering potential therapeutic drug candidate molecules. These in silico methodologies have the capability to predict the toxicity and effectiveness of a molecule even before progressing to in vivo experimental trials. This predictive ability allows for the identification and elimination of highly toxic molecules at an early stage, thus contributing to a reduction in the utilization of animals for testing purposes and saving valuable time in the drug development process.

Furthermore, Quantitative Structure-Activity Relationship (QSAR) modeling stands out as a dependable and efficient instrument currently employed to bridge the gaps present in in vivo and in vitro investigations, while also aiming to curtail the necessity for biological assays. QSAR is a technique that seeks to establish a quantitative correlation between the activity measured through experimentation and the theoretical or empirical descriptors associated with the molecular structure.

Molecular docking is a computational technique employed to predict the most optimal conformation, which refers to the relative position and orientation, of two molecules interacting with each other to form a stable complex. Once the preferred conformation is determined, it facilitates the estimation of the binding affinity between these molecules, indicating the strength of their association. This method plays a crucial role in understanding the molecular interactions and potential outcomes of these interactions.

The utilization of in silico methods has significantly contributed to the exploration of the pharmacokinetic characteristics of pharmaceuticals, encompassing the stages of absorption, distribution, metabolism, excretion, as well as the toxicological assessment. These computational approaches are particularly valuable during the preclinical stages of drug development, aiding in the evaluation of potential drug interactions and the identification of late-onset side effects, even for drugs already on the market.

The current importance of Density Functional Theory (DFT) in exploring experimental data and studying biological systems is immense. DFT's use spans diverse calculations, aiding in constructing a theoretical framework that offers valuable insights into molecular geometry, molecular orbitals, and spectroscopic properties. Recent advancements have made these techniques highly reliable, accurately predicting molecular properties with exceptional precision, as evidenced by numerous studies. Given the promising advantages of compounds in various biological applications, they've garnered significant attention, prompting investigations into designing new and potentially active compounds<sup>5</sup>. Hence, the culmination of our research involves utilizing DFT for analyzing candidate drug compounds, aiming to delve deeper into their characteristics for a more comprehensive understanding.

The primary objective of this research is to employ virtual screening techniques like Quantitative Structure-Activity Relationship (QSAR), molecular docking, Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) prediction, as well as Density

Functional Theory (DFT) calculations on pyrimidine derivatives. The goal is to identify promising compounds that exhibit more favorable reactivity towards the treatment of neuropathic pain conditions. By leveraging these computational tools, the study aims to enhance the selection process for potential drug candidates that could offer improved therapeutic outcomes in managing neuropathic pain.

Our work is divided into four chapters:

The first chapter a general overview of the pyrimidine and inhibition of human  $\sigma_1$  receptor for neuropathic pain syndrome treatment.

In the second chapter we gave an overview of the main virtual screening strategies that are the most widely used predictive methods in the design of new molecules of therapeutic interest: both the strategy based on target structure and that based on ligand structure and properties.

In the third chapter, we present the different materials and methods used in this study.

In the fourth chapter, we present the main results and a discussion.

Finally, we offer a general conclusion summarizing the work carried out.

### References

1. Chiari, L. P., da Silva, A. P., de Oliveira, A. A., Lipinski, C. F., Honorio, K. M., & da Silva, A. B. (2021). Drug design of new sigma-1 antagonists against neuropathic pain: A QSAR study using partial least squares and artificial neural networks. *Journal of Molecular Structure*, 1223, 129156.
2. A.M. Abramyan, H. Yano, M. Xu, et al., The Glu102 mutation disrupts higher-order oligomerization of the sigma 1 receptor, *ComputStructBiotechnolJ* (18) (2020) 199–206. <https://doi.org/10.1016/j.csbj.2019.12.012>.
3. Meng, F., Xiao, Y., Ji, Y., Sun, Z., & Zhou, X. (2022). An open-like conformation of the sigma-1 receptor reveals its ligand entry pathway. *Nature Communications*, 13(1), 1267.
4. Dalwadi, D. A., Kim, S., Schetz, J., Schreihof, D. A., & Kim, S. (2022). Brain-derived neurotrophic factor for high-throughput evaluation of selective Sigma-1 receptor ligands. *Journal of pharmacological and toxicological methods*, 113, 107129.
5. Tlidjane, H., Chafai, N., Chafaa, S., Bensouici, C., & Benbouguerra, K. (2022). New thiophene-derived  $\alpha$ -aminophosphonic acids: Synthesis under microwave irradiations, antioxidant and antifungal activities, DFT investigations and SARS-CoV-2 main protease inhibition. *Journal of molecular structure*, 1250, 131853.

# *Chapter I*

## **The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment**

**Chapter I: The pyrimidine and inhibition of human  $\sigma_1$  receptor for neuropathic pain syndrome treatment**

1. The pyrimidine.....	8
1.1 Introduction.....	8
1.2 Pyrimidine ring structure.....	9
1.3 Medicinal Properties of Pyrimidines.....	10
2. Pain: A simplified model.....	11
2.1 Nociceptive input.....	12
2.2 Modulatory output.....	13
2.3 Local control.....	14
2.4 Neuropathic pain.....	16
3. The sigma-1 receptor.....	16
3.1 History and discovery of Sigma-1 Receptor.....	16
3.2 The pharmacology of Sigma-1 Receptor.....	18
References.....	20

## 1. The pyrimidine

### 1.1 Introduction

The escalation of resistance to antimicrobial agents has emerged as a significant global issue of concern. Out of the annual 2 million individuals contracting bacterial infections in US medical facilities, approximately 70% of cases now involve strains that exhibit resistance to at least one medication. The prevalence of patients with antibiotic-resistant infections is steadily increasing in both community and hospital settings worldwide. A prominent worry in the United Kingdom is the surge of methicillin-resistant *Staphylococcus aureus* (MRSA), which previously had minimal presence but currently represents around 50% of all *S. aureus* isolates. Urgent substantial funding and research efforts in the field of anti-infectives are imperative to prevent a looming public health crisis. The etiology of antimicrobial resistance is complex and multifaceted. The persistence of resistance to antibiotics is predominantly attributed to the continued excessive reliance on and inappropriate utilization of these antibacterial drugs, with mounting indications suggesting a similar scenario for the emergence of biocide resistance. A critical issue is the potential co-resistance of antibiotics and biocides due to shared resistance mechanisms. The rise of metal resistance is being observed as a consequence of contaminated environments. Prolonged exposure to antibacterial environments leads to the proliferation of bacteria inherently resistant to antimicrobials or that have developed mechanisms to resist these substances<sup>1</sup>.

The modification of the structure of antimicrobial drugs facing resistance has been demonstrated as an effective strategy to prolong the efficacy of antifungal agents like the azoles, antiviral agents such as nonnucleoside reverse transcriptase inhibitors, and a variety of antibacterial agents including  $\beta$ -lactams and quinolones. Consequently, it is not unexpected that, in response to antimicrobial resistance, major pharmaceutical corporations have tended to focus on enhancing existing classes of antimicrobial agents. Nonetheless, given the current array of therapeutic options, it is recognized that researchers are approaching a critical juncture in terms of modifying parent structures. As a result, there is a growing demand for the development of novel drug classes that target different sites from those targeted by existing drugs.

Heterocyclic compounds, which are prevalent in nature, hold significant importance in the realm of life due to the presence of their structural components in various natural substances



## The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment

like vitamins, hormones, and antibiotics. Consequently, they have garnered substantial interest in the realm of designing biologically active molecules and in the field of advanced organic chemistry. Nitrogen-containing heterocycles, a vital subgroup within the family of heterocyclic compounds, play a crucial role in medicinal chemistry and have made notable contributions to both biological and industrial sectors, aiding in the comprehension of life processes. An unsaturated six-membered ring that contains nitrogen is termed as azine or pyridine (1); when it contains two nitrogen atoms, it is referred to as diazine. Additionally, if a nitrogen is located at the 1,2-position, it is identified as pyridine (2), at the 1,3-position as pyrimidine (3), and at the 1,4-position as pyrazine (4) (Fig. 1). The present review, however, aims to concentrate on the importance of antimicrobial agents belonging to the pyrimidine class, elucidating their clinical and in vitro applications in order to facilitate the enhancement of more potent and efficacious antimicrobial agents<sup>1</sup>.

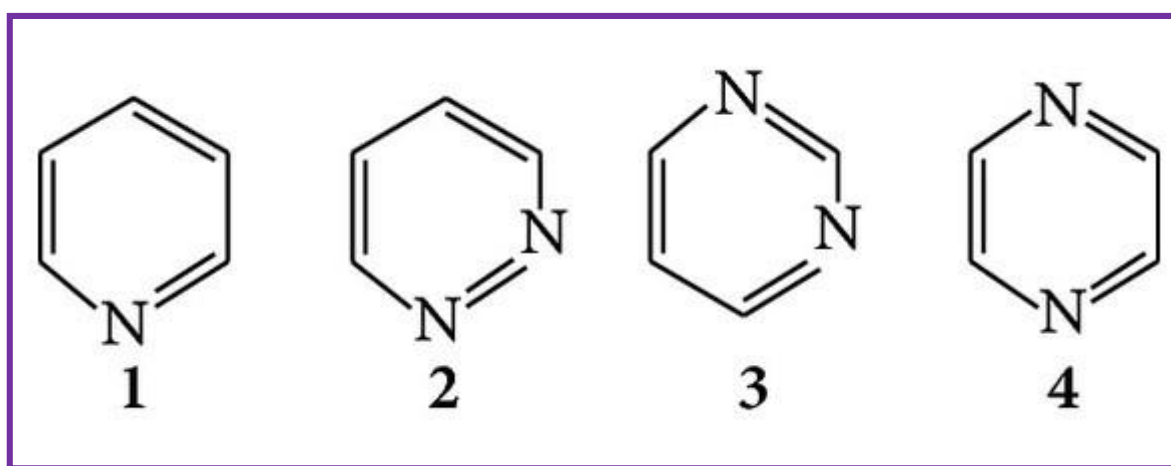


Fig. 1. Pyridine and different isomeric forms of diazine family.

### 1.2 Pyrimidine ring structure

Pyrimidines are aromatic heterocyclic compounds resembling benzene and pyridine with two nitrogen atoms positioned at 1 and 3 within the six-membered rings, thus garnering significant scientific interest due to their natural and synthetic diversity, many of which demonstrate noteworthy biological activities and clinical relevance. The presence of pyrimidine moiety in heterocycles is of particular importance, as they constitute a vital category of various natural and synthetic compounds, some of which are prominently featured in biological systems and were among the initial subjects of exploration by organic chemists. Purines and

## The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment

pyrimidines, when substituted, are prevalent in biological systems and hold historical significance as some of the first compounds scrutinized by organic chemists (Fig. 2(a))<sup>1</sup>.

Pyrimidines, as a class of heterocycles, hold immense biological significance, being the most widespread entities within the diazine family, notably represented by uracil and thymine found in RNA and DNA, alongside cytosine as depicted in Fig. 2(b). Moreover, the structural framework of pyrimidines is encountered in various natural products like thiamine (vitamin B1) and in synthetic compounds such as barbituric acid and Veranal, both utilized for their hypnotic properties (Fig. 2(b))<sup>1</sup>.

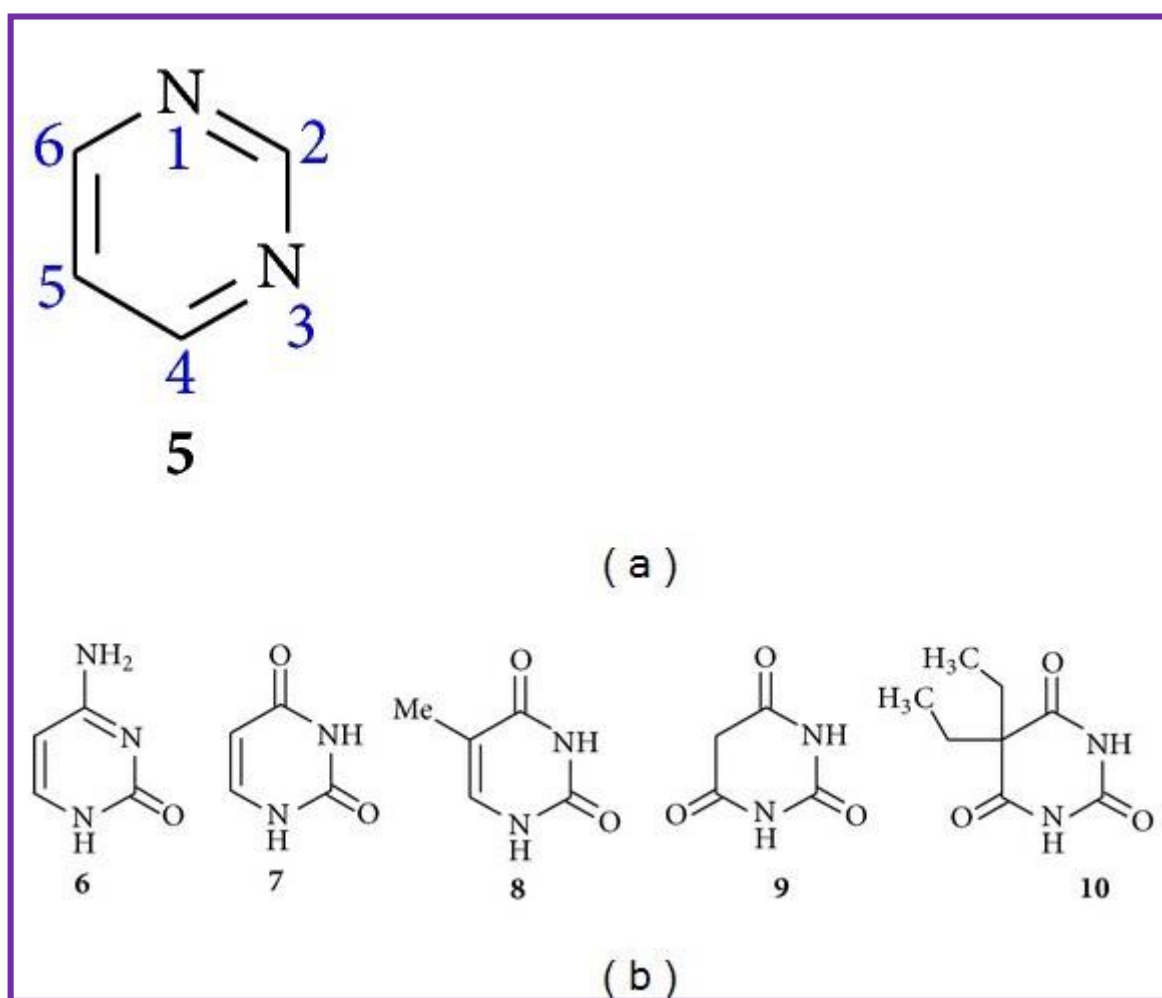


Fig. 2. (a) Pyrimidine. (b) Pyrimidine containing natural and synthetic products.

### 1.3 Medicinal Properties of Pyrimidines

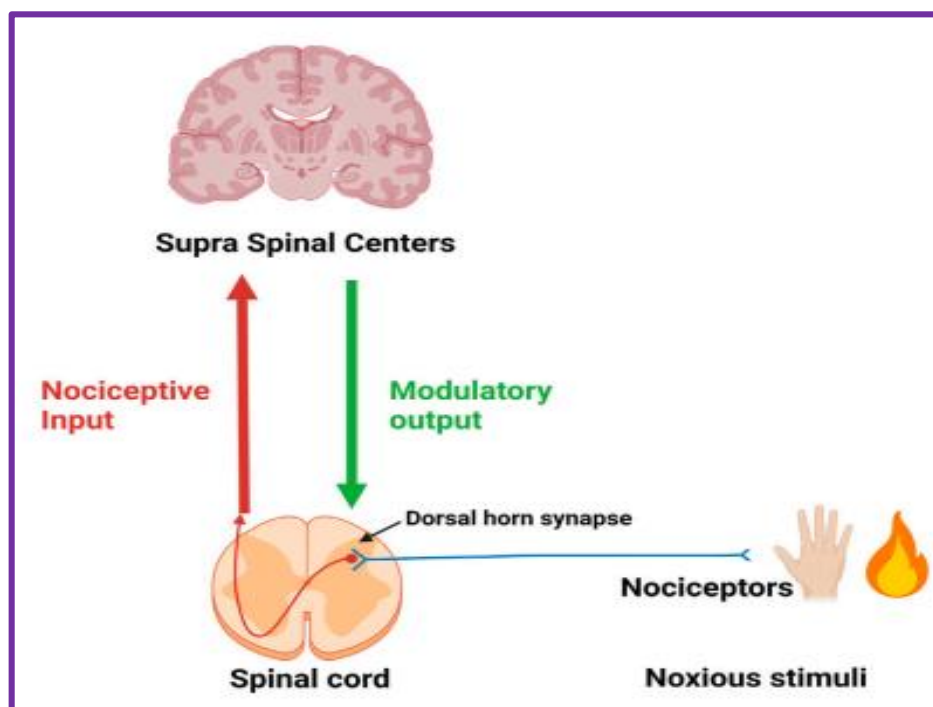
The existence of pyrimidine base in thymine, cytosine, and uracil, components crucial for the formation of nucleic acids DNA and RNA, may be attributed to their extensive therapeutic

## The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment

uses. Pyrimidines are known to belong to a highly active category of compounds with a broad range of biological functions, demonstrating notable in vitro efficacy against diverse DNA and RNA, viruses such as polioherpes viruses, diuretic, antitumor, anti-HIV, and cardiovascular. Research literature has revealed that compounds containing the pyrimidines nucleus display a diverse array of pharmacological effects. Moreover, a variety of analogs of pyrimidines have exhibited antibacterial, antifungal, antileishmanial, anti-inflammatory, analgesic, antihypertensive, antipyretic, antiviral, antidiabetic, antiallergic, anticonvulsant, antioxidant, antihistaminic, herbicidal, and anticancer properties, with several derivatives of pyrimidines identified for their potential as central nervous system (CNS) depressants<sup>1</sup>.

### 2. Pain: A simplified model

Post-synaptically, the transmission of nociceptive input occurs towards superior (supraspinal) centers. The regulation of descending output from supra-spinal centers is perceived at the dorsal horn level. The local environment (inflammatory soup) at the original site of noxious stimuli (injury) and surrounding the dorsal horn plays a significant role in influencing the strength and duration of the signal (Fig. 3)<sup>2</sup>.

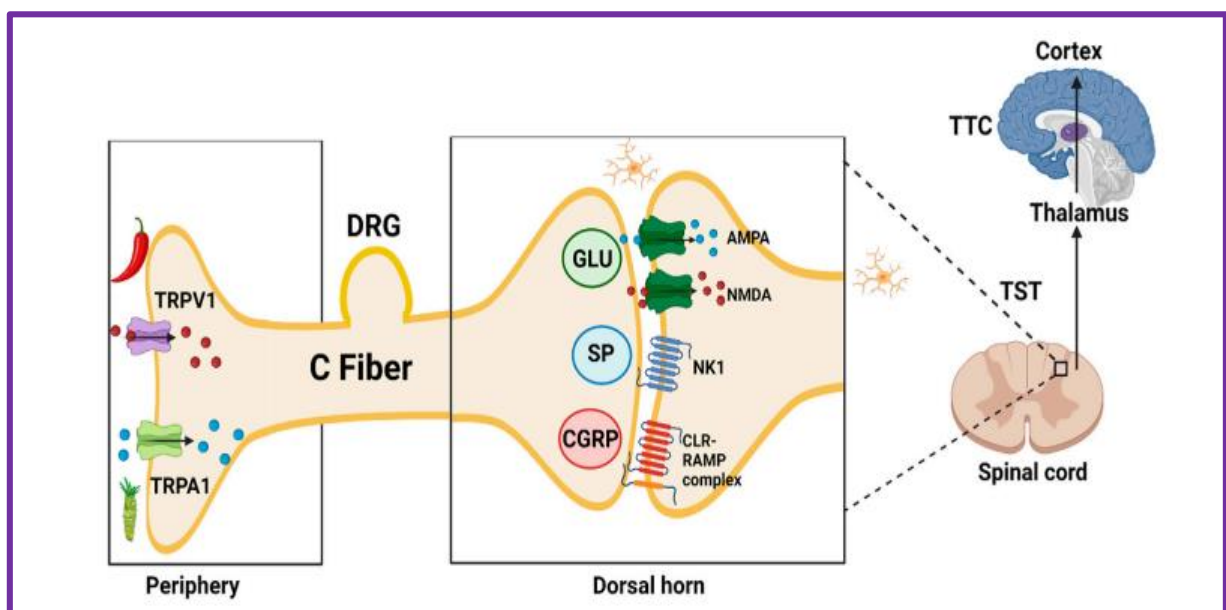


**Fig. 3.** First neuron (in blue) of the ascending pain pathway. Red-brown arrow: Second ascending neuron leading to supra-spinal centers. Green arrow: Descending modulatory output from supra-spinal centers<sup>2</sup>.






### 2.1 Nociceptive input

Injury initiates the activation of Transient Receptor Potential (TRP) nociceptors in the periphery, leading to the subsequent opening of cation channels causing depolarization and propagation of action potential along afferent sensory fibers to the synapse in the dorsal horn. Excitatory non-peptide transmitters (glutamate, AMPA, and NMDA receptor agonists) and peptide transmitters [Substance P (SP) (NK1 neurokinin receptor 1 agonist) and Calcitonin Gene-Related Peptide (CGRP) (agonist of Calcitonin like Receptor and Receptor Activity Modifying Protein complex)] are released by presynaptic vesicles. Neuropeptide Y (NPY) and CGRP receptors have extensive co-localization with mostly opposing effects. NPY acts on Y2 receptors at the central terminals of primary afferents to inhibit SP release. CGRP acts as a potent vasodilator in blood vessels when compared to known vasodilators like histamine, prostaglandin E2, and SP. Inhibitors of the CGRP receptor are distinguished by the suffix-gepant (ubrogepant; atogepant). AMPA receptor antagonists are recognized by the suffix-ampanel (perampanel). Kynurenic acid acts as an endogenous antagonist at ionotropic AMPA, NMDA (glycine-site ligand), and kainate glutamate receptors. The second neuron: Upon crossing to the contralateral side, the signal progresses through the ascending lateral spinothalamic tract (TST) to the thalamus<sup>3</sup>.

The third neuron: The signal travels from the thalamus to the sensory cortex in the parietal lobe through the thalamo-cortical tract TTC (Brodmann areas 1, 2, and 3) enabling pain localization (**Fig. 4**).

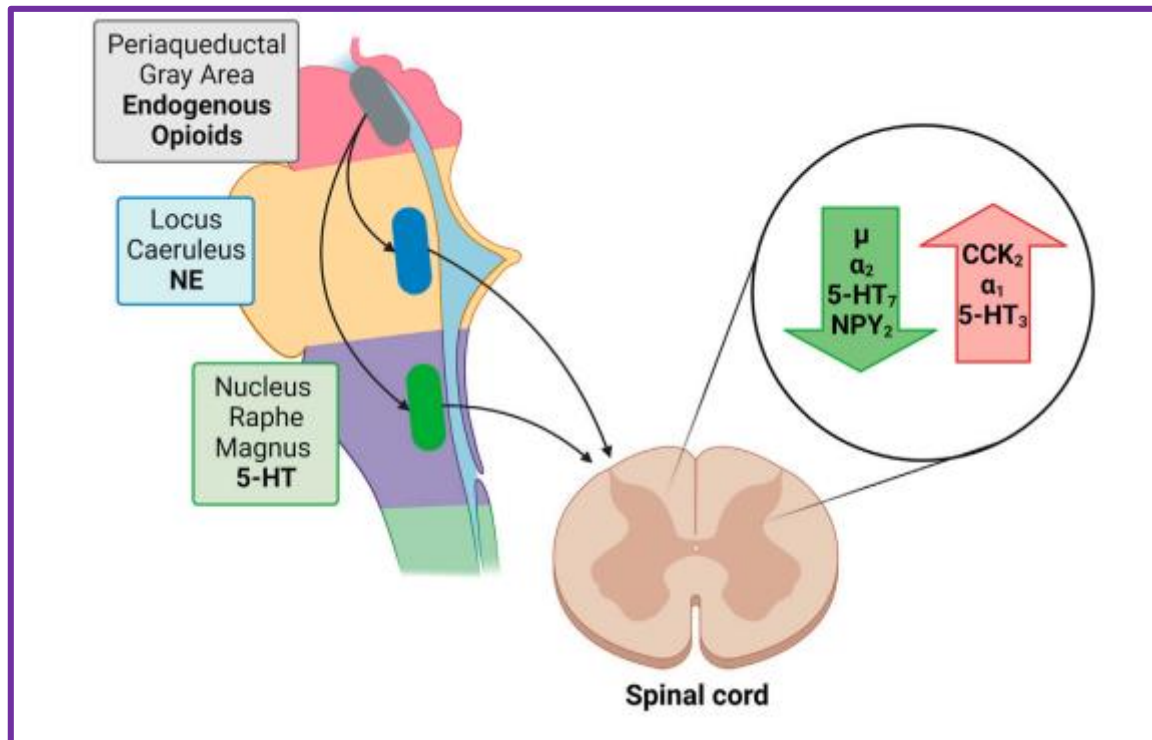


## The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment

**Fig. 4.** TRPV1, transient receptor potential vanilloid subfamily, member 1, capsaicin or hot chili pepper receptor; TRPA1, transient receptor potential ankyrin subfamily, member 1, allyl isothiocyanate or wasabi receptor; DRG, dorsal root ganglion; GLU, glutamate; SP, substance P (11 AA); AMPA,  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; NMDA, N-methyl-D-aspartate; NK1, Neurokinin 1 receptor; CGRP, calcitonin gene-related peptide (37 AA); CLR-RAMP, calcitonin like receptor—receptor activity modifying protein complex; TST, Tractus spino-thalamicus; TTC, tractus thalamo-corticalis; Ca<sup>2+</sup>; Na<sup>+</sup>; Microglia ; Chili pepper: Capsaicin; Wasabi: Allyl isothiocyanate<sup>2</sup>.

### 2.2 Modulatory output

Descending noradrenergic (norepinephrine; NE) and serotonergic (5-HT) fibers exert an impact on pain perception. The majority of noradrenergic fibers stem from the pontine locus coeruleus (LC; blue spot) while descending serotonergic pathways originate from the floor of the medulla oblongata from the nucleus raphe magnus (NRM). NE and 5-HT induce membrane hyperpolarization, reduce the release of excitatory transmitters from primary A $\delta$  and C afferent fibers pre-synaptically, and enhance the release of inhibitory GABA and glycine from interneurons. Administration of 5-HT results in membrane hyperpolarization in approximately 50% of dorsal horn neurons, while NE hyperpolarizes over 80% of them, indicating the necessity to enhance concentrations of both NE and 5-HT to suppress algesia. Neither atomoxetine (selective noradrenaline reuptake inhibitor) nor an SSRI alone display such a pronounced effect. Norepinephrine inhibits neuropathic pain via  $\alpha 2$ -adrenoceptors, while  $\alpha 1$ -adrenoceptors exacerbate it. Regarding the receptors involved in serotonergic pain modulation, research suggests a role for 5-HT<sub>7</sub> receptors in antinociception and a role for 5-HT<sub>3</sub> in pro-nociceptive facilitation. Activation of 5-HT<sub>7</sub> receptors does not directly inhibit nociceptive dorsal horn neurons as these receptors are positively linked to adenylate cyclase and their stimulation is excitatory. Activation of 5-HT<sub>7</sub> receptors situated on spinal inhibitory enkephalinergic or GABAergic interneurons leads to the release of enkephalins or GABA, resulting in the inhibition of nociceptive transmission<sup>4</sup> (Fig. 5).



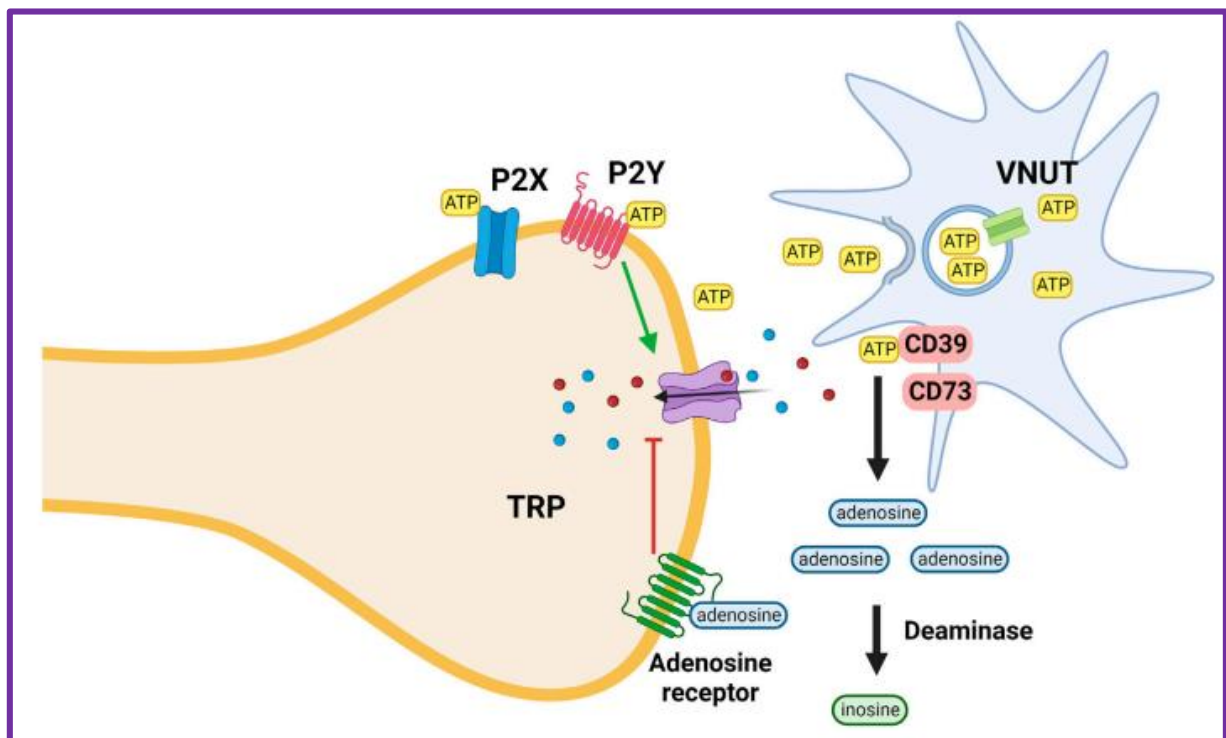
**Fig. 5.** Green arrow: Receptors reducing nociception:  $\mu$  opioid receptor;  $\alpha_2$ , adrenergic receptor; 5-HT<sub>7</sub>, serotonin receptor; NPY<sub>2</sub>, neuropeptide Y receptor type 2. Red arrow: Receptors enhancing nociception: CCK<sub>2</sub>, cholecystokinin 2(B) receptor;  $\alpha_1$ , adrenergic receptor; 5-HT<sub>3</sub>, ionotropic serotonin receptor; ● Periaqueductal Gray: Endogenous Opioids; ● Locus Caeruleus: NE; ● Nucleus Raphe Magnus: 5-HT<sup>2</sup>.

### 2.3 Local control

The intricate feedback loop involving (afferent) nociceptive input and modulatory (efferent) output is significantly impacted by the composition of the "cytokine soup" at the origin site and synaptic relay stations. An environment rich in ATP promotes inflammation and nociception, whereas an adenosine-rich environment exerts the opposite effect. Following injury, glia cells release intracellular ATP through the vesicular nucleotide transporter (VNUT) into the extracellular space. ATP activates purinergic P2X ionotropic and metabotropic P2Y receptors, enhancing the sodium and calcium conductance of various TRP receptors. The degradation of nucleotides by cell surface (ecto) nucleotidases terminates purinergic signaling, akin to cholinesterase activity at cholinergic synapses. Adenosine produced extracellularly from the breakdown of nucleotides activates adenosine receptors (A<sub>1</sub>, A<sub>2A</sub>, A<sub>2B</sub>, and A<sub>3</sub>) that reduce the sodium and calcium permeability of TRP receptors<sup>5</sup> (Fig. 6). The analgesic effects of adenosine primarily occur through the activation of adenosine A<sub>1</sub> receptors, with a potential role also attributed to A<sub>3A</sub> receptors. However, the role of adenosine A<sub>2A</sub> and A<sub>2B</sub> receptors

## The pyrimidine and inhibition of human $\sigma 1$ receptor for neuropathic pain syndrome treatment

remains controversial, as their activation elicits both pro-nociceptive and antinociceptive outcomes. The purinergic modulation is accompanied by a variety of pro-and anti-nociceptive substances, including endogenous opioid peptides, anti-opioid cholecystokinin (CCK), bradykinin, and cytokines, as well as non-peptides like histamine, prostanoids, and leukotrienes. Non-steroidal anti-inflammatory drugs (NSAIDs) are effective in managing nociceptive pain and inflammation-induced pain by inhibiting cyclooxygenase-2. While commonly prescribed for neuropathic pain, the evidence supporting the use of NSAIDs in this context is insufficient and thus they are not included in major guidelines for neuropathic pain treatment. It is important to acknowledge that pain typically involves a combination of neuropathic and nociceptive components<sup>2</sup>.



**Fig. 6.** ATP is released from microglia, keratinocytes and other cells via the activity of the VNUT; inhibition of the transporter reduces the ATP concentration and thus its pro-inflammatory effect mediated via interaction with both ionotropic (P2X) and metabotropic (P2Y) purinergic receptors. ATP is metabolized by ectoenzymes (CD; cluster of differentiation 73 and 39). Extracellular adenosine generated by metabolic break-down of nucleotides activates adenosine receptors that decrease the sodium and calcium permeability of TRP receptors. Adenosine is metabolized by deaminases<sup>2</sup>.

### 2.4 Neuropathic pain

Nociceptive pain arises as a physiological response to nociceptor activation and serves a protective function due to its role in eliciting appropriate responses. On the other hand, neuropathic pain, which occurs independently of nociceptor activation, is unlikely to be beneficial and results from sensitization and ectopic activity of various pronociceptive entities, primarily sodium and calcium channels. The nature of dull nociceptive pain differs from the sharp characteristics of neuropathic pain. Chronic and neuropathic pain exhibit some overlap, with neuropathic pain being chronic but not all chronic pain being neuropathic. The inefficacy of opioids in managing neuropathic pain is thought to be linked to increased activity of the endogenous pronociceptive neuropeptide CCK. In the central nervous system, the predominant form of CCK is an octapeptide. Opioid-induced release of CCK is a proposed mechanism for opioid tolerance and hyperalgesia. The recent findings have demonstrated the heterodimerization of opioid and CCK receptors following the binding of CCK octapeptide. This specific interaction has been recognized as the underlying factor responsible for the inhibition of opioid signal transduction and the counteraction of morphine analgesia. Studies have shown that CCK antagonists can amplify the analgesic effectiveness of endogenous opioids in animal pain models. NPY exhibits pro-analgesic properties, and the administration of naloxone diminishes the analgesic effects of NPY. Despite the historical success of opioids in pain management, there has been a notable increase in opioid overdoses posing a significant challenge in clinical pain treatment. Notably, the most devastating drug overdose epidemic in the history of the United States has been linked to opioids. The introduction of a slow-release version of oxycodone in 1996 is believed to have played a crucial role in the emergence of this overdose crisis, commonly referred to as the US opioid epidemic<sup>6</sup>.

### 3. The sigma-1 receptor

#### 3.1 History and discovery of Sigma-1 Receptor

The term sigma receptor is derived historically from the sigma/opioid receptor proposed by Martin et al. (1976) to mediate the psychotomimetic action seen with the benzomorphan, N-allylnormetazocine, and its analogs. The receptor was called an opioid receptor because the effect of N-allylnormetazocine was reported to be antagonized by the universal opioid antagonist, naloxone. Attempting to demonstrate the sigma/opioid receptor, Su identified a binding site that, although labeled by the prototypic sigma/opioid receptor ligand ( $\pm$ )SKF-



## The pyrimidine and inhibition of human $\sigma_1$ receptor for neuropathic pain syndrome treatment

10,047 (( $\pm$ )N-allylnormetazocine), was however insensitive to naloxone<sup>7</sup>. Mistakenly, the binding site then identified was termed the sigma/opioid receptor, which in fact is not the sigma/opioid receptor proposed by Martin et al. (1976), as the latter is sensitive to naloxone. It must be mentioned that the SKF-10047 experiment was repeated by the same laboratory at a later date and it was found that the “psychotomimetic” effect caused by SKF-10047 was not blocked by naltrexone, another potent analog of naloxone. Thus, the name of the binding site identified by Su (1982) was changed to sigma receptor to differentiate it from the sigma/opioid receptor. The sigma receptor was confused with the PCP/NMDA receptor for a while as some ligands for each receptor cross-react with the other receptor. The confusion was dispelled after it was realized that the confusion arose from the fact that numerous ligands from the two receptors could cross-react. The psychotomimetic effect of SKF-10047 is now believed to be mediated via NMDA receptors, kappa opioid receptors, or sigma-1 receptors<sup>7</sup>.

Based on the ligand selectivity in the receptor binding assay, as seen in different tissues, the sigma receptor was later found to consist of two subtypes, the sigma-1 and sigma-2 receptors. Apparently, some ligands could bind both subtypes of the receptor. Nevertheless, the sigma receptor identified by Su (1982) is in fact the sigma-1 receptor because the ligand selectivity of the sigma receptor identified by Su (1982; Su et al., 1988) is exactly the same as the sigma-1 receptor demonstrated by Bowen’s lab. The exact reason why sigma-2 receptors were not identified in Su’s 1982 study could have been that the ligand used in the study did not have sufficient affinity for sigma-2 receptors. The sigma-1 receptor was first cloned in 1996. The sigma-2 receptor has not yet been cloned. This review will focus mainly on the sigma-1 receptor, but will review some data on the sigma-2 receptor as well as it is relevant to the content of the review.

Since the discovery of the sigma-1 receptor, many preclinical studies have implicated the receptor in many diseases. The contents of this review will first examine the role of sigma-1 receptors in different diseases followed by a discussion of potential mechanistic explanations. Because the sigma receptor has been the subject of many excellent reviews in the past, we will focus on discoveries made mainly over the past 10 years<sup>7</sup>.

### 3.2 The pharmacology of Sigma-1 Receptor

The involvement of the sigma-1 receptor in the analgesic effects mediated by various opioid receptors such as mu-, delta-, kappa1, and kappa3 was identified by Mei and Pasternak in 2002. This research serves as a continuation of their seminal discovery from the 1990s regarding the role of sigma-1 receptors in opioid-induced analgesia as reported by King et al. (1997). The study by Mei and Pasternak in 2002 highlighted that the function of sigma-1 receptors in facilitating the analgesic effects of opioids primarily occurs at the supraspinal level. Consistent with their previous findings, it was observed that the sigma-1 receptor antagonist, haloperidol, enhances opioid-induced analgesia, whereas the sigma-1 receptor agonist diminishes opioid analgesia. Additionally, the decrease in sigma-1 receptors in the supraspinal region intensifies opioid analgesia. Recent investigations by the research group have revealed that the brain regions crucial for the modulation of morphine-induced analgesia by sigma-1 receptor ligands are located in the brainstem, specifically in the periaqueductal gray and the rostroventral medulla, and the locus coeruleus<sup>8</sup>. The series of fascinating reports by the group indicate the existence of an endogenous anti-opioid sigma-1 receptor system in the CNS. Their study also provides a tentative molecular explanation, afforded by the haloperidol-sigma-1 receptor interaction, of the long-held mystery of why haloperidol potentiates morphine analgesia which has puzzled medical experts for several decades.

Results from the investigations conducted by Pasternak's group have been corroborated by numerous studies; however, the precise molecular mechanism through which sigma-1 receptors influence opioid analgesia is currently the subject of intense scrutiny. Cendan et al. (2005) demonstrated a reduction in formalin-induced pain in sigma-1 receptor knockout mice. The administration of the sigma-1 receptor antagonist, BD1047, via intrathecal treatment in mice not only decreased formalin-induced pain but also mitigated the phosphorylation of the N-methyl-D-aspartate (NMDA) receptor subunit 1 induced by formalin. Moreover, the intrathecal administration of the sigma-1 receptor agonists, PRE-084 and carbetapentane, in mice led to an augmentation in the protein kinase C- and protein kinase A-dependent phosphorylation of the NR1 subunit of the NMDA receptor. This enhancement was impeded by the sigma-1 receptor antagonist, BD1047. The research team also expanded this observation to neuropathic pain<sup>7</sup>. The neurosteroid, DHEA, has been demonstrated to be a sigma-1 receptor agonist by many mnemonic studies. DHEA induced a rapid pronociceptive action in sciatic-neuropathic rats, consistent with it being a sigma-1 receptor agonist. Further, the sigma-1 receptor antagonist,

## The pyrimidine and inhibition of human $\sigma_1$ receptor for neuropathic pain syndrome treatment

BD1047, blocked the transient pronociceptive effect provoked by DHEA, which is a sigma-1 agonist<sup>9</sup>. Further, Ronsisvalle's group synthesized a new sigma-1 receptor agonist (1R,2S/1S,2R)-2-[4-hydroxy-4-phenylpiperidin-1-yl]methyl]-1-(4-methylphenyl)cyclopropanecarboxylate and found that it antagonized kappa opioid receptor-mediated antinociceptive effects<sup>10</sup>.

The exact molecular mechanism of the action of sigma-1 receptors in the modulation of opioid-induced pain and neuropathic pain needs to be definitively established and certainly warrants further investigation.

## References

1. Sharma, V., Chitranshi, N., & Agarwal, A. K. (2014). Significance and biological importance of pyrimidine in the microbial world. *International journal of medicinal chemistry*, 2014.
2. Petroianu, G. A., Aloum, L., & Adem, A. (2023). Neuropathic pain: Mechanisms and therapeutic strategies. *Frontiers in Cell and Developmental Biology*, 11, 1072629.
3. De Caro, C., Cristiano, C., Avagliano, C., Cuzzo, M., La Rana, G., Aviello, G., ... & Russo, R. (2021). Analgesic and anti-inflammatory effects of perampanel in acute and chronic pain models in mice: interaction with the cannabinergic system. *Frontiers in Pharmacology*, 11, 620221.
4. Liu, Q. Q., Yao, X. X., Gao, S. H., Li, R., Li, B. J., Yang, W., & Cui, R. J. (2020). Role of 5-HT receptors in neuropathic pain: potential therapeutic implications. *Pharmacological Research*, 159, 104949.
5. Battastini, A. M. O., Figueiró, F., Leal, D. B. R., Doleski, P. H., & Schetinger, M. R. C. (2021). CD39 and CD73 as promising therapeutic targets: what could be the limitations?. *Frontiers in pharmacology*, 12, 633603.
6. Alpert, A., Evans, W. N., Lieber, E. M., & Powell, D. (2022). Origins of the opioid crisis and its enduring impacts. *The Quarterly Journal of Economics*, 137(2), 1139-1179.
7. Maurice, T., & Su, T. P. (2009). The pharmacology of sigma-1 receptors. *Pharmacology & therapeutics*, 124(2), 195-206.
8. Mei, J., & Pasternak, G. W. (2007). Modulation of brainstem opiate analgesia in the rat by  $\sigma 1$  receptors: a microinjection study. *Journal of Pharmacology and Experimental Therapeutics*, 322(3), 1278-1285.
9. Kibaly, C., Meyer, L., Patte-Mensah, C., & Mensah-Nyagan, A. G. (2008). Biochemical and functional evidence for the control of pain mechanisms by dehydroepiandrosterone endogenously synthesized in the spinal cord. *The FASEB Journal*, 22(1), 93-104.
10. Prezzavento, O., Parenti, C., Marrazzo, A., Ronsisvalle, S., Vittorio, F., Aricò, G., ... & Ronsisvalle, G. (2008). A new sigma ligand,  $(\pm)$ -PPCC, antagonizes kappa opioid receptor-mediated antinociceptive effect. *Life sciences*, 82(11-12), 549-553.

# *Chapter II*

## **Overview on Computational methods in drug discovery**

**Chapter II: Overview on Computational methods in drug discovery**

1. Introduction.....	25
2. Computer-aided drug design.....	25
2.1 CADD applications in drug discovery and development.....	26
2.2 Classification of CADD.....	28
2.2.1 Ligand-based drug design (LBDD).....	28
2.2.2 Structure-based drug discovery (SBDD).....	28
2.3 Molecular mechanics.....	29
3. Quantitative structure–activity relationship.....	30
3.1 3D-QSAR.....	31
3.1.1 The CoMFA method.....	32
3.1.2 The CoMSIA method.....	32
3.1.3 Alignment of the structures.....	32
3.1.4 Calculation of molecular interaction fields.....	33
3.1.5 Interpretation and validation of a QSAR model.....	34
3.1.5.1 Internal validation.....	34
3.1.5.2 Standard statistical tests and coefficients.....	35
3.1.5.3 External validation.....	36
3.1.6 Applications of the QSAR study .....	37
3.2 Structure-based virtual screening.....	39

4. Molecular docking.....	39
4.1 General protocol.....	40
4.1.1 Ligand preparation.....	41
4.1.2 Target preparation.....	41
4.1.3 Binding site detection.....	42
4.1.4 Docking validation.....	42
4.2 Types of molecular docking.....	43
4.2.1 Rigid docking.....	43
4.2.2 Semi-flexible docking.....	43
4.2.3 Flexible-flexible docking.....	43
4.3 Scoring Functions (SFs).....	45
4.3.1 Physics-based SFs.....	45
4.3.2 Empirical SFs.....	46
4.3.3 Knowledge-based SFs.....	47
4.3.4 Consensus scoring.....	48
4.3.5 Machine-learning-based SFs.....	49
5. Pharmacokinetics Properties and Computational tools employed in ADMET.....	50
6. Quantum mechanics.....	51
7. Density Functional Theory (DFT).....	51
7.1 Conceptual Density Functional Theory (DFT).....	51

7.1.1 Fundamental and Computational Aspects of DFT.....52

References.....54



### 1. Introduction

Drug discovery plays a pivotal role in the expansion of pharmaceutical companies and society at large, as it leads to the introduction of newer and safer drugs in the market. The primary objective of drug discovery is to enhance the therapeutic value and safety of medications. Over the years, the pharmaceutical industry has consistently demonstrated its ability to discover and develop innovative medicines for a wide range of diseases <sup>1</sup>. The journey of drug research, as we know it today, began at a time when chemistry had reached its zenith, enabling the application of chemical principles and theories to problems beyond the realm of chemistry. This coincided with the emergence of pharmacology as an independent scientific discipline in its own right. By the year 1870, the foundations of chemistry theory had been firmly established <sup>2, 3</sup>. The twentieth century witnessed the remarkable influence of biochemistry on drug research in numerous ways. It was during this period that the concept of targeting enzymes and designing drugs as inhibitors came into existence <sup>4</sup>. However, it is important to note that the current drug-discovery process is not without its challenges. It is a time-consuming and expensive endeavor that can span over a decade, requiring exhaustive research, substantial financial investment, and rigorous clinical trials before a molecule can be recognized as a drug. Despite the various research and development (R&D) approaches adopted by pharmaceutical companies, the attrition rate remains unacceptably high. One of the factors contributing to this high attrition rate is the presence of active compounds that exhibit unfavorable absorption, distribution, metabolism, excretion, and toxicity (ADMET) effects, necessitating their withdrawal from further development. This factor alone accounts for approximately 50% of all costly failures in drug development <sup>5</sup>. Consequently, it has become widely acknowledged that these areas should be addressed as early as possible in the drug-discovery process <sup>6,7</sup>. Given the pitfalls of the current drug-discovery process, there is a pressing need for an unconventional approach that not only reduces the R&D time but also minimizes the associated costs <sup>8</sup>.

### 2. Computer-aided drug design

In response to these challenges, the field of computer-aided drug design (CADD) has emerged as a multidisciplinary endeavor that attracts researchers from various fields, including information technology, medicine, and pharmacology. CADD leverages computational tools to enhance the drug discovery and design processes (**Figure 1**). Computational chemistry methods are routinely employed to study drug-receptor complexes at an atomic level of detail and to

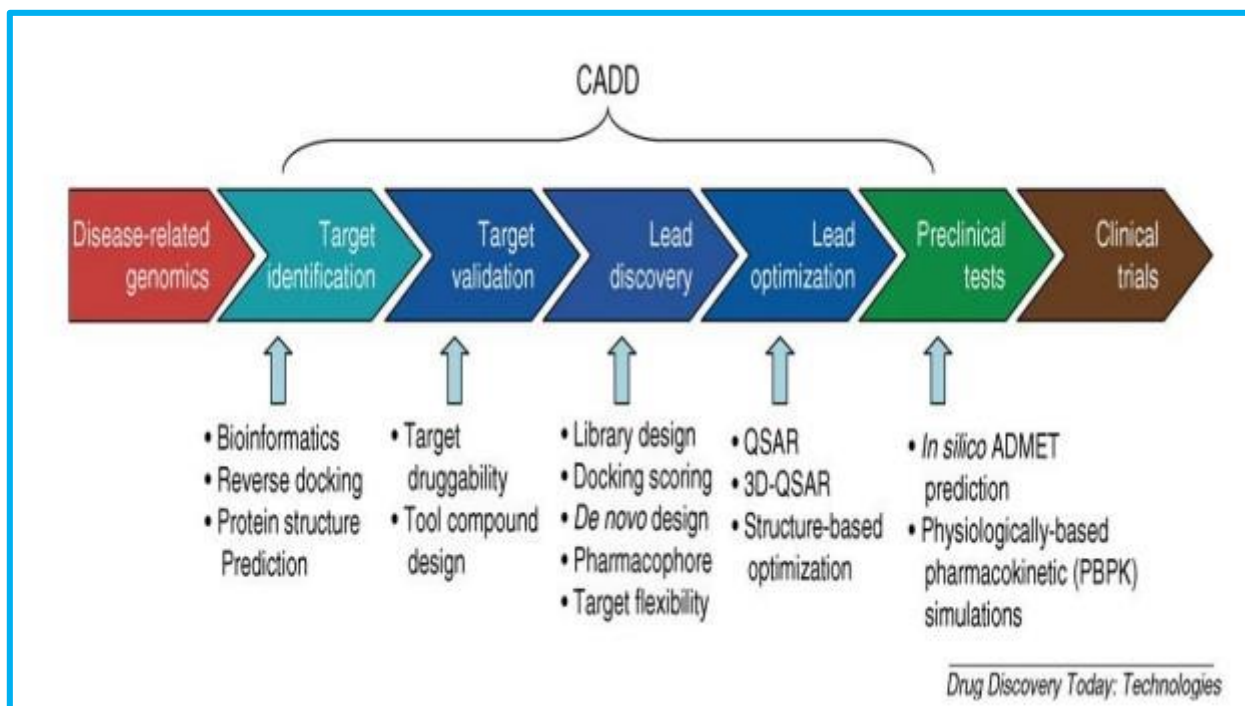
calculate properties of small-molecule drug candidates. On the other hand, tools from information sciences and statistics are increasingly crucial for organizing and managing the vast chemical and biological activity databases that pharmaceutical companies now possess, thereby optimizing their utilization <sup>9</sup>.

### 2.1 CADD applications in drug discovery and development

- There are several key areas where CADD plays a crucial role in the design of effective drugs. Virtual high-throughput screening (vHTS), which is a method used to search for new lead molecules that have the potential to be developed into promising drugs for specific disease targets. In vHTS, small molecules similar to drug compounds that are stored in a database are screened against protein targets to identify molecules that exhibit strong binding affinity to the target protein <sup>10</sup>. Lead molecules for a specific disease are referred to as such. These lead molecules are subsequently extracted from the database for further experimentation. In today's day and age, with the availability of efficient computer-aided drug design (CADD) screening tools, the time and expenditure required to discover a promising lead molecule is significantly reduced compared to traditional methods.
- Sequence analysis plays a crucial role in the design of successful drugs as it provides valuable insights into the amino acid sequence of protein molecules in various organisms. To facilitate this process, CADD researchers have developed numerous sequence analysis tools and algorithms that enable the determination of similarity among species based on proteomic and genomic sequences. This information on sequence similarity proves useful in assuming the relationships among the different organisms under study.
- Homology Modeling. In drug design, having meticulous knowledge about the three-dimensional structure of protein molecules, which serve as the majority of drug targets, is essential. However, the number of experimentally determined protein structures is limited. To overcome this limitation, CADD techniques can be employed to predict the three-dimensional structures of protein molecules. This prediction process, known as homology modeling, relies on the fact that many protein molecules share similar amino acid sequences. If the three-dimensional structure of a known protein molecule is available, it can be used to predict the structure of other protein molecules that exhibit high similarity scores with the reference protein. Various databases, such as the SWISS

MODEL repository, have been created using CADD homology modeling techniques to house these predicted three-dimensional protein structures.

- **Similarity Searches.** Another important aspect of drug discovery is the search for drug analogues. By utilizing CADD tools, one can search for chemical compounds with similar protein structures (in either two or three dimensions), common amino acid sequences, or similar electrostatic properties from existing proteomic and genomic databases. These drug analogues can then be further tested to identify an improved drug candidate that could potentially replace an existing drug.
- **Physicochemical Modeling:** Understanding the physicochemical properties of both the drug and its target is crucial in drug-receptor interactions, as these interactions occur at atomic scales. Properties such as hydrophobicity and polarity of the drug and target play a significant role in determining how candidate drugs bind to protein targets. Therefore, studying the biochemical and biophysical properties of the drug and its target provides a comprehensive understanding of these interactions.
- **Drug Optimization:** Once a promising candidate drug has been identified during the drug discovery process, the next step is to optimize it to enhance its affinity and binding towards the target protein. This optimization process involves modifying the structure of the drug and evaluating alternative templates or scaffolds that resemble the newly discovered drug. The goal is to find a promising drug candidate for the targeted disease. Additionally, the metabolic and toxic properties of the candidate drug are also optimized to maximize its potential.
- **ADMET properties** of a drug, which encompass Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET), are crucial characteristics that determine the bioavailability and bioactivity of a drug. In clinical trials, many candidate drugs fail due to issues related to their toxicity and metabolism in human beings, resulting in the wastage of billions of dollars and years of research. To mitigate these challenges, it is imperative to measure these properties in the laboratory. However, it is also possible to predict these properties in advance using Computer-Aided Drug Design (CADD) tools, which not only save valuable time but also reduce the financial burden associated with experimental studies on candidate drugs <sup>11,12</sup>.



**Fig.1.** CADD in drug discovery

## 2.2 Classification of CADD

CADD can be classified into two general categories: structure-based and ligand-based (Figure 2).

### 2.2.1 Ligand-based drug design (LBDD)

Ligand-based drug design, also known as indirect drug design, relies on the knowledge of active molecules that have demonstrated potential against specific biological targets of interest<sup>13</sup>. Pharmacophore models, derived from these known molecules, define the essential structural characteristics required for effective binding to the biological target<sup>14</sup>. On the other hand, Quantitative Structure-Activity Relationship (QSAR) involves establishing a correlation between the calculated molecular properties of a compound and its experimentally determined biological activity<sup>15</sup>. These QSAR correlations can then be utilized to predict the activity of novel analogs<sup>16</sup>.

### 2.2.2 Structure-based drug discovery (SBDD)

Structure-based drug discovery (SBDD) is another approach used in CADD when the three-dimensional structure of a disease-related drug target is known. This technique involves

designing therapeutics based on the knowledge of the target's structure. Molecular docking approaches and de novo ligand design are two commonly employed methods in SBDD. Molecular dynamics (MD) simulations play a significant role in SBDD as they provide insights into the binding mechanisms of ligands with target proteins, including the pathways of interaction and the flexibility of the target. This becomes particularly crucial in the case of drug targets that are membrane proteins, as membrane permeability is a vital consideration for the effectiveness of drugs <sup>17, 18</sup>. The simulation of a biomolecular system can be achieved using molecular mechanics (MM), quantum mechanics (QM), or a hybrid method (QM/MM), depending on the specific research problem at hand.

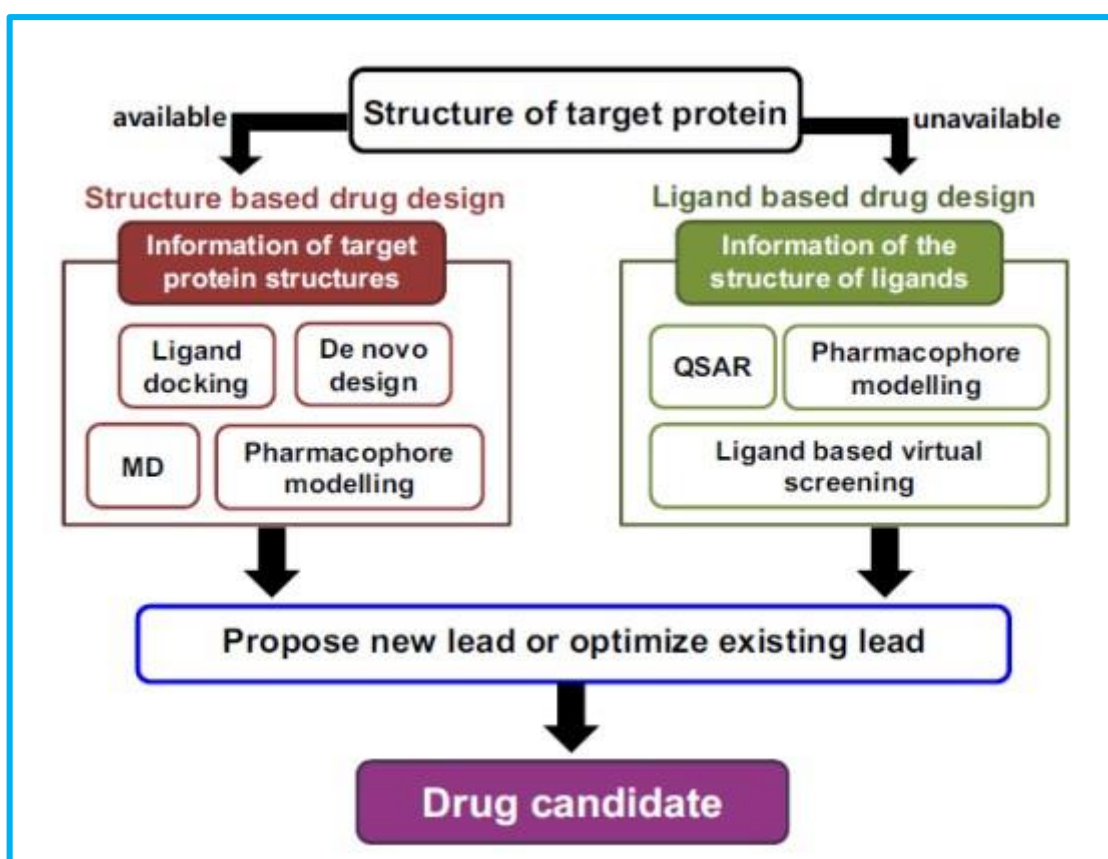


Fig.2. Flow chart of CADD processes

### 2.3 Molecular mechanics

Molecular mechanics (MM) is a commonly applied technique in large-scale systems for calculating molecular structures and determining the relative potential energies of different conformations or atom arrangements <sup>19-21</sup>. In MM, the electrons within the system are not explicitly considered; rather, each atom, including its atomic nucleus and associated electrons,

is treated as a single particle. This exclusion of electrons is justified by the Born-Oppenheimer approximation, which states that electronic and nuclear motions can be decoupled and considered separately <sup>22</sup>. The focus of MM calculations is on energy differences between conformations, rather than absolute values of potential energies. MM can be likened to a ball-and-spring model of atoms and molecules, where classical forces act between them <sup>23</sup>. These forces are accounted for by potential energy functions that take into account various structural features such as bond length, bond angles, and torsional angles. The parameters used in these potential energy functions are designed to replicate experimental properties <sup>20</sup>. The total potential energy of a molecule, known as the MM or molecular mechanics, encompasses several energy components. These components include the energy associated with bond stretching (referred to as  $E_{str}$ ), the energy resulting from bond angle bending (known as  $E_{bend}$ ), the torsion energy arising from twisting motions ( $E_{tor}$ ), as well as the energy of interactions among unbound atoms ( $E_{nb}$ ). The energy contributions from the latter include both van der Waals forces and electrostatic interactions.

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdw} + E_{elec} \quad (1)$$

$$E_{tot} = \sum_{bonds} K_r (r - r_e)^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} (1 - \cos(n\phi - \tau)) + \sum_{i < j} [A_{ij}/r_{ij}^{12} + B_{ij}/r_{ij}^6 + q_i q_j / \epsilon r_{ij}] \quad (2)$$

In the context of MM, the overall potential energy, denoted as  $E_{tot}$ , can be understood as the sum of various energy terms. The stretch terms pertain to the energy associated with bond stretching ( $E_{str}$ ), the bend terms refer to the bond angle-bending energy ( $E_{bend}$ ), the torsional terms relate to the twisting energy ( $E_{tor}$ ), and the unbound interactions encompass the van der Waals and electrostatic forces between atoms that lack chemical bonding. Additionally, the MM framework also incorporates energy contributions from special treatments of hydrogen bonding and stretch-bend coupling interactions.

### 3. Quantitative structure–activity relationship

Quantitative structure-activity relationship (QSAR) modeling, which was initially proposed by Hammett in the 1930s and later developed by Hansch and Fujita in the mid-1960s, is a ligand-based drug design method <sup>24</sup>. The methodology of QSAR is rooted in the utilization of chemometric techniques to generate statistically-derived models that establish correlations between the independent variables, also known as descriptors, of systems and their dependent

variables<sup>25</sup>. In the context of drug design, the independent variables encompass all the structural attributes of chemical entities, including their physicochemical and biological properties, while the dependent variables refer to the functions of these entities, such as binding affinity, activity, toxicity, rate constants, and more<sup>25</sup>. The general formula of a QSAR model can be expressed as follows: Predicted Biological Activity = Function (Chemical Structure).

QSAR has emerged as a vital analytical tool in various fields of chemistry, encompassing medicinal chemistry, agricultural chemistry, environmental chemistry, and toxicology. Particularly within pharmaceutical chemistry, QSAR analysis has become an indispensable means for drug discovery and is now a standard component of all industrial drug design software packages<sup>26</sup>. The recent trends in the drug discovery and development process can be summarized into two fundamental points. Firstly, there is an emphasis on the development of reliable models that possess the ability to accurately predict and classify the biological responses of potential leads. Secondly, these models are being applied in the design of new chemical entities (NCEs) and the screening of extensive libraries or datasets of compounds to identify new hits with desired attributes. However, it is crucial to validate these predictions experimentally<sup>27, 28</sup>. By selecting promising hits, QSAR analysis allows for a reduction in the number of costly experiments, thereby minimizing the expenses associated with candidate drug failures<sup>29</sup>.

### 3.1 3D-QSAR

3D-QSAR approaches have been extensively developed in order to establish a correlation between the biological activity of a series of reference active compounds and the spatial arrangement of numerous properties exhibited by the molecule, including steric, lipophilic, and electronic properties. This correlation analysis is crucial as it provides valuable indications for the optimization of compounds through pharmacomodulation and the design of novel compounds with enhanced activity profiles. The first 3D-QSAR approach, proposed back in 1979, focused on describing the molecular field properties of compounds by calculating them on a regular grid and subsequently correlating them with their biological activity using principal component analysis (PCA). This approach, initially known as DYLOMMS (Dynamic Lattice-Oriented Molecular Modeling System), gained momentum only when Partial Least Squares (PLS) was introduced for the correlation of properties with biological activity. It is worth noting that there are various 3D-QSAR methods currently in use, such as CoMFA (Comparative Molecular Field Analysis), CoMSIA (Comparative Molecular Similarity Indices Analysis),

GRID/GOLPE, and Phase. An important point to highlight is that all these methods necessitate a meticulous alignment of the reference ligands. Once the 3D structure of the biological target has been determined, receptor-dependent 3D-QSAR (RD-QSAR) models can be implemented <sup>30</sup>.

### 3.1.1 The CoMFA method

Which stands for Comparative Molecular Field Analysis, is based on the fundamental concept that differences in activity observed among molecules in a series are directly associated with dissimilarities in the shape of non-covalent Molecular Interaction Fields (MIFs). This method carefully analyzes the steric and electrostatic interaction fields exhibited by the molecules. The shape of these interaction fields is described by sampling their magnitude at regular intervals within the space surrounding the molecules, resulting in the generation of a three-dimensional matrix. Once the interaction fields have been described for each molecule, they are compared using a statistical analysis similar to Partial Least Squares (PLS). This analysis is crucial as it allows for the establishment of correlations between variations in the position and intensity of the potentials and the corresponding variations in activity. Consequently, valuable insights can be obtained regarding trends that either favor or hinder the target property. However, in order to accurately compare these three-dimensional objects, they need to be aligned in a common frame of reference, which entails aligning the molecules themselves <sup>31</sup>.

### 3.1.2 The CoMSIA method

On the other hand, is an extension of CoMFA that incorporates additional fields, including a lipophilic interaction field, a "hydrogen bond acceptor" field, and a "hydrogen bond donor" field, in addition to the steric and electrostatic interaction fields. While both techniques generally yield comparable results, CoMSIA models are often considered to be more comprehensive and easier to interpret <sup>32</sup>.

### 3.1.3 Alignment of the structures

Another important aspect in the development of a 3D-QSAR model of the CoMFA or CoMSIA type is the alignment of the structures in their active conformation. Ideally, the structure of the receptor should be co-crystallized with one or more ligands that closely resemble the series of compounds being studied. In this case, the co-crystallized ligand serves



as the reference molecule onto which the structures in the series can be aligned. Moreover, having the structure of the receptor also enables the docking of the compounds, which can potentially provide a plausible alignment.

Unfortunately, the availability of the structure of the receiver is lacking in the majority of cases.

Consequently, the alignment process will have to rely solely on the structures present in the series. This alignment process heavily depends on the experience and expertise of the modeller. While rational and automated methods have been developed to aid in the alignment process, these methods still present challenges in terms of accessibility.

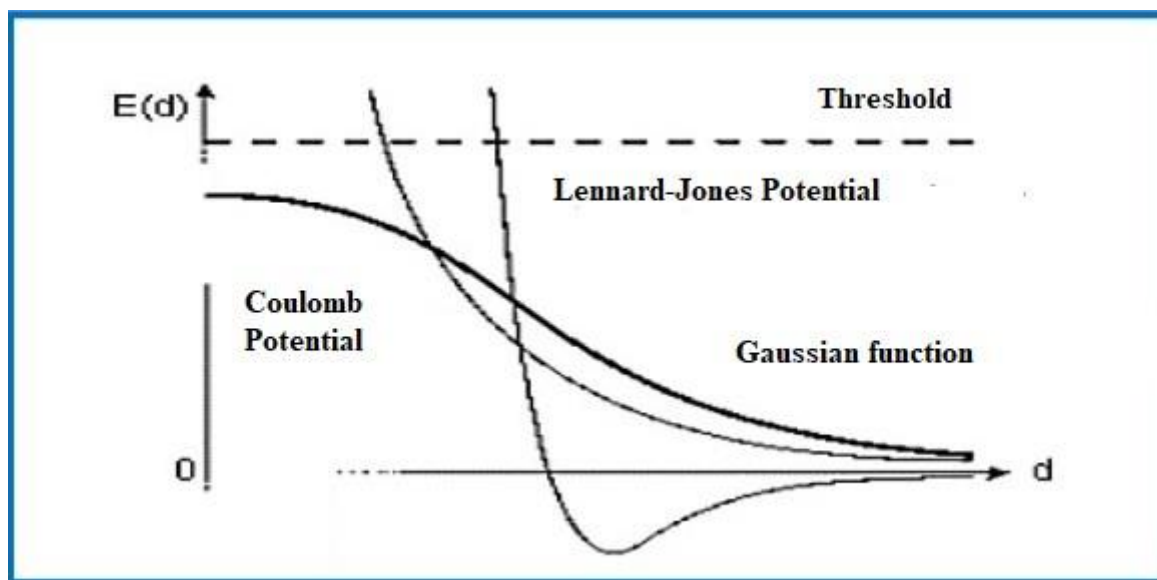
- One approach to automated alignment involves the execution of algorithms that call on different independent programs and scripts. However, this method may require a significant amount of user intervention and is not fully automated.
- Another approach involves integrating the entire alignment procedure into a single software. However, this approach is typically found in commercial software and may not be readily available to all researchers <sup>33-35</sup>.

### 3.1.4 Calculation of molecular interaction fields

When it comes to the calculation of molecular interaction fields, the CoMFA method utilizes a Lennard-Jones potential to calculate the steric field and a Coulomb potential to calculate the electrostatic field <sup>31</sup>. While this approach is widely accepted and efficient, it does present certain challenges. Specifically, both of these potential functions exhibit a steep slope near the Van der Waals surface of the molecule. This steep slope leads to abrupt changes in the description of the surfaces, necessitating the use of threshold values to avoid calculating potentials within the molecule. Additionally, a scaling factor is applied to the steric field to allow for comparison and utilization alongside the electrostatic field in the same Partial Least Squares (PLS) analysis. Furthermore, if the measurement matrix's orientation is altered relative to the set of aligned molecules, significant variations in the analysis results can be observed. These discrepancies are likely attributable to the use of strict threshold values <sup>36</sup>.

In the case of CoMSIA, five distinct similarity fields are calculated, including the steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields. These fields encompass the major types of interaction involved in ligand-receptor binding <sup>37</sup>. Similarity indices are computed in a three-dimensional matrix, which is comparable to that used

in CoMFA. The potentials, which depend on the distance between the probe and the molecule, are modeled using a Gaussian function. The shape of this function differs from conventional potential functions and enables the calculation of similarity indices for all points in the matrix, both within and outside the Van der Waals surface (Figure 3).



**Fig.3.** General shape of the classic potentials used by CoMFA (thin lines) and the potential used by CoMSIA (bold line).

### 3.1.5 Interpretation and validation of a QSAR model

Interpretation and validation of a QSAR model entails various important considerations. Once the model equation has been obtained, it becomes imperative to assess not only the stability and goodness of fit of the model, but also to estimate its power and validity before utilizing it for the prediction of biological activity. The validity of a QSAR model involves establishing the reliability and significance of the method for a specific purpose. Hence, it is essential to validate a QSAR model <sup>38</sup>.

The process of validation encompasses three main types: internal validation, external validation, and standard statistical tests and coefficients.

#### 3.1.5.1 Internal validation

Internal validation primarily focuses on ensuring the stability of the model using the dataset from which it was built. A commonly employed technique for internal validation is cross-

validation (CV), which is frequently used in statistical modeling. The CV method involves the extraction of one compound from the set each time, followed by the recalculation of the model using the remaining compounds ( $n-1$ , where  $n$  is the number of compounds) as the training set. This allows for the prediction of the biological activity value for the extracted compound across all compounds in the set. This iterative process is repeated for each compound in the initial set, resulting in a prediction for every object <sup>39</sup>.

The LOO (leave-one-out) method is a specific form of cross-validation that is often utilized in QSAR modeling. Alternatively, the CV leave-out method, also known as Leave-many-Out (LMO) or leave-group-out (LGO) <sup>40</sup>, can be employed by omitting multiple compounds from the dataset during each iteration. The calculation of the correlation coefficient for the  $Q^2$  cross-validation procedure is crucial, as it provides an indication of the model's validity. By definition, the  $Q^2$  correlation coefficient is smaller than or equal to the  $R^2$  set (correlation coefficient) for a QSAR equation. Another procedure for testing the validity of the model is the randomization test, which involves randomizing the compounds. This is particularly useful in situations where a large number of compounds and a small number of descriptors are present, as an equation can still possess low predictive power <sup>41</sup>.

### 3.1.5.2 Standard statistical tests and coefficients

Apart from internal validation, standard statistical tests and coefficients provide additional means of judging and validating QSAR models. A range of tests and coefficients can be employed for this purpose, and **Table 1** presents the most significant ones that serve as a standard for model validation <sup>42</sup>. These tests and coefficients play a pivotal role in ensuring the reliability and robustness of QSAR models, thereby contributing to their overall validation process.

**Table 1.** Standard tests for model validation.

Test	Definition
Fischer Test (F)	$F \text{ (observé)} = \frac{\sum(\hat{y}_i - \hat{y})^2}{\sum(y_i - \hat{y}_i)^2} \frac{n-p-1}{P}$
Mean Square Error (MSE)	$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$
Correlation Coefficient (r)	$r = \sqrt{1 - \frac{\sum(\hat{y}_i - \hat{y})^2}{\sum(y_i - \hat{y}_i)^2}}$
The Residual Standard Error (s)	$s = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-p-1}}$

$\hat{y}$  and  $y_i$  the observed and calculated values of the dependent variable.

n: number of observations (molecules).

P: degrees of freedom.

### 3.1.5.3 External validation

In order to ensure the dependability of evaluating the forecasting capability, it becomes imperative to employ an external validation set that has not been used for the development of the model. This external validation set is necessary to ensure that the initial data set is of sufficient size. The validation of the model is done using specific parameters, namely  $R^2$  (test) and  $R_{CV}^2$  (test). Moreover, it is crucial to examine other parameters for the purpose of external validation. These parameters, commonly known as "external validation criteria" or often referred to as "Tropsha criteria", play a significant role in assessing the model's predictive power. The "Tropsha criteria" <sup>43</sup> serve as the guiding principles for the evaluation of the model's performance during the external validation process.

External validation criteria, also known as test series, serve as the foundation for evaluating the reliability and accuracy of the model. These criteria play a pivotal role in determining the model's ability to make accurate predictions when confronted with new and unseen data. By employing an external validation set that has not been utilized during the model development phase, it ensures that the model's performance is tested on an independent and unbiased dataset. This external validation process is crucial in order to validate the model's generalizability and

to ensure that the model is not over fitting to the initial data set. Therefore, the utilization of external validation criteria is imperative for accurately assessing the model's predictive power and its ability to perform well on unseen data. Additionally, these criteria help in establishing the model's credibility and applicability in real-world scenarios, making them an essential component of any predictive modeling process.

### External validation criteria (test series)

- $R^2 > 0.7$
- $R_{CV}^2 > 0.6$
- $R^2 - R_0^2 < 0.1$  and  $0.85 \leq k \leq 1.15$
- $R^2 - R_0'^2 < 0.1$  and  $0.85 \leq k' \leq 1.15$
- $|R^2 - R_0^2| \leq 0.3$

With:

$R^2$ : Correlation coefficient for the molecules in the test series.

$R_0^2$ : Correlation coefficient between predicted and experimental values for the test series.

$R_0'^2$ : Correlation coefficient between experimental and predicted values for the test series.

K: is the constant of the correlation line (at the origin) (predicted values as a function of experimental values)

K': is the constant of the correlation line (at the origin) (experimental values as a function of predicted values).

### 3.1.6 Applications of the QSAR study

The applications of the Quantitative Structure-Activity Relationship (QSAR) study are vast and diverse. While some QSAR studies may appear to be more academically-oriented, these models have found numerous practical applications <sup>44</sup>. These applications encompass a wide range of fields and industries, including but not limited to:

- The optimization of pharmacological activity: QSAR models can be used to predict and optimize the activity of drug compounds. By understanding the relationship between the structure of a molecule and its biological activity, researchers can design more effective and potent drugs.
- The rational design of other products: QSAR models can also be applied to the design of various non-pharmaceutical products. For example, they can be used to optimize the properties of surfactants, fragrances, dyes, and fine chemicals, ensuring their desired functionality and performance.
- The identification of hazardous compounds: QSAR models can aid in the early detection and identification of potentially hazardous compounds during the product development process. By predicting the toxicity of compounds, researchers can prioritize safety and mitigate potential risks.
- Predicting toxicity and side effects: QSAR models can help in predicting the toxicity and side effects of new compounds. This information is crucial in drug development, as it allows researchers to identify potential safety concerns before extensive testing in animals or humans.
- Predicting toxicity to environmental species: QSAR models can also be used to predict the toxicity of chemicals to various environmental species. This information is valuable in assessing the potential environmental impact of new compounds and guiding regulatory decisions.
- The selection of compounds with optimal pharmacokinetic properties: QSAR models can assist in identifying compounds with favorable pharmacokinetic properties. This includes factors such as stability and availability in biological systems, which are important considerations in drug development.
- Predicting physico-chemical properties: QSAR models can be utilized to predict a variety of physico-chemical properties of molecules. This includes properties such as solubility, partition coefficient, and molecular weight, which are essential for understanding a compound's behavior and interactions.
- Predicting combined effects of molecules: QSAR models can predict the combined effects of molecules, whether in mixtures or formulations. This information is valuable in fields such as environmental toxicology and pharmaceutical formulation development.

### 3.2 Structure-based virtual screening

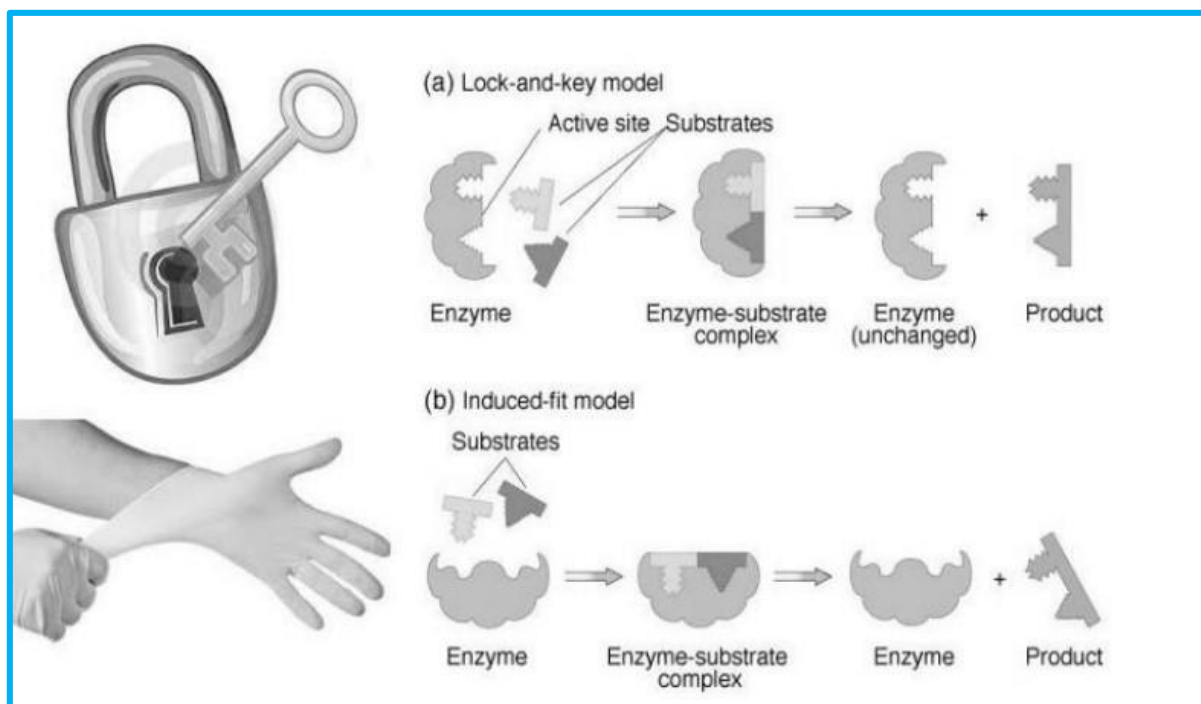
Moving on to another aspect of drug discovery, structure-based virtual screening plays a crucial role in identifying potential drug candidates. This approach relies on the knowledge of the target's three-dimensional (3D) structure and evaluates the ability of ligands to interact with the binding site <sup>45</sup>. Structure-based virtual screening can be performed using two experimental methods: X-ray crystallography and Nuclear Magnetic Resonance (NMR). These methods provide detailed information about the 3D structure of the target, which can then be used to discover new active compounds.

When the 3D structure of the target is available, various structure-based methods can be employed. These methods include the construction of pharmacophore models based on the target's structure, the establishment of RD-QSAR models, de novo design, and docking methods. These techniques rely on the identification of the binding site and aim to identify molecules that can effectively interact with the target.

In summary, QSAR studies have a wide range of applications in the fields of drug discovery and product development. These models can optimize pharmacological activity, aid in the rational design of various products, identify hazardous compounds, predict toxicity and side effects, select compounds with optimal properties, predict physico-chemical properties, and assess the combined effects of molecules. Additionally, structure-based virtual screening, based on the 3D structure of the target, allows for the identification of potential drug candidates through the assessment of ligand-binding interactions <sup>45, 46</sup>.

### 4. Molecular docking

In an organism, there are numerous ways in which two chemical elements can interact, with the most common interactions being protein-protein or protein-small molecule interactions. The field of molecular docking enables us to estimate the intermolecular framework that is generated between a macromolecule and a small molecule or between two macromolecules. It also allows us to identify the binding modes that are necessary for target regulation. Traditionally, the concept of molecular recognition of a ligand/target complex was compared to a "lock and key" mechanism. However, this terminology has now been refined to take into account the target-flexibility and the mutual adaptation with the ligand, resembling a "hand and a glove" interaction (**Figure 4**).



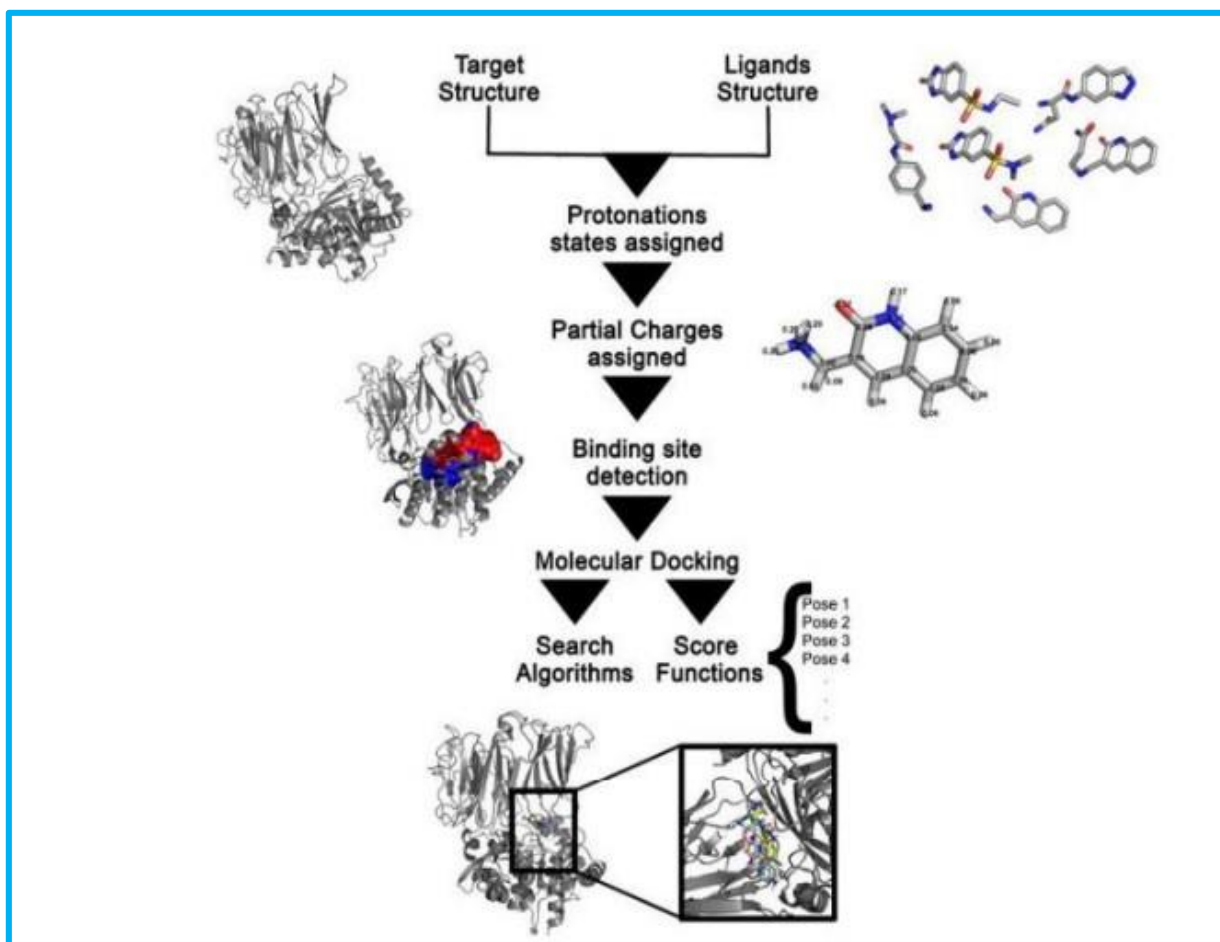
**Fig.4** The classical analogy of ligand/target complex as "lock and key" and the current analogy as "hand and glove" <sup>47</sup>.

To conduct reliable docking experiments, it is essential to have a high-resolution X-ray, NMR, or homologymodeled structure with a known or expected binding site in the target. Docking approaches involve fitting a ligand into a binding site by considering factors such as steric, hydrophobic, and electrostatic complementarity. Additionally, the evaluation of binding free energy (scoring) is also crucial <sup>48–50</sup>.

#### 4.1 General protocol

The process of molecular docking is carried out through two main pathways. Firstly, all possible poses of the ligand within the target binding site are sampled, and each pose is associated with a score value that approximates its free energy landscape. Then, at the end of the simulations, the best binding modes are ranked based on these score values, and the most suitable complexes are selected. Ideally, the sampling algorithms should be able to reproduce the experimental binding mode, and the scoring function (SF) should rank it as the best among all the generated conformations <sup>51</sup>. **Figure 5** provides an overview of the general workflow of a molecular docking simulation.





**Fig. 5.** General workflow of molecular docking simulation <sup>52</sup>.

#### 4.1.1 Ligand preparation

Ligand preparation involves the generation, optimization, and analysis of its three-dimensional (3D) structure. There are several drawing software options available such as ChemSketch, ChemDraw, Avogadro, and others, which can be used to manually build the 3D structures of ligands. These structures can be created either from their 2D representations or from simpler schemes like SMILES. Additionally, ligand structures can also be obtained directly from virtual databases such as the Cambridge Structural Database (CSD), Available Chemical Directory (ACD), MDL Drug Data Report (MDDR), PubChem, and more. These databases provide a vast number of conformations that can be used for virtual screening <sup>50-52</sup>.

#### 4.1.2 Target preparation

Target preparation involves the minimization, correction, and protonation of its 3D structure. The structures of targets can be obtained in PDB file format from various protein databases like

the RCSB-Protein Data Bank (<https://www.rcsb.org/>). These protein databases provide access to the 3D atomic coordinates of target conformations, which are characterized using experimental and/or computational prediction methods such as X-ray crystallography, nuclear magnetic resonance (NMR), infrared spectroscopy, homology modeling, electron density, and more <sup>50-52</sup>.

### 4.1.3 Binding site detection

In the context of binding site detection, it is common for the binding site that is of interest for the docking simulation to already be known and assigned by the docking software. However, in cases where the information regarding the binding location is unavailable, it is possible to predict the most likely locus algorithmically or by implementing a so-called "blind docking" approach. The blind docking method covers the entire surface of the target, but it comes with a high computational cost <sup>50, 52</sup>.

### 4.1.4 Docking validation

If the process of connecting two entities together, known as docking, is able to regenerate the identical manner of binding that was originally established and defined through the utilization of experimental and/or computational approaches, it is acknowledged as being valid. The Root-Mean-Square Deviation (RMSD) serves as a valuable tool for evaluating the degree of structural similarity between two configurations that have been superimposed upon one another. This calculation is derived from the equation presented as (\*) <sup>53</sup>. Furthermore, the RMSD, in combination with the values obtained from the scoring function, known as Sscore, are employed for the purpose of evaluating and ranking the stability and attraction of complexes consisting of a ligand and a target. These assessments are made possible through the values generated from both RMSD and Sscore, as highlighted in references <sup>50, 52</sup>.

$$RMSD = \sqrt{\frac{1}{n} \sum_{x=1}^n d_x^2} \quad (*)$$

### 4.2 Types of molecular docking

#### 4.2.1 Rigid docking

In this particular type of molecular docking, both the ligand and the protein are treated as rigid objects, meaning that their structures are not allowed to change during the docking process. This approach is employed in order to limit the search space and reduce the computational complexity by considering only three translational and three rotational degrees of freedom for both the ligand and the protein. The flexibility of the ligand, however, is taken into account through the use of a pre-adjusted list of ligand conformations or through the inclusion of atom-atom interactions between the target and the ligand. This approximation allows for the simulation of the "lock-key" binding mechanism, which is often observed in cases where the number of conformational degrees of freedom is too high to be sampled, particularly in protein-protein docking. In this method, the target site and the ligand are represented as "hot" points, and the superposition of these matching points is assessed to determine the best binding conformation <sup>54</sup>.

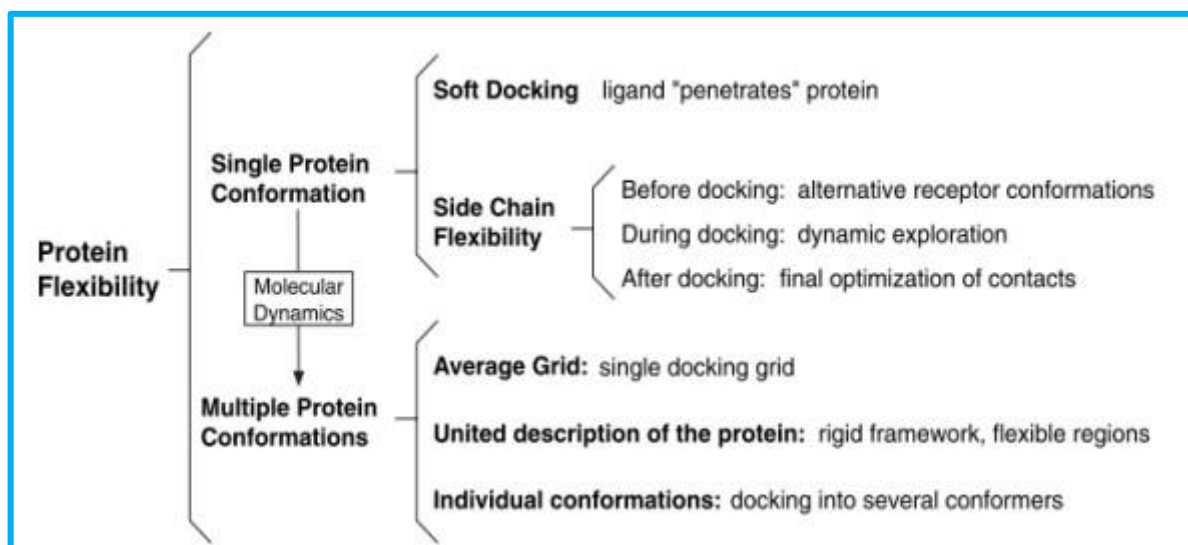
#### 4.2.2 Semi-flexible docking

Considering the inherent flexibility of molecular systems, it is crucial to incorporate this flexibility into the docking process to accurately model the interactions between ligands and receptors. This flexibility allows both ligands and receptors to adapt their conformations and form the most energetically favorable complex. However, due to the computationally expensive nature of fully flexible docking, most docking software employ a semi-flexible docking approach. In this approach, the protein conformation is assumed to be fixed and capable of recognizing the ligands to be anchored. Although this assumption is not always supported, it is a practical compromise to reduce the computational cost and time. In semi-flexible docking, the ligands are treated as flexible entities, while the receptor remains rigid throughout the docking process. The search space is expanded to include six translational and six rotational degrees of freedom, and all the conformational degrees of freedom of the ligand are sampled <sup>55</sup>.

#### 4.2.3 Flexible-flexible docking

On the other hand, fully flexible docking takes into account the flexibility of both the ligand and the receptor. This technique allows the flexible ligand to freely anchor into a flexible

receptor, taking into consideration that the intrinsic kinetics of the protein are strongly influenced by the orientation of the ligand in the binding site. Implementing receptor flexibility in molecular docking represents a significant advancement in the field, as it enables the modeling of all degrees of freedom in the ligand-receptor complex. However, this approach is not without its limitations. One of the main challenges is the problem of insufficient sampling, which can lead to inaccurate results. Additionally, the computational costs associated with fully flexible docking can be prohibitively high, especially when dealing with large chemical libraries. The size and complexity of protein structures further complicate the inclusion of full receptor flexibility in the docking process. Therefore, most studies on molecular docking are typically limited to specific residues or regions of interest within the protein. Nonetheless, there have been recent developments in docking programs that partially incorporate receptor flexibility, offering more flexibility in terms of target conformation. These programs provide various strategies for implementing receptor flexibility, allowing researchers to explore different conformational states and potentially improve the accuracy of molecular docking predictions. **Figure 6** provides a visual representation of the different strategies for implementing receptor flexibility in molecular docking.



**Fig. 6.** The described strategies for including receptor flexibility in docking simulation <sup>56</sup>.

### 4.3 Scoring Functions (SFs)

Scoring functions (SFs) serve as crucial pose selectors, effectively distinguishing between the most efficient biological binding modes and binders from inactive ligands within the set of poses acquired through the sampling algorithm (Figure 7). In contrast, SFs rely on various approximations and simplifications to estimate, rather than precisely calculate, the binding affinity of the target and the ligand<sup>55, 57</sup>. Consequently, an accurate and reliable SF should possess three essential capabilities: firstly, the competence to select the optimal binding mode of a ligand from a collection of computationally simulated poses; secondly, the competence to accurately rank a given set of ligands with known binding modes when attached to the same target; and thirdly, the competence to generate binding scores that are linearly associated with experimentally measured binding affinities of target-ligand complexes possessing known three-dimensional structures<sup>58</sup>.

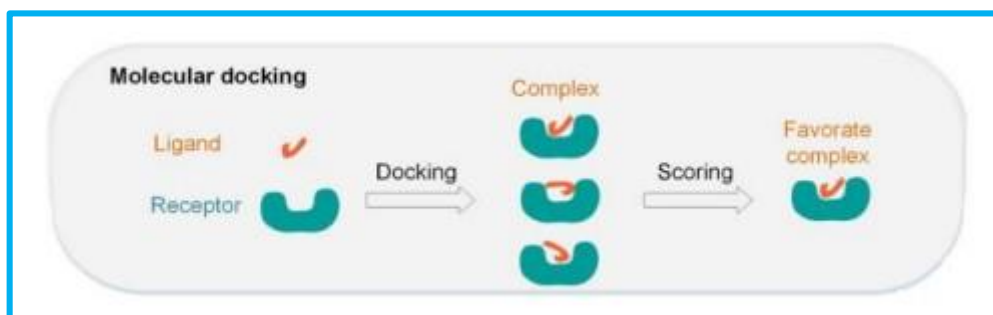
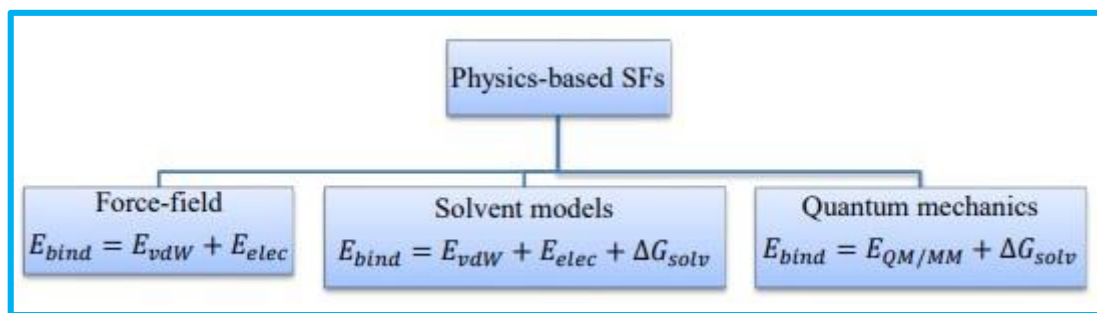


Fig. 7. The SF role as pose selector<sup>59</sup>.

Recently, the field of SFs has witnessed the development of five distinct classes: physics-based, empirical, knowledge-based, consensus, and machine-learning-based SFs<sup>59</sup>.

#### 4.3.1 Physics-based SFs

Physics-based SFs encompass SFs that are founded upon force field principles, solvation models, and quantum mechanics methods (Figure 8)<sup>59</sup>. The concept of force field, rooted in molecular mechanics, represents a traditional approach that entails utilizing a combination of the bonded (intramolecular) and non-bonded (intermolecular) components of a system to estimate its potential energy. In the context of the docking approach, non-bonded elements are frequently taken into consideration, with the possibility of incorporating ligand-bonded terms, particularly the torsional elements<sup>55</sup>.



**Fig. 8.** The description for physics-based SF <sup>59</sup>.

The estimation of binding energy in classical force-field-based scoring functions (SFs) often relies on the sum of non-bonded interactions, which are represented by electrostatic and van der Waals energy terms. However, in addition to these interactions, other factors such as hydrogen bonds, solvation effects, and entropy contributions can also be taken into account in force-field-based SFs. The calculation of electrostatic forces involves the use of the coulomb formula. However, due to the limitations of point charge computations in accurately describing the surrounding environment of the target, a distance-dependent dielectric function is typically employed to adjust the effect of charge-charge interactions. Similarly, the van der Waals terms are defined by the Lennard-Jones potential function. The hardness of this potential, which governs the permissible cut-off distances between the target and ligand atoms, can be adjusted by varying the parameter settings for the Lennard-Jones potential <sup>51</sup>.

Different software programs have varying approaches to handling hydrogen bonding, the form of the energy function, and other aspects of SFs. To improve the accuracy of binding energy prediction, the results of docking simulations using force-field-based functions can be further refined using other methods such as linear interaction energy and free-energy perturbation methods (FEP). Several examples of force-field-based SFs include D-Score, G-Score, GOLD, AutoDock, MOE, DOCK, HADDOCK Score, ICM SF, QXP SF, GBVI/WSA, and many more <sup>48,55</sup>.

#### 4.3.2 Empirical SFs

On the other hand, empirical SFs consider simpler energy factors, resulting in faster binding score calculations and reasonably accurate predicted binding energies. These SFs incorporate various energy components, including van der Waals, electrostatic, hydrogen bond, ionic interaction, desolvation, hydrophobic effect, and binding entropy. To obtain the final score,

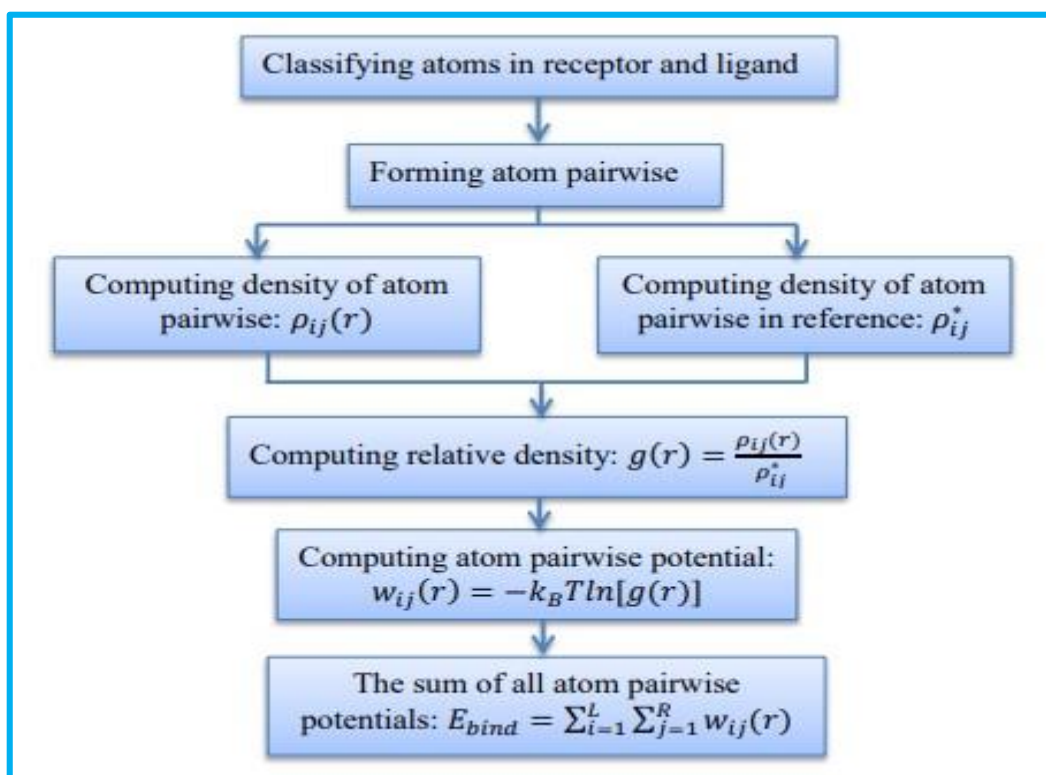
each energy term is multiplied by a coefficient and then summed together. These coefficients are determined through regression analysis on a training set of experimental ligand-target complexes with known binding affinities. However, it is important to note that these SFs are heavily dependent on the molecular data sets used for regression analysis and fitting, which often leads to different weighting factors for individual terms. Consequently, merging terms from different SFs into a new SF can be challenging. Some examples of SFs in this category include LUDI, GlideScore, ChemScore, PlantsChemlp, SCORE, RankScore, LigScore, HINT, F-Score, Fresno, X-Score, and more <sup>48, 55</sup>.

### 4.3.3 Knowledge-based SFs

In contrast to force-field-based SFs, knowledge-based SFs aim to replicate experimental structures rather than binding energies. They are based on the premise that more favorable interactions are more likely to occur. To achieve this goal, statistical analysis is applied to a dataset of crystal 3D structures of ligand-target complexes to determine the frequencies of interatomic interactions and distances between the ligand and target. These frequency distributions are then used to derive paired atom-type potentials. The score is subsequently calculated by prioritizing favorable connections and eliminating repulsive interactions between ligand-target atoms within a predefined cutoff <sup>48</sup>. The general pathway of knowledge-based SFs is illustrated in [Figure 9](#).

Knowledge-based functions provide a convenient means of conducting thorough searches in vast molecular databases, making them highly valuable tools in the field of drug discovery. In addition to their ability to efficiently scan through these databases, knowledge-based functions also possess the capability to model specific types of interactions that are often overlooked in empirical scoring functions (SFs). These interactions include sulphur-aromatic or  $\pi$ -cation interactions, which are known to play crucial roles in ligand binding. Furthermore, one of the major advantages of knowledge-based SFs is that their training sets primarily focus on the structural details of the molecules, completely disregarding any experimental binding affinity. This is particularly significant as it eliminates any uncertainty or bias that may arise from the experimental environment, thus enhancing the reliability of the predicted binding affinity. Consequently, knowledge-based SFs are better suited for accurately predicting binding poses rather than binding affinities. However, it is important to note that these SFs are not without their challenges. Two key challenges that researchers face when utilizing knowledge-

based SFs are the accuracy of estimating the reference state and the under-representation of interactions with halogens and metals in the limited training sets.



**Fig. 9.** The general pathway of knowledge-based SFs.  $\rho_{ij}(\mathbf{r})$ : number density of the target-ligand atom pair  $i\_j$  at distance  $r$ .  $\rho_{ij}^*$  pair density in a reference state.  $g(r)$ : relative number density of atom pairwise  $i\_j$  at distance  $r$ .  $\mathbf{K}_B$ : Boltzmann constant.  $T$ : absolute temperature <sup>59</sup>.

Within the realm of knowledge-based SFs, there exist several notable examples that fall under this category, each with its own unique set of features and characteristics. Some of these SFs include DrugScore, GOLD/ASP, PMF, SMOG, Bleep, MScore, and ITScore/SE. These SFs have been extensively utilized in various drug discovery studies and have proven to be effective in their respective applications. Their diverse range of functionalities allows researchers to choose the most suitable SF for their specific needs and objectives.

#### 4.3.4 Consensus scoring

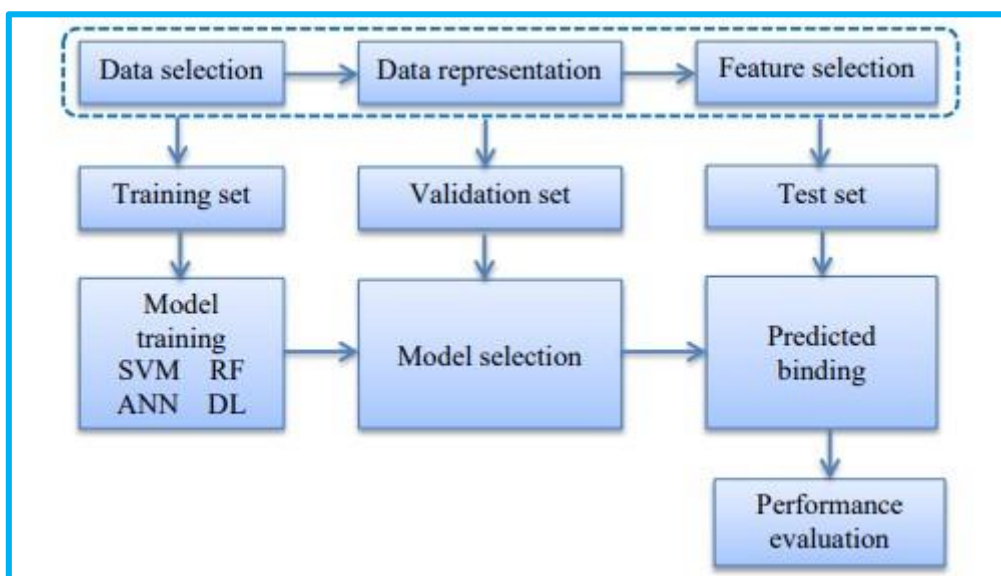
To overcome the inherent limitations and constraints associated with the aforementioned classes of SFs, the scientific community has embraced the concept of combining multiple SFs to achieve more accurate and comprehensive results. This approach, known as "consensus scoring," involves integrating information from multiple scoring schemes to balance out any



potential errors in individual scores and improve the likelihood of accurately detecting the optimal ligand conformation. In essence, a suggested ligand pose is more likely to be accepted if it performs well across a variety of scoring schemes. One prominent example of the consensus scoring approach is X-CSCORE, which combines the functionalities of GOLD-like, DOCK-like, ChemScore, PMF, and FlexX SFs. By leveraging the strengths of these different SFs, X-CSCORE aims to provide a robust and reliable scoring system that can enhance the accuracy of ligand binding prediction.

#### 4.3.5 Machine-learning-based SFs

In recent years, machine learning algorithms have emerged as powerful tools in the field of drug discovery. Machine-learning-based SFs have been developed and applied using various algorithms, including support vector machine, random forest, neural network, and deep learning, among others. These SFs have showcased superior performance compared to classical SFs, particularly in the area of rescoring. Rather than being integrated directly into docking software, machine-learning-based SFs are often utilized to rescore the results obtained from classical docking techniques. This two-step process, involving initial docking using classical methods followed by rescoring using machine learning SFs, has been shown to significantly improve the accuracy of ligand binding predictions. By harnessing the information contained within the training dataset, machine-learning-based SFs are able to effectively capture the complex relationships between ligand-target interactions and generate more accurate predictions <sup>60</sup> (Figure 10).



**Fig. 10.** Workflow of training machine-learning-based SFs <sup>59</sup>.

### 5. Pharmacokinetics Properties and Computational tools employed in ADMET

During the early stages of drug development, the initial assessment of candidate drugs' activities and specificities is typically followed by evaluations of pharmacokinetics and toxicities. However, it has been observed that a significant number of candidate drugs fail in the final stage due to poor efficacy and safety, primarily attributed to their absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics <sup>61</sup>. Researchers have identified that poor safety and toxicity are the most prominent factors contributing to the failure of more than half of all project closures. Consequently, recognizing the importance of filtering and optimizing the ADMET features of pharmaceuticals at an early stage has become widely acknowledged and employed in order to reduce the attrition rate in drug research and development, similar to the process of drug discovery. In recent years, *in vitro* and *in vivo* ADMET prediction approaches have gained popularity. However, conducting sophisticated and expensive ADMET tests on a large number of drugs is not feasible. Therefore, there is a growing interest in an *in silico* approach to predict ADMET characteristics as a cost-effective and high-throughput alternative to experimental testing methods <sup>62</sup>.

In order to effectively predict ADMET characteristics, there are two essential components to consider: data modelling and molecular modelling, each requiring its own set of tools. Data modelling commonly employs Quantitative Structure Activity Relationship (QSAR) techniques. The QSAR method aims to identify connections between a specific property and a series of chemical and structural descriptors for the molecules under investigation. Over the past six decades, a wide range of descriptors suitable for QSAR research have been generated, such as those available in the software program Dragon. From this vast pool of descriptors, it is possible to select a subset that could prove useful in future predictions of ADME features. On the other hand, molecular modelling utilizes quantum mechanical methods to analyze the potential for interaction between small compounds and proteins known to be involved in ADME processes, such as cytochrome P450s. In cases where the structure of the human protein is unknown, homology modelling of related structures can be employed to construct three-dimensional structural information of the protein <sup>60</sup>.

### 6. Quantum mechanics

Quantum mechanics (QM) serves as an essential method for comprehending the behavior of systems at the atomic level. In QM, molecules are treated as collections of nuclei and electrons, devoid of any explicit reference to chemical bonds. By applying the principles of QM, approximations are made to the wave function and the Schrödinger equation is solved. The solutions to the Schrödinger equation provide insights into the motions of electrons, which subsequently determine molecular structure, energy, and various other observables, including bonding information. However, it is important to note that the Schrödinger equation cannot be solved exactly for systems with more than one electron, except for the hydrogen atom, necessitating the use of approximations <sup>63</sup>.

$$H=T+V \quad (3)$$

Where H is the Hamiltonian operator (sum of kinetic energy), T the potential energy, and V the operator. H can also be defined as:

$$H=[-\frac{\hbar^2}{2m} \sum_i (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})] + \sum_i \sum_{j < i} (e_i e_j / r_{ij})$$

QM methods encompass various approaches, such as ab initio calculations, density functional theory (DFT), and semi-empirical methods. For more accurate QM calculations, electron correlation methods, including CCSDT and MP2, are often employed. These methods are also valuable in identifying the activated complex in chemical reactions, thereby aiding in the determination of reaction pathways. Due to the inherent complexity of solving the Schrödinger equation for large molecular systems, semi-empirical and ab initio DFT methods have been developed as approximations to obtain precise QM solutions. Although QM models offer the highest level of accuracy, they are also the most computationally demanding and time-consuming, typically applied to smaller systems due to resource limitations <sup>64</sup>.

## 7. Density Functional Theory (DFT)

### 7.1 Conceptual Density Functional Theory (DFT)

The utilization of methodologies in Computational Chemistry plays an extremely crucial role in the application of modern medicinal chemistry, providing a vast potential for enhancing various stages of drug research, with a particular focus on time and cost effectiveness <sup>65</sup>. The recent impact of density functional theory (DFT) has had a considerable influence on the advancement of quantum chemistry, which can be attributed to the achievement of what is

known as "chemical accuracy" in the late 1980s through the introduction of gradient-corrected and hybrid functional methods. DFT is primarily based on the renowned Hohenberg and Kohn theorems established in 1964, which concentrate on the electron density,  $p(\mathbf{r})$ , itself as the carrier of molecular (or atomic) properties at significantly lower costs compared to traditional ab initio wave function techniques. The incorporation of orbitals into the conceptual framework of DFT was accomplished through the Kohn-Sham formalism. The Kohn-Sham methodology encompasses the determination of molecular energy and density for a given system, as well as the orbital energies that are explicitly associated with the frontier orbitals, including the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO) <sup>60</sup>.

### 7.1.1 Fundamental and Computational Aspects of DFT

#### ➤ The Basics of DFT: The Hohenberg–Kohn Theorems

In the realm of chemical reactions, the formation and breaking of bonds occur due to the accumulation and depletion of electron density between nuclei. Hence, understanding how the electron density in a molecule redistributes during a chemical reaction is crucial in the field of chemistry. For a system consisting of  $N$  electrons bound by an external potential  $v(\mathbf{r})$ , the Hamiltonian  $H$  is fully defined by  $N$  and  $v(\mathbf{r})$ . By solving the Schrödinger equation with  $H$ , one can obtain the many-electron wave function  $\psi(r_1, r_2, \dots, r_N)$ , which contains all the physical information about the system. Integrating over the coordinates of  $(N-1)$  electrons, the single-particle density or the electron density  $p(\mathbf{r})$  can be obtained <sup>60</sup>.

$$p(\mathbf{r}) = N \int \dots \int \psi^*(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) d\mathbf{r}_2 \dots d\mathbf{r}_N \quad (3)$$

Which integrates to the total number of electrons,

$$\int p(\mathbf{r}) d\mathbf{r} = N \quad (4)$$

Therefore  $N$  and  $v(\mathbf{r})$  determine  $p(\mathbf{r})$ . That is, there is a mapping from  $v(\mathbf{r})$  to  $p(\mathbf{r})$ .

#### ➤ DFT as a Tool for Calculating Atomic and Molecular Properties: The Kohn–Sham Equations

When examining the quantitative aspects related to Conceptual DFT descriptors, the Kohn-Sham approach proves to be advantageous. However, there is currently significant controversy surrounding the utilization of a range-separated exchange-correlation density functional in Kohn-Sham DFT. The partitioning of the exchange and correlation operators into long- and short-ranged components, along with the inclusion of a range-separation parameter that governs the rate at which long-range behavior is achieved, is vital in constructing these density functionals. Through a molecule-by-molecule approach and following certain tuning criteria, the estimation of the range-separation parameter can be fixed or "tuned." The optimal tuning process relies on the relationship between the KS HOMO energy and the vertical ionization potential (IP), which calculates the energy difference,  $E(N-1) - E(N)$ . In the case of an N-electron molecular system, the application of Generalized KS theory is necessary <sup>60</sup>.

$$-\text{IP}(N) = \epsilon_{\text{H}}(N)$$

One might consider the Generalized KS theory as the DFT equivalent of the well-known Koopmans' theorem. However, it is important to note that only the exact density functional is truly valid. In situations where an approximated density functional must be utilized for practical reasons, there may be a substantial discrepancy between  $-\text{IP}(N)$  and  $\epsilon_{\text{H}}(N)$ . Consequently, achieving perfect tuning involves establishing a range-separation parameter that is specific to the system at hand <sup>60</sup>.

### References

1. E. A. Ratti and D. G. Trist, "The continuing evolution of the drug discovery process in the pharmaceutical industry.," *Farmaco*, vol. 56, no. 1–2, pp. 13–19, 2001.
2. J. Drews, "Drug discovery: a historical perspective," *Science*, vol. 287, no. 5460, pp. 1960–1964, Mar. 2000.
3. J. Drews, *In Quest of Tomorrow's Medicines*. New York: Springer-Verlag, 1999.
4. N. U. Meldrum and F. J. W. Roughton, "Carbonic anhydrase. Its preparation and properties," *J Physiol*, vol. 80, no. 2, pp. 113–142, Dec. 1933.
5. E. N. Bharath, S. N. Manjula, and A. Vijaychand, "In silico drug design: tool for overcoming the innovation deficit in the drug discovery process," *Int J Pharm Pharm Sci*. vol. 3, no. 2, pp. 5, Janry.2011.
6. M. J. Alomar, "Factors affecting the development of adverse drug reactions (Review article)," *Saudi Pharm J*, vol. 22, no. 2, pp. 83–94, Apr. 2014.
7. C. Jochum, and J. Gasteiger Canonical, "numbering and constitutional symmetry," *J Chem Inf Comput Sci.*, vol. 17, no. 12, pp. 113–117, May 1977.
8. A. Baldi, "Computational approaches for drug design and discovery: An overview." *Sys. Rev. Pharm.*, vol. 1, no. 1, pp. 99–105, 2010.
9. J. J. Irwin et al., "Automated docking screens: a feasibility study." *J. Med. Chem.*, vol. 52, no. 18, pp. 5712–5720, Sep. 2009.
10. C. McInnes, "Virtual screening strategies in drug discovery," *Curr Opin Chem Biol*, vol. 11, no. 5, pp. 494–502, Oct. 2007.
11. G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, "Computational methods in drug discovery," *Pharmacol. Rev.*, vol. 66, no. 1, pp. 334–395, 2014.
12. H. J. Huang et al., "Structure-based and ligand-based drug design for HER 2 receptor," *J. Biomol. Struct. Dyn.*, vol. 28, no. 1, pp. 23–37, Aug. 2010.
13. S. Wold and W. J. Dunn, "Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability," *J. Chem. Inf. Comput. Sci.*, vol. 23, no. 1, pp. 6–13, Feb. 1983.

14. G. Klebe, U. Abraham, and T. Mietzner, "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity," *J. Med. Chem.*, vol. 37, no. 24, pp. 4130–4146, Nov. 1994.
15. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *J Chem Inf Comput Sci*, vol. 43, no. 6, pp. 1947–1958, Dec. 2003.
16. Y. Wang, S. A. Shaikh, and E. Tajkhorshid, "Exploring transmembrane diffusion pathways with molecular dynamics," *Physiology (Bethesda)*, vol. 25, no. 3, pp. 142–154, Jun. 2010.
17. S. M. Hanson, S. Newstead, K. J. Swartz, and M. S. P. Sansom, "Capsaicin interaction with TRPV1 channels in a lipid bilayer: molecular dynamics simulation," *Biophys. J.*, vol. 108, no. 6, pp. 1425–1434, Mar. 2015.
18. Atkins Physical Chemistry" 8TH EDITION, 8th edition. WH Freeman , 2006.
19. D.W. Rogers. Computational Chemistry Using the PC. 3rd ed. Hoboken (NJ): Wiley; 2003.
20. E. G. Lewars, Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, 2nd ed. Springer Netherlands, 2011.
21. J. C. Tully, "Perspective on 'Zur Quantentheorie der Molekeln,'" *Theor Chem Acc*, vol. 103, no. 3, pp. 173–176, Feb. 2000.
22. M. Hotokka, Calculation of vibrational frequencies by molecular mechanics. In: Chalmers JM, Griffiths PR, editors. Handbook of Vibrational Spectroscopy. Hoboken (NJ): Wiley; 2006.
23. D. J. Tannor, Introduction to Quantum Mechanics, 2007 edition. Sausalito, Calif: University Science Books, 2007.
24. C. Hansch, T. Fujita,  $\rho$ - $\sigma$ - $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, *J. Am. Chem. Soc.* 86 (1964) 1616–1626.  
<https://doi.org/10.1021/ja01062a035>.
25. J. Verma, V. Khedkar, E. Coutinho, 3D-QSAR in Drug Design - A Review, *Curr. Top. Med. Chem.* 10 (2010) 95–115. <https://doi.org/10.2174/156802610790232260>.

- 26.** H. Tandon, T. Chakraborty, V. Suhag, A Concise Review on the Significance of QSAR in Drug Design, *Chem. Biomol. Eng.* 4 (2019) 45–51.  
<https://doi.org/10.11648/j.cbe.20190404.11>.
- 27.** K. Roy, On some aspects of validation of predictive quantitative structure-activity relationship models, *Expert Opin. Drug Discov.* 2 (2007) 1567–1577.  
<https://doi.org/10.1517/17460441.2.12.1567>.
- 28.** B.J. Neves, R.C. Braga, C.C. Melo-Filho, J.T. Moreira-Filho, E.N. Muratov, C.H. Andrade, QSAR-based virtual screening: Advances and applications in drug discovery, *Front. Pharmacol.* 9 (2018) 1–7. <https://doi.org/10.3389/fphar.2018.01275>.
- 29.** K.Z. Myint, X.Q. Xie, Recent advances in fragment-based QSAR and multidimensional QSAR methods, *Int. J. Mol. Sci.* 11 (2010) 3846–3866.  
<https://doi.org/10.3390/ijms11103846>.
- 30.** Cronin, Mark TD. "Quantitative structure–activity relationships (QSARs)—applications and methodology." *Recent advances in QSAR studies: methods and applications* (2010): 3-11.
- 31.** Cramer, Richard D., David E. Patterson, and Jeffrey D. Bunce. "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins." *Journal of the American Chemical Society* 110.18 (1988): 5959-5967.
- 32.** Klebe, Gerhard, Ute Abraham, and Thomas Mietzner. "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity." *Journal of medicinal chemistry* 37.24 (1994): 4130-4146.
- 33.** Lemmen, C. and T. Lengauer, Computational methods for the structural alignment of molecules. *Journal of computer-aided molecular design* FIELD Publication Date: (2000). 14(3): p. 215-32. FIELD Reference Number: 142 FIELD Journal Code: 8710425.
- 34.** Melani, Fabrizio, et al. "Field interaction and geometrical overlap: a new simplex and experimental design based computational procedure for superposing small ligand molecules." *Journal of medicinal chemistry* 46.8 (2003): 1359-1371.
- 35.** Arakawa, Masamoto, Kiyoshi Hasegawa, and Kimito Funatsu. "Novel alignment method of small molecules using the Hopfield Neural Network." *Journal of chemical information and computer sciences* 43.5 (2003): 1390-1395.
- 36.** Cho, Sung Jin, and Alexander Tropsha. "Cross-validated R<sup>2</sup>-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results." *Journal of medicinal chemistry* 38.7 (1995): 1060-1066.



37. Cichero, Elena, et al. "CoMFA and CoMSIA analyses on 1, 2, 3, 4-tetrahydropyrrolo [3, 4-b] indole and benzimidazole derivatives as selective CB2 receptor agonists." *Journal of molecular modeling* 16 (2010) : 1481-1498.
38. Phuong, HUYNH Thi Ngoc. "Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo [c] phénanthridiniques." (2007).
39. Debnath, Asim Kumar. "Quantitative structure-activity relationship (QSAR) paradigm--Hansch era to new millennium." *Mini reviews in medicinal chemistry* 1.2 (2001): 187-195.
40. Shao, Jun. "Linear model selection by cross-validation." *Journal of the American statistical Association* 88.422 (1993): 486-494.
41. Golbraikh, Alexander, et al. "Rational selection of training and test sets for the development of validated QSAR models." *Journal of computer-aided molecular design* 17 (2003): 241-253.
42. Clark, Robert D., and Peter C. Fox. "Statistical variation in progressive scrambling." *Journal of computer-aided molecular design* 18.7-9 (2004): 563-576.
43. Bahat, Mehmet, and Emre Yörük. "A computational study on structural, electronic and nonlinear optical properties of fury pyridine molecules." *Proceedings of the ninth WSEAS international conference on applied computer science.* (2009).
44. Wold S, Eriksson L, Clementi S. Statistical validation of QSAR results. *Chemo metric methods in molecular design.* (1995) :309-38.
45. Ziani. T. "Prédiction de l'effet inhibiteur des dérivés d'imidazole sur l'enzyme Cox2". *University Med Khider Biskra,* (2013).
46. Playe, Benoit. "Méthodes d'apprentissage statistique pour le criblage virtuel de médicament." *" Diss. Paris Sciences et Lettres (ComUE),* (2019).
47. M. Mukhopadhyay, a Brief Survey on Bio Inspired Optimization Algorithms for Molecular Docking, *Int. J. Adv. Eng. Technol.* 7 (2014) 868–878.  
[https://doi.org/10.7323/ijaet/v7\\_iss3](https://doi.org/10.7323/ijaet/v7_iss3).
48. A. Sethi, K. Joshi, K. Sasikala, M. Alvala, *Molecular Docking in Modern Drug Discovery: Principles and Recent Applications,* in: *Drug Discov. Dev. - New Adv.,* 2020: pp. 1–21.  
<https://doi.org/10.5772/intechopen.85991>.
49. D.J. Diller, K.M. Merz, High throughput docking for library design and library prioritization, *Proteins Struct. Funct. Genet.* 43 (2001) 113–124. [https://doi.org/10.1002/1097-0134\(20010501\)43:23.0.CO;2-T](https://doi.org/10.1002/1097-0134(20010501)43:23.0.CO;2-T).

- 50.** A. Stefaniu, Introductory Chapter: Molecular Docking and Molecular Dynamics Techniques to Achieve Rational Drug Design, in: *Mol. Docking Mol. Dyn.*, 2019: pp. 1–5. <https://doi.org/10.5772/intechopen.84200>.
- 51.** X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery., *Curr. Comput. Aid. Dru. Des.* 7 (2011) 146–157. <https://doi.org/10.2174/157340911795677602>.
- 52.** P.H.M. Torres, A.C.R. Sodero, P. Jofily, F.P. Silva-Jr, Key topics in molecular docking for drug design, *Int. J. Mol. Sci.* 20 (2019) 1–29. <https://doi.org/10.3390/ijms20184574>.
- 53.** S.P. Leelananda, S. Lindert, Computational methods in drug discovery, *Beilstein J. Org. Chem.* 12 (2016) 2694–2718. <https://doi.org/10.3762/bjoc.12.267>.
- 54.** R.D. Taylor, P.J.Jewsbury, J.W. Essex1, A review of protein-small molecule docking methods, *J. OfComputer-Aided Mol. Des.* 16 (2002) 151–166. <https://doi.org/doi:10.1023/A:1020155510718>.
- 55.** V. Salmaso, S. Moro, Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview, *Front. Pharmacol.* 9 (2018) 1–16. <https://doi.org/10.3389/fphar.2018.00923>.
- 56.** H. Alonso, A.A. Bliznyuk, J.E. Gready, Combining docking and molecular dynamic simulations in drug design, *Med. Res. Rev.* 26 (2006) 531–568. <https://doi.org/10.1002/med.20067>.
- 57.** F.D. Prieto-Martínez, M. Arciniega, J.L. Medina-Franco, Molecular docking: current advances and challenges, *TIP Rev. Espec. En Ciencias Químico-Biológicas.* 21 (2018) 1–23. <https://doi.org/10.22201/fesz.23958723e.2018.0.143>.
- 58.** H.M. Ashtawy, N.R. Mahapatra, Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins, *BMC Bioinformatics.* 16 (2015) 1–17. <https://doi.org/10.1186/1471-2105-16-S6-S3>.
- 59.** J. Li, A. Fu, L. Zhang, An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking, *Interdiscip. Sci. Comput. Life Sci.* 11 (2019) 320– 328. <https://doi.org/10.1007/s12539-019-00327-w>.
- 60.** Heyd J, Scuseria GE. Efficient hybrid density functional calculations in solids: Assessment of the Heyd-Scuseria-Ernzerhof screened Coulomb hybrid functional. *J Chem Phys.* 2004;121(3):1187–92.
- 61.** Caldwell G, Yan Z, Tang W, Dasgupta M, Hasting B. ADME Optimization and Toxicity Assessment in Early- and Late-Phase Drug Discovery. *Curr Top Med Chem.* 2009;9(11):965–80.

62. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discov Today* [Internet]. 2012;17(1–2):44
63. E.G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Heidelberg: Springer; 2011.
64. K. Burke, “Perspective on density functional theory,” *The Journal of Chemical Physics*, vol. 136, no. 15, p. 150901, Apr. 2012.
65. Flores-Holguín N, Frau J, Glossman-Mitnik D. Conceptual DFT-based computational peptidology of marine natural compounds: Discodermins A–H. *Molecules*. 2020;25(18):1–20.

# *CHAPTER III*

## **Materials and methods**

**CHAPTER III: Materials and methods**

1. Introduction.....62

2. Computational methods.....62

3. Data sets.....65

4. Equilibrium structure optimizations.....66

5. Generation of 3D-QSAR studies and CoMSIA analysis.....66

6. Molecular docking.....67

6.1. Preparation of the membrane protein 5HK1.....67

6.2. Docking of the proposed molecules into the 5HK1 binding site.....67

7. Prediction of synthetic accessibility and ADMET characteristics.....67

8. DFT calculations.....68

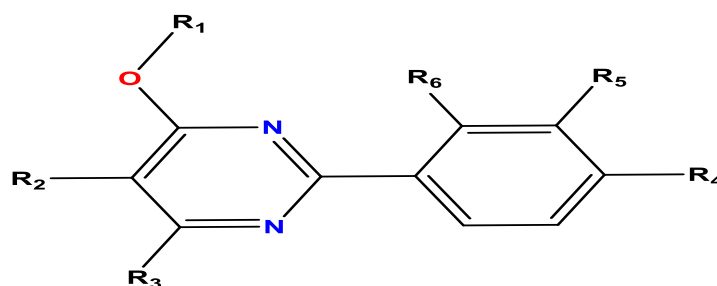
References.....70

## 1. Introduction

In this section, a comprehensive description of the materials and methods employed for the examination of 54 pyrimidine derivative compounds will be provided. Initially, the analysis involves the utilization of CoMSIA technique within the framework of a 3D-QSAR investigation. Subsequently, an extensive exploration is conducted to introduce a maximal number of novel pyrimidine derivative compounds. The validation of the novelty and feasibility of this proposed initiative is corroborated through the meticulous examination of molecular docking patterns exhibited by these compounds. Furthermore, the assessment extends to the prediction of ADMET properties, serving as a crucial adjunct to the investigative process. Ultimately, the investigation culminates in the implementation of DFT calculations, underscoring the comprehensive and multifaceted nature of the research endeavor being undertaken.

## 2. Computational methods

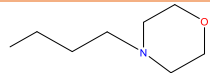
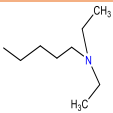
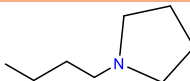
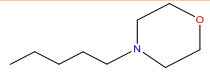
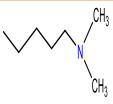
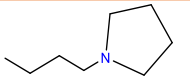

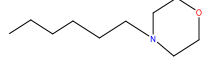
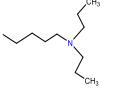
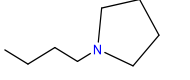
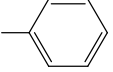
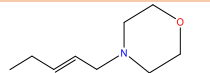
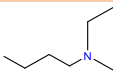
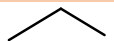
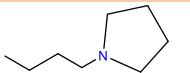
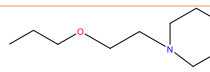
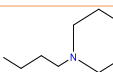

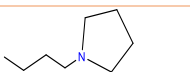
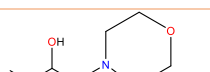
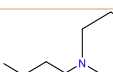
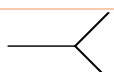
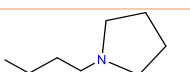
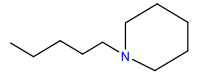
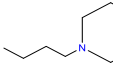
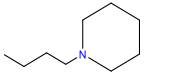
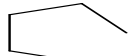
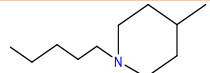
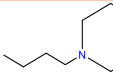
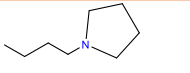
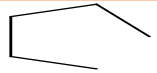
This research is performed by using several packages: HyperChem software<sup>1</sup> for geometry pre-optimizations and SYBYL X-2.1<sup>2</sup> for 3D-QSAR studies. Molecular docking studies are carried out using the Chemical Computing Group's Molecular Operating Environment (MOE) software (2014.09)<sup>3</sup> and the analysis of the interactions are done by Discovery Studio (2019)<sup>4</sup>. The ADMET pharmacokinetic parameters are determined by the pkCSM online server<sup>5</sup>. Also, the SwissADME online tool<sup>6</sup> is used to create a drawing of a boiled-egg.



**Fig. 1.** Structure of the studied compounds. See Table 1 for the designation of R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, R<sub>4</sub>, R<sub>5</sub> and R<sub>6</sub> substituents.

**Table 1**

Series of 54 compounds (Cp.) containing pyrimidine of S1R antagonists investigated presently is for the test set.

Cp.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	Cp.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	Cp.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>
C1 <sup>t</sup>		H	CH <sub>3</sub>	H	H	H	C18		H	CH <sub>3</sub>	H	H	H	C35		H	CF <sub>3</sub>	H	H	H
C2		H	CH <sub>3</sub>	H	H	H	C19		H	CH <sub>3</sub>	H	H	H	C36		H		H	H	H
C3		H	CH <sub>3</sub>	H	H	H	C20		H	CH <sub>3</sub>	H	H	H	C37		H		H	H	H
C4		H	CH <sub>3</sub>	H	H	H	C21		H		H	H	H	C38		CH <sub>3</sub>	CH <sub>3</sub>	H	H	H
C5		H	CH <sub>3</sub>	H	H	H	C22		H		H	H	H	C39		F	CH <sub>3</sub>	H	H	H
C6 <sup>t</sup>		H	CH <sub>3</sub>	H	H	H	C23		H		H	H	H	C40		Cl	CH <sub>3</sub>	H	H	H
C7		H	CH <sub>3</sub>	H	H	H	C24		H	OCH <sub>3</sub>	H	H	H	C41			H	H	H	
C8		H	CH <sub>3</sub>	H	H	H	C25		H	CF <sub>3</sub>	H	H	H	C42			H	H	H	

## Materials and methods

<b>C9</b>		H	CH 3	H	H	H	<b>C26</b>		H		H	H	H	<b>C43</b>			H	H	H	
<b>C10</b> t		H	CH 3	H	H	H	<b>C27</b>		H		H	H	H	<b>C44</b>		CH 3	Cl	CH <sub>3</sub>	H	H
<b>C11</b> t		H	CH 3	H	H	H	<b>C28</b> t		CH 3	CH <sub>3</sub>	H	H	H	<b>C45</b> t		Cl	CH <sub>3</sub>	OCH <sub>3</sub>	H	H
<b>C12</b>		H	CH 3	H	H	H	<b>C29</b>		F	CH <sub>3</sub>	H	H	H	<b>C46</b>		Cl	CH <sub>3</sub>	CF <sub>3</sub>	H	H
<b>C13</b>		H	CH 3	H	H	H	<b>C30</b>		CH 3	Cl	H	H	H	<b>C47</b> t		Cl	CH <sub>3</sub>	F	H	H
<b>C14</b>		H	CH 3	H	H	H	<b>C31</b>		H		H	H	H	<b>C48</b>		Cl	CH <sub>3</sub>	F	F	H
<b>C15</b>		H	CH 3	H	H	H	<b>C32</b>		H		H	H	H	<b>C49</b>		Cl	CH <sub>3</sub>	Cl	H	H
<b>C16</b> t		H	CH 3	H	H	H	<b>C33</b>		H		H	H	H	<b>C50</b> t		Cl	CH <sub>3</sub>	Cl	Cl	H
<b>C17</b>		H	CH 3	H	H	H	<b>C34</b>		H	OCH <sub>3</sub>	H	H	H	<b>C51</b> t		Cl	CH <sub>3</sub>	H		
<b>C52</b> t		Cl	CH 3	H	H	H	<b>C53</b> t		Cl	CH <sub>3</sub>	CH 3	H	H	<b>C54</b> t		Cl	CH <sub>3</sub>		H	H



### 3. Data sets

**Table 2**

Series of 54 compounds (Cp.) containing pyrimidine as S1R antagonists with their experimental, predicted and residual values of pKi (CoMSIA model). **t** is for the test set. The experimental values are from Ref [7].

Cp.	Exp.pKi	Pred.pKi	Resid.pKi	Cp.	Exp.pKi	Pred.pKi	Resid.pKi
C1 <sup>t</sup>	7.440	7.100	0.340	C28 <sup>t</sup>	8.020	7.717	0.303
C2	7.020	7.668	-0.648	C29	8.430	7.760	0.670
C3	6.410	7.314	-0.904	C30	8.630	8.055	0.575
C4	6.940	7.239	-0.299	C31	7.650	7.780	-0.130
C5	6.370	7.379	-1.009	C32	7.410	7.426	-0.016
C6 <sup>t</sup>	6.310	7.334	-1.024	C33	6.660	7.419	-0.759
C7	7.930	7.871	0.059	C34	7.710	7.702	0.008
C8	7.780	7.808	-0.028	C35 <sup>t</sup>	7.560	7.818	-0.258
C9	7.520	7.895	-0.375	C36 <sup>t</sup>	6.560	7.506	-0.946
C10 <sup>t</sup>	7.420	7.964	-0.544	C37 <sup>t</sup>	6.130	7.111	-0.981
C11 <sup>t</sup>	6.340	7.334	-0.994	C38 <sup>t</sup>	7.940	7.819	0.121
C12	6.500	7.427	-0.927	C39	8.450	7.832	0.618
C13	6.820	7.246	-0.426	C40	8.400	7.984	0.416
C14	7.640	7.823	-0.183	C41	7.350	7.573	-0.223
C15	7.510	7.609	-0.099	C42 <sup>t</sup>	7.230	7.537	-0.307
C16 <sup>t</sup>	7.730	7.587	0.143	C43	5.720	7.260	-1.540
C17	7.870	8.129	-0.259	C44	8.710	8.087	0.623
C18	7.330	7.613	-0.283	C45 <sup>t</sup>	8.410	7.960	0.450
C19	7.080	7.612	-0.532	C46	8.350	8.002	0.348
C20	7.410	7.517	-0.107	C47 <sup>t</sup>	8.940	8.024	0.916
C21	7.690	7.706	-0.016	C48	9.010	7.895	1.115
C22	7.460	7.802	-0.342	C49	8.970	7.998	0.972
C23	6.730	7.441	-0.711	C50 <sup>t</sup>	7.050	7.979	-0.929
C24	7.710	7.671	0.039	C51 <sup>t</sup>	6.690	8.234	-1.544
C25	7.560	7.860	-0.300	C52 <sup>t</sup>	6.730	7.866	-1.136
C26	6.600	7.330	-0.730	C53 <sup>t</sup>	7.230	7.964	-0.734
C27	6.020	6.904	-0.884	C54 <sup>t</sup>	7.400	8.003	-0.603

In this study, we consider 54 compounds containing pyrimidine scaffold (**Fig. 1** and **Table 1**) as characterized by Lan et al. <sup>7</sup>. We quote in **Table 2** their pKi values as S1R antagonists as determined by Lan et al. First, we created 3D structures for these 54 compounds and optimized their energies using Gasteiger-Hückel charge in the Tripos force field<sup>8</sup>. Then, we

randomly selected 36 compounds to construct the quantitative model as a training group and the rest of the compounds as a test group consisting of 18 compounds as specified in **Tables 1 and 2**.

### 4. Equilibrium structure optimizations

Reliable predictions of molecular geometries are sensitive to the choice of electronic structure method and of the set of atomic bases used for the description of atoms. We started our work by selecting an appropriate methodology to be used for the determination of the equilibrium structures of the pyrimidine derivatives under study. Indeed, these molecular structures were constructed using the sketch module and then optimized under the tripos force standard field utilizing the matching Gasteiger–Huckel atomic partial charges implemented on SYBYL-X 2.1 software<sup>2</sup>. Additionally, 0.005 kcal/mol was chosen as the Powell gradient algorithm's convergence threshold, with a maximum of 10 000 iterations required to obtain a stable conformation<sup>5,9</sup>. Prior to that, we pre-optimized these molecules using HyperChem<sup>1</sup>.

All compounds in the series were carefully converted and saved as separate Mol2 files in *MOL2* format. The SYBYL X2.1.1 software was used to generate fifty-four molecular structures and to minimize the energy of each 3D structure created with the standard Tripos Powell force field (100 iterations)<sup>9</sup>. Then, their Gasteiger-Hückel atomic partial charges to construct 3D-QSAR models were determined<sup>10</sup>. All molecular structures were analyzed with a distance-dependent buffer function until a root mean square (RMS) deviation of 0.05 kcal/(mol) was reached by the SYBYL X2.1.1 software.

### 5. Generation of 3D-QSAR studies and CoMSIA analysis

The selected compounds cover the full range of targeted biological activities. Afterwards, we created the data set in the molecular data table. Then we made several attempts (till 300), while leaving the value of the Powell gradient energy at 0.005 kcal/mol. In order to obtain the best 3D-QSAR model and the molecular morphology, we used multiple ways to search. For instance, we created the CoMSIA model with the training set, which consists of 36 compounds and using 18 molecules as a test set as specified in **Table 2**. The CoMSIA model was validated after analysis of the partial least squares (PLS) regression with a grid spacing of 2 Å. We calculated the field dimensions of the CoMSIA model for both steric and electrostatic fields, hydrophobic and H-bonds donor and acceptor, and we also explained the main structural properties of the work of antagonists of S1R against neuropathic pain<sup>11</sup>. For the calculations, we used the PLS of the limit factor  $R^2$ , the cross-validation factor  $Q^2$  and the cross-validation standard error of the estimate SEE, and for evaluating the model we calculated the value of

F. Based on the high values of F,  $R^2$  and  $Q^2$  whose values should represent ( $R^2 > 0.6$ ) and ( $Q^2 > 0.5$ ) and the low value of SEE, we choose the best model among others that respect the tests.

### 6. Molecular docking

We screened sixteen compounds and the active compound **C48** using molecular docking in order to identify their binding mechanisms to the Human S1R protein (PDB ID: 5HK1)<sup>12</sup>. We also considered **61W** inhibitor, which is assumed to be a standard reference.

#### 6.1. Preparation of the membrane protein 5HK1

We downloaded the 5HK1 structure with inhibitor **61W** from the PDB file (Protein Data Bank)<sup>12,13</sup>. We performed several steps to prepare the selected transmembrane protein for the molecular docking study. We added hydrogen atoms to the system. Then, we corrected all errors in the connections or in the types of atoms. Afterwards, we searched an active site of the selected protein that fits with the **61W** inhibitor through the application of a site finder. Briefly, this consisted in mapping the structure of the protein and the identification of this active site using virtual elements by creating dummy atoms for the docking step. Finally, we minimized the energy of the protein<sup>14,15</sup>.

#### 6.2. Docking of the proposed molecules into the 5HK1 binding site

We first loaded the active protein file prepared in the previous step and then started the docking process<sup>16</sup>. We initially performed a validation of the active site by binding the reference compound **61W** inhibitor. The accuracy of the docking program is thus checked. Then we applied the docking process for the series of compounds under study. Indeed, we modified the program where we set the docking site as dummy atoms. For the refinement methodology, we chose the rigid receptors to find the best sites. GBVI / WSA dG is the scoring method for this selection, and the corresponding triangle method as the positioning methodology. Specifically, the scoring methodology that we implemented is London dG. Afterwards, we downloaded the MDB file for our proposed 33 molecules with damper **61W** and automated the global docking calculations. After computations, we investigated the binding sites. The best ones were chosen according to the RMSD refine values, which highlight their best degrees and the types of interactions of the linking site<sup>17</sup>.

### 7. Prediction of synthetic accessibility and ADMET characteristics

ADMET properties correspond to absorption, distribution, metabolism, elimination and toxicity, respectively. Because of the criteria for potentially unfavorable therapeutic agents for

ADMET, many compounds fail to reach clinical trials<sup>18,19</sup>. The most common strategy to discover pharmaceuticals chemicals is drug-likeness.

To anticipate the *in silico* properties of ADMET for the proposed new compounds, we implemented an online pkCSM server. We considered the most absorbable compounds, which possess greater than 30% in the intestinal absorption function (human) (percentage absorbed). For a stable volume distribution, VD<sub>ss</sub> (human) less than 1.209 log L/kg means that it is a low value. Besides, we checked whether a potential P450 inhibitor is metabolized or not. This is to prove the possibility of the action of these compounds in order to inhibit some of the known isoforms of cytochrome P450, because the latter can lead to inhibition of the pharmacokinetic-drug interactions. Therefore, the drugs that are taken together in the diet fail and accumulate in toxic forms<sup>20</sup>.

In order to achieve steady-state concentrations that agree with bioavailability according to the values of total clearance doses values were set. The proposed new compounds are expected to have toxic properties that we observed through hepatotoxicity, skin sensitivity and predicted AMES toxicity<sup>21</sup>.

### 8. DFT calculations

The Gaussian 09 package<sup>22</sup> was employed to carry out the three-dimensional (3D) geometry optimizations of the ligand of reference 61w and its mol2, mol3, and mol4 compounds. Shown in the [supplementary in Table S1](#) with the 2D structures with their chemical nomenclature and as well as their masses with close values. This was done using the DFT/B3LYP functional with a 6-31++G(d, p) basis set. The calculations began with the optimization of the geometry and subsequent evaluation of the energy. Orientations were then investigated based on the analysis of the frontier molecular orbitals, specifically the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO). Furthermore, quantum chemical parameters were derived from the HOMO and LUMO energy values. These parameters included the energy gap ( $\Delta E$ ), electronegativity ( $\chi$ ), chemical potential ( $\mu$ ), chemical hardness ( $\eta$ ), softness ( $\sigma$ ), and electrophilicity index ( $\omega$ )<sup>23</sup>, can be expressed through equations (1)–(8) in the following manner:

$$\text{Energy gap } (\Delta E) = \text{LUMO} - \text{HOMO} \quad (1)$$

$$\text{Ionization potential (IP)} = - E_{\text{HOMO}} \quad (2)$$

$$\text{Electron affinity (EA)} = - E_{\text{LUMO}} \quad (3)$$

$$\text{Hardness } (\eta) = (\text{IP} - \text{EA})/2 \quad (4)$$

$$\text{Electronegativity } (\chi) = (\text{IP} + \text{EA})/2 \quad (5)$$

## Materials and methods

$$\text{Softness } (\sigma) = 1/2\eta \quad (6)$$

$$\text{Chemical potential } (\mu) = -\chi \quad (7)$$

$$\text{Electrophilicity index } (\omega) = \mu^2/2\eta \quad (8)$$

To calculate the molecular orbital contribution of different constituting elements to the total system for the compound mol2, the Density of States (DOS) was determined using the GaussSum software<sup>24</sup>.

## References

1. HyperChem (TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
2. [www.tripossoftware.com](http://www.tripossoftware.com)
3. Dahlgren, D., & Lennernäs, H. (2019). Intestinal permeability and drug absorption: predictive experimental, computational and in vivo approaches. *Pharmaceutics*, *11*(8), 411.
4. Discovery Studio Visualizer. San Diego: DassaultSystèmes BIOVIA; (2019).
5. <http://biosig.unimelb.edu.au/pkcsmprediction>.
6. <http://www.swissadme.ch/index.php>.
7. Lan, Y., Chen, Y., Cao, X., Zhang, J., Wang, J., Xu, X., ... & Zhang, G. (2014). Synthesis and biological evaluation of novel sigma-1 receptor antagonists based on pyrimidine scaffold as agents for treating neuropathic pain. *Journal of Medicinal Chemistry*, *57*(24), 10404-10423.
8. Gasteiger, J., & Marsili, M. (1980). Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, *36*(22), 3219-3228.
9. Clark, M., Cramer III, R. D., & Van Opdenbosch, N. (1989). Validation of the general purpose tripos 5.2 force field. *Journal of computational chemistry*, *10*(8), 982-1012.
10. Aparoy, P., Suresh, G. K., Reddy, K. K., & Reddanna, P. (2011). CoMFA and CoMSIA studies on 5-hydroxyindole-3-carboxylate derivatives as 5-lipoxygenase inhibitors: Generation of homology model and docking studies. *Bioorganic & medicinal chemistry letters*, *21*(1), 456-462.
11. Chtita, S., Belhassan, A., Bakhouch, M., Taourati, A. I., Aouidate, A., Belaidi, S., ... & Lakhlifi, T. (2021). QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods. *Chemometrics and Intelligent Laboratory Systems*, *210*, 104266.
12. <http://www.rcsb.org>.
13. Schmidt, H. R., Zheng, S., Gulpinar, E., Koehl, A., Manglik, A., & Kruse, A. C. (2016). Crystal structure of the human  $\sigma$ 1 receptor. *Nature*, *532*(7600), 527-530.
14. Ghanem, A., Emara, H. A., Muawia, S., Abd El Maksoud, A. I., Al-Karmalawy, A. A., & Elshal, M. F. (2020). Tanshinone IIA synergistically enhances the antitumor activity of doxorubicin by interfering with the PI3K/AKT/mTOR pathway and inhibition of

- topoisomerase II: in vitro and molecular docking studies. *New Journal of Chemistry*, 44(40), 17374-17381.
15. Al-Karmalawy, A. A., & Khattab, M. (2020). Molecular modelling of mebendazole polymorphs as a potential colchicine binding site inhibitor. *New Journal of Chemistry*, 44(33), 13990-13996.
  16. Daoui, O., Elkhatabi, S., Chtita, S., Elkhlabi, R., Zgou, H., & Benjelloun, A. T. (2021). QSAR, molecular docking and ADMET properties in silico studies of novel 4, 5, 6, 7-tetrahydrobenzo [D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase. *Heliyon*, 7(7).
  17. Ouassaf, M., Belaidi, S., Lotfy, K., Daoud, I., & Belaidi, H. (2018). Molecular docking studies and ADMET properties of new 1,2,3-triazole derivatives for anti-breast cancer activity. *Journal of Bionanoscience*, 12(1), 26-36.
  18. Wang, Y., Guo, H., Tang, G., He, Q., Zhang, Y., Hu, Y., ... & Lin, Z. (2019). A selectivity study of benzenesulfonamide derivatives on human carbonic anhydrase II/IX by 3D-QSAR, molecular docking and molecular dynamics simulation. *Computational Biology and Chemistry*, 80, 234-243.
  19. Hadni, H., & Elhallaoui, M. (2020). 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents. *Heliyon*, 6(4).
  20. Lynch, T., & Price, A. M. Y. (2007). The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *American family physician*, 76(3), 391-396.
  21. Osman, E. A., Abdalla, M. A., Abdelraheem, M. O., Ali, M. F., Osman, S. A., Tanir, Y. M., ... & Alzain, A. A. (2021). Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR, molecular docking and molecular dynamics. *Informatics in Medicine Unlocked*, 26, 100765.
  22. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Pople, J. A. (2008). Gaussian 09, revision B. 01, Gaussian, Inc., Wallingford CT, 2009 Search PubMed;(b) NM O'Boyle, AL Tenderholt and KM Langner. *J. Comput. Chem*, 29, 839.
  23. Abdou, A. (2022). Synthesis, structural, molecular docking, DFT, vibrational spectroscopy, HOMO-LUMO, MEP exploration, antibacterial and antifungal activity of new Fe (III), Co (II) and Ni (II) hetero-ligand complexes. *Journal of Molecular Structure*, 1262, 132911.

24. Rijal, R., Sah, M., & Lamichhane, H. P. (2023). Molecular simulation, vibrational spectroscopy and global reactivity descriptors of pseudoephedrine molecule in different phases and states. *Heliyon*, 9(3).



# *CHAPTER IV*

## **Results and discussion**

**CHAPTER IV: Results and discussion**

1. Introduction.....	75
2. Optimized geometries of the compounds of the series.....	75
3. CoMSIA results.....	75
3.1. CoMSIA model.....	75
3.2. CoMSIA graphical interpretation and contour analysis.....	76
4. Design and selection of new human $\sigma_1$ receptor inhibitors.....	78
5. Docking results and analysis.....	79
6. ADMET prediction and synthetic accessibility.....	84
7. DFT calculations.....	88
7.1. The trend of reactivity and FMO study.....	88
7.2. Density of states (DOS).....	90
References.....	92

## 1. Introduction

In this section, the outcomes derived from the implementation of different methodologies specified in the third chapter will be thoroughly examined and analyzed in order to gain a comprehensive understanding of the findings.

## 2. Optimized geometries of the compounds of the series

Table S1 gives the optimized 3D structures of the series of the 54 compounds containing pyrimidine of S1R antagonists currently studied. This table shows that the phenyl pyrimidine subunit as specified in (Fig. 1 (chapter III)) is planar in all compounds which favor their stability by electronic delocalization. For most of the compounds, the R<sub>1</sub> group is also lying on the plan of this subunit in spite that the *a priori* free rotations around the simple bonds to the C - O atom connecting this group to the phenyl pyrimidine subunit. This may be due to steric effects that favor the deployment of the whole compound instead of its folding. These structural findings may be connected to the efficiency of these compounds against S1R favoring thus the interaction of these inhibitors with the active site of the enzyme. This is confirmed by the docking studies we performed (*vide infra*).

## 3. CoMSIA results

### 3.1. CoMSIA model

COMSIA is applied in order to predict accurate values of the field by calculating each point around the studied molecular structure from a 3D lattice. The COMSIA algorithm is designed by calculating five fields including electrostatic interactions, steric probes, H-bond acceptor and donor, and hydrophobic properties. Next, we perform a molecular least squares (PLS) analysis by exploiting the data represented by COMSIA<sup>1</sup>.

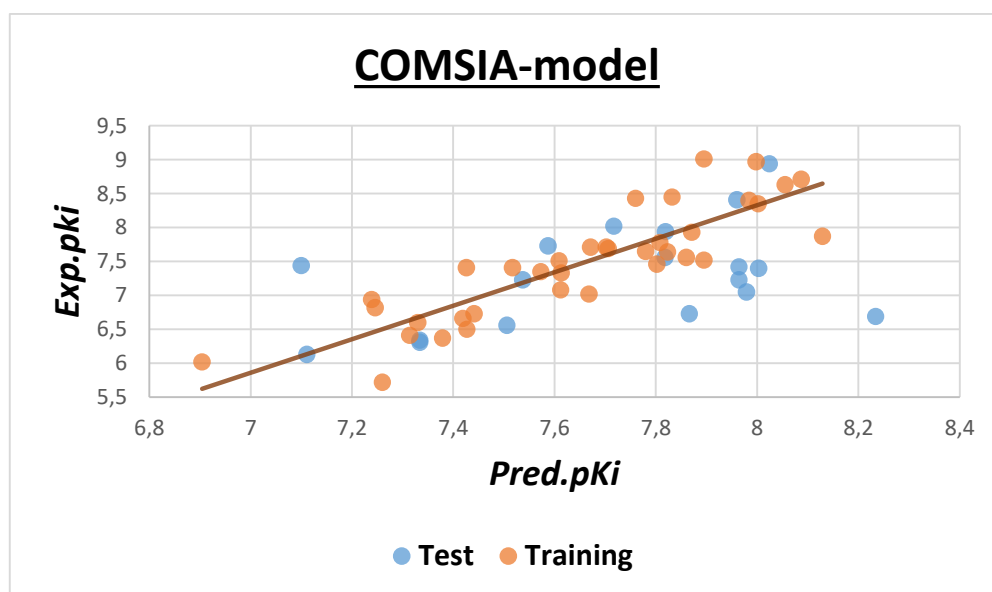
**Table 1**

PLS statistics of CoMSIA model. **R<sup>2</sup>**: Non-cross- validated determination coefficient. **Q<sup>2</sup>**: Cross-validated determination coefficient. **SEE**: Standard error of the estimate. **F**: F-test value. **Str**: Steric field, **Ele**: Electrostatic field, **Hyd**: Hydrophobic field, **HBA** and **HBD** correspond to H-bonds Acceptor and Donor.

Model	R <sup>2</sup>	Q <sup>2</sup>	SEE	F	Fractions				
					Str	Ele	Hyd	HBA	HBD
CoMSIA	0.957	0.653	0.185	108.280	0.141	0.455	0.232	0.134	0.038

Using CoMSIA model, we obtained the predicted activity values for the training and test groups as presented in (Table 2 (chapter III)), as well as their residual values. The result of the CoMSIA model is given in (Table 1 (chapter III)). They show good values for the statistical parameters:  $R^2=0.957$ ,  $Q^2=0.653$  and test value  $F=108.280$  and a lower standard error of the estimate  $SEE=0.185$ . This means that the CoMSIA model has a good predictive power, as well as a high level of stability. Also, Table 3 shows that the contributions of hydrophobic and electrostatic fields amount to 23.2% and 45.5%, respectively. Such large percentages indicate their great importance in this model.

The plot depicted in Fig. 1 show cases the relationship between the actual pKi values and the predicted pKi values determined by the CoMSIA model. Upon observing the plot, it becomes evident that the blue circle solid point and the orange circle solid point closely align with the line  $Y = X$ . This alignment signifies a robust linear correlation between the actual and predicted activities across the entirety of the data set<sup>2</sup>.

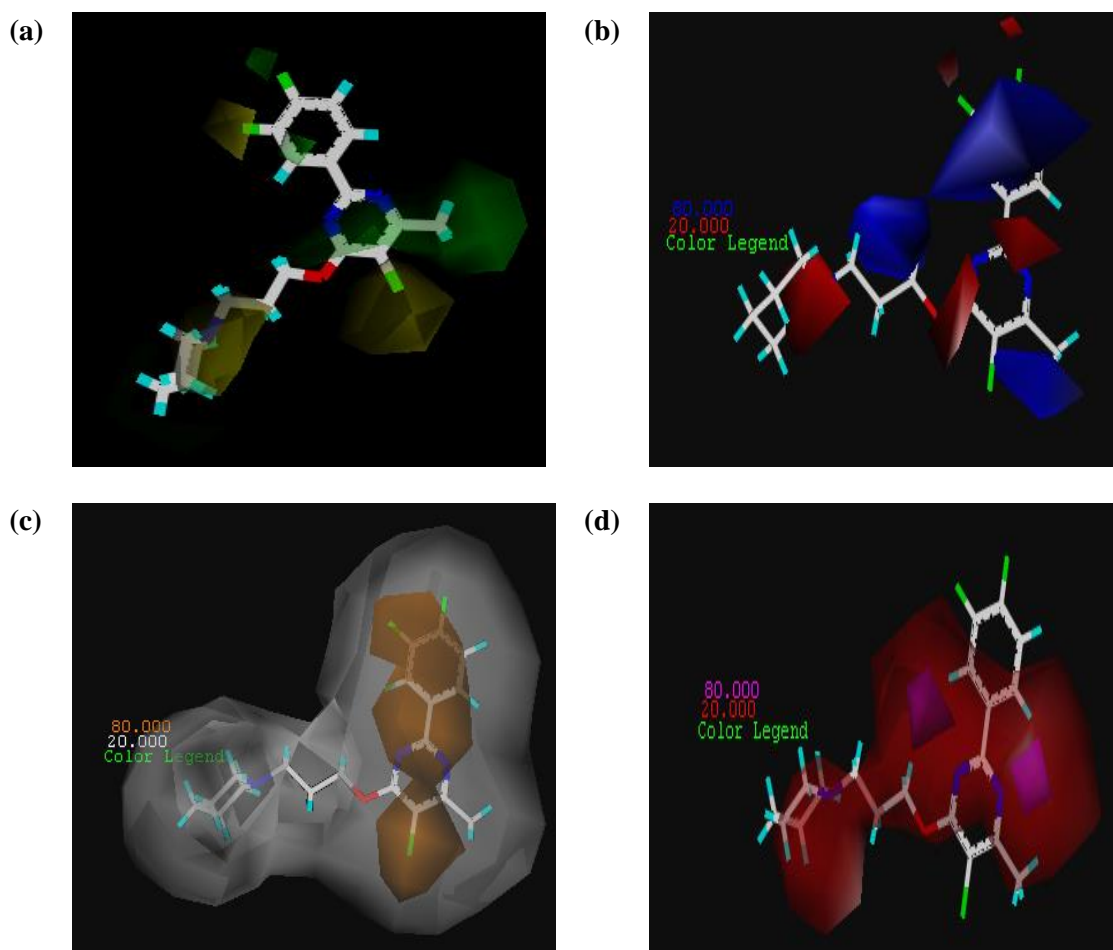


**Fig. 1.** The plot showcases the correlation between the experimental activities and the predicted activities as determined by the CoMSIA model.

### 3.2. CoMSIA graphical interpretation and contour analysis

We analyzed the CoMSIA model in order to determine the nature of the alternatives that increase the activity of the studied molecules. Through the analysis of the respective generated contour maps, we derived information explaining the appropriate and unsuitable regions of the investigated molecules regarding their biological activity. For illustration, we considered **C48** compound as a reference that has a large activity. Therefore, we designed 3D contour maps,

which are used to highlight the complexation induced transformations that occur in the physical and chemical properties of this compound. Indeed, changes may occur in its structure, which can increase or decrease its biological activity.



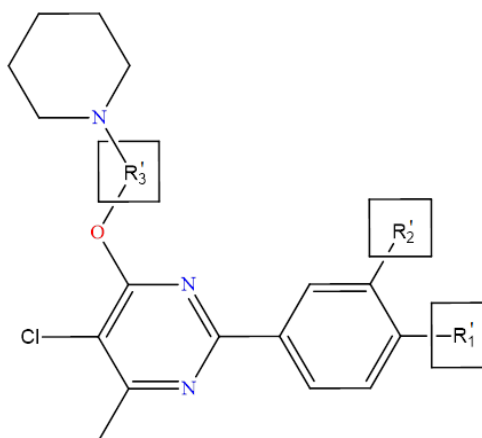
**Fig. 2.** CoMSIA contour maps of **C48** compound. (a) steric field; (b) electrostatic field; (c) hydrophobic field; (d) hydrogen bond acceptor field. See text for more details.

**Fig. 2** shows the CoMSIA contour maps of compound **C48**, which is the most active in the series. In this figure, we specify the steric, electrostatic, hydrophobic and hydrogen bond acceptor fields. For the steric field, the green color appears around the positions starting from  $R_3$  group extending to  $R_1$  substituent. This indicates that the increase in the activity of this molecule lies in the addition of substituents of larger size, whereas the yellow color can be found at the  $R_1$  position and around the chlorine and fluorine atoms at  $R_2$  and  $R_5$  positions, respectively. This reveals that the addition of larger groups and of chlorine and fluorine atoms is suitable for improving the biological activity. For the electrostatic field, the blue color can be seen around  $R_4$ ,  $R_2$  and  $R_5$ , which indicates that these areas have an increase in the positive

charge, while the areas in which the negative charges increase are in red. This explains the large activity of compound **C48** ( $pK_i=9.010$ ) compared to the other compounds of this series.

For the hydrophobic field, the brown color indicates that the hydrophobic substituents around  $R_2$ ,  $R_4$  and  $R_5$  are preferred and are favorable in these positions, while the gray color shows that the hydrophilic substituents are granular and improve the activity of the compound with an inhibitory character. For the hydrogen bond acceptor domain, the purple contour represents the H-acceptor region, while the unfavorable region of the H-acceptor is represented in red. For instance, the comparison of the most active compound **C48** with the rest of the compounds in the series, reveals that compounds **C18**, **C19** and **C20**, which exhibit a less efficient biological activity than **C48**, lack large-volume substituents at position  $R_1$ . As for many compounds of the series that do not contain chlorine or fluorine atoms, let us mention some of them from compound **C1** to compound **C28** in positions  $R_2$ ,  $R_4$  and  $R_5$ . The latter have also smaller  $pK_i$  than that of **C48**. In these same areas, some compounds such as **C31** and **C32** do not have the positive charge.

#### 4. Design and selection of new human $\sigma_1$ receptor inhibitors



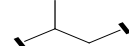
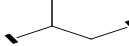
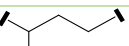


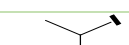
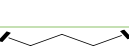
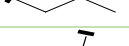
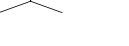


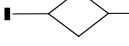

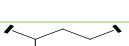



**Fig. 3.** Chemical structures of the proposed new human  $\sigma_1$  receptor inhibitors. See Table 2 for the designation of the  $R'_1$ ,  $R'_2$  and  $R'_3$  substituents.

**Table 2**

Designation of the  $R'_1$ ,  $R'_2$  and  $R'_3$  substituents,  $pK_i$  (CoMSIA model) and total scoring values of the proposed new human  $\sigma_1$  receptor inhibitors and of the active compound **C48**.

Cp.	$R'_1$	$R'_2$	$R'_3$	Pred. $pK_i$ CoMSIA	Total scoring	Cp.	$R'_1$	$R'_2$	$R'_3$	Pred. $pK_i$ CoMSIA	Total scoring
				A						A	

<b>C48</b>	F	F		7.895	-9.5122	<b>Mol9</b>	Cl	H		10.4	-8.9117
<b>Mol 1</b>	F	F		9.75	-9.3772	<b>Mol10</b>	H	OCH <sub>3</sub>		9.8	-9.2238
<b>Mol 2</b>	F	F		9.67	-9.4249	<b>Mol11</b>	H	OCH <sub>3</sub>		9.67	-8.9416
<b>Mol 3</b>	F	F		9.19	-9.3492	<b>Mol12</b>	H	CF <sub>3</sub>		9.29	-8.8011
<b>Mol 4</b>	F	F		9.67	-7.7285	<b>Mol13</b>	H	CF <sub>3</sub>		9.32	-8.7852
<b>Mol 5</b>	OCH <sub>3</sub>	H		10.18	-8.8107	<b>Mol14</b>	H	CF <sub>3</sub>		9.62	-9.3833
<b>Mol 6</b>	OCH <sub>3</sub>	H		9.97	-9.2341	<b>Mol15</b>	H	Cl		9.79	-9.9964
<b>Mol 7</b>	CF <sub>3</sub>	H		10.47	-8.8538	<b>Mol16</b>	H	Cl		9.4	-9.4849
<b>Mol 8</b>	CF <sub>3</sub>	H		9.76	-8.9814						

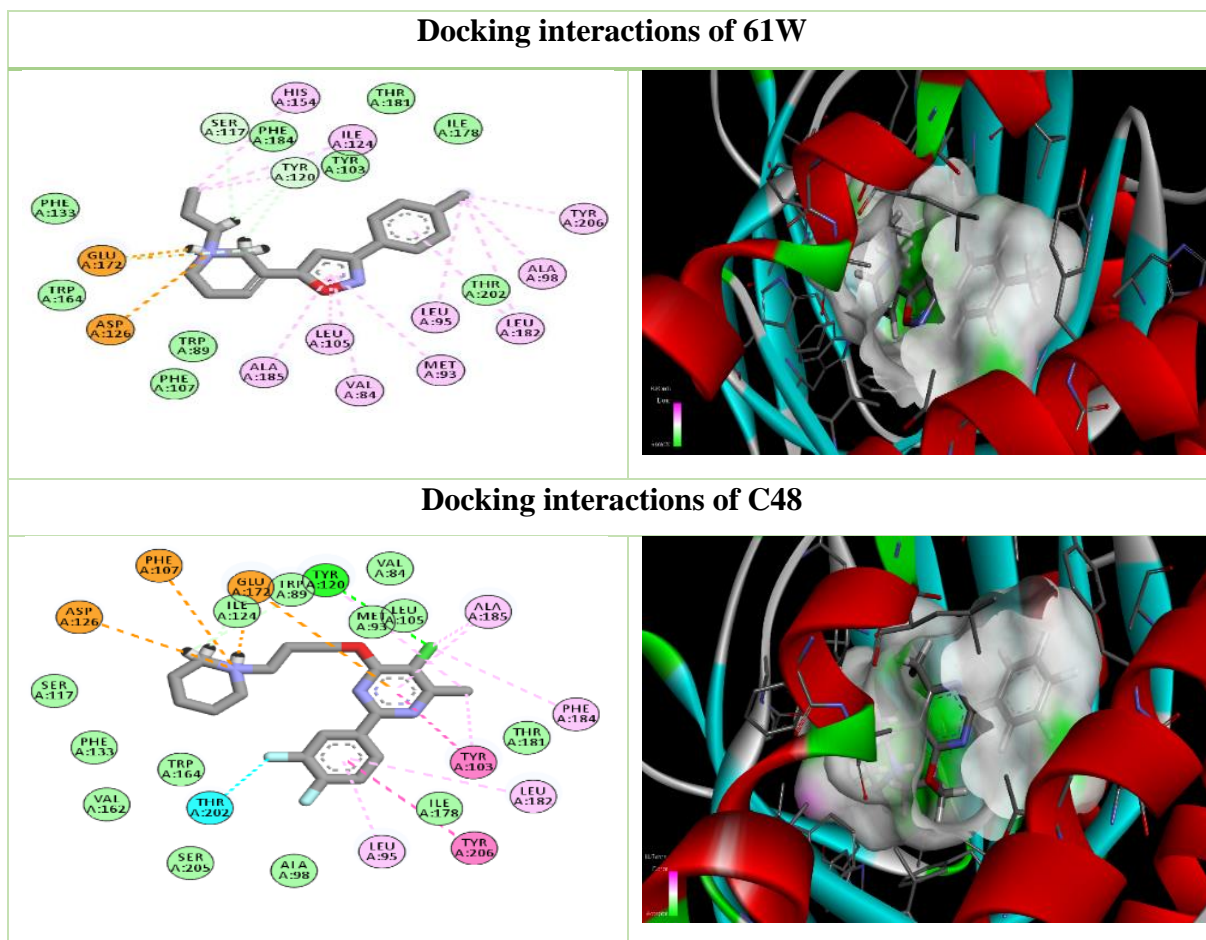
Based on the results of the study of the bioactive series and the model obtained, we initially designed one hundred and twenty-seven new molecules as potent human  $\sigma_1$  receptor inhibitors. Indeed, these molecules were created based on the study and analysis of 3D-QSAR that were inferred through the analysis and information collected from the fields we studied for the most active molecule **C48**. Then we made changes in the molecular structure of **C48** at the level of radicals that we named **R<sub>1</sub>'**, **R<sub>2</sub>'** and **R<sub>3</sub>'** (**Fig. 3**) in order to increase the respective biological activity. Then we selected the best sixteen compounds in terms of activity value (pKi) for their in-depth investigation. The molecular structures of these new compounds are given in **Fig. 3** and **Table 2**. Then, the activity of these proposed new molecules was predicted using the CoMSIA model, from this we obtained their pKi values as well as their total score values as listed in **Table 2**. As expected, this table shows that these compounds exhibit an improved inhibitory activity compared to compound **C48**, since their CoMSIA predicted pKi values are larger. In particular, the most active ones are **Mol2**, **Mol15** and **Mol16** as confirmed by their high total scoring value of docking compared to the rest of the designed compounds (**Table 2**).

## 5. Docking results and analysis

Molecular docking is a powerful method to identify and characterize the interactions that occur within the active site of a protein (5HK1 here) and a ligand (an inhibitor here). First we analyzed the interactions of this active site with the reference ligand **61W** and the previously

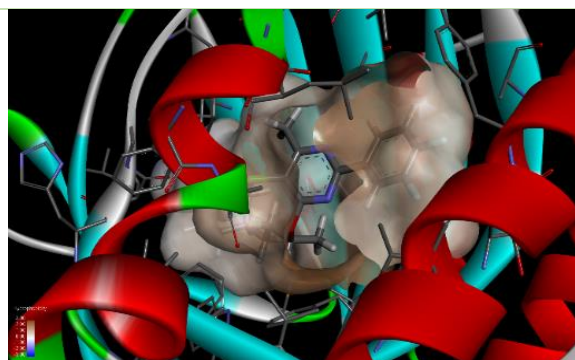
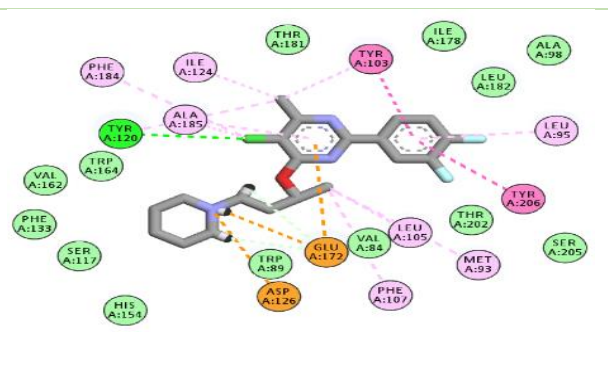
identified as the most active compound **C48**. We also investigated the three most active inhibitors proposed in this work, namely **Mol2**, **Mol15** and **Mol16** compounds. These data are displayed in Fig. 5 for these compounds and in **Figures S2-S4** (In the **Supplementary**) for those of the other compounds of **Table 2**.

By comparing the interactions of the proposed active molecules **Mol12**, **Mol15**, **Mol16** and the most active compound **C48** with the reference compound **61W** (**Table S3** (In the **Supplementary**)), we found that there are, in all cases, hydrogen bonds with the same amino acids: Interactions of salt bridge with the common amino acid GLU A:172 and of alkyl and  $\pi$ -alkyl interactions type. In particular, the proposed compound **Mol15** - as well as the reference compound **61W** - interacts with the largest number of amino acids (i.e. ILE A:124, TYR A:120, ALA A:185, LEU A:105, MET A:93, LEU A:95, LEU A:182). So, this documents that compound **Mol15** is the most active among all the proposed compounds, as its activity increases in the presence of these interactions, as well as the Score value, which is -9.9964 (**Table 2**). The latter value is larger, in absolute value, than that of the reference compound **61W** (of -8.7178) and of **C48** (of -9.5122). This may explain its stability as well its strong inhibition effect on the receptor protein.

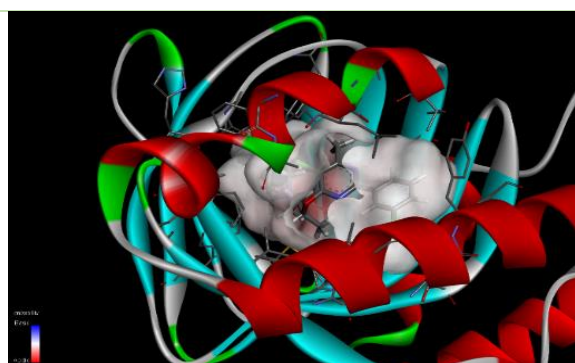
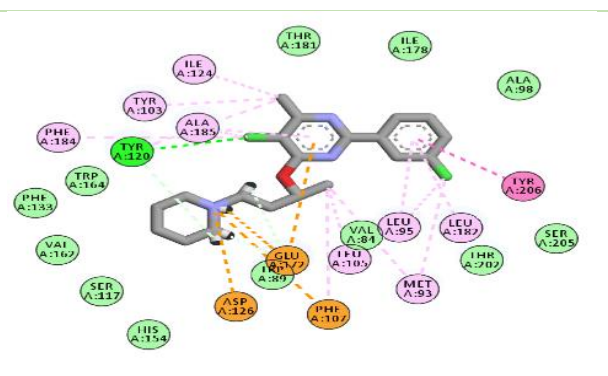




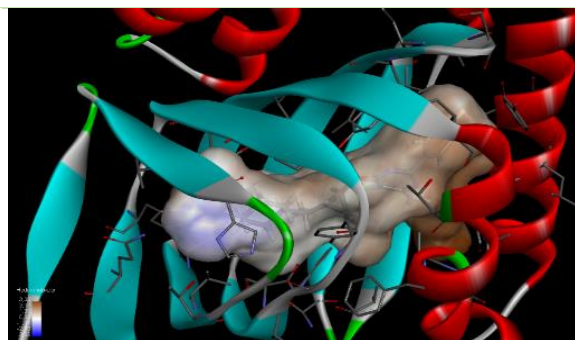
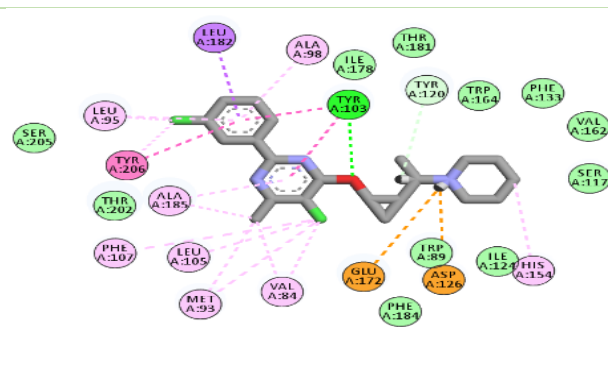
### Docking interactions of Mol2



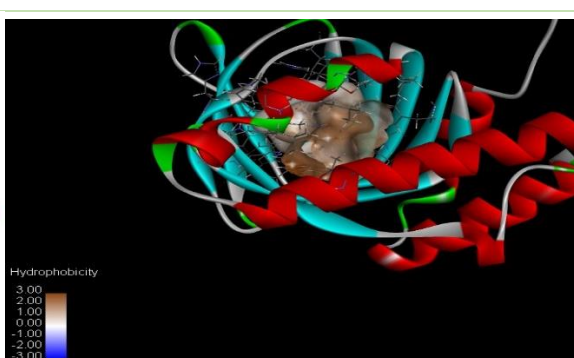
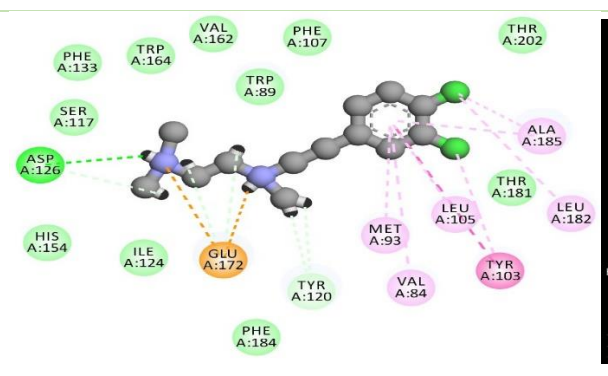
### Docking interactions of Mol15



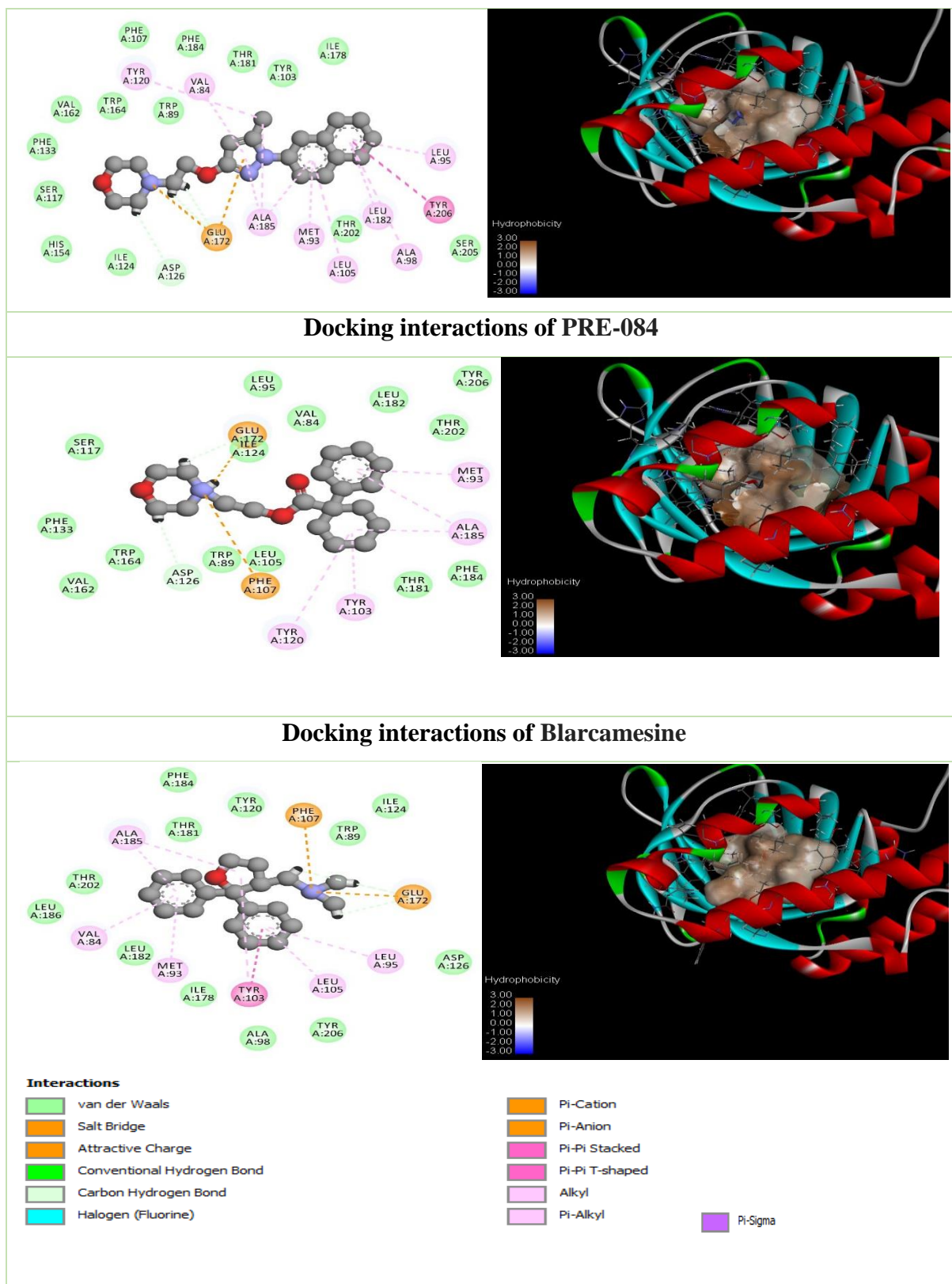
### Docking interactions of Mol16



### Docking interactions of BD-1047



### Docking interactions of S1RA



**Fig. 4.** Docking interactions of the 5HK1 enzyme with the three best inhibitors **Mol2**, **Mol15** and **Mol16** as well with the reference ligand **61W** and **C48** compound, as well as examining the intricate interactions for the antagonists **BD-1047** and **S1RA**, as well as the agonists **PRE-084** and **Blarcamesine**. These interactions are also specified in **Table S4** (In the **Supplementary**). Other views of compound **C48**; **61W**, **Mol2** and **Mol3** in **Figures S2-S4** (In the **Supplementary**).

In addition to the previous interactions mentioned, we found the establishment of a number of interactions with the enzyme 5HK1 while **C48**, **Mol12**, **Mol15** and **Mol16** are complexed with this enzyme such as conventional hydrogen bonds,  $\pi$ - $\pi$  T-shaped and also  $\pi$ -cation (except **Mol12** and **Mol16**),  $\pi$ -anion (except **Mol16**), attractive charge, as for the latter, (except for **Mol16**), halogen (Fluorine) interactions with THE A:202 and  $\pi$ - $\pi$  stacking with TYR A:103, TYR A:206 (for **C48**). Besides, the proposed compound **Mol16** is characterized by its own  $\pi$ - $\sigma$  interaction with LEU A: 182. Interestingly, these interactions are not present in the reference compound **61W** interacting with 5HK1. In sum, the exploration of all these interactions between the inhibitors and 5HK1 confirm the stability of the newly proposed compounds with respect to the active site of this protein. Also, we conclude that the addition of chlorine and fluorine atoms and the large size of the radicals effectively increase the inhibitory activity of these molecules against 5HK1.

To perform a thorough validation of the docking results for the proposed new compounds Mol12, Mol15, and Mol16, we conducted a redocking experiment involving both activators (such as PRE-084 and Blarcamesine) and antagonists (such as BD-1047 and S1RA). The purpose of this experiment was to assess whether the latter compounds share the same action as our proposed compounds. As a result of our experiment, we obtained a comprehensive set of results, which can be found in the supplementary part of this study, specifically in the table provided (**Table S5** (In the **Supplementary**)).

In order to verify the validity of our proposed compounds, we utilized a comparative approach, focusing on the interactions exhibited by these compounds. Interestingly, we found that these interactions align closely with those observed with the antibodies BD-1047 and S1RA. For instance, in the case of BD-1047, which possesses a Score value of -7.9371 as indicated in the **Table S5** (In the **Supplementary**), we observed the presence of Alkyl and Pi-Alkyl interactions with specific amino acids shared by our proposed compounds, namely VAL A:84, ALA A:185, LEU A:105, MET A:93, and LEU A:182. Moreover, BD-1047 also engaged in Pi-Pi T-shaped interactions with the amino acid TYR A:103, Carbon Hydrogen Bond interaction with TYR A:120, salt bridge and attractive charge interactions with the common amino acid GLU A:172, and Conventional Hydrogen Bond with ASP A:126.

Similarly, when examining the results pertaining to S1RA, which possesses a Score value of -9.0658 (**Table S5** (In the **Supplementary**)), we found the presence of Alkyl and Pi-Alkyl interactions with the same amino acids shared by our proposed compounds, including VAL

A:84, TYR A:120, ALA A:185, LEU A:105, MET A:93, LEU A:95, LEU A:182, and ALA A:98. Additionally, S1RA exhibited Pi-Pi T-shaped interactions with the amino acid TYR A:206, Carbon Hydrogen Bond interaction with ASP A:126, and interactions of Pi-Anion and attractive charge with the common amino acid GLU A:172.

On the other hand, show a significant difference when comparing the interactions of compounds Mol12, Mol15, and Mol16 with those of the activators (PRE-084 and Blarcamesine). Specifically, when examining Blarcamesine, which displayed a relatively weak Score value of -4.4224 (**Table S5** (In the **Supplementary**)), we observed that it lacked several fundamental interactions, including Carbon Hydrogen Bond, interactions of Pi-Anion, Conventional Hydrogen Bond, and interactions of salt bridge. Consequently, Blarcamesine only exhibited Alkyl and Pi-Alkyl interactions with amino acids such as VAL A:84, ALA A:185, LEU A:105, MET A:93, and LEU A:95. Furthermore, it participated in Pi-Pi T-shaped interactions with the amino acid TYR A:103 and Pi-Cation interactions with the amino acid GLU A:172 and PHE A:107.

Similarly, in the case of PRE-084, which possessed a Score value of -7.4501 (**Table 3**), we observed a lack of crucial Pi-Pi T-shaped interactions, interactions of attractive charge, Conventional Hydrogen Bond, and interactions of Pi-Anion. Consequently, the only interactions exhibited by PRE-084 were Alkyl and Pi-Alkyl interactions with amino acids such as ALA A:185, MET A:93, TYR A:103, and TYR A:120. Additionally, it engaged in Carbon Hydrogen Bond with ASP A:126 and Pi-Cation interactions with the amino acid GLU A:172 and PHE A:107. In conclusion, our comprehensive analysis of the docking results for the proposed new compounds Mol12, Mol15, and Mol16, along with the comparative assessment of the interactions exhibited by both activators (PRE-084 and Blarcamesine) and antagonists (BD-1047 and S1RA), shed light on the similarities and differences among these compounds. The detailed findings, which are outlined in the supplementary table, provide valuable insights into the molecular interactions and potential mechanisms of action associated with these compounds<sup>2,3</sup>.

### 6. ADMET prediction and synthetic accessibility

PkCSM<sup>4</sup> is an online tool. It was used to calculate the ADMET pharmacokinetic parameters of the proposed new compounds in order to verify their applicability. The results are listed in **Table 3**. They consist on important computed ADMET properties and computed safety end points of the screened best sixteen compounds and of the most active compound. This table shows that the intestinal absorption for the sixteen compounds range from 88.067% to 93.832%. Thus, their absorption characteristics are very high, where we find the highest

value for compound **Mol10**. If the composite values are more than 0.45, it reveals a high volume distribution VDss. Indeed, **Table 3** shows that and proves a high volume distribution of VDss because its values exceed the specified value and range from 0.697 to 1.209 log L/kg.

More than 90% of drugs enter the first stage of metabolism<sup>5,6</sup>. The primary role of enzymatic metabolism in the body is to exchange drug molecules, which is referred to as the biochemical conversion of a drug. Medicines create different enzymatic receptors in the body, which in turn work at concentrations with multiple pharmacological properties to activate the reaction<sup>7</sup>. For compounds with predictors of the trimer substrate indicates YES, we can predict their metabolism using cytochrome P450. So that it has the primary role in drug metabolism because the metabolic task is led by the main liver enzyme system. But some inhibitors are strong inhibitors to cytochrome P450. This is the case, for instance of **Mol11** which carries a categorical inhibitor YES prediction for CYP1A2/ CYP2C19/ CYP2D6 and of **Mol10** for CYP1A2/ CYP2C19/ CYP3A4. In 2020, out of 17 families only 57 genes of CYP were identified in humans. Assuming that only CYP is responsible for biotransformation, we treated the cases of 1A2, 2C19, 2D6, 3A4 and 2C9. Our study revealed that the most important inhibitor is CYP3A4 among all these families<sup>8</sup>. Indeed, Table 6 shows that the new proposed molecules are substrates as well as CYP3A4 inhibitors.

**Table3**

Important computed ADMET properties and computed safety end points of the screened best sixteen compounds of and of the most active compound **C48**.

<b>Cp.</b>	<b>Caco2 permeability (log P<sub>app</sub> in 10<sup>-6</sup> cm/s)</b>	<b>Intestinal absorption (human) (% Absorbed)</b>	<b>V<sub>Dss</sub>(human) (log L/kg)</b>	<b>Fraction unbound (human) (Fu)</b>	<b>P-gp substrate (Yes/No)</b>	<b>BBB Permeability (log BB)</b>	<b>Synthetic accessibility</b>	<b>Total clearance (log ml/min/kg)</b>	<b>Renal OCT2 substrate (Yes/No)</b>	<b>Skin sensitization (Yes/No)</b>
<b>C48</b>	1.476	92.393	0.728	0.258	no	0.403	3.10	0.753	no	no
<b>Mol1</b>	1.431	92.530	0.697	0.280	no	0.495	3.53	0.583	no	no
<b>Mol2</b>	1.478	91.664	0.776	0.245	no	0.434	3.66	0.743	no	no
<b>Mol3</b>	1.475	92.333	0.764	0.248	no	0.415	3.66	0.559	no	no
<b>Mol4</b>	1.521	90.593	0.749	0.158	no	0.410	3.89	0.570	yes	no
<b>Mol5</b>	1.272	91.704	1.160	0.205	yes	0.433	3.35	1.013	yes	no
<b>Mol6</b>	1.177	93.820	1.023	0.237	yes	0.521	3.97	0.902	yes	no
<b>Mol7</b>	1.228	88.678	0.827	0.163	no	0.865	3.63	0.460	no	no
<b>Mol8</b>	1.274	88.919	1.010	0.149	no	0.815	4.01	0.608	yes	no
<b>Mol9</b>	1.31	89.374	1.209	0.179	no	0.713	3.78	0.971	yes	no
<b>Mol10</b>	1.173	93.832	0.935	0.248	yes	0.392	3.61	0.633	yes	no
<b>Mol11</b>	1.189	93.785	1.073	0.242	yes	0.383	4.19	0.686	yes	no
<b>Mol12</b>	1.286	88.067	0.924	0.122	no	0.798	4.23	0.485	no	no
<b>Mol13</b>	1.222	88.445	0.974	0.140	no	0.904	4.17	0.502	no	no
<b>Mol14</b>	1.222	88.793	0.892	0.159	no	0.891	4.07	0.646	no	no
<b>Mol15</b>	1.262	90.420	1.053	0.192	no	0.762	3.59	0.732	no	no
<b>Mol16</b>	1.275	91.113	1.083	0.179	no	0.788	4.16	0.691	no	no

**Table4**Metabolism and toxicity predictions for the best sixteen compounds and for the active compound **C48**.

<b>Cp.</b>	<b>CYP1A2 inhibitor (Yes/No)</b>	<b>CYP2C19 inhibitor (Yes/No)</b>	<b>CYP2C9 inhibitor (Yes/No)</b>	<b>CYP2D6 inhibitor (Yes/No)</b>	<b>CYP3A4 inhibitor (Yes/No)</b>	<b>Hepatotoxicity (Yes/No)</b>	<b>AMES toxicity (Yes/No)</b>	<b>hERG I inhibitor (Yes/No)</b>	<b>Oral rat acute toxicity (LD<sub>50</sub>) (mol/kg)</b>	<b>Oral rat chronic toxicity (LOAEL) (log mg/kg_bw/day)</b>
<b>C48</b>	no	no	no	no	no	yes	no	no	3.061	0.830
<b>Mol1</b>	no	no	no	no	no	yes	no	no	3.03	0.991
<b>Mol2</b>	no	no	no	no	no	no	no	no	3.064	1.007
<b>Mol3</b>	no	no	no	no	no	no	no	no	3.087	0.873
<b>Mol4</b>	no	no	no	no	yes	no	no	no	2.911	1.358
<b>Mol5</b>	no	no	no	no	yes	yes	no	no	2.981	1.186
<b>Mol6</b>	yes	no	no	no	no	yes	no	no	3.035	1.323
<b>Mol7</b>	no	no	no	no	no	yes	no	no	3.185	0.675
<b>Mol8</b>	no	no	no	no	no	yes	no	no	3.299	0.696
<b>Mol9</b>	no	no	no	no	yes	yes	no	no	3.043	1.300
<b>Mol10</b>	yes	yes	no	no	yes	yes	no	no	2.962	1.332
<b>Mol11</b>	yes	yes	no	yes	no	yes	no	no	2.954	1.322
<b>Mol12</b>	no	no	no	no	yes	yes	no	no	3.23	0.695
<b>Mol13</b>	no	no	no	no	yes	yes	no	no	3.308	0.789
<b>Mol14</b>	no	no	no	no	no	yes	no	no	3.282	0.744
<b>Mol15</b>	no	no	no	no	no	yes	no	no	3.06	0.710
<b>Mol16</b>	no	no	no	no	no	yes	no	no	3.101	0.802

By using the online SwissADME tool, we detected the similarity features of the drugs (**Table 3**). For a particular molecule, the blood-brain barrier (BBB) permeability coefficients should be in the range  $-1 < \text{Log BB} < 0.3$ . When  $\text{Log BB} < -1$ , this means that the distribution is weak on the brain, while it is essential for the passage of the BBB at  $\text{Log BB} > 0.3$ <sup>9</sup>. Table 5 shows that all values of the proposed compounds achieve  $\text{Log BB} > 0.3$ , which evidence their great potential to penetrate the barriers mainly. Also, we have reached the significant and complex reduction of the structure based on the values of the potentials of new compounds for synthetic accessibility, which ranges between the lowest value of 3.10 for the most active molecule (i.e. **C48**) and the value of 4.23 for compound **Mol12**, and thus proving its industrial potential applications.

**Table 4** shows that we achieve good results for skin sensitization, as indicated by the hERGI inhibitor and for AMES toxicity with NO for all the proposed molecules. The values of Oral Rat Acute Toxicity in the field (2.911-3.308) mol/kg and Oral Rat Chronic Toxicity in the field (0.675-1.358), and for hepatotoxicity are also good for compounds **Mol2**, **Mol3** and **Mol4**. Comparing it with the poor YES score for the most active molecule **C48**, this is the best indicator of the effectiveness of the new designed molecules. We can also deduce that the latter facilitates the basic and normal function of the liver (**Table 4**).

**Table 4** clearly revealed that Mol2 and Mol3 exhibit non-hepatotoxicity and non-AMES toxicity, confirming that ADMET results reveal potential new outcomes with improved efficacy for sigma-1 antagonists<sup>2</sup>. In the **Supplementary Information** section, we have added Smiles icons (**Table S2**) and in **Fig. S1** a boiled-egg drawing of the proposed new compounds. For instance, **Fig. S1** shows the plot between TPSA and WLOGP for boiled eggs in order to predict penetration into the brain and to predict better absorption at the gastrointestinal level of the proposed new compounds. Commenting on this graph, we note that compound **Mol13**, which has the highest degree of BBB within the yellow region, exhibits a good absorption at the level of the gastrointestinal tract. The white region contains higher HIA values for two compounds. For substrate prediction (PGP+) and non-substrates (PGP-), this chart predicts them in order to know the permeability of glycoprotein (PGP).

## 7. DFT calculations

### 7.1. The trend of reactivity and FMO study

The frontier molecular orbitals (FMOs), namely the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), play a crucial role in comprehending the chemical characteristics of a molecule. In order to gain a deeper



understanding of the HOMO-LUMO orbitals of the compounds Mol2, Mol3, and Mol4 with the ligand of reference, an illustrative representation was provided in Fig. 5. Exploiting the LUMO-HOMO energies, several important parameters such as the energy gap ( $\Delta E$ ), ionization potential (IP), electron affinity (EA), electronegativity ( $\chi$ ), chemical potentials ( $\mu$ ), chemical hardness ( $\eta$ ), softness ( $\sigma$ ), and electrophilicity index ( $\omega$ ) were meticulously calculated and presented in Table 5. The HOMO and LUMO orbitals serve as distinctive markers for determining the susceptibility of a molecule towards an attack by electrophiles. Based on the obtained  $E_{\text{HOMO}}$  and  $E_{\text{LUMO}}$  values, as shown in Table 5, it is noteworthy that a higher  $E_{\text{HOMO}}$  signifies the ease of electron donation between the substrate and the target proteins. Conversely, a lower  $E_{\text{LUMO}}$  indicates the ease of electron acceptance between the substrate and the target proteins<sup>10</sup>. Consequently, the order of increasing electron donation/acceptance ease can be established as follows: Mol2 > Mol3 > Mol4.

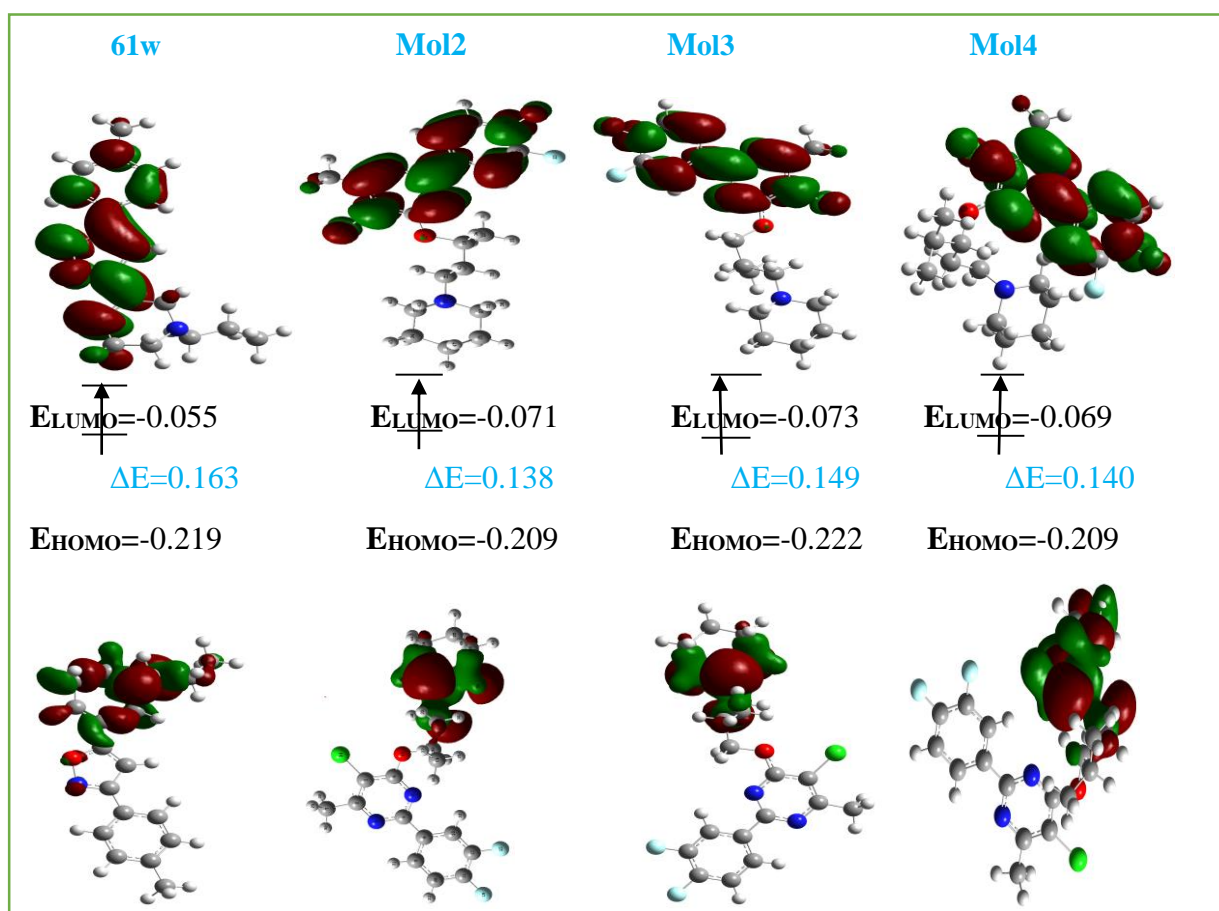


Fig. 5. HOMO-LUMO structures of the compound Mol2, Mol3 and Mol4 with ligand of reference 61w.

The energy gap ( $\Delta E$ ) essentially represents the disparity between the LUMO and HOMO values. It is worth mentioning that a smaller  $\Delta E$  value signifies a more reactive molecule

towards docking. Hence, the global reactivity of the compounds under investigation can be ranked as **Mol2** > **Mol4** > **Mol3**. The intermolecular behavior of one molecule towards another molecule can be effectively analyzed by employing the hard-soft-acid-base rule. In the realm of biological targets such as cells and proteins, the softness factor plays a pivotal role. Consequently, it is observed that soft molecules can effortlessly interact with biological targets compared to hard molecules. This implies that as softness increases and hardness decreases, the biological activity also increases, as stated in reference<sup>10</sup>. Therefore, the reactivity ranking should adhere to the following order: **Mol2** > **Mol4** > **Mol3**, as depicted in **Table 5**. Furthermore, the negative value of the chemical potential serves as an indicator of the stability of the compounds **Mol2**, **Mol3**, and **Mol4**, as elaborated in reference<sup>11</sup>. It is also worth noting that a higher value of the electrophilicity index ( $\omega$ ) and a lower value of the chemical potential favor the electrophilic behavior of the compounds, as highlighted in reference<sup>12</sup>.

**Table5**

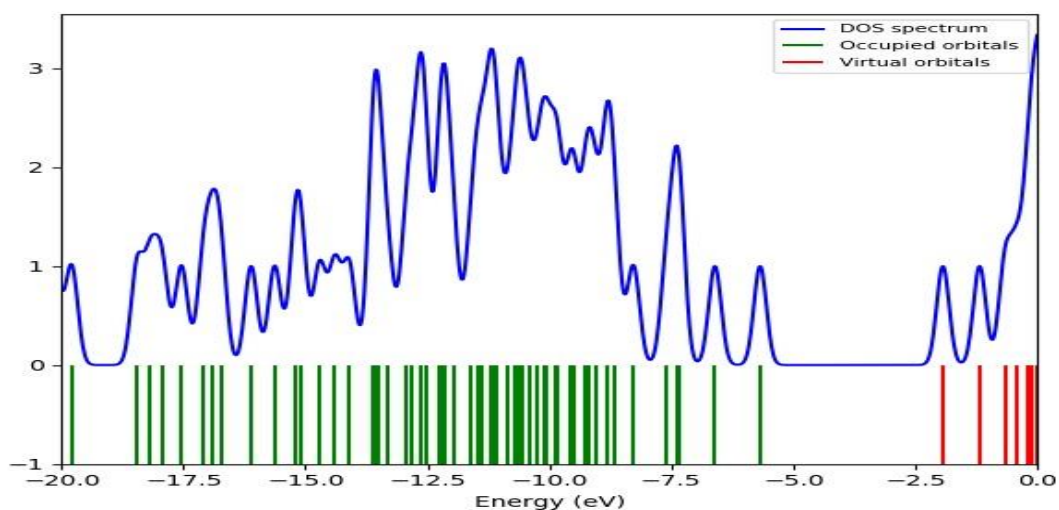
The  $E_{\text{HOMO}}$ ,  $E_{\text{LUMO}}$ , energy gap ( $\Delta E$ ), ionization potential (IP), electron affinity (EA), electronegativity ( $\chi$ ), chemical potentials ( $\mu$ ), Chemical hardness ( $\eta$ ), Softness ( $\sigma$ ), and electrophilicity index ( $\omega$ ) of the subject ligands have been accurately calculated in electron volt (eV) units.

	$E_{\text{HOMO}}$	$E_{\text{LUMO}}$	$\Delta E$	IP	EA	$\chi$	$\mu$	$\eta$	$\sigma$	$\omega$
<b>61w</b>	-0.219	-0.055	0.163	0.219	0.055	0.137	-0.137	0.082	6.121	0.115
<b>Mol2</b>	-0.209	-0.071	0.138	0.209	0.071	0.140	-0.140	0.069	7.263	0.142
<b>Mol3</b>	-0.222	-0.073	0.149	0.222	0.073	0.147	-0.147	0.075	6.706	0.145
<b>Mol4</b>	-0.209	-0.069	0.140	0.209	0.069	0.139	-0.139	0.070	7.138	0.139

## 7.2. Density of states (DOS)

The DOS spectrum, which showcases the distribution of states for a given compound, was generated through the utilization of the highly advanced and sophisticated Gauss Sum 3.0 program. This program possesses the remarkable capability of providing an accurate representation of the states by employing a full width at half maximum (FWHM) of 0.3 eV, ensuring a precise and comprehensive analysis. In the energy range spanning from -20 eV to 0 eV, the DOS spectra for the compound **Mol2** are depicted in the captivating **Fig. 6**, captivatingly capturing the essence of the compound's electronic structure. The intricately designed plot

masterfully unravels the nature of the interaction between two orbitals, allowing for a profound understanding of whether the interaction is bonding or anti-bonding in nature. Through careful examination of the plot, one can discern the presence of a high-intensity DOS at specific energy levels, signifying the existence of multiple states that are readily available for occupation. Conversely, a state of zero intensity within the plot signifies the absence of any available states, thereby suggesting a lack of occupation possibilities. Interestingly, the presence of negative intensity within the plot serves as a clear indication of an anti-bonding interaction, further enhancing the depth of the analysis. It is truly fascinating to observe that the energy values presented in the HOMO-LUMO analysis are intricately coordinated with the energy gap that is depicted within the DOS spectrum, providing a cohesive and comprehensive understanding of the compound's electronic properties<sup>13</sup>.



**Fig. 6.** DOS plot of the compound **Mol2**.

## References

1. Shao, S., Sun, H., Muhammad, Y., Huang, H., Wang, R., Nie, S., ... & Zhao, Z. (2021). Accurate prediction for adsorption rate of peptides with high ACE-inhibitory activity from sericin hydrolysate on thiophene hypercross-linked polymer using CoMSIA in 3D-QSAR model. *Food research international*, *141*, 110144.
2. Khamouli, S., Belaidi, S., Ouassaf, M., Lanez, T., Belaaouad, S., & Chtita, S. (2022). Multi-combined 3D-QSAR, docking molecular and ADMET prediction of 5-azaindazole derivatives as LRRK2 tyrosine kinase inhibitors. *Journal of Biomolecular Structure and Dynamics*, *40*(3), 1285-1298.
3. Osman, E. A., Abdalla, M. A., Abdelraheem, M. O., Ali, M. F., Osman, S. A., Tanir, Y. M., ... & Alzain, A. A. (2021). Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR, molecular docking and molecular dynamics. *Informatics in Medicine Unlocked*, *26*, 100765.
4. <http://biosig.unimelb.edu.au/pkcsml/prediction>.
5. Daoui, O., Elkhatabi, S., Chtita, S., Elkhlabi, R., Zgou, H., & Benjelloun, A. T. (2021). QSAR, molecular docking and ADMET properties in silico studies of novel 4, 5, 6, 7-tetrahydrobenzo [D]-thiazol-2-Yl derivatives derived from dimedone as potent anti-tumor agents through inhibition of C-Met receptor tyrosine kinase. *Heliyon*, *7*(7).
6. Hadni, H., & Elhallaoui, M. (2020). 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents. *Heliyon*, *6*(4).
7. Edache, E. I., Uzairu, A., Mamza, P. A., & Shallangwa, G. A. (2022). Structure-based simulated scanning of rheumatoid arthritis inhibitors: 2D-QSAR, 3D-QSAR, docking, molecular dynamics simulation, and lipophilicity indices calculation. *Scientific African*, *15*, e01088.
8. Shi, J., Zhao, L. X., Wang, J. Y., Ye, T., Wang, M., Gao, S., ... & Fu, Y. (2022). The novel 4-hydroxyphenylpyruvate dioxygenase inhibitors in vivo and in silico approach: 3D-QSAR analysis, molecular docking, bioassay and molecular dynamics. *Arabian Journal of Chemistry*, *15*(7), 103919.
9. Khaldan, A., Bouamrane, S., En-Nahli, F., El-Mernissi, R., Hmamouchi, R., Maghat, H., ... & Lakhlifi, T. (2021). Prediction of potential inhibitors of SARS-CoV-2 using 3D-QSAR, molecular docking modeling and ADMET properties. *Heliyon*, *7*(3).
10. Ismael, M., Abdel-Mawgoud, A. M. M., Rabia, M. K., & Abdou, A. (2021). Ni (II) mixed-ligand chelates based on 2-hydroxy-1-naphthaldehyde as antimicrobial agents: Synthesis, characterization, and molecular modeling. *Journal of Molecular Liquids*, *330*, 115611.

11. Al-Wabli, R. I., Resmi, K. S., Mary, Y. S., Panicker, C. Y., Attia, M. I., El-Emam, A. A., & Van Alsenoy, C. (2016). Vibrational spectroscopic studies, Fukui functions, HOMO-LUMO, NLO, NBO analysis and molecular docking study of (E)-1-(1, 3-benzodioxol-5-yl)-4, 4-dimethylpent-1-en-3-one, a potential precursor to bioactive agents. *Journal of Molecular Structure*, 1123, 375-383.
12. Chaudhary, A. P., Bharti, S. K., Kumar, S., Ved, K., & Padam, K. (2017). Study of molecular structure, chemical reactivity and first hyperpolarizability of a newly synthesized N-(4-oxo-2-phenylquinazolin-3 (4H)-yl)-1H-indole-2-carboxamide using spectral analysis. *Journal of Molecular Structure*, 1148, 356-363.
13. Rijal, R., Sah, M., & Lamichhane, H. P. (2023). Molecular simulation, vibrational spectroscopy and global reactivity descriptors of pseudoephedrine molecule in different phases and states. *Heliyon*, 9(3).

# *General Conclusion*

With the aim of discovering new anti-neuropathic pain drugs, we performed a 3D-QSAR study of a series of 54 S1R antagonists, which are known to have a good inhibitory activity against this receptor. Because of the excellent predictive ability of the CoMSIA model, highly statistical results were obtained. Afterwards, we analyzed the contour maps resulting from the CoMSIA model, which allowed identifying the appropriate and unsuitable alternatives for the structure that can produce a distinctive activity. Based on that, sixteen new active compounds were successfully designed. They exhibit a more efficient biological activity. Afterwards, interactions between the transmembrane protein 5HK1 with the newly designed most active molecules **Mol2**, **Mol15**, **Mol16** and the active compound **C48** were studied. The three newly proposed active compounds reveal a stability that exceeded those of the reference compound **61W** and of compound **C48**. The ADMET model, which showed us the features and characteristics of all the proposed compounds, proved the effectiveness and the essential role of the normal function of the liver for these molecules. We hope that our work will encourage the synthesis in the laboratory of the new series of sixteen proposed compounds and their pharmacological investigations. DFT calculations were expertly executed, and an in-depth analysis was conducted to examine the discrepancies in the HOMO-LUMO gap present within the finest **Mol2**, **Mol3**, and **Mol4** compounds from prior research in comparison to the reference compound **61w**. This ground breaking technique has bestowed upon us the remarkable ability to accurately forecast various properties of molecules and delve into the intricate realm of geometric properties and energy evaluation. Moreover, it facilitates a meticulous examination of the border molecular orbitals, shedding light on their characteristics, while simultaneously permitting the determination of the density of states (DOS) of the exceptionally stable compound known as **Mol2**. As a result, this research could contribute to the development of drugs for neuropathic pain with high inhibitory power.

# *Appendices*



Table S1

Structures of the best newly compounds, Chemical nomenclature and their masses (Mol2, Mol3 and Mol4).

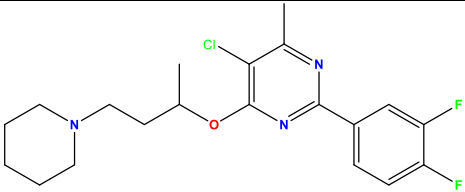
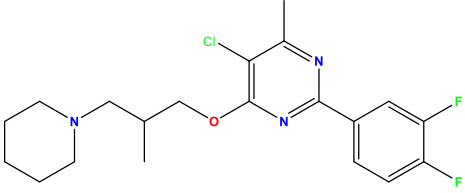
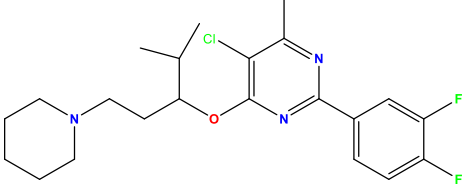
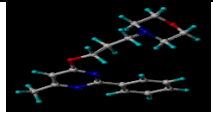
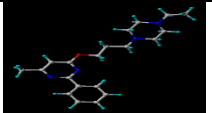

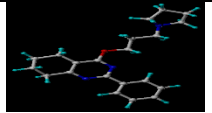
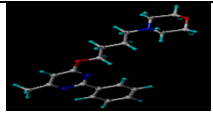

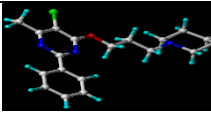
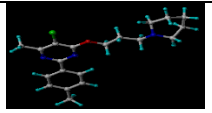
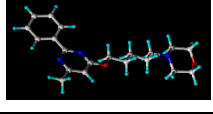



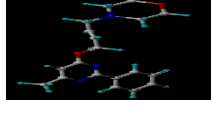
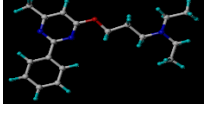

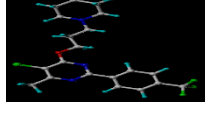
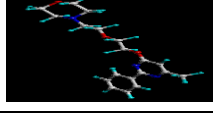
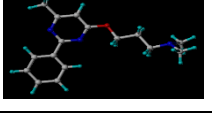
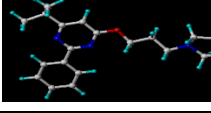
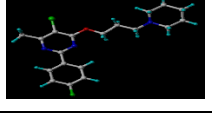
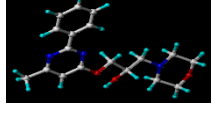
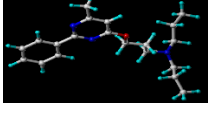
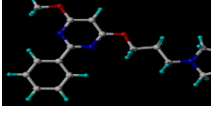

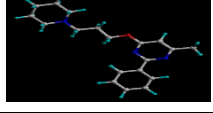
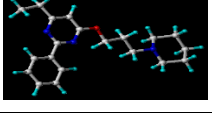
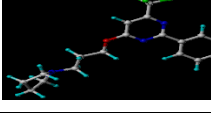
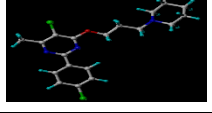
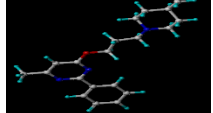
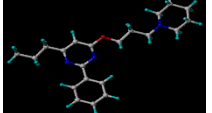
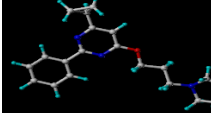

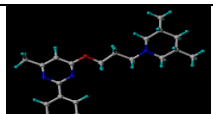
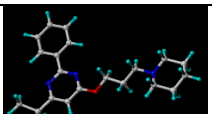
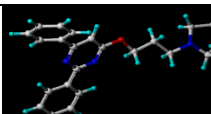
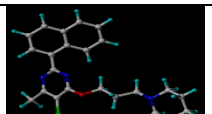
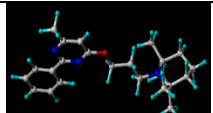
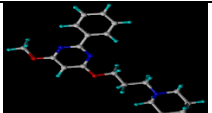
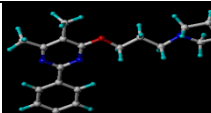
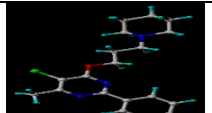
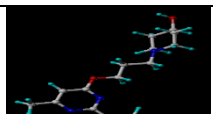
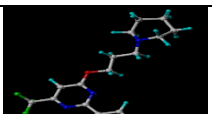
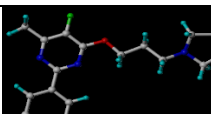
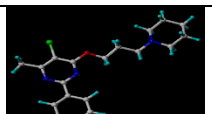
Structure	Chemical nomenclature	MW amu
<p><b>Mol2</b></p> 	<p>5-chloro-2-(3,4-difluorophenyl)-4-methyl-6-((4-(piperidin-1-yl)butan-2-yl)oxy)pyrimidine</p>	395.88
<p><b>Mol3</b></p> 	<p>5-chloro-2-(3,4-difluorophenyl)-4-methyl-6-(2-methyl-3-(piperidin-1-yl)propoxy)pyrimidine</p>	395.88
<p><b>Mol4</b></p> 	<p>5-chloro-2-(3,4-difluorophenyl)-4-methyl-6-((4-methyl-1-(piperidin-1-yl)pentan-3-yl)oxy)pyrimidine</p>	423.93

Table S2

Optimized 3D structures (Cp.) of the series of 54 optimized compounds containing pyrimidine of S1R antagonists currently studied.

Cp.	3D structure	Cp.	3D structure	Cp.	3D structure	Cp.	3D structure
C1 <sup>t</sup>		C15		C29		C43	
C2		C16 <sup>t</sup>		C30		C44	
C3		C17		C31		C45 <sup>t</sup>	
C4		C18		C32		C46	
C5		C19		C33		C47 <sup>t</sup>	
C6 <sup>t</sup>		C20		C34		C48	
C7		C21		C35 <sup>t</sup>		C49	
C8		C22		C36 <sup>t</sup>		C50 <sup>t</sup>	
C9		C23		C37 <sup>t</sup>		C51 <sup>t</sup>	
C10 <sup>t</sup>		C24		C38 <sup>t</sup>		C52 <sup>t</sup>	
C11 <sup>t</sup>		C25		C39		C53 <sup>t</sup>	



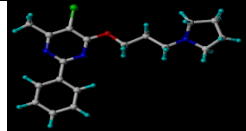
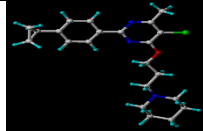
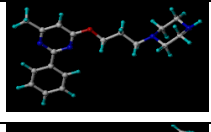

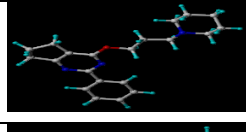
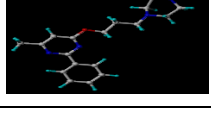
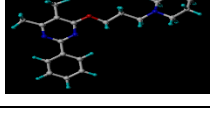

C12		C26		C40		C54'	
C13		C27		C41			
C14		C28'		C42'			

Table S3

SMILES notation for the proposed new molecules.

<i>Comp.</i>	<i>Smile</i>
<i>C48</i>	<chem>Cc:1:n:c(:n:c(OCCCN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:c(F):c(F):[cH]:3</chem>
<i>Mol1</i>	<chem>Cc:1:n:c(:n:c(OC(C)CN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:c(F):c(F):[cH]:3</chem>
<i>Mol2</i>	<chem>Cc:1:n:c(:n:c(OC(C)CN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:c(F):c(F):[cH]:3</chem>
<i>Mol3</i>	<chem>Cc:1:n:c(:n:c(OCC(C)CN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:c(F):c(F):[cH]:3</chem>
<i>Mol4</i>	<chem>Cc:1:n:c(:n:c(OC(CCN2CCCCC2)C(C)C):c:1Cl)c:3:[cH]:[cH]:c(F):c(F):[cH]:3</chem>
<i>Mol5</i>	<chem>COc:1:[cH]:[cH]:c(:[cH]:[cH]:1)c:2:n:c(C):c(Cl):c(:n:2)OCC3(CC3)CN4CCCC4</chem>
<i>Mol6</i>	<chem>COc:1:[cH]:[cH]:c(:[cH]:[cH]:1)c:2:n:c(C):c(Cl):c(:n:2)OC3CC(C3)N4CCCC4</chem>
<i>Mol7</i>	<chem>Cc:1:n:c(:n:c(OC(C)CN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:c(:[cH]:[cH]:3)C(F)(F)F</chem>
<i>Mol8</i>	<chem>Cc:1:n:c(:n:c(OC2CC2CN3CCCCC3):c:1Cl)c:4:[cH]:[cH]:c(:[cH]:[cH]:4)C(F)(F)F</chem>
<i>Mol9</i>	<chem>Cc:1:n:c(:n:c(OCC2(CC2)C(C)N3CCCCC3):c:1Cl)c:4:[cH]:[cH]:c(Cl):[cH]:[cH]:4</chem>
<i>Mol10</i>	<chem>COc:1:[cH]:[cH]:[cH]:c(:[cH]:1)c:2:n:c(C):c(Cl):c(:n:2)OC(C)CN3CCCC3</chem>
<i>Mol11</i>	<chem>COc:1:[cH]:[cH]:[cH]:c(:[cH]:1)c:2:n:c(C):c(Cl):c(:n:2)OC3C(C)C3N4CCCC4</chem>
<i>Mol12</i>	<chem>Cc:1:n:c(:n:c(OCC(C)C(C)N2CCCCC2):c:1Cl)c:3:[cH]:[cH]:[cH]:c(:[cH]:3)C(F)(F)F</chem>
<i>Mol13</i>	<chem>Cc:1:n:c(:n:c(OC2CCCC2N3CCCCC3):c:1Cl)c:4:[cH]:[cH]:[cH]:c(:[cH]:4)C(F)(F)F</chem>
<i>Mol14</i>	<chem>Cc:1:n:c(:n:c(OC2CC(C2)N3CCCCC3):c:1Cl)c:4:[cH]:[cH]:[cH]:c(:[cH]:4)C(F)(F)F</chem>
<i>Mol15</i>	<chem>Cc:1:n:c(:n:c(OC(C)CN2CCCCC2):c:1Cl)c:3:[cH]:[cH]:[cH]:c(Cl):[cH]:3</chem>
<i>Mol16</i>	<chem>Cc:1:n:c(:n:c(OC2CC2C(C)N3CCCCC3):c:1Cl)c:4:[cH]:[cH]:[cH]:c(Cl):[cH]:4</chem>

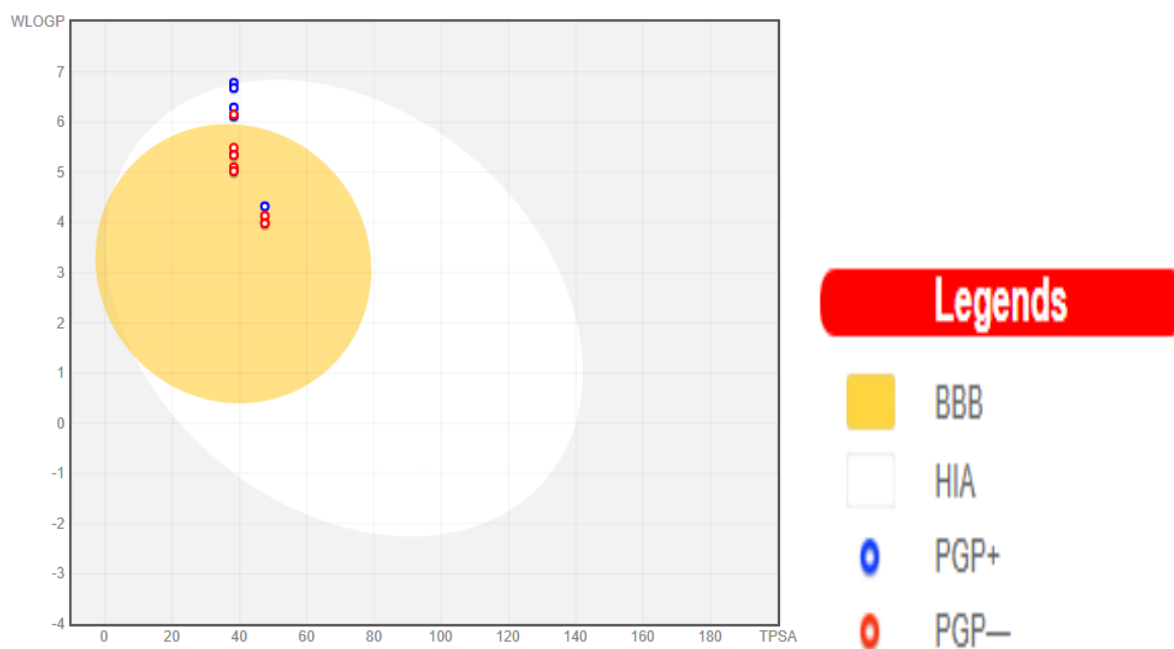
**Table S4**

Interactions of the enzyme 5HK1 with the three best inhibitors and reference ligand 61W and C48 as presented in Figure 4.

Complex (Ligands-5HK1)	Amino Acids	Categories	Interactions
61W-5HK1	HIS A:154	Hydrophobic	Alkyl and Pi-Alkyl
	ILE A:124	Hydrophobic	
	TYR A:120	Hydrophobic	
	ALA A:185	Hydrophobic	
	LEU A:105	Hydrophobic	
	VAL A:84	Hydrophobic	
	MET A:93	Hydrophobic	
	LEU A:95	Hydrophobic	
	THR A:202	Hydrophobic	
	LEU A:182	Hydrophobic	
	ALA A:98	Hydrophobic	
	TYR A:206	Hydrophobic	
	GLU A:172	Hydrogen	Carbon Hydrogen Bond
	SER A:117	Hydrogen	
	TYR A:120	Hydrogen	
GLU A:172	Other	Salt Bridge	
ASP A:126	Other		
C48-5HK1	ALA A:185	Hydrophobic	Alkyl and Pi-Alkyl
	TYR A: 120	Hydrophobic	
	PHE A: 184	Hydrophobic	
	TYR A:103	Hydrophobic	
	LEU A: 182	Hydrophobic	
	LEU A:95	Hydrophobic	
	ILE A:124	Hydrogen	Carbon Hydrogen Bond
	TRPA A:89	Hydrogen	
	TYR A:120	Hydrogen	Conventional Hydrogen Bond
	THR A:202	Halogen	Halogen (Fluorine)
	TYR A:103	Hydrophobic	Pi-Pi Stacked
	TYR A:206	Hydrophobic	Pi-Pi T-shaped
	GLU A:172	Other	Pi-Cation
	PHE A:107	Other	Pi-Anion
ASP A:126	Other	Salt Bridge Attractive Charge	
Mol2-5HK1	PHE A:107	Hydrophobic	Alkyl and Pi-Alkyl

	LEU A:105	Hydrophobic	
	MET A:93	Hydrophobic	
	LEU A :95	Hydrophobic	
	TYR A :103	Hydrophobic	
	ILE A :124	Hydrophobic	
	ALA A: 185	Hydrophobic	
	PHE A: 184	Hydrophobic	
	TYR A:120	Hydrophobic	
	GLU A:172	Hydrogen	Carbon Hydrogen Bond
	TYR A: 120	Hydrogen	Conventional Hydrogen Bond
	GLU A:172	Other	Pi-Anion
	ASP A:126	Other	Salt Bridge Attractive Charge
	TYR A:103	Hydrophobic	Pi-Pi T-shaped
	TYR A:206	Hydrophobic	
Mol15-5HK1	ILE A:124	Hydrophobic	Alkyl and Pi-Alkyl
	TYR A:103	Hydrophobic	
	ALA A:185	Hydrophobic	
	PHE A:184	Hydrophobic	
	TYR A:120	Hydrophobic	
	PHE A:107	Hydrophobic	
	LEU A:105	Hydrophobic	
	MET A:93	Hydrophobic	
	LEU A:95	Hydrophobic	
	LEU A:182	Hydrophobic	
	TYR A: 120	Hydrogen	Carbon Hydrogen Bond
	GLU A:172	Hydrogen	
	TYR A: 120	Hydrogen	Conventional Hydrogen Bond
	TYR A:206	Hydrophobic	Pi-Pi T-shaped
	GLU A:172	Other	Pi-Cation Pi-Anion
	PHE A:107	Other	Salt Bridge
	ASP A:126	Other	Attractive Charge
Mol16-5HK1	ALA A:98	Hydrophobic	Alkyl and Pi-Alkyl
	LEU A:95	Hydrophobic	
	ALA A:185	Hydrophobic	
	PHE A:107	Hydrophobic	
	LEU A:105	Hydrophobic	
	MET A:93	Hydrophobic	

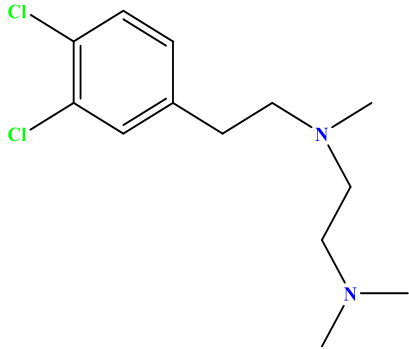
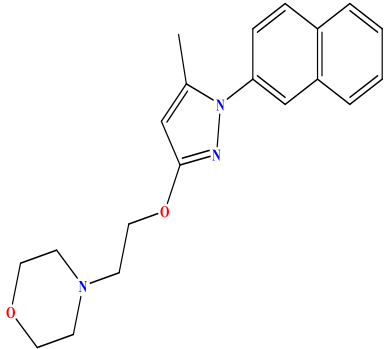
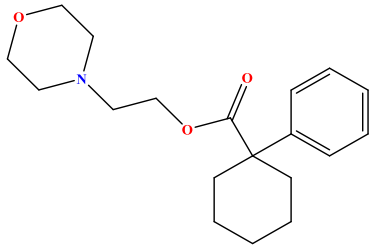
	VAL A:84	Hydrophobic	
	HIS A:154	Hydrophobic	
	TYR A: 120	Hydrogen	Carbon Hydrogen Bond
	TYR A:103	Hydrogen	Conventional Hydrogen Bond
	LEU A: 182	Hydrophobic	Pi-Sigma
	TYR A:103	Hydrophobic	Pi-Pi T-shaped
	TYR A:206	Hydrophobic	
	GLU A:172	Other	Salt Bridge
	ASP A:126	Other	



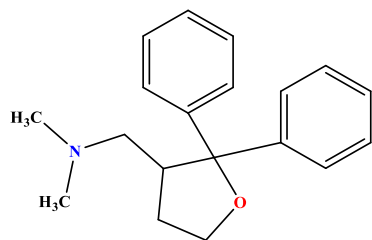
**Fig. S1.** Boiled-egg drawing of the proposed new compounds.

Table S5

Structures of the antagonists BD-1047 and S1RA, as well as the agonists PRE-084 and Blarcamesine, Chemical nomenclature and their Scores.

Generic Name	Structure	Chemical nomenclature	Total scoring
BD-1047		N-(2-(3,4-Dichlorophenyl)ethyl)- N-methyl-2- (dimethylamino)ethylamine	-7.9371
S1RA		4-(2-((5-methyl-1-(naphthalen-2- yl)-1H-pyrazol-3- yl)oxy)ethyl)morpholine	-9.0658
PRE-084		2-(4-Morpholino)ethyl-1- phenylcyclohexane-1-carboxylate	-4.4224

Blarcamesine



1-(2,2-diphenyltetrahydrofuran-3-yl)-N,N-dimethylmethanamine

-7.4501

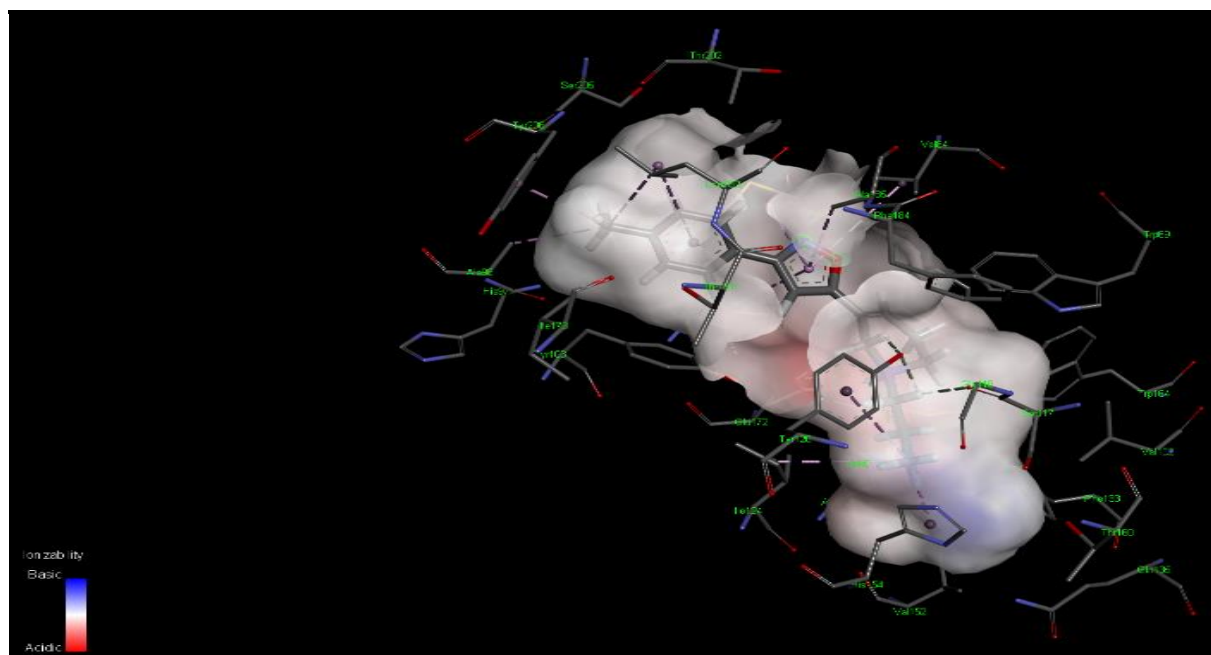


Fig.S2. Docking interactions of 61W





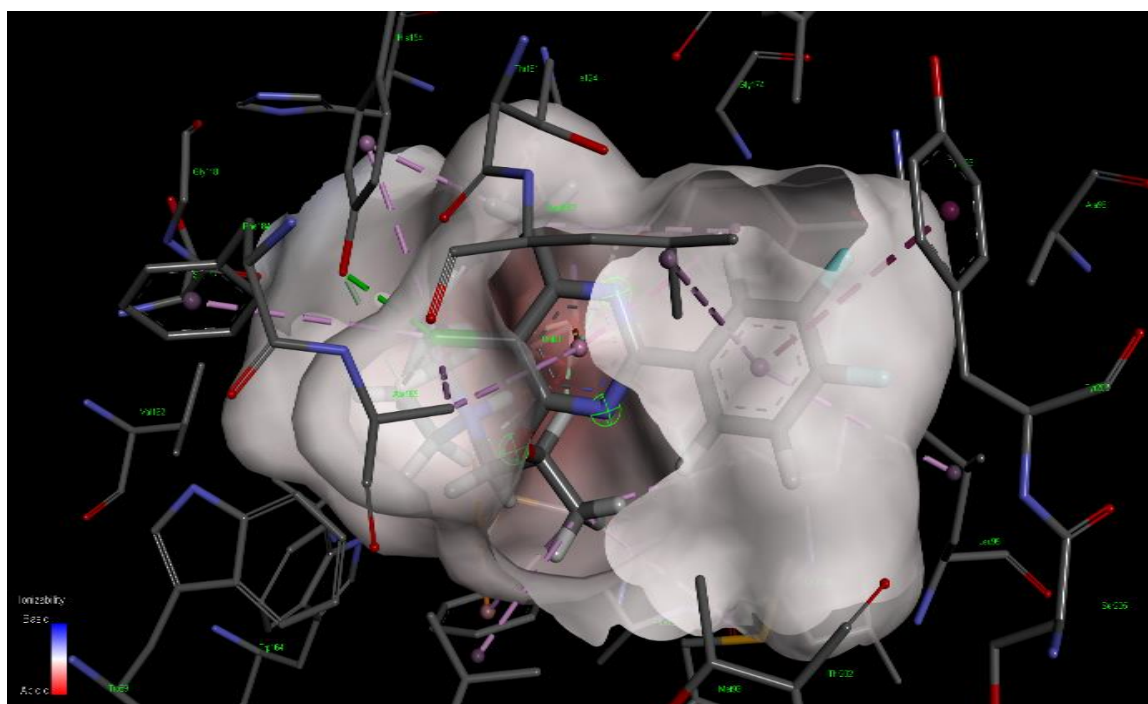
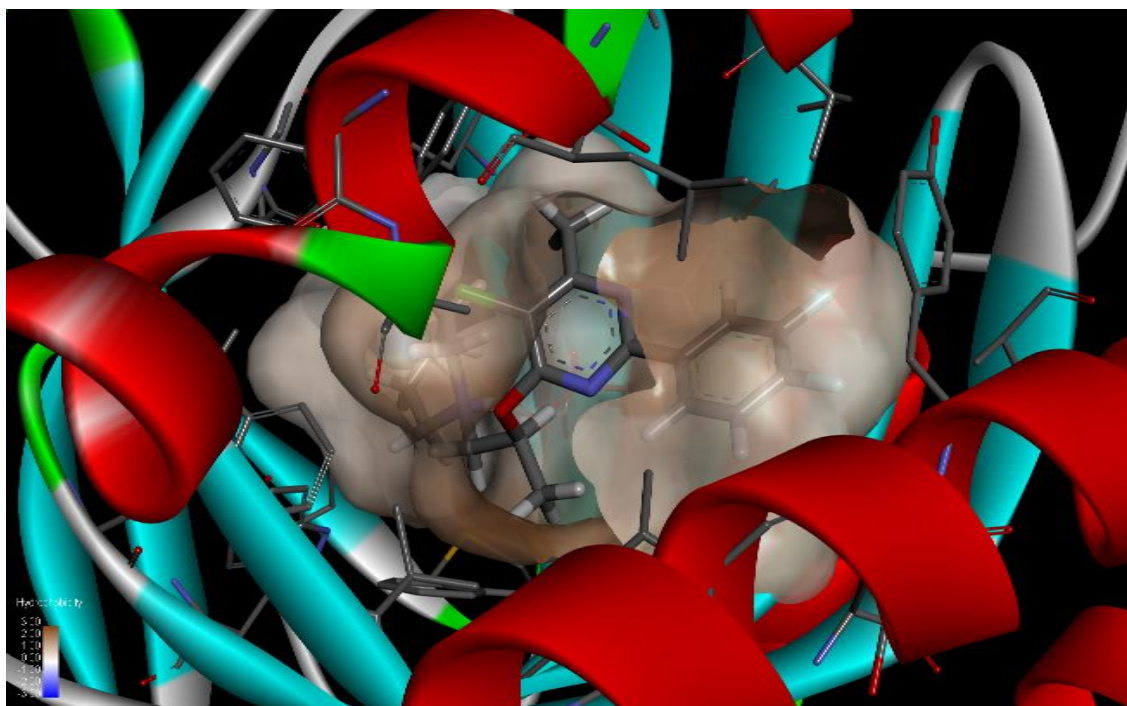


Fig.S4. Docking interactions of Moll1.



ELSEVIER

Contents lists available at ScienceDirect

Chemical Physics Impact

journal homepage: [www.sciencedirect.com/journal/chemical-physics-impact](http://www.sciencedirect.com/journal/chemical-physics-impact)

Full Length Article

## Computational studies of pyrimidine derivatives as inhibitors of human $\sigma_1$ receptor using 3D-QSAR analysis, molecular docking, ADMET properties and DFT investigation

Maroua Fattouche<sup>a</sup>, Salah Belaidi<sup>a</sup>, Mebarka Ouassaf<sup>a,\*</sup>, Samir Chtita<sup>b</sup>,  
Muneerah Mogren Al-Mogren<sup>c</sup>, Majdi Hochlaf<sup>d</sup>

<sup>a</sup> Group of Computational and Medicinal Chemistry, LMCE Laboratory, University of Biskra, BP 145, Biskra, 07000, Algeria

<sup>b</sup> Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sick, Hassan II University of Casablanca, B.P. Casablanca, 7955, Morocco

<sup>c</sup> Department of Chemistry, College of Sciences, King Saud University, PO Box 2455, Riyadh, 11451, Saudi Arabia

<sup>d</sup> COSYS/IMSE, Université Gustave Eiffel, 5 Bd Descartes, Champs sur Marne, 77454, France

## ARTICLE INFO

## Keywords:

$\sigma_1$  receptor antagonists  
Pyrimidine derivatives  
3D-QSAR  
Molecular docking  
ADMET  
DFT  
FMOS  
DOS

## ABSTRACT

Neuropathic pain syndrome has a profoundly negative and agonizing impact on the lives of the individuals it afflicts. In order to find an effective treatment for this condition, extensive and thorough scientific studies have demonstrated that the  $\sigma_1$  receptor serves as an exceptional target for therapeutic compounds. The 3D-QSAR studies were constructed using the technique of comparative molecular similarity indice analysis (CoMSIA). The outcomes of these studies demonstrated the reliability of CoMSIA model (with  $R^2$  train value of 0.96 and  $Q^2$  value of 0.54) in accurately predicting the activity of various compounds. By assimilating the valuable insights gathered from the field contributors of the 3D-QSAR models and conducting molecular docking studies on the highly potent compound C48, a total of sixteen new compounds were successfully designed to exhibit enhanced efficacy against neuropathic pain. In addition to the comprehensive 3D-QSAR analysis, the newly synthesized compounds were subjected to an absorption, distribution, metabolism, excretion, and toxicity evaluation. This evaluation aimed to assess the pharmacokinetic and toxicological properties of the compounds, providing valuable insight for future in vitro investigations. Calculations DFT of the new compounds mol2, mol3, and mol4, including the analysis of their molecular properties, geometric optimization, frontier molecular orbital (FMOs), density states (DOS), and energy evaluation, In order to compare these compounds with the reference ligand 61 w, their respective properties were thoroughly investigated. The most stable orientations for the compounds Mol2, Mol4, and Mol3 were determined to be the ones that yielded the highest stability and efficiency. The significant advancements made in this study, should serve as a strong motivation for future in vitro investigations on these compounds.

## 1. Introduction

Peripheral or central nervous system injury produces a syndrome called neuropathy, which greatly affects about 6–10 % of the world's population. Recent studies expect that this syndrome will increase in prevalence due to aging in the coming years, and this leads to an increase in the survival of patients with chronic diseases related to neuropathic pains, which are associated also with the emergence of specific symptoms in patients presenting this syndrome. For the treatment of these diseases, some medications are prescribed, such as opiates,

anticonvulsants and antidepressants, which have a certain ability to block ion channels responsible for these neuropathic pains. However, the effectiveness of these drugs is weak in addition to their confirmed significant side effects. Therefore, it is necessary to reduce the symptoms of this syndrome by discovering new effective drugs [1].

Due to its ease of transport to many parts of the cell, various studies have shown that human  $\sigma_1$  receptor, S1R, uniquely treats neuropathic pain. S1R is a transmembrane protein, which is predominantly located in the endoplasmic reticulum and the mitochondrial reticulum [2]. Such receptors were widely investigated because of their important role in

\* Corresponding author.

E-mail address: [nouassaf@univ-biskra.dz](mailto:nouassaf@univ-biskra.dz) (M. Ouassaf).

<https://doi.org/10.1016/j.chphi.2024.100463>

Received 5 December 2023; Received in revised form 2 January 2024; Accepted 7 January 2024

Available online 14 January 2024

2667-0224/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).