جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم ظهر المهراز فاس
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

**N° d'ordre : 03/2024**

Centre des Etudes : *Sciences et Techniques et Sciences Médicales*

Formation Doctorale : *Ressources Naturelles Et Développement Durable*

Discipline : *Chimie*
Spécialité : *Chimie-Physique Appliquée*
Laboratoire : *Laboratoire d'Ingénierie, Modélisation et Analyse des Systèmes*

## Thèse pour l'obtention de DOCTORAT

# Modélisation moléculaire d'inhibiteurs du récepteur NMDA en utilisant les techniques in-silico : 2D/3D-QSAR, ADME-Tox, Docking moléculaire et dynamique moléculaire

Préparée par : **EL FADILI Mohamed**

Soutenue le 12/01/2024 devant le jury :

| Nom et Prénom | Grade | Etablissement | Qualité |
|---|---|---|---|
| Pr TOUIMI BENJELLOUN Adil | **PES** | Faculté des Sciences Dhar El Mahraz, Fès | **Président** |
| Pr BOUACHRINE Mohammed | **PES** | Faculté des Sciences, Meknès | **Rapporteur** |
| Pr SENHAJI Omar | **PES** | Faculté des Sciences, Meknès | **Rapporteur** |
| Pr ELKHATTABI Souad | **PH** | École Nationale des Sciences Appliquées, Fès | **Rapporteuse** |
| Pr SFAIRA Mouhcine | **PES** | Faculté des Sciences Dhar El Mahraz, Fès | **Examinateur** |
| Pr BENZAKOUR Mohammed | **PES** | Faculté des Sciences Dhar El Mahraz, Fès | **Examinateur** |
| Pr ELHALLAOUI Menana | **PES** | Faculté des Sciences Dhar El Mahraz, Fès | **Directrice de Thèse** |

**Année Universitaire : 2023-2024**

# Remerciements

***« Je remercie Dieu tout-Puissant de m'avoir donné le courage et la patience à réaliser ce travail »***

Tout d'abord, je souhaite adresser mes sincères remerciements à mon directeur de thèse, le Professeur ***EL HALLAOUI Menana*** ; enseignant chercheur à la *Faculté des Sciences Dhar El Mahraz de Fès*, d'avoir accepté de diriger ce travail. Ce fut un grand plaisir de poursuivre mon parcours doctoral sous votre supervision. Vous m'avez toujours réservé le meilleur accueil, malgré vos obligations professionnelles. Vos encouragements constants, votre humilité et gentillesse méritent toute admiration. Grâce à vos suggestions constructives, remarques et critiques, ce travail a pu être mis en œuvre, et sans vous ce travail n'aurait pu être effectué. Merci de m'avoir fait confiance pendant toutes ces trois années de recherche scientifique. Veuillez trouver ici toute ma gratitude et ma profonde reconnaissance.

Je suis sensible à l'honneur que me fait le Professeur ***TOUIMI BENJELLOUN Adil ;*** enseignant chercheur à la *Faculté des Sciences Dhar El Mahraz de Fès*, d'assumer la responsabilité de présider le Jury d'évaluation de cette thèse avec la plus grande spontanéité malgré ses occupations, vous avez d'emblée montré un intérêt certain pour le sujet dénoté travail, veuillez recevoir l'expression de ma profonde reconnaissance et de mon grand respect.

Avec gratitude et sincérité, je voudrais remercier le Professeur ***BOUACHRINE Mohammed*** ; directeur de l'*Ecole Supérieure de Technologie de Khenifra,* et le Professeur ***SENHAJI Omar*** ; enseignant chercheur à la *Faculté des Sciences de Meknès*, d'avoir accepté d'être les rapporteurs de cette thèse, vos qualités humaines et professionnelles exceptionnelles ont toujours suscité estime et admiration, je vous prierais de trouver ici l'expression de ma très haute considération.

Mes remerciements s'adressent également au Professeur ***ELKHATTABI Souad*** ; enseignant chercheur à l'*École Nationale des Sciences Appliquées de Fès***,** d'avoir accepté de faire un rapport sur cette thèse, Je tiens à vous exprimer ma sincère gratitude pour avoir aimablement accepté de juger ce travail.

Mes vifs remerciements s'adressent également aux examinateurs de cette thèse, le professeur ***SFAIRA Mouhcine*** ; directeur du Laboratoire d'Ingénierie, Modélisation et Analyse des Systèmes (LIMAS), et le professeur ***BENZAKOUR Mohammed*** ; enseignant chercheur à la *Faculté des Sciences Dhar El Mahraz de Fès*, pour avoir accepté la lourde tâche de lire l'intégralité de ce manuscrit et d'examiner ces travaux de recherche, veuillez accepter mon profond respect et gratitude chers professeurs.

# Dédicaces

*A ma chère mère*

Aucune dédicace ne pourrait exprimer mon attachement à toi. Tu représentes pour moi la source de tendresse, le symbole de toute bonté, et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.

*A mon cher Père*

Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation.

*A ma chère épouse* ***Leila***

Aucune dédicace ne peut rendre justice à ce que tu mérites pour les énormes sacrifices que tu as faits pour moi. Je te prie d'accepter le grand respect, la gratitude et l'amour que j'ai toujours pour toi.

*A ma fille* ***Amal***

En effet, tu es l'espoir qui m'a accompagné tout au long des années de préparation de cette thèse, et qui m'a poussé d'un moment à l'autre à surmonter tous les obstacles qui ont failli m'empêcher à avancer. Que Dieu vous protège de tout mal.

*A mes chères sœurs* ***Lamyae, Wafae,*** *et* ***Hafssa***

Preuve de l'attachement, de l'amour et de l'affection que je vous porte, malgré la distance, votre place restera toujours dans mon cœur. Merci pour votre sincère amour, votre soutien et vos encouragements continus. Merci pour la chaleur que vous m'apportez chaque jour.

*A mon cher frère* ***Abdelhamid***

Les mots ne suffisent guère pour exprimer l'attachement, l'amour et l'affection que je porte pour toi. Je te dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite dans ta vie personnelle et professionnelle.

*A ceux, qui me sont chers.*

« **Il faut viser la lune, car même en cas d'échec, on atterrit dans les étoiles** »

**Oscar Wilde**, 12 Février 2009

## Liste des productions scientifiques

1. ***M. El fadili***, M. Er-Rajy, M. Kara, A. Assouguem, A. Belhassan, A. Alotaibi, N.N. Mrabti, H. Fidan, R. Ullah, S. Ercisli, S. Zarougui, M. Elhallaoui, QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia, ***Pharmaceuticals***. 15 (2022) 670. **https://doi.org/10.3390/ph15060670.**

2. ***M. El fadili***, M. Er-rajy, H. Imtara, M. Kara, S. Zarougui, N. Altwaijry, O.M. Al kamaly, A. Al Sfouk, M. Elhallaoui, 3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists, ***Processes***. 10 (2022) 1462. **https://doi.org/10.3390/pr10081462.**

3. ***M. El fadili***, M. Er-rajy, H. Imtara, O.M. Noman, R.A. Mothana, S. Abdullah, S. Zerougui, M. Elhallaoui, QSAR, ADME-Tox, molecular docking and molecular dynamics simulations of novel selective glycine transporter type 1 inhibitors with memory enhancing properties, ***Heliyon***. (2023) e13706. **https://doi.org/10.1016/j.heliyon.2023.e13706.**

4. ***M. El fadili***, M. Er-rajy, W. Ali Eltayb, M. Kara, A. Assouguem, A. Saleh, O. Al Kamaly, S. Zerougui, M. Elhallaoui, In-silico Screening Based on Molecular Simulations of 3,4-disubstituted Pyrrolidine Sulfonamides as Selective and Competitive GlyT1 Inhibitors, ***Arabian Journal of Chemistry***. (2023) 105105. **https://doi.org/10.1016/j.arabjc.2023.105105.**

5. ***M. El fadili***, M. Er-rajy, W. Ali Eltayb, M. Kara, H. Imtara, S. Zarougui, N. Al-Hoshani, A. Hamadi, M. Elhallaoui, An in-silico investigation based on molecular simulations of novel and potential brain-penetrant GluN2B NMDA receptor antagonists as anti-stroke therapeutic agents, ***Journal of Biomolecular Structure and Dynamics***. (2023) 1–15. **https://doi.org/10.1080/07391102.2023.2232024.**

6. ***M. El fadili,*** M. Er-rajy, M. Abdalla, H.A. Abuelizz, S. Zarougui, F.M. Alkhulaifi, N.F. Alahmady, A. Shami, M. Elhalaoui, In-silico investigations of novel tacrine derivatives potency against Alzheimer's disease, ***Scientific African***. (2023) e02048. **https://doi.org/10.1016/j.sciaf.2023.e02048.**

7. M. Er-rajy, ***M.E. Fadili***, S. Mujwar, F.Z. Lenda, S. Zarougui, M. Elhallaoui, QSAR, molecular docking, and molecular dynamics simulation–based design of novel anti-cancer drugs targeting thioredoxin reductase enzyme, ***Struct Chem***. (2023). **https://doi.org/10.1007/s11224-022-02111-x.**

8. M. Er-rajy, ***M. El Fadili***, H. Hadni, N.N. Mrabti, S. Zarougui, M. Elhallaoui, 2D-QSAR modeling, drug-likeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives, ***Struct Chem***. 33 (2022) 973–986. **https://doi.org/10.1007/s11224-022-01913-3.**

9. M. Er-rajy, ***M. El fadili***, S. Mujwar, S. Zarougui, M. Elhallaoui, Design of novel anti-cancer drugs targeting TRKs inhibitors based 3D QSAR, molecular docking and molecular dynamics simulation, ***Journal of Biomolecular Structure and Dynamics***. (2023) 1–14. **https://doi.org/10.1080/07391102.2023.2170471.**

10. M. Er-Rajy, ***M. El Fadili***, N.N. Mrabti, S. Zarougui, M. Elhallaoui, QSAR, molecular docking, ADMET properties in silico studies for a series of 7-propanamide benzoxaboroles as potent anti-cancer agents, ***Chinese Journal of Analytical Chemistry***. 50 (2022) 100163. **https://doi.org/10.1016/j.cjac.2022.100163.**

11. M. Er-rajy, ***M. El fadili***, S. Mujwar, H. Imtara, O. Al kamaly, S. Zuhair Alshawwa, F.A. Nasr, S. Zarougui, M. Elhallaoui, Design of novel anti-cancer agents targeting COX-2 inhibitors based on computational studies, ***Arabian Journal of Chemistry***. 16 (2023) 105193. **https://doi.org/10.1016/j.arabjc.2023.105193.**

12. M. Er-rajy, ***M. El fadili***, H. Imtara, A. Saeed, A. Ur Rehman, S. Zarougui, S.A. Abdullah, A. Alahdab, M.K. Parvez, M. Elhallaoui, 3D-QSAR Studies, Molecular Docking, Molecular Dynamic Simulation, and ADMET Proprieties of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer, ***Life***. 13 (2023) 127. **https://doi.org/10.3390/life13010127.**

13. S. Zarougui, M. Er-rajy, A. Faris, H. Imtara, ***M. El fadili***, O. Al kamaly, S. Zuhair Alshawwa, F.A. Nasr, M. Aloui, M. Elhallaoui, QSAR, DFT studies, docking molecular and simulation dynamic molecular of 2-styrylquinoline derivatives through their anticancer activity, ***Journal of Saudi Chemical Society***. 27 (2023) 101728. **https://doi.org/10.1016/j.jscs.2023.101728.**

## Liste des présentations orales

1. ***M. El fadili*****, "**QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia", ***6th International education and innovative sciences congress***. Burdur Mehmet Akif Ersoy University, Türkiye, November 24-25, 2022

2. ***M. El fadili*****, "**QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia", ***7th International Conference on Pharmaceutics and Advanced Drug Delivery Systems held during***. Essex Partnership University NHS Foundation Trust, United Kingdom, March 27-28, 2023

3. ***M. El fadili*****, "**QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia", ***Première Edition des Jeunes Chercheurs de L'USMBA***, Centre de formation de l'USMBA, December 21-22, 2022

4. ***M. El fadili*****, "**3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-

Selective NMDA Receptor Antagonists", ***Azerbaijan international congress of scientific research.*** Sheki branch of Azerbaijan State Pedagogical University, Azerbaijan/Sheiki, August 16-17, 2023

5. ***M. El fadili*****,** **"**QSAR, ADME-Tox, molecular docking and molecular dynamics simulations of novel selective glycine transporter type 1 inhibitors with memory enhancing properties", ***III. International Architectural Sciences And Applications Symposium.*** University of Naples Federico-II, Journal of Architectural Sciences and Applications and Institute of Economic Development and Social Research, September 14-15, 2023

6. ***M. El fadili*****,** **"**3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists", ***2nd International Cankaya Scientific Studies Congress. Ankaraa, Turkiye,*** September 28-29, 2023

7. ***M. El fadili*****,** **"**3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists", ***2nd International Scientific Research Conference,*** September 26-27, 2023

8. ***M. El fadili*****,** "3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists", ***2nd International Congress TOPKAPI. Istanbul, Turkiye,*** October 20-21, 2023

9. ***M. El fadili*****,** "3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists", ***International Ege Agriculture Congress. Rize, Turkiye,*** November 01-03, 2023

10. ***M. El fadili*****,** "3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists", ***5th International Black Sea Modern Scientific Research. Rize, Turkiye,*** November 08-10, 2023

## Liste des présentations affichées (Poster)

1. ***M. El fadili*****,** **"**QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia", ***6th International Conference on Materials and Environment: 'ICME 6-2022'***, Laboratory of Chemistry-Biology Applied to the Environment "LCBAE, URL-CNRST N°13, in collaboration with the Moroccan Association of Catalysis and Environment "AMCE", Moulay Ismail University, Meknes, Maroc, December 19-20, 2022

## Résumé de thèse

Les récepteurs ionotropiques de type NMDA, constituent une famille de neurotransmetteurs excitateurs majeurs dans le cerveau humain, en raison de leur implication significative dans la transmission synaptique et la plasticité neuronale au sein du système nerveux central (SNC), comme dans le processus d'excitotoxicité responsable de la physiopathologie de diverses maladies telles que la schizophrénie, l'épilepsie, la maladie d'Alzheimer, la douleur neuropathique, les accidents vasculaires cérébraux (AVC), et les perturbations de la mémoire.
La glycine et le glutamate plasmatique agissent en tant qu'agonistes clés, qui se fixent spontanément à différents sites de liaison, pour contrôler l'activité biologique des récepteurs NMDA dans les érythrocytes humains responsables de la survie de la cellule. Cependant, de nombreuses perturbations neurodégénératives se produisent à la suite de leur activation incontrôlée, qui affecte davantage les cellules âgées que les jeunes. Par conséquent, les chimistes et les biologistes travaillent continument à la conception et à l'évaluation de centaines d'inhibiteurs compétitifs et non compétitifs qui affectent tous les récepteurs NMDA disponibles, et d'autres qui agissent de manière sélective et spécifique sur les transporteurs de glycine de type I (GlyT1), grâce aux tests expérimentaux In-vivo et/ou In-vitro, sur lesquels on s'est basé pour réaliser des évaluations In-silico, en exploitant les techniques de modélisation moléculaire dans le but de proposer des modèles 2D et 3D. Ainsi, pour atteindre cet objectif, un certain nombre de descripteurs moléculaires physico-chimiques, topologiques et électroniques ont été déterminés avec les méthodes MM2 et DFT, ainsi que les champs d'interaction moléculaire, afin d'explorer les relations quantitatives potentielles entre la structure moléculaire d'inhibiteurs étudiés et leurs activités biologiques envers des récepteurs NMDA. Ensuite, on a examiné les propriétés physicochimiques et pharmacocinétiques en vue de découvrir des molécules bioactives prédites avec des profils ADMET normalisées à des médicaments-candidats. Le mécanisme d'inhibition des molécules candidates vis-à-vis la cible protéique a été étudié en détectant les types d'interactions intermoléculaires résultant des complexes (ligands-protéine) au moyen de la technique du Docking moléculaire. Enfin, la stabilité des complexes produits au cours du temps de simulation moléculaire estimé en nanosecondes, a été examinée et analysée en utilisant la technique de la dynamique moléculaire couplée aux calculs MM-GBSA.



**Mots clés :** NMDA ; GlyT1 ; Modélisation moléculaire ; ADMET ; Docking moléculaire ; Dynamique moléculaire.

## Thesis summary

Ionotropic receptors of the N-Methyl-D-Aspartate type, represent a family of major excitatory neurotransmitters in the human brain, due to their significant involvement in synaptic transmission and neuronal plasticity within the central nervous system, as well as in the process of excitotoxicity responsible for the pathophysiology of various diseases such as Schizophrenia, Epilepsy, Alzheimer, Neuropathic pain, Strokes, and Memory disorders.

Glycine and plasmatic glutamate act as key agonists, spontaneously fixing to different binding sites, to control the biological activity of NMDA receptors in human erythrocytes responsible for cell survival. However, many neurodegenerative disorders occur as a result of their uncontrolled activation, which affects older cells more than younger ones. Consequently, chemists and biologists are continuously working on the design and evaluation of hundreds of competitive and non-competitive inhibitors that act universally on all NMDA receptors and selectively on the glycine type 1 transporters (GlyT1), through experimental In-vivo and/or In-vitro tests, on which we based to carry out In-silico evaluations, exploiting molecular modeling techniques to propose 2D and 3D models. To achieve this objective, several physicochemical, topological, and electronic molecular descriptors were calculated using MM2 and DFT methods, as well as molecular interaction fields, to explore the potential quantitative relationships between the molecular structure of the studied inhibitors and their biological activities towards NMDA receptors. Next, physicochemical and pharmacokinetic properties were examined to discover predicted bioactive molecules with ADMET profiles normalized to drug candidates. The mechanism of inhibition of the candidate molecules towards the protein target was studied by detecting the types of intermolecular interactions resulting in complexes (ligands-protein) using the molecular docking technique. Finally, the stability of the complexes produced over the molecular simulation time, estimated in nanoseconds, was examined and analyzed using the molecular dynamics technique combined with MM-GBSA calculations.



**Keywords:** NMDA receptors; GlyT1; Molecular modeling; ADMET; Molecular docking; Molecular dynamics.

## ملخص الأطروحة

تشكل المستقبلات الأيونية من صنف NMDA، عائلة من الناقلات العصبية المثيرة والرئيسية في دماغ الإنسان، نظرًا لتأثيرها الكبير على النقل التشابكي واللدونة العصبية في الجهاز العصبي المركزي، وكذلك في عملية السمية المثيرة المسؤولة عن الفيزيولوجيا المرضية لاضطرابات عقلية مختلفة مثل الفصام، الصرع، مرض الزهايمر، آلام الأعصاب، السكتات الدماغية، واضطرابات الذاكرة.

يعمل الجليسين والغلوتامات في البلازما كمنبهات رئيسية، ترتبط تلقائيًا بمواقع ربط مختلفة للتحكم في النشاط البيولوجي لمستقبلات NMDA في كريات الدم الحمراء البشرية المسؤولة عن حياة الخلية، ومع ذلك فإن العديد من الاضطرابات العصبية تحدث نتيجة لنشاطها المضطرب، مما يؤثر بشكل كبير على الخلايا المسنة أكثر من الخلايا الشابة. لذلك يعمل الكيميائيون وعلماء الأحياء بشكل مستمر على تصميم وتقييم مئات المثبطات التنافسية وغير التنافسية والتي من شأنها أن تؤثر على جميع مستقبلات NMDA المتاحة، وأخرى تعمل بشكل انتقائي على ناقلات الجليسين من النوع الأول، وذلك بفضل الاختبارات التجريبية In-vivo و/أو In-vitro، والتي اعتمدنا عليها لإجراء التقييمات السيليكونية In-silico من خلال استغلال تقنيات النمذجة الجزيئية في اقتراح نماذج ثنائية وثلاثية الأبعاد. ولتحقيق هذا المبتغى، قمنا في بادئ الأمر بحساب عدد من الواصفات الجزيئية الفيزيائية والكيميائية، الطوبولوجية، والإلكترونية باستخدام تقنيات MM2 وDFT، بالإضافة إلى حقول التفاعل الجزيئي، لاستكشاف العلاقات الكمية المحتملة بين البنية الجزيئية للمثبطات المدروسة ونشاطها البيولوجي تجاه مستقبلات NMDA. بعد ذلك، تم فحص الواصفات الفيزيائية والكيميائية والحركية الدوائية لاكتشاف الجزيئات النشطة بيولوجيًا المتوقعة ذات ملفات تعريف ADMET مرغوب فيها والتي تشبه الأدوية المرشحة، ثم دراسة آلية تثبيط الجزيئات المرشحة فيما يتعلق بالهدف البروتيني من خلال الكشف عن أنواع التفاعلات بين الجزيئات الناتجة عن المركبات (ligands- بروتين) باستخدام تقنية الالتحام الجزيئي. ختاما، تم فحص وتحليل استقرار المجمعات الناتجة خلال وقت المحاكاة الجزيئية المقدر بالنانو ثانية باستخدام تقنية الديناميكيات الجزيئية المقترنة بحسابات MM-GBSA.



**الكلمات المفتاح :** مستقبلات NMDA ؛ ناقلات الجليسين من النوع الأول ؛ النمذجة الجزيئية ؛ ADMET ؛ الإلتحام الجزيئي ؛ الديناميكيات الجزيئية.

## Liste des figures

Figure 1. Structure de l'acide (2R)-2-(méthyl amino) butane dioïque (NMDA)..... 6
Figure 2. Représentation schématique d'une synapse glutamatergique..... 7
Figure 3. Antagonistes compétitifs du récepteur NMDA..... 8
Figure 4. Antagonistes non compétitifs du récepteur NMDA..... 9
Figure 5. Structure cristalline de la protéine préparée codée 3QEL.pdb, en complexe avec ligand co-cristallisé (partie gauche), et les sites actifs produits en chaine D de la cible protéique (partie droite) ..... 29
Figure 6. Structure cristalline de la protéine préparée codée 4M48.pdb, en complexe avec ligand co-cristallisé (partie gauche), et les sites actifs produits en chaine A de la cible protéique (partie droite) ..... 30
Figure 7. Structural formula of isoxazole (1–6), piperidine linker moieties (6–14) and phenylpiperazine (12, 15–32) derivatives ..... 42
Figure 8. 3D structure of the molecule 22 as template (left), 32 aligned and superposed molecules on the template (right) and the common core (middle). ..... 47
Figure 9. Std*coeff. contour maps corresponding to steric (A), electrostatic (B), hydrogen bond acceptor (C) and hydrogen bond donor (D) fields explained for the compound 22 as a template ..... 50
Figure 10. BOILED-Egg predictive model of the co-crystallized ligand and the thirty-two inhibitors..... 51
Figure 11. Experimental pose view of ifenprodil with the protein's active sites ..... 53
Figure 12. 3D (a) and 2D (b) intermolecular interactions established between the targeted protein and docked ifenprodil with a binding energy equal to −5.45 kcal/mol, and redocking pose (c) with a RMSD value of 0.395 Å (original ifenprodil in cyan, and docked ifenprodil in green)..... 54
Figure 13. 3D (left) and 2D (right) visualization of intermolecular interactions between L1, L2 and L3 ligands and the transport protein encoded as 3QEL, corresponding to the following binding energies in kcal/mol: −6.45, −5.57 and −6.32, respectively ..... 56
Figure 14. Structural formula of 3-ethylisobenzofuran-1(3H)-one derivatives. ..... 66
Figure 15. BOILED-Egg model of thirty inhibitors and ifenprodil as co-crystallized ligand. 70
Figure 16. Bioavailability radars of ifenprodil and thirty compounds, taking into account six physicochemical properties ideal for oral bioavailability, namely lipophilicity (LIPO), polarity (POLAR), size (SIZE), solubility (INSOLU), saturation (INSATU), and flexibility (FLEX). 76
Figure 17. 2D and 3D visualizations of intermolecular interactions established between 3QEL.pdb protein and C22, C13, and C23 ligands, with binding energy of -7.27 Kcal/mol, -7.89 Kcal/mol, and -9.17 Kcal/mol, respectively. ..... 79
Figure 18. Changes in root mean square deviation (RMSD) and root mean square fluctuation (RMSF) during 200 nanoseconds of molecular dynamics (MD) simulation time for C22, C13, and C23 ligands complexed to 3QEL.pdb protein. ..... 81
Figure 19. Variations of root mean square deviation (RMSD), radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA) and polar surface area (PSA) during 200 ns of MD simulation time for C22, C13, and C23 ligands complexed to 3QEL.pdb protein. ..... 82
Figure 20. Frontier molecular orbitals (FMO) of the most active compounds C22 (A), C13 (B), and C23 (C) ..... 84
Figure 21. Data visualization on the first two principal components..... 94

Figure 22. Influence degree of the descriptors on the biological activity. ..... 97
Figure 23. William's diagram of the MLR model established by Equation (1). 1 and 32 are outliers in the test set and 33 is a non-aberrant molecule in the training test..... 99
Figure 24. Correlation between the observed and predicted activities using MLR technique. ..... 100
Figure 25. Correlation between the observed and predicted activities using MNLR technique. ..... 100
Figure 26. BOILED-egg predictive model of the most active ligands..... 106
Figure 27. Experimental pose view of Nortriptyline with the protein's active sites..... 107
Figure 28. 2D and 3D visualization of intermolecular interactions between DAT (PDB code: 4M48) and the more active ligands (L6, L9, L30, L31 and L37), with binding energies of −10.74 kcal/mol, −10.22 kcal/mol, −8.46 kcal/mol, −8.78 kcal/mol and −8.59 kcal/mol, respectively. ..... 109
Figure 29. Re-docking pose with RMSD equal to 0.022 Å (original nortriptyline of cyan color, and docked nortriptyline of green color). ..... 110
Figure 30. 2D (a) and 3D (b) visualization of intermolecular interaction between the docked nortriptyline and the protein target (binding energy of −9.11 kcal/mol). ..... 110
Figure 31. RMSD and RMSF graphs for the ligand 9 complexed with the dopamine transporter membrane protein during 100 ns. ..... 111
Figure 32. Rg, MolSA, SASA and PSA during 100 ns of MD simulation, and the variation of total free energy for the ligand 9 complexed with DAT protein. ..... 113
Figure 33. Impact of molecular descriptors on the biological activity..... 130
Figure 34. William's diagram of the predictive QSAR models ..... 131
Figure 35. Correlation among the predicted and observed activities using multiple linear regression (MLR) ..... 132
Figure 36. Correlation among the predicted and observed activities using multiple nonlinear regression (MNLR) ..... 133
Figure 37. BOILED-Egg model of nortriptyline and forty pyrrolo[3,4-c]pyrazole and azetidine derivative ..... 136
Figure 38. Active sites of co-crystalized ligand bound to 4M48.pdb protein in A-chain ..... 138
Figure 39. 2D (left) and 3D (right) inter-molecular interactions, produced between L28 and L30 ligands and DAT protein of 4M48.pdb code with binding energies of −8.72 kcal/mol and −8.70 kcal/mol, respectively. ..... 139
Figure 40. 2D (A) and 3D (B) visualization of intermolecular interactions created between the docked nortriptyline and 4M48.pdb protein with a binding energy of −8.36 kcal/mol, and re-docking pose (c) defined by a RMSD value of 0.552 Å (docked nortriptyline in cyan and co-crystallized nortriptyline colored by the atomic elements) ..... 140
Figure 41. Evolution of RMSD. RMSF. Rg. MolSA. SASA. PSA and total free energy during 50 ns. for (L28-DAT protein) complex ..... 142
Figure 42. Evolution of RMSD, RMSF, Rg, MolSA, SASA, PSA, and total free energy during 50 ns, for (L30-DAT protein) complex ..... 143
Figure 43. Impact of selected descriptors on human GluT1 activity ..... 157
Figure 44. Quantitative correlation between observed and predicted GluT1 activity using MLR technique ..... 159
Figure 45. Quantitative correlation between observed and predicted GluT1 activity using MNLR technique ..... 159

Figure 46. Predictive model of BOILED-Egg for 3,4-disubstituted pyrrolidine sulfonamides ..... 163
Figure 47. 2D (left) and 3D (right) intermolecular interactions, produced for (C19 ligand - DAT protein) complex with binding energy of -8.96 Kcal/mol. ..... 166
Figure 48. A: Active sites of co-crystallized ligand of DAT protein in A-chain. B: 2D intermolecular interactions produced for the (docked nortriptyline - DAT protein) complex with a binding energy of −9.02 kcal/mol. C: Re-docking pose with a RMSD value of 0.128 Å (co-crystallized nortriptyline colored in cyan superposed on docked nortriptyline) ..... 167
Figure 49. Changes of Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) during 100 ns, for the dopamine transporter protein complexed with C19 ligand. ..... 168
Figure 50. Changes in radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA), and polar surface area (PSA), over 100 ns of molecular dynamics simulation time ..... 168
Figure 51. Histogram of interaction fractions between DAT protein and C19 ligand (left), and the timeline depiction of intermolecular interactions and chemical contacts of candidate ligand in the active sites of targeted protein (right) ..... 169

## Liste des tableaux

Table 1. Critères statistiques de Golbreikh et Tropsha ..... 26
Table 2 . Studied compounds and their observed activities of pIC50 order ..... 42
Table 3. Statistical results of CoMFA and CoMSIA models in various molecular field combinations using partial least squares analysis ..... 45
Table 4. Results of external validation corresponding to CoMFA and CoMSIA/EDA models. ..... 46
Table 5. Physicochemical parameters prediction of ifenprodil, L1, L2 and L3 ligands on the basis of Lipinski, Veber, Egan, Muegge and Ghose regulations ..... 52
Table 6. ADMET properties prediction of ifenprodil, L1, L2 and L3 ligands ..... 52
Table 7. Neuroprotection of thirty chemical compounds against NMDA-induced neurotoxicity in hippocampal neurons at 10 µM. ..... 67
Table 8. Physicochemical properties of ifenprodil and thirty GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists, based on Lipinski, Veber, Egan, and Ghose violations. ..... 71
Table 9. ADMET in-silico pharmacokinetic properties of ifenprodil and thirty GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists. ..... 72
Table 10. Binding energies in Kcal/mol of the studied complexes using MMGBSA approach. ..... 79
Table 11. DFT Calculations for the candidate drugs ..... 85
Table 12. Values of selected descriptors for 44 molecules. ..... 95
Table 13. Significance test of the slopes. ..... 97
Table 14. Variance analysis. ..... 97
Table 15. External validation results of the MLR and MNLR models. ..... 99
Table 16. Observed and predicted activity values from the QSAR models. ..... 101
Table 17. Y-randomization test results. ..... 102
Table 18. Golbraikh and Tropsha statistical criteria to validate the designed QSAR model. 103
Table 19. Prediction of the physicochemical properties of nortriptyline and more active ligands, based on Lipinski, Veber, Egan and Ghose violations. ..... 105

Table 20. Prediction of ADMET pharmacokinetic properties of nortriptyline and more active ligands. ..... 105
Table 21. Significant weights of the selected descriptors ..... 129
Table 22. Analysis of variance using the fisher test ..... 130
Table 23. Observed and predicted activities of pIC50 order using MLR and MNLR QSAR models ..... 132
Table 24. Observed and predicted activities using MLR. MNLR and CVLOO techniques .. 133
Table 25. Tropsha and Golbraikh conditions to examine the accuracy of generated MLR QSAR model ..... 135
Table 26. Prediction of physico-chemical parameters of L28, L30, and nortriptyline ligands based on Lipinski, Egan, Veber, Ghose, and Muegge regulations ..... 137
Table 27. ADMET-Tox pharmacokinetic properties of L28, L30, and Nortriptyline ligands 137
Table 28. Significant weights of the selected descriptors ..... 156
Table 29. Analysis of variance using the fisher test ..... 157
Table 30. Observed and predicted activities of pKi order using MLR and MNLR models... 158
Table 31. Observed and predicted activities using MLR, MNLR and CVLOO techniques .. 160
Table 32. Tropsha and Golbraikh conditions to examine the accuracy of generated MLR QSAR model ..... 162
Table 33. Prediction of physico-chemical parameters of central nervous system agents ..... 164
Table 34. Prediction of ADMET-Toxicity pharmacokinetic properties of CNS agents ..... 164
Table 35. Binding energies in Kcal/mol of the studied complex using MMGBSA approach. ..... 170

## Liste des abréviations

| | |
|---|---|
| ADMET | Absorption, distribution, metabolism, excretion, and toxicity |
| BBB | Blood-brain barrier |
| B3LYP | Becke 3-Parameter Lee-Yang-Parr |
| CNS | Central nervous system |
| CoMFA | Comparative molecular field analysis |
| CoMSIA | Comparative molecular similarity index analysis |
| CV | Cross validation |
| DA | Domaine of applicability |
| DFT | Density functional theory |
| E-B | Bend energy |
| E-d,d | Dipol-dipol energy |
| E-GAP | Gap energy |
| E-S | Strech energy |
| E-S,B | Strech-bend energy |
| E-T | Torsion energy |
| E-VDW | Van der Waals energy |
| FDA | Food and drug administration |
| GlyT1 | Glycine transporter type 1 |
| HOMO | Highest occupied molecular orbital |
| LOO | Leave-one-out |
| LUMO | Lowest unoccupied molecular orbital |
| MD | Molecular dynamics |
| MLR | Multiple linear regression |
| MM-GBSA | Molecular mechanics with generalized born and surface area |
| MNLR | Multiple nonlinear regression |
| MolSA | Molecular surface area |
| MR | Molar refractivity |
| MV | Molar volume |
| MW | Molecular weight |
| NHA | Number of Hydrogen-bond acceptors |
| NHD | Number of Hydrogen-bond donors |

| | |
|---|---|
| NMDAR | N-methyl-D-aspartate receptor |
| NRB | Number of rotatable bonds |
| OECD | Organization for economic cooperation and development |
| PCA | Principal components analysis |
| PDB | Protein data bank |
| pH | Potentiel of Hydrogen |
| PSA | Polar surface area |
| QSAR | Quantitative structure-activity relationship |
| RMSD | Root mean square deviation |
| RMSE | Root mean square error |
| RMSF | Root mean square fluctuation |
| SASA | Solvent accessible surface area |

## Liste des symboles

| | |
|---|---|
| Æ | Polarizability |
| d | Electronic density |
| J | Balban index |
| Log P | Partition coefficient |
| n | Index of refraction |
| P | Parachor |
| χ | Electronegativity |
| η | Hardness |
| ɣ | Surface tension |
| ω | Electrophilicity |
| %Br | Bromine mass percentage |
| %C | Carbon mass percentage |
| %Cl | Chlorine mass percentage |
| %F | Fluorine mass percentage |
| %H | Hydrogen mass percentage |
| %N | Nitrogen mass percentage |
| %O | Oxygen mass percentage |
| %S | Sulfur mass percentage |

## Table de matières

Introduction générale ..... 1
Chapitre I : Généralités sur les récepteurs NMDA et leur effet thérapeutique sur les maladies neurologiques ..... 5
1. Introduction ..... 6
2. Notion de NMDA ..... 6
3. Mécanisme d'activation ..... 7
4. Inhibiteurs des récepteurs NMDA ..... 8
4.1. Inhibiteurs compétitifs ..... 8
4.2. Inhibiteurs non compétitifs ..... 9
5. Activités biologiques étudiées ..... 9
5.1. Activité analgésique contre la douleur neuropathique ..... 10
5.2. Activité inhibitrice du transporteur de glycine de type 1 ..... 10
Références ..... 11
Chapitre II : Approches de base concernant les techniques In-silico destinées à la modélisation moléculaire ..... 14
1. Introduction ..... 15
2. Relations quantitatives structure-activité ..... 15
2.1. Définition ..... 15
2.2. Principe ..... 15
2.3. Objectif ..... 15
2.4. Méthodologie suivie dans l'étude 2D/3D-QSAR ..... 16
2.5. Méthodes de calcul théorique ..... 18
2.5.1. Théorie de la fonctionnelle de la densité ..... 18
2.5.2. Théorie de la mécanique moléculaire ..... 19
2.6. Descripteurs moléculaires ..... 19
2.6.1. Généralités ..... 19
2.6.2. Types des descripteurs ..... 20
2.6.2.1. Descripteurs 1D ..... 20
2.6.2.2. Descripteurs 2D ..... 20
2.6.2.3. Descripteurs 3D ..... 21
2.6.2.4. Descripteurs 4D ..... 21
2.7. Méthodes statistiques ..... 21
2.7.1. Analyse en composantes principales ..... 21
2.7.2. Régression linéaire multiple ..... 21
2.7.3. Régression non linéaire multiple ..... 22
2.8. Techniques de validation ..... 22
2.8.1. Test de de Fisher ..... 22
2.8.2. Test de Student ..... 23

2.8.3. Validation interne ..... 23
2.8.4. Test de Y-randomisation ..... 24
2.8.5. Validation externe ..... 24
2.8.6. Domaine d'applicabilité ..... 25
2.8.7. Critères statistiques de Golbreikh et Tropsha ..... 25
3. Prédiction des propriétés physicochimiques et pharmacocinétiques ADME-Toxicité . 26
4. Docking moléculaire ..... 27
4.1. Généralités ..... 27
4.2. Principe ..... 28
4.3. Protocole de validation ..... 28
4.4. Objectif ..... 28
4.5. Cibles protéiques ..... 28
5. Dynamique moléculaire ..... 30
5.1. Définition ..... 30
5.2. Principe ..... 31
5.3. Objectif ..... 31
6. Serveurs, plates-formes, et logiciels utilisés ..... 31
Références ..... 33
Chapitre III. Modélisation moléculaire d'inhibiteurs sélectifs du récepteur NMDA, exerçant un effet analgésique contre la douleur neuropathique ..... 38
Résumé ..... 39
1. Introduction ..... 41
2. Materials and Methods ..... 42
2.1. Experimental Database ..... 42
2.2. Optimisation and Alignment ..... 44
2.3. Development of 3D-QSAR Models ..... 45
2.4. Partial Least Squares (PLS) Analysis ..... 45
2.5. ADMET in Silico Pharmacocinetics ..... 46
2.6. Molecular Docking Modeling ..... 46
3. Results and Discussion ..... 47
3.1. Molecular Alignement ..... 47
3.2. 3D-QSAR Analysis ..... 47
3.3. A Graphical Explanation of CoMSIA/EDA and CoMFA Models ..... 48
3.4. ADMET in Silico Pharmacocinetics ..... 50
3.5. Molecular Docking ..... 52
3.6. Docking Validation Protocol ..... 53
3.7. Molecular Docking Modeling for L1, L2 and L3 Ligands ..... 54
4. Conclusion ..... 56
References ..... 57

Chapitre IV. Étude in-silico de nouveaux antagonistes potentiels du récepteur NMDA pénétrant dans le cerveau en tant qu'agents thérapeutiques contre les accidents vasculaires cérébraux (AVC) ..... 62
Résumé ..... 63
1. Introduction ..... 65
2. Materials and Methods ..... 66
2.1. Experimental database ..... 66
2.2. In-silico pharmacokinetics, Drug-Likeliness, and ADME-Toxicity Prediction ..... 68
2.3. Molecular Docking ..... 68
2.4. Molecular dynamics ..... 69
3. Results and discussion ..... 70
3.1. Drug Kinetics Assessment, Drug-likeliness, and ADME-Toxicity Profile Analysis 70
3.2. Molecular Docking Simulation ..... 76
3.3. Molecular Dynamics Simulation ..... 79
3.4. Chemical reactivity using DFT calculations ..... 82
4. Conclusion ..... 85
References ..... 86
Chapitre V. Modélisation moléculaire de nouveaux bicyclo (aryl methyl) benzamides en tant qu'inhibiteurs puissants de GlyT1 destinés au traitement de la schizophrénie ..... 90
Résumé ..... 91
1. Introduction ..... 93
2. Results and Discussion ..... 94
2.1. Pricipal Component Analysis ..... 94
2.2. Statistical Database ..... 95
2.3. Multiple Linear Regression ..... 95
2.4. Multiple Non-Linear Regression ..... 98
2.5. QSAR Model Validation ..... 98
2.5.1. Applicability Domain ..... 98
2.5.2. External Validation ..... 99
2.5.3. Internal Validation ..... 100
2.5.4. Validation Using Y-Randomisation Test ..... 102
2.5.5. Golbreikh and Tropsha Criteria ..... 103
2.6. In Silico Pharmacokinetics ADMET Prediction ..... 104
2.7. Molecular Docking ..... 106
2.8. Docking Validation Protocol ..... 109
2.9. Molecular Dynamics Simulations ..... 110
3. Materials and Methods ..... 113
3.1. Database ..... 113
3.2. Molecular Descriptors Calculation ..... 113

3.3. Statistical Methods ..... 114
3.4. Drug Likeness and In Silico Pharmacokinetics ADMET Prediction ..... 114
3.5. Molecular Docking Modeling ..... 114
3.6. Molecular Dynamics ..... 115
4. Conclusion ..... 115
References ..... 117
Chapitre VI. Modélisation moléculaire d'inhibiteurs sélectifs du transporteur de glycine type 1, destinés pour le traitement des perturbations de la mémoire ..... 122
Résumé ..... 123
1. Introduction ..... 125
2. Materials and methods ..... 126
2.1. Experimental Database ..... 126
2.2. Molecular Descriptors Calculation ..... 126
2.3. Statistical Methods ..... 126
2.4. In silico pharmacokinetics, drug-likeliness, and ADMET depiction exploration 127
2.5. Molecular Docking Simulation ..... 127
2.6. Molecular Dynamic ..... 127
3. Results and discussion ..... 128
3.1. Principal Component Analysis ..... 128
3.2. Experimental Database ..... 128
3.3. Multiple Linear Regression ..... 128
3.4. Multiple Non-Linear Regression ..... 130
3.5. Validation of predictive QSAR Models ..... 131
3.5.1. Applicability Domain ..... 131
3.5.2. External Validation ..... 131
3.5.3. Internal Validation ..... 133
3.5.4. Validation Using Y-Randomization Test ..... 134
3.5.5. Statistical Criteria of Golbreikh and Tropsha's Study ..... 135
3.6. In Silico Pharmacokinetics ADMET Prediction ..... 135
3.7. Molecular Docking ..... 137
3.8. Docking Validation Protocol ..... 139
3.9. Molecular Dynamic simulation ..... 140
4. Conclusion ..... 143
References ..... 145
Chapitre VII. Étude in-silico d’inhibiteurs sélectifs et compétitifs du transporteur de glycine type 1 destinés au traitement des perturbations associés à l’hypofonctionnement du récepteur NMDA ..... 150
Résumé ..... 151
1. Introduction ..... 153
2. Materials and methods ..... 154

2.1. Experimental Database ..... 154
2.2. Molecular Descriptors Calculation ..... 154
2.3. Statistical Methods ..... 154
2.4. In-silico pharmacokinetics, drug-likeliness, and ADMET depiction exploration 155
2.5. Molecular Docking Simulation ..... 155
2.6. Molecular Dynamic ..... 155
3. Results and discussion ..... 156
3.1. Quantitative Structure-Activity Relationships (QSARs) ..... 156
3.1.1. Principal Components Analysis ..... 156
3.1.2. Multiple Linear Regression ..... 156
3.2. Multiple Non-Linear Regression ..... 157
3.3. Validation of predictive QSAR Models ..... 158
3.3.1. External Validation ..... 158
3.3.2. Internal Validation ..... 159
3.3.3. Validation Using Y-Randomisation Test ..... 162
3.3.4. Statistical Criterias of Golbreikh and Tropsha's study ..... 162
3.4. In-silico Pharmacokinetics ADMET Prediction ..... 163
3.5. Molecular Docking ..... 165
3.6. Docking Validation Protocol ..... 166
3.7. Molecular Dynamic simulation ..... 167
4. Conclusion ..... 170
Références ..... 171
Conclusion générale ..... 177
Annexes ..... 183

# Introduction générale

Le secteur pharmaceutique s'est considérablement développé au cours des dernières décennies grâce à l'évolution remarquable et transitoire des dispositifs expérimentaux vers des outils computationnels destinés à la conception de nouveaux médicaments. Même si l'expérience est forte et solide en termes de résultats scientifiques obtenus, énormément de temps et d'argent ont été perdus sans atteindre les objectifs souhaités, par ce que les tests expérimentaux constituent une arme à double tranchant, puisqu'elles peuvent être couronnées de succès à un moment donné, et vouées à l'échec à un autre. Cependant, l'expérience ne peut jamais être remplacée, car les technologies basées sur l'ordinateur ne font que des prédictions concernant la découverte des médicaments efficaces, plus actives, et non toxiques, sur la base des molécules précédemment synthétisées, plutôt que de tester au hasard des centaines de molécules, de sorte à minimiser au maximum possible le temps perdu au niveau de laboratoires pour synthétiser ou examiner une molécule donnée, en plus pour économiser les coûts massifs gaspillés à l'échelle industriel lors de la production. Par conséquent, l'expérience est le premier et le dernier examen des candidats-médicaments, puisque les molécules recommandées par des prédictions In-silico, et validées par l'ordinateur, doivent être évaluée au stade final par les tests expérimentaux In-vivo et/ou In-vitro, avant de procéder à la production et ensuite à la distribution des médicaments aux êtres humains et/ou animaux.

En effet, la chimie computationnelle revêt une énorme importance dans la conception de médicaments assistée par ordinateur, grâce au développement continu de nouvelles technologies par des ingénieurs, chimistes, et programmeurs à l'échelle national et international, ce qui nous a fourni un terrain fertile pour la mise en œuvre de la présente thèse, tout en appliquant des techniques scientifiques informatisées innovantes et récentes en termes de la modélisation moléculaire, à savoir les études bidimensionnels et tridimensionnels des relations quantitatives probables entre la structure chimique d'inhibiteurs étudiés et leur activité biologique (2D/3D-QSAR), à travers lesquelles nous avons appliqué des méthodes purement statistiques sur une base de données expérimentales, comme l'analyse en composantes principales (ACP), la régression linéaire multiple (MLR), la régression non linéaire multiple (MNLR), la régression aux moindres carrés partiels (PLSR), ce qui nous a permis par la suite de générer des modèles mathématiques validés, robustes, fiables, et applicables sur nouvelles molécules, ayant le potentiel de prédire des activités biologiques des composés chimiques synthétisés ou non encore synthétisés sur la base des descripteurs moléculaires unidimensionnels (1D), bidimensionnels (2D), tridimensionnels (3D), et quadridimensionnels (4D), qui ont été calculés pour justifier une activité biologique donnée (analgésique, antidiabétique, antifongique, anticancéreuse, antibactérienne …).

Ensuite, vient en deuxième intention l'étude prédictive des propriétés physico-chimiques et pharmacocinétiques sur la base des cinq règles de Lipinski, et nombres de violations désirées, assumées par Veber, Egan, Muegge, et Ghose, qui se combinent pour prédire de très bons profils ADME-Tox adaptés à un nombre précis d'inhibiteurs, afin de garantir leur aptitude à traverser la barrière hémato-encéphalique, sans engendrer d'effets secondaires indésirables tels que la toxicité hépatique, l'allergie cutanée et la toxicité de l'AMES, ce qui en fait des agents puissants du système nerveux central (SNC), et donc des cibles thérapeutiques pour une classe de troubles neurologiques associés au système nerveux central.

A la phase suivante, viennent les simulations du Docking moléculaire, pour découvrir le mécanisme d'inhibition des molécules prédites au préalable par un bon profil ADME-Tox, tout en indiquant la nature des interactions intermoléculaires produites entre les ligands actifs (petites molécules) vis-à-vis les cibles protéiques (macromolécules). Par le biais de visualisations bidimensionnelles (2D) et tridimensionnelles (3D), on entend vérifier que les molécules candidates sont dockées dans les sites actifs de la protéine, avec un minimum d'énergie de liaison en Kcal/mol, afin d'assurer la stabilité moléculaire d'inhibiteurs examinés.

En dernière phase, la technique de la dynamique moléculaire a été appliquée avec une résolution temporelle extrêmement détaillée, et une échelle atomique minutieuse, afin d'évaluer la stabilité du complexe résultant (ligand-protéine), tout au long du temps de simulation moléculaire, estimé en nanosecondes. Par ailleurs, les changements conformationnels en déviation et fluctuation du complexe, ainsi que les propriétés du ligand actif lors de sa réaction avec la cible protéique ont été bien analysés, et finalement les énergies de liaison du complexe résultant ont été calculées pendant le temps de simulation de la dynamique moléculaire, grâce aux approches MM-GBSA.

Les travaux de recherche présentés dans cette thèse sont structurés en sept chapitres :

❖ Le premier chapitre est consacré spécifiquement à une recherche bibliographique portant sur le récepteur NMDA, en commençant par sa nature chimique et biologique, passant par son mécanisme d'activation grâce aux deux agonistes principaux ; la glycine et le glutamate, ainsi que les problèmes résultant de cette activation, révélant la nécessité absolue du développement de nouveaux inhibiteurs compétitives ou non compétitifs, ayant le potentiel d'améliorer le fonctionnement des récepteurs NMDA, en produisant des effets thérapeutiques contre les maladies neurodégénératives liées au système nerveux central, en particulier la schizophrénie, les accidents vasculaires cérébraux (AVC), et les perturbations de la mémoire.

❖ Le deuxième chapitre présente une recherche bibliographique sur les techniques In-silico appliquées sur la base de la modélisation moléculaire, entre autres 2D/3D-QSAR, puis les prédictions des propriétés physico-chimiques et pharmacocinétiques ADME-Toxicité, et finalement les simulations du Docking moléculaire, et celles de la dynamique moléculaire.

❖ Le troisième chapitre est dédié à l'étude de nouveaux antagonistes sélectifs du récepteur NMDA, étudiés pour leurs activités analgésiques contre la douleur neuropathique.

❖ Le quatrième chapitre porte sur l'étude in-silico de nouveaux antagonistes potentiels du récepteur NMDA pénétrant dans le cerveau en tant qu'agents thérapeutiques contre les accidents vasculaires cérébraux.

❖ En ce qui concerne le cinquième chapitre, il est consacré tout particulièrement à l'étude de nouveaux bicyclo (aryl methyl) benzamides en tant qu'inhibiteurs puissants de GlyT1destinés au traitement de la schizophrénie en assurant la régulation de la fonction des récepteurs NMDA.

❖ Le sixième chapitre porte sur l'étude d'inhibiteurs compétitifs du transporteur de glycine type 1, destinés au traitement des perturbations de la mémoire générées par l'hypofonctionnement des récepteurs NMDA.

❖ Dans le dernier chapitre, on a mené une étude In-silico sur une famille de trente dérivés de pyrrolidine sulfonamides 3,4-disubstitués afin d'identifier un médicament candidat pour traiter la schizophrénie et les déficiences associées à l'hypofonctionnement des récepteurs NMDA, grâce à l'inhibition sélective et compétitive du transporteur de glycine type 1.

# Chapitre I : Généralités sur les récepteurs NMDA et leur effet thérapeutique sur les maladies neurologiques

## 1. Introduction

De nombreux troubles neurodégénératifs et cognitifs résultent d'un déficit structurel et fonctionnel des récepteurs NMDA au niveau des synapses excitatrices centrales. Cet hypofonctionnement est essentiellement dû à une réduction des sites de liaison au ligand des récepteurs NMDA, qui peut entraîner par la suite une diminution de leurs performances avec l'âge. C'est pourquoi des récentes études se sont orientées vers le développement d'inhibiteurs puissants pour le traitement des déficiences issues d'une activation incontrôlée des récepteurs NMDA.

## 2. Notion de NMDA

N-Méthyl-D-Aspartate (NMDA) est un récepteur de l'acide glutamique dont la formule brute est $C_5H_9NO_4$. Il est connu par l'acide (2R)-2-(méthyl amino)butane dioïque selon l'union international de chimie pure et appliquée (UICPA), comme il est structuré dans la Figure 1.

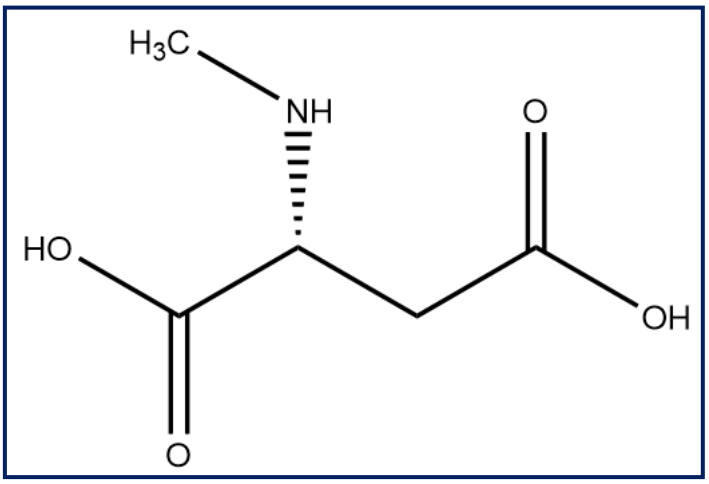



**Figure 1. Structure de l'acide (2R)-2-(méthyl amino) butane dioïque (NMDA)**

Les récepteurs glutamatergiques de type N-méthyl-D-aspartate (NMDA) sont des canaux ioniques qui jouent un rôle crucial dans le système nerveux central (SNC) des vertèbres [1], car ils sont impliqués dans la transmission synaptique et la plasticité neuronale dans tout le SNC, et ils contribuent également à l'activité épileptiforme et à la mort des cellules neuronales dans un certain nombre de conditions pathologiques [2]. Pour cela, Ils sont considérés comme des cibles thérapeutiques de haut niveau dans le traitement de la douleur et des maladies neurodégénératives telles que la maladie d'Alzheimer, Huntington, Parkinson, Epilepsie, Ischémie cérébrale et de Schizophrénie [3]. Ce type des récepteurs se compose de deux familles principales de sous-unités; les sous-unités NR1 liées à la glycine (NR1a-h) [4], et les sous-unités NR2 attachées au glutamate (NR2A-D) [5–8], avec quatre sites de liaison bien identifiées, un site central pour les anesthésiques dissociatifs, dont la Kétamine et le MK-801 qui réalise un blocage de longue durée, deux sites pour l'acide Glutamique, et un site sur la sous

unité NR2B sur lequel se fixe des ligands endogènes comme les polyamines [9]. Les récepteurs NMDA contenant GluN2B sont fortement impliqués dans des maladies neurodégénératives comme l'épilepsie, l'ischémie cérébrale et la schizophrénie [10].

### 3. Mécanisme d'activation

Les récepteurs NMDA, comme d'autres familles des récepteurs ionotropiques (AMPA, et Kaïnate) sont responsables de la transmission synaptique excitatrice rapide, agissant à l'échelle de la milliseconde au niveau des synapses du SNC, comme il est montré dans la Figure 2. Leur activation nécessite la liaison de deux agonistes différentes, le glutamate et la glycine (ou D-sérine), ce qui confère une perméabilité importante aux cations divalents de calcium ($Ca^{2+}$) et ceux monovalents de sodium et de potassium ($Na^+$ et $K^+$) [11]. Cette activation génère une pré-dépolarisation de la membrane cellulaire qui est nécessaire à l'ouverture du canal ionique, ce qui contrôle le flux des ions de calcium, de sodium et de potassium sur la membrane des cellules nerveuses [3]. Ces ions libérés sont à la base des phénomènes de plasticité synaptique les plus répandus, dont la potentialisation à long terme (PLT) et la dépression à long terme (DLT), qui représentent les mécanismes cellulaires responsables de mémoire et d'apprentissage [1,12]. Toute modification de l'activité de ce récepteur par un agoniste ou un antagoniste provoque une altération du flux ionique à travers le canal. Par conséquent, l'activation incontrôlée du récepteur entraîne une accumulation transitoire de $Ca^{2+}$ intracellulaire induisant une contraction cellulaire et une modification du potentiel d'Hydrogène (pH) intracellulaire [13].

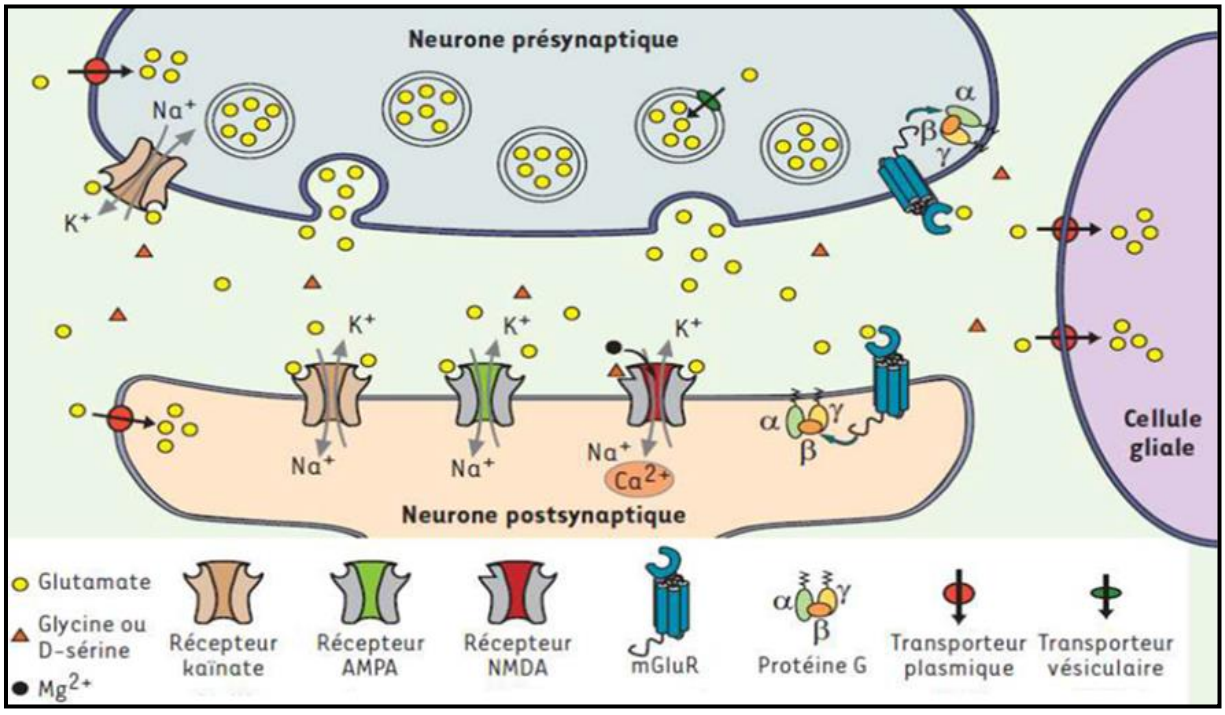



**Figure 2. Représentation schématique d'une synapse glutamatergique**

## 4. Inhibiteurs des récepteurs NMDA

Les antagonistes du récepteur N-méthyl-D-aspartique (anti-NMDA) constituent une classe de médicaments qui sert à antagoniser ou inhiber l'action du récepteur N-méthyl-D-aspartique (R-NMDA). Ils sont couramment utilisés comme anesthésiques pour les animaux et les humains [10]. Les dernières études démontrent que les antagonistes du récepteur NMDA peuvent améliorer les symptômes de l'anhédonie, de l'anxiété et des déficits cognitifs dans un modèle de dépression chez les rats [14]. Deux familles d'inhibiteurs des récepteurs NMDA peuvent agir de manière compétitive ou non compétitive sur les sites de liaison du glutamate et celui de glycine, comme des agonistes principaux du canal ionique.

### 4.1. Inhibiteurs compétitifs

Les inhibiteurs compétitifs du récepteur NMDA sont fréquemment utilisés comme antagonistes généraux et non sélectifs, car ils agissent sur tous les récepteurs NMDA disponibles [2,15]. Les premières études expérimentales ont montré que la plupart des antagonistes compétitifs des récepteurs NMDA, produisent au mieux une faible excitation comportementale lorsqu'ils sont administrés par voie systémique, cependant ils indiquent cliniquement un risque minimal d'effets secondaires psychologiques par rapport les antagonistes non compétitifs [16]. Nous citons à titre d'exemple l'acide (2R)-2-amino-5-phosphonopentanoïque (APV), l'acide (+)-cis-4-phosphonométhyl-2-pipéridine carboxylique (CGS 19755), l'acide 2-amino-3-(2-(2 phosphonoéthyl)cyclohexyl)propanoïque (SDZ 215-439), et l'acide 6-(phosphonométhyl)décahydroisoquinoline-3-carboxylique (LY 2746 14) [17], comme schématisés respectivement dans la Figure 3.

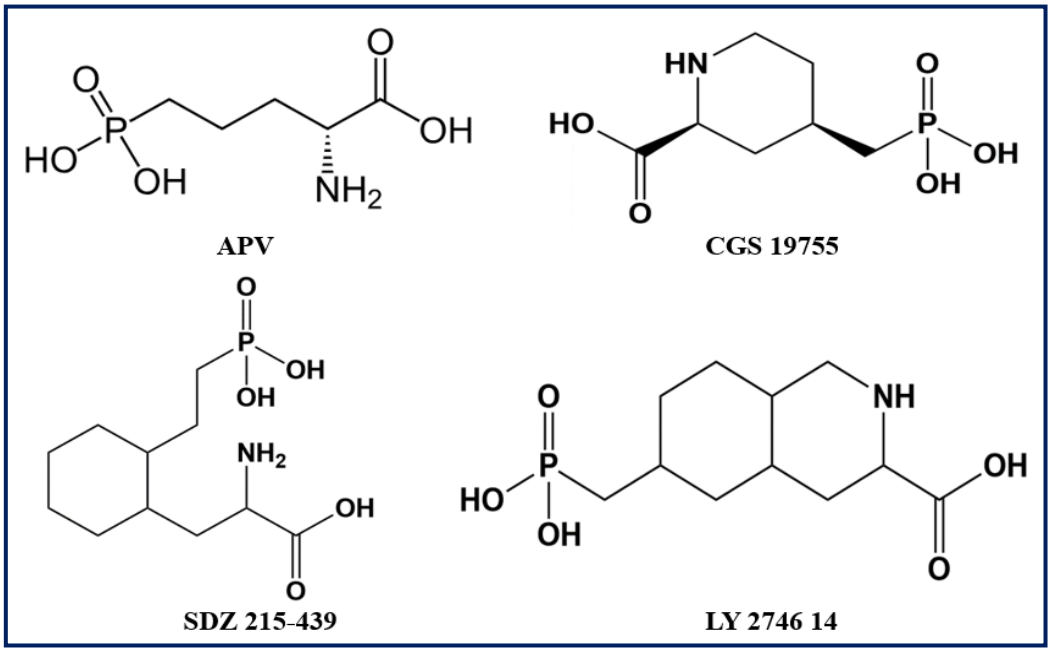



**Figure 3. Antagonistes compétitifs du récepteur NMDA**

### 4.2. Inhibiteurs non compétitifs

Les antagonistes non compétitifs des récepteurs NMDA sont des bloqueurs de canaux ouverts. Par conséquent, ces antagonistes bloquent les récepteurs NMDA actifs [15]. Le blocage du canal NMDA par des antagonistes non compétitifs permet d'inhiber ou bien de retarder l'influx postsynaptique des cations $Ca^{2+}$, qui pourrait à son tour de retarder et inhiber la dépolarisation neuronale de longue durée [18]. Les antagonistes non compétitifs du récepteur N-méthyl-D-aspartate (R-NMDA) imitent les symptômes de la schizophrénie et produisent des effets antidépresseurs persistants [19]. Certains de ces antagonistes, tels que la dizocilpine (MK-801), la mémantine, la kétamine, et la phencyclidine (PCP), sont des drogues récréatives utilisées pour leurs propriétés dissociatives, hallucinogènes et euphorisantes [20], comme schématisés dans la Figure 4.

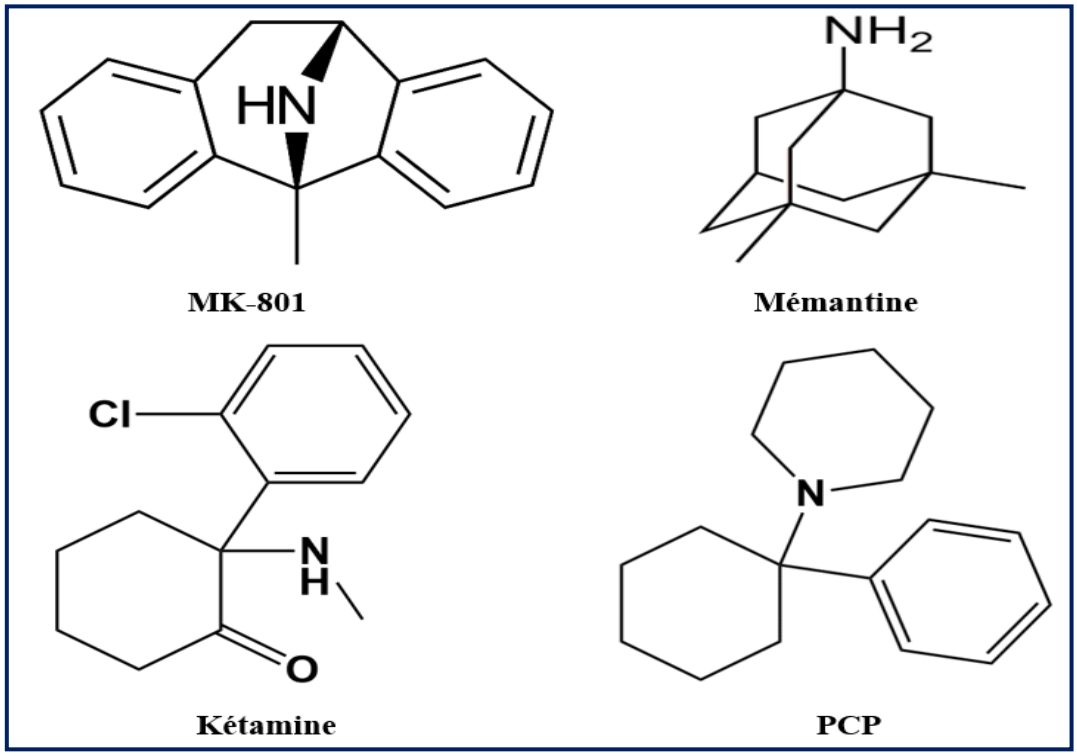



**Figure 4. Antagonistes non compétitifs du récepteur NMDA**

### 5. Activités biologiques étudiées

La présente étude vise à examiner une variété de classes structurelles d'inhibiteurs du récepteur NMDA, qui ont été synthétisées avec deux activités biologiques différentes. Une classe structurelle d'inhibiteurs sélectifs présentant des activités analgésiques contre la douleur neuropathique [21]. La deuxième constitue une famille d'inhibiteurs compétitifs présentant des activités inhibitrices contre le transporteur de glycine de type 1 (TGly-1) [22,23], et la dernière constitue une famille d'inhibiteurs sélectifs et compétitifs de TGly-1 [24].

### 5.1. Activité analgésique contre la douleur neuropathique

Une classe structurelle d'inhibiteurs sélectifs du R-NMDA de sous-unité NR2B, ont été synthétisés chimiquement et évalués biologiquement en utilisant les tests expérimentaux In-vivo et In-vitro [21], en présentant des activités analgésiques contre la douleur neuropathique exprimées en concentrations inhibitrices à cinquante pourcent ($IC_{50}$), dans lequel nous avons réalisé nos études in-silico basées sur la modélisation moléculaire afin de découvrir les meilleurs inhibiteurs qui peuvent agir en toute sécurité sur les cytochromes de l'organisme humain, en produisant des interactions intermoléculaires assez stables en ordre énergétique vis-à-vis les sites actifs de la cible protéique encodée par 3QEL.pdb [25], qui a été extraite de la banque de données des protéines en se basant sur la technique de la diffraction des rayons X, avec une résolution de 2.60 Å. (https://www.rcsb.org/structure/3QEL).

### 5.2. Activité inhibitrice du transporteur de glycine de type 1

Une seconde classe d'inhibiteurs ont été synthétisés et évalués par des tests In-vivo et In-vitro par leurs activités inhibitrices sur le transporteur de glycine de type 1 (TGly-1), tout en améliorant les propriétés de la mémoire chez l'homme par l'activation de la fonction des récepteurs NMDA bloqués [22,23]. La même activité biologique a été également examinée pour des familles de benzamides et celles de sulfonamides, en agissant comme des inhibiteurs sélectifs et compétitifs puissants de TGly-1, connus par leur potentiel de traiter les cas de la schizophrénie et des déficiences associées à un dysfonctionnement des récepteurs NMDA [24]. A travers ces travaux expérimentaux nous avons appliqué nos investigations computationnelles basées sur les simulations moléculaires pour identifier les inhibiteurs efficaces prédites par leur sécurité, leur similitude aux médicaments, et leur stabilité moléculaire vis-à-vis les sites actifs de la protéine visée codée par 4M48.pdb [26,27], extraite avec une résolution de 2.96 Å. (https://www.rcsb.org/structure/4M48).

**Références**

[1] S.H.R. Oliet, T. Papouin, De l'importance de la localisation des récepteurs du glutamate NMDA, Med Sci (Paris). 29 (2013) 260–262. https://doi.org/10.1051/medsci/2013293011.

[2] B. Feng, R. Morley, D. Jane, D. Monaghan, The effect of competitive antagonist chain length on NMDA receptor subunit selectivity, Neuropharmacology. 48 (2005) 354–359. https://doi.org/10.1016/j.neuropharm.2004.11.004.

[3] V.P. Zambre, R.B. Patil, J.N. Sangshetti, S.D. Sawant, Comprehensive QSAR studies reveal structural insights into the NR2B subtype selective benzazepine derivatives as N-Methyl--Aspartate receptor antagonists, Journal of Molecular Structure. 1197 (2019) 617–627. https://doi.org/10.1016/j.molstruc.2019.07.068.

[4] H. Sugihara, K. Moriyoshi, T. Ishii, M. Masu, S. Nakanishi, Structures and properties of seven isoforms of the NMDA receptor generated by alternative splicing, Biochemical and Biophysical Research Communications. 185 (1992) 826–832. https://doi.org/10.1016/0006-291X(92)91701-Q.

[5] H. Meguro, H. Mori, K. Araki, E. Kushiya, T. Kutsuwada, M. Yamazaki, T. Kumanishi, M. Arakawa, K. Sakimura, M. Mishina, Functional characterization of a heteromeric NMDA receptor channel expressed from cloned cDNAs, Nature. 357 (1992) 70–74. https://doi.org/10.1038/357070a0.

[6] H. Monyer, R. Sprengel, R. Schoepfer, A. Herb, M. Higuchi, H. Lomeli, N. Burnashev, B. Sakmann, P.H. Seeburg, Heteromeric NMDA Receptors: Molecular and Functional Distinction of Subtypes, Science. 256 (1992) 1217–1221. https://doi.org/10.1126/science.256.5060.1217.

[7] T. Ishii, K. Moriyoshi, H. Sugihara, K. Sakurada, H. Kadotani, M. Yokoi, C. Akazawa, R. Shigemoto, N. Mizuno, M. Masu, Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits, J Biol Chem. 268 (1993) 2836–2843.

[8] B. Laube, H. Hirai, M. Sturgess, H. Betz, J. Kuhse, Molecular Determinants of Agonist Discrimination by NMDA Receptor Subunits: Analysis of the Glutamate Binding Site on the NR2B Subunit, Neuron. 18 (1997) 493–503. https://doi.org/10.1016/S0896-6273(00)81249-0.

[9] W.D. Willis Jr., Central Nervous System Mechanisms for Pain Modulation, Stereotact Funct Neurosurg. 48 (1985) 153–165. https://doi.org/10.1159/000101121.

[10] M. Amat-Foraster, P. Celada, U. Richter, A.A. Jensen, N. Plath, F. Artigas, K.F. Herrik, Modulation of thalamo-cortical activity by the NMDA receptor antagonists ketamine and phencyclidine in the awake freely-moving rat, Neuropharmacology. 158 (2019) 107745. https://doi.org/10.1016/j.neuropharm.2019.107745.

[11] R. Labas, G. Gilbert, O. Nicole, M. Dhilly, A. Abbas, O. Tirel, A. Buisson, J. Henry, L. Barré, D. Debruyne, F. Sobrio, Synthesis, evaluation and metabolic studies of radiotracers containing a 4-(4-[18F]-fluorobenzyl)piperidin-1-yl moiety for the PET imaging of NR2B NMDA receptors, European Journal of Medicinal Chemistry. 46 (2011) 2295–2309. https://doi.org/10.1016/j.ejmech.2011.03.013.

[12] Z. Li, G. Wang, S. Zhong, X. Liao, S. Lai, Y. Shan, J. Chen, L. Zhang, Q. Lu, S. Shen, H. Huang, Y. Zhang, L. Zhang, Y. Jia, Alleviation of cognitive deficits and high copper levels by an NMDA receptor antagonist in a rat depression model, Comprehensive Psychiatry. 102 (2020) 152200. https://doi.org/10.1016/j.comppsych.2020.152200.

[13] P. Zambrano, M. Suwalsky, M. Jemiola-Rzeminska, K. Strzalka, Studies on the interaction of NMDA receptor antagonist memantine with cell membranes: A mini-review, Chemico-Biological Interactions. 283 (2018) 47–50. https://doi.org/10.1016/j.cbi.2018.01.022.

[14] Z. Li, G. Wang, S. Zhong, X. Liao, S. Lai, Y. Shan, J. Chen, L. Zhang, Q. Lu, S. Shen, H. Huang, Y. Zhang, L. Zhang, Y. Jia, Alleviation of cognitive deficits and high copper levels by an NMDA receptor antagonist in a rat depression model, Comprehensive Psychiatry. 102 (2020) 152200. https://doi.org/10.1016/j.comppsych.2020.152200.

[15] K. Blot, J. Bai, S. Otani, The effect of non-competitive NMDA receptor antagonist MK-801 on neuronal activity in rodent prefrontal cortex: an animal model for cognitive symptoms of schizophrenia, Journal of Physiology-Paris. 107 (2013) 448–451. https://doi.org/10.1016/j.jphysparis.2013.04.003.

[16] M.M. Marcus, J.M. Mathé, G.G. Nomikos, T.H. Svensson, Effects of competitive and non-competitive NMDA receptor antagonists on dopamine output in the shell and core subdivisions of the nucleus accumbens, Neuropharmacology. 40 (2001) 482–490. https://doi.org/10.1016/S0028-3908(00)00199-4.

[17] N.S. Cheung, D. O'Callaghan, M.C. Ryan, R. Dutton, M.G. Wong, P.M. Beart, Structure-activity relationships of competitive NMDA receptor antagonists, European Journal of Pharmacology. 313 (1996) 159–162. https://doi.org/10.1016/0014-2999(96)00608-5.

[18] Z. Tóth, A. Mihály, A. Mátyás, B. Krisztin-Péva, Non-competitive antagonists of NMDA and AMPA receptors decrease seizure-induced c-fos protein expression in the cerebellum and protect against seizure symptoms in adult rats, Acta Histochemica. 120 (2018) 236–241. https://doi.org/10.1016/j.acthis.2018.02.004.

[19] M. Amat-Foraster, P. Celada, U. Richter, A.A. Jensen, N. Plath, F. Artigas, K.F. Herrik, Modulation of thalamo-cortical activity by the NMDA receptor antagonists ketamine and phencyclidine in the awake freely-moving rat, Neuropharmacology. 158 (2019) 107745. https://doi.org/10.1016/j.neuropharm.2019.107745.

[20] Antagoniste des récepteurs NMDA - NMDA receptor antagonist - qaz.wiki, (n.d.). https://fr.qaz.wiki/wiki/NMDA_receptor_antagonist (accessed January 21, 2021).

[21] K. Anan, M. Masui, A. Tazawa, M. Tomida, Y. Haga, M. Kume, S. Yamamoto, S. Shinohara, H. Tsuji, S. Shimada, S. Yagi, N. Hasebe, H. Kai, Discovery of NR2B-selective antagonists via scaffold hopping and pharmacokinetic profile optimization, Bioorganic & Medicinal Chemistry Letters. 29 (2019) 1143–1147. https://doi.org/10.1016/j.bmcl.2019.02.017.

[22] V.J. Santora, T.A. Almos, R. Barido, J. Basinger, C.L. Bellows, B.C. Bookser, J.G. Breitenbucher, N.J. Broadbent, C. Cabebe, C.-K. Chai, M. Chen, S. Chow, D.M. Chung, L.

Crickard, A.M. Danks, G.C. Freestone, D. Gitnick, V. Gupta, C. Hoffmaster, A.R. Hudson, A.P. Kaplan, M.R. Kennedy, D. Lee, J. Limberis, K. Ly, C.C. Mak, B. Masatsugu, A.C. Morse, J. Na, D. Neul, J. Nikpur, M. Peters, R.E. Petroski, J. Renick, K. Sebring, S. Sevidal, A. Tabatabaei, J. Wen, Y. Yan, Z.W. Yoder, D. Zook, Design and Synthesis of Novel and Selective Glycine Transporter-1 (GlyT1) Inhibitors with Memory Enhancing Properties, J. Med. Chem. 61 (2018) 6018–6033. https://doi.org/10.1021/acs.jmedchem.8b00372.

[23] A.R. Hudson, V.J. Santora, R.E. Petroski, T.A. Almos, G. Anderson, R. Barido, J. Basinger, C.L. Bellows, B.C. Bookser, N.J. Broadbent, C. Cabebe, C.-K. Chai, M. Chen, S. Chow, D.M. Chung, L. Heger, A.M. Danks, G.C. Freestone, D. Gitnick, V. Gupta, C. Hoffmaster, A.P. Kaplan, M.R. Kennedy, D. Lee, J. Limberis, K. Ly, C.C. Mak, B. Masatsugu, A.C. Morse, J. Na, D. Neul, J. Nikpur, J. Renick, K. Sebring, S. Sevidal, A. Tabatabaei, J. Wen, S. Xia, Y. Yan, Z.W. Yoder, D. Zook, M. Peters, J.G. Breitenbucher, Azetidine-based selective glycine transporter-1 (GlyT1) inhibitors with memory enhancing properties, Bioorganic & Medicinal Chemistry Letters. 30 (2020) 127214. https://doi.org/10.1016/j.bmcl.2020.127214.

[24] Y. Wang, H. Zhao, J.T. Brewer, H. Li, Y. Lao, W. Amberg, B. Behl, I. Akritopoulou-Zanze, J. Dietrich, U.E.W. Lange, F. Pohlki, C. Hoft, W. Hornberger, S.W. Djuric, J. Sydor, M. Mezler, A.L. Relo, A. Vasudevan, De Novo Design, Synthesis, and Biological Evaluation of 3,4-Disubstituted Pyrrolidine Sulfonamides as Potent and Selective Glycine Transporter 1 Competitive Inhibitors, J. Med. Chem. 61 (2018) 7486–7502. https://doi.org/10.1021/acs.jmedchem.8b00295.

[25] M. El fadili, M. Er-rajy, H. Imtara, M. Kara, S. Zarougui, N. Altwaijry, O. Al kamaly, A. Al Sfouk, M. Elhallaoui, 3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists, Processes. 10 (2022) 1462. https://doi.org/10.3390/pr10081462.

[26] M. El fadili, M. Er-rajy, H. Imtara, O.M. Noman, R.A. Mothana, S. Abdullah, S. Zerougui, M. Elhallaoui, QSAR, ADME-Tox, molecular docking and molecular dynamics simulations of novel selective glycine transporter type 1 inhibitors with memory enhancing properties, Heliyon. 9 (2023) e13706. https://doi.org/10.1016/j.heliyon.2023.e13706.

[27] M. El fadili, M. Er-rajy, W. Ali Eltayb, M. Kara, A. Assouguem, A. Saleh, O. Al Kamaly, S. Zerougui, M. Elhallaoui, In-silico Screening Based on Molecular Simulations of 3,4-disubstituted Pyrrolidine Sulfonamides as Selective and Competitive GlyT1 Inhibitors, Arabian Journal of Chemistry. (2023) 105105. https://doi.org/10.1016/j.arabjc.2023.105105.

# Chapitre II : Approches de base concernant les techniques In-silico destinées à la modélisation moléculaire

## 1. Introduction

Au cours des deux dernières décennies, la modélisation moléculaire a suscité une évolution prodigieuse en divers domaines d'application, comme la chimie computationnelle, la biologie structurale, et l'industrie pharmaceutique. Grâce aux techniques in-silico basées sur la modélisation moléculaire, il est désormais envisageable de modéliser les relations quantitatives prédites en deux et trois dimensions entre la structure moléculaire des composés chimiques et leur activité biologique examinée, ainsi que les prédictions in-silico des propriétés physico-chimiques et pharmacocinétiques ADME-Toxicité d'inhibiteurs examinés, en étudiant par la suite leur mécanisme d'inhibition vis-à-vis la cible protéique, comme on peut simuler le comportement dynamique des molécules bioactives et leur mouvement interne vis-à-vis la protéine responsable le long du temps de simulation moléculaire estimé en nanosecondes.

## 2. Relations quantitatives structure-activité

### 2.1. Définition

Les relations quantitatives structure-activité (QSAR) font partie des méthodes de relations quantitatives structure-propriété (QSPR), dont la propriété étudiée est spécifiquement l'activité biologique. Les techniques QSAR se reposent donc principalement sur l'hypothèse que les propriétés étudiées de la structure moléculaire sont fortement liées à l'activité biologique à travers un certain nombre des relations quantitatives expliquées par des modèles prédictifs qui peuvent être des équations mathématiques linéaires ou non linéaires.

### 2.2. Principe

La technique de modélisation moléculaire basée sur les relations quantitatives structure-activité (QSAR), a pour principe de générer un modèle mathématique applicable et validé par des méthodes purement statistiques, afin de prédire la relation quantitative établie entre l'activité biologique et les descripteurs structuraux d'une famille des composés chimiques biologiquement actifs, et synthétisés dans les mêmes conditions expérimentales, tout en se basant sur l'équation suivante :

$$\textit{Activité biologique} = f(\textit{descripteurs moléculaires}). \qquad \textit{Equation 1}$$

### 2.3. Objectif

L'objectif principal de l'étude des relations quantitatives structure-activité (QSAR) est la prédiction de l'activité biologique de certaines nouvelles molécules qui n'ont pas encore été synthétisées en raison du coût élevé et du temps perdu à synthétiser toutes les molécules souhaitées, grâce aux modèles mathématiques validés pour une base de données expérimentales portant sur un nombre assez important de composés chimiques.

### 2.4. Méthodologie suivie dans l'étude 2D/3D-QSAR

L'étude bidimensionnelle des relations quantitatives structure-activité (2D-QSAR), commence par la recherche d'une série de molécules présentant des activités biologiques bien définies, dans lesquelles nous pouvons calculer une variété de descripteurs moléculaires (géométriques, topologiques, constitutionnels, physicochimiques, thermodynamiques, et quantiques), pour avoir une base de données expérimentales qu'on la considère comme point de départ de nos évaluations computationnelles basées sur la machine. Tout d'abord, on cherche à réduire la dimension des variables calculées en se concentrant sur les descripteurs moléculaires faiblement corrélés, tout en appliquant la technique de l'analyse en composantes principales (ACP). Par la suite, on cherche à développer les meilleurs modèles mathématiques basés sur la régression linéaire multiple (RLM), ou bien sur la régression non linéaire multiple (RNLM), ou d'autres modèles à condition qu'ils soient générés pour un ensemble d'entrainement contenant 80% des composés chimiques et validés pour un ensemble de test contenant 20% des molécules actives à partir des centaines de randomisations aléatoires. Les modèles développés sont testés et évalués par un ensemble des critères de validation, comme la validation externe et validation interne croisée de type CVLOO, test de randomisation, tests statistiques de Fisher et de Student, ainsi que le domaine d'applicabilité (DA).

Concernant l'étude tridimensionnelle des relations quantitatives structure-activité (3D-QSAR), on suit presque la même procédure sauf que pour les descripteurs moléculaires, qui se remplacent par des descripteurs tridimensionnelles extrêmement limités, comme les champs stériques, électrostatiques, et hydrophobiques, dont lesquels on cherche à les corréler vis-à-vis l'activité biologique en développant des modèles prédictifs à trois dimension (3D), récemment connus par des modèles CoMFA et CoMSIA notés en anglais par « Comparative Molecular Field Analysis » et « Comparative Molecular Similarity Indices Analysis », respectivement. On peut résumer cette démarche dans le Schéma 1.

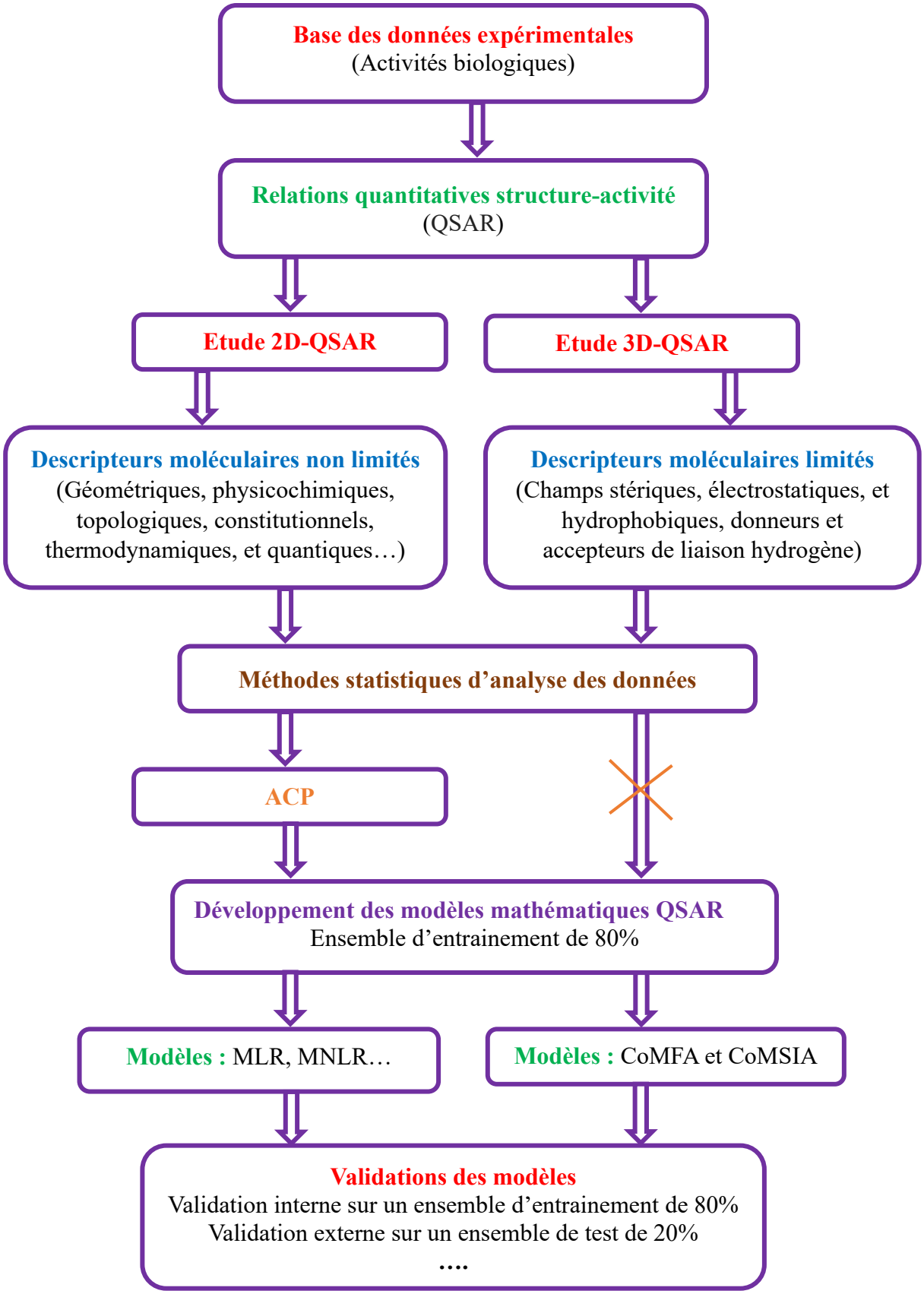



**Schéma 1. Stratégie suivie dans l'étude 2D/3D-QSAR**

## 2.5. Méthodes de calcul théorique

### 2.5.1. Théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité, connue en anglais par Density Functional Theory (DFT) est une approche purement théorique, basée sur les calculs quantiques visant à étudier la structure des systèmes électroniques de différentes tailles, variant de quelques atomes à des centaines d'éléments chimiques, tout en déterminant les propriétés physico-chimiques des molécules examinées comme le moment dipolaire, l'électronégativité, l'électrophilicité, la dureté, et les énergies des orbitales frontières ; à savoir l'énergie d'orbitale moléculaire la plus haute occupée (HOMO), et l'énergie d'orbitale moléculaire la plus basse non occupée (LUMO).

En principe, la DFT est fondée sur des raisonnements de Kohn et Hohenberg, qui ont énoncé que la densité électronique d'une molécule donnée correspond à une seule fonction d'onde, et l'état du système électronique peut être complètement défini par la densité électronique de l'état le plus stable (l'état fondamental) associé à un minimum d'énergie. Ainsi, l'approche DFT est fondée sur la célèbre équation du physicien Schrödinger Erwin, qui relie la fonctionnelle de la densité avec l'énergie du système électronique étudié, comme exprimée avec la forme indépendante du temps par : $\hat{H}\psi=E\psi$, et par sa forme dépendante du temps : $\hat{H}\psi(t)=E\psi(t)$ ; dont $\psi$ représente la fonction d'onde, E montre la valeur propre correspondante à l'énergie du système électronique étudié, et $\hat{H}$ est l'opérateur hamiltonien moléculaire, comportant cinq termes différents, qui sont respectivement : le terme d'énergie cinétique pour N électrons, terme d'énergie cinétique pour A noyaux, potentiel d'interaction électron-noyau, potentiel d'interaction électron-électron et potentiel d'interaction noyau-noyau, comme exprimé dans l'équation suivant :

$$\hat{H} = \hat{\boldsymbol{T}}_{\boldsymbol{e}} + \hat{\boldsymbol{T}}_{\boldsymbol{N}} + \hat{\boldsymbol{v}}_{\boldsymbol{NN}} + \hat{\boldsymbol{v}}_{\boldsymbol{ee}} + \hat{\boldsymbol{v}}_{\boldsymbol{eN}} \text{ . Equation 2}$$

Tel que :

L'opérateur mathématique d'énergie cinétique pour N électrons est : $\hat{\boldsymbol{T}}_{\boldsymbol{e}} = -\sum_{k=1}^{n} \frac{\hbar^2}{2m_e} \nabla_k^2$

L'opérateur mathématique d'énergie cinétique pour A noyaux est : $\hat{\boldsymbol{T}}_{\boldsymbol{N}} = -\sum_{A=1}^{N} \frac{h^2}{8\pi^2 M_A} \nabla_A^2$

L'opérateur du potentiel d'interaction électron-noyau est : $\hat{\boldsymbol{v}}_{\boldsymbol{NN}} = \sum_{A=1}^{N} \sum_{B>A}^{N} \frac{Z_A Z_B e^2}{4\pi\varepsilon_0 R_{AB}}$

L'opérateur du potentiel d'interaction électron-électron est : $\hat{\boldsymbol{v}}_{\boldsymbol{ee}} = \sum_{k=1}^{n} \sum_{l>k}^{n} \frac{e^2}{4\pi\varepsilon_0 r_{kl}}$

L'opérateur du potentiel d'interaction noyau-noyau est : $\hat{\boldsymbol{v}}_{\boldsymbol{eN}} = -\sum_{A=1}^{N} \sum_{k=1}^{n} \frac{Z_A e^2}{4\pi\varepsilon_0 r_{kA}}$

Sachant que : h et ħ : représentent respectivement ; la constante et la constante réduite de Planck. me et MA : représentent respectivement ; la masse d'électron et celle de noyau. ZA et

ZB: représentent respectivement ; la charge électrique du noyau d'atome A et celle de B. rKA et RAB : représentent respectivement ; distance entre le noyau A et l'électron K, et la distance entre les noyaux d'atomes A et B. $\nabla_k^2$ est Laplacien propre d'électron K.

Par conséquent, la structure électronique du système à base de plusieurs noyaux et d'électrons peut être décrite, en recherchant des solutions possibles à l'équation de Schrödinger qui sont basées en principe sur un certain nombre des approximations comme celles de Born Max et Oppenheimer Robert, qui ont allé plus loin pour simplifier la résolution de cette forme assez compliquée, tout en proposant que les énergies cinétiques correspondant aux noyaux des atomes soient négligées, en considérant que les positions des noyaux comme constantes [1].

Dans la présente thèse, les systèmes électroniques examinés ont été soumis à l'optimisation par DFT, en utilisant la fonctionnelle hybride précise, largement utilisée, notée par B3LYP (Becke 3-Parameter Lee-Yang-Parr), couplée à la base 6-31+G(d,p), en tant qu'une base étendue aux sous-couches électroniques d et p.

### 2.5.2. Théorie de la mécanique moléculaire

La mécanique moléculaire (MM) est une méthode de calcul théorique, dont l'énergie du système électronique étudié se minimise en se basant sur le principe de la mécanique classique, en considérant que la molécule examinée est composée de plusieurs atomes sur lesquels s'exercent des champs de forces de type harmoniques et élastiques, chacune est associée à une fonction d'énergie potentielle qui varie en fonction de la géométrie probable du système électronique tout au long du temps de simulation moléculaire. Différentes classes de champs de forces sont adaptées aux certains nombres des systèmes électroniques ; à savoir les champs de force d'Allinger de type MM2, MM3, et MM4 ; bien adaptés pour des petites molécules, champs de force de type Amber et Tripos : spécialisées aux grosses molécules, et champs de force unifiés (united force field : UFF) : mieux adaptés aux composés organométalliques [2].

Dans la présente Thèse, nous avons utilisé la théorie de la mécanique moléculaire avec champ de force de type MM2 afin d'optimiser la géométrie de chaque système électronique, tout en déterminant une famille des descripteurs moléculaires, notamment les descripteurs géométriques, thermodynamiques, stériques, et physicochimiques, à travers l'utilisation des programmes ACD/ChemSketch, ChemBioOffice, et ChemBio3D.

## 2.6. Descripteurs moléculaires

### 2.6.1. Généralités

Le développement du modèle QSAR nécessite le calcul des descripteurs moléculaires, ce qui exige principalement l'optimisation de la géométrie de la structure moléculaire en

recherchant le minimum d'énergie pour chaque composé chimique à l'aide de DFT, de sorte que les géométries optimisées des molécules actives sont d'une importance critique pour le calcul précis des différents descripteurs moléculaires [3]. En principe, les descripteurs moléculaires sont des variables généralement quantitatives qui caractérisent la structure moléculaire d'un composé chimique. Des milliers de descripteurs moléculaires pourraient être exploités pour prédire un événement biologique comme l'activité biologique [4].

### 2.6.2. Types des descripteurs

Plus de six milles descripteurs moléculaires pourraient être calculés afin de prédire l'activité biologique d'un nombre donné de molécules synthétisées [5]. Cependant, les méthodes de QSAR ont été conçues à minimiser au maximum le nombre de descripteurs moléculaires, de sorte que les modèles développés puissent être facilement contrôlés, maîtrisés et généralisés à de nouvelles molécules. À cette fin, nos travaux publiés ont été strictement limités à un certain nombre de descripteurs en faisant appel aux descripteurs les plus pertinents de différentes dimensions.

#### 2.6.2.1.Descripteurs 1D

Le calcul des descripteurs moléculaires à une dimension (1D) repose principalement sur les formules brutes des molécules, qui révèlent en général leur composition chimique, en précisant le genre et le nombre d'atomes, ce qui permet par la suite d'obtenir un ensemble de caractéristiques connues par les descripteurs constitutionnels telles que le poids moléculaire, la masse molaire, et les pourcentages massiques des atomes. Cependant, les descripteurs unidimensionnels ont l'inconvénient de poser des problèmes d'interprétation des mécanismes d'interaction, car ils ne prennent pas en considération les isomères constitutionnels des molécules étudiées. Il est donc crucial de faire intervenir en plus des descripteurs moléculaires d'autres dimensions [6].

#### 2.6.2.2.Descripteurs 2D

Les descripteurs moléculaires bidimensionnels sont issus de la structure planaire de la molécule, et fournissent des informations au sujet de la forme globale du système électronique et de ses ramifications [7]. Les indices topologiques font partie des descripteurs 2D qui caractérisent la connectivité des atomes dans une molécule donnée. On cite parmi les plus courants, l'indice de Balaban [8], l'indice de Wiener [9], l'indice topologique moléculaire , la somme des degrés de valence [10], et la superficie de la surface polaire [11].

Ainsi, le calcul des descripteurs moléculaires à deux dimensions (2D) donne globalement des valeurs quantitatives précises qui pourraient avoir un effet significatif sur les modèles

prédictifs QSAR. Cependant, ils restent insuffisants pour prédire l'activité biologique d'une manière assez satisfaisante [6].

#### 2.6.2.3.Descripteurs 3D

Les descripteurs moléculaires à trois dimensions (3D) sont déterminés en fonction de la position relative des espèces chimiques dans un espace tridimensionnel, en décrivant des caractéristiques assez complexes liées à la géométrie 3D. Le calcul de cette catégorie des descripteurs coute beaucoup du temps, car nécessite l'utilisation des méthodes de la modélisation moléculaire de type ab-initio ou bien empiriques, dans lesquels on distingue des descripteurs géométriques, physico-chimiques, thermodynamiques, et quantiques [1].

#### 2.6.2.4.Descripteurs 4D

Les descripteurs moléculaires à quatre dimensions (4D) sont inclus dans l'étude tridimensionnelle des relations quantitatives structure-activité (3D-QSAR) en calculant les champs d'interaction moléculaire grâce à deux modèles prédictifs, classiquement connus en tant qu'analyse comparative du champs moléculaire notée en anglais par « Comparative Molecular Field Analysis : CoMFA » [12,13] et analyse comparative des indices de similarité moléculaire, désignée par « Comparative Molecular Similarity Indices Analysis : CoMSIA » [14]. Ces descripteurs moléculaires sont limités en cinq ; les champs stériques, électrostatiques, hydrophobiques, donneurs et accepteurs de liaison Hydrogène.

## 2.7. Méthodes statistiques

### 2.7.1. Analyse en composantes principales

La base de données statistiques qui regroupe en valeurs quantitatives les activités biologiques des molécules étudiées et différents descripteurs moléculaires calculés, est traitée pour la première fois par l'analyse en composantes principales (ACP) ; une technique largement appliquée pour réduire la dimension des variables (descripteurs moléculaires calculés) en un nombre limité des nouveaux variables, appelées les axes factoriels ou bien les composantes principales en se basant sur la matrice de corrélation, dont l'un des deux descripteurs fortement corrélés ayant le coefficient corrélation de Person supérieur à 0.9 doit être supprimé et l'autre est retenue, en insistant sur les descripteurs faiblement corrélés de coefficient de corrélation qui tend vers 0.

### 2.7.2. Régression linéaire multiple

La régression linéaire multiple (RLM) est une méthode purement statistique destinée à établir une relation quantitative et linéaire, expliquée par un modèle QSAR conçu pour prédire une variable dépendante (l'activité biologique) en fonction d'un nombre réduit des descripteurs

moléculaires non corrélées (variables indépendantes) pour un ensemble d'apprentissage contenant trois quarts (80%) des molécules actives, comme il est traduit dans l'équation 3.

$$\boldsymbol{Activité} = \boldsymbol{a0} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{ai} \times \boldsymbol{Xi}). \qquad Equation\ 3$$

### 2.7.3. Régression non linéaire multiple

La régression non linéaire multiple (RNLM) est une méthode non linéaire visant à développer un modèle mathématique non linéaire entre l'activité biologique et les mêmes descripteurs moléculaires précédemment sélectionnés dans la régression linéaire multiple (RLM) pour le même ensemble d'entrainement, grâce à la fonction programmée donnée par l'équation 4.

$$\boldsymbol{Activité} = \boldsymbol{a0} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{ai} \times \boldsymbol{Xi} + \boldsymbol{bi} \times \boldsymbol{Xi^2}). \qquad Equation\ 4$$

## 2.8. Techniques de validation

### 2.8.1. Test de de Fisher

Le test statistique de Fisher est utilisé dans cette thèse pour vérifier l'homogénéité de variance appliquée sur le modèle mathématique développé de QSAR, en se basant sur le test d'hypothèse, tout en rejetant l'hypothèse nulle (H0) et en acceptant l'hypothèse inverse H1 qui postule que les variances sont homogènes à condition que la valeur calculée de Fisher soit suffisamment supérieure que la valeur critique pour un seuil de confiance de 95%.

- Le test d'hypothèse de Fisher :

H0 : **F** calculée < **F** critique ($n_1$, $n_2$) : Les variances ne sont pas homogènes.
(Seuil de confiance de 95%, Probabilité de 5%)
H1 : **F** calculée > **F** critique ($n_1$, $n_2$) : Les variances sont homogènes.

La valeur critique de l'indice de Fisher à 5% est extraite des tableaux statistiques de Student, où $n_1$ est le nombre des molécules composant l'ensemble d'entrainement et $n_2$ représente le nombre de descripteurs sélectionnés, tandis que la valeur calculée de l'indice de Fisher, est exprimée dans l'équation 5.

$$\boldsymbol{F}\ calculée = \times \frac{\sum_{\boldsymbol{i}}^{\boldsymbol{n}} \left(\hat{\mathrm{A}}\boldsymbol{i} - \bar{\mathrm{A}}\right)^{\boldsymbol{2}}}{\sum_{\boldsymbol{i}}^{\boldsymbol{n}} \left(\boldsymbol{Ai} - \hat{\mathrm{A}}\boldsymbol{i}\right)^{\boldsymbol{2}}} \times \frac{(\boldsymbol{n-p-1})}{\boldsymbol{p}}. \qquad Equation\ 5$$

Tel que : Ai et Âi représentent respectivement, les valeurs quantitatives de l'activité biologique observé et l'activité biologique estimée (ou prédite par le modèle QSAR) ; Ā est la moyenne des activités biologiques prédites ; (n-p-1) représente le degré de liberté, dont : n est

le nombre de molécules étudiées et p représente le nombre des descripteurs moléculaires introduits dans le modèle QSAR [15–17].

### 2.8.2. Test de Student

Le test de significativité des pentes établi par Student est l'un des tests statistiques fondamentaux destinés à examiner le poids significatif de chaque descripteur sélectionné sur l'activité biologique, tout en vérifiant que la valeur calculée de l'indice de Student est supérieure que sa valeur critique pour chaque descripteur moléculaire, ça veut dire que la probabilité par pente de chaque variable ne dépasse pas la probabilité critique de 5%, en travaillant à un niveau de confiance de 95%.

- Le test d'hypothèse de Student :

H0 : **T**i calculée < **T** critique [(1- $\frac{\alpha}{2}$), (n-p-2)], Pi value > P critique de 0.05 : Le descripteur moléculaire étudié n'a pas un poids significatif sur l'activité biologique.

H1 : **T**i calculée > **T** critique [(1- $\frac{\alpha}{2}$), (n-p-2)], Pi value < P critique de 0.05 : Le descripteur moléculaire étudié a un poids significatif sur l'activité biologique.

La valeur critique de l'indice de Student est déterminée pour un seuil de confiance α et un degré de liberté de (n-p-2), à savoir que n représente le nombre des échantillons (molécules étudiées), et p est le nombre de variables indépendantes (descripteurs moléculaires non corrélées).

En parallèle, La valeur calculée de l'indice de Student pour chaque descripteur moléculaire « i » est donnée par l'équation 6. Tel que : **T**i représente l'indice de Student calculé pour le descripteur moléculaire « i » ; ai est le coefficient associé à chaque descripteur moléculaire « i » ; et S(ai) représente l'erreur type associé au descripteur « i » [15–17].

$$\textit{Ti calculée} = \frac{a_i}{S(a_i)} = \frac{ai}{\sqrt{\frac{\sum_i^n (\hat{A}i - Ai)^2}{(n-p-1)}}}. \qquad \textit{Equation 6}$$

### 2.8.3. Validation interne

Le modèle mathématique des relations quantitatives structure-activité (QSAR) se valide pour la première fois par une validation croisée en utilisant la procédure "Leave-one-out (LOO)" [18]. Il s'agit d'une technique de validation interne du modèle prédictif, développé sur un ensemble d'apprentissage de 80% de la totalité des molécules étudiées, sans faire intervenir de nouvelles molécules de l'ensemble de test.

Cette technique fait partie des méthodes de validation croisée avec la procédure notée en anglais par "Leave-Many-Out (LMO)" qui consiste à diviser l'ensemble d'entrainement en K

groups, puis choisir l'un des k molécules comme ensemble de validation et les autres (k-1) molécules comme l'ensemble d'entrainement. Ainsi, le modèle QSAR sera développé K fois d'après l'ensemble d'entrainement, et les activités biologiques seront prédites par la suite d'une manière automatique de l'ensemble de validation. La procédure suivie est légèrement différente pour la validation croisée de type LOO, car elle est spécifiquement adaptée à l'examen d'une molécule donnée à chaque fois, alors que le modèle se génèrera sur N-1 molécules, et sera validé par la suite sur la seule molécule examinée, et l'astuce se répète donc n fois en générant n modèles de manière à extraire l'activité prédite correspondante à chaque molécule testée pour qu'en fin de compte on arrive à la valeur désirée du coefficient de corrélation de la validation croisée exprimé dans l'équation 7, ce qui ne doit pas dépasser le seuil de 0.5, afin que le modèle original soit robuste et que son pouvoir prédictif interne soit élevé [1,6,19].

$$\boldsymbol{r^2cv = 1 - \frac{\sum_i^n \left( \hat{A}\boldsymbol{i}\ (entrainement) - \boldsymbol{Ai}\ (entrainement) \right)^2}{\sum_i^n \left( \boldsymbol{Ai}\ (entrainement) - \bar{A}\ (entrainement) \right)^2}}. \quad Equation\ 7$$

Tel que : Âi et Ai représentent respectivement, l'activité biologique estimée et observée de la molécule « i » de l'ensemble d'entrainement ; et Ā représente la moyenne des activités biologiques des molécules de l'ensemble d'entrainement.

### 2.8.4. Test de Y-randomisation

Une valeur élevée du coefficient de corrélation de la validation croisée peut être due à une surestimation ou encore au hasard [20]. Pour cette raison, la validation croisée est nécessaire mais pas suffisante, ce qui nous oblige à mettre en œuvre un test de randomisation ou d'hasardisation, basé sur des centaines de distributions aléatoires sur l'ensemble d'entrainement, comme une autre technique de validation du modèle QSAR [21,22], dans laquelle nous cherchons à vérifier ultérieurement que le modèle original est le meilleur modèle établi en termes de critères statistiques, de telle sorte que le coefficient de détermination, coefficient de corrélation, et particulièrement le coefficient de corrélation de la validation croisée soient les plus élevés par rapport à l'ensemble des autres randomisations.

### 2.8.5. Validation externe

Le modèle QSAR développé et validé internement sur l'ensemble d'entrainement de 80%, se valide également en externe sur un ensemble de test contenant 20% de toutes les molécules étudiées, afin d'examiner le pouvoir prédictif externe du modèle établi [23]. Par conséquent, la puissance prédictive externe du modèle QSAR repose sur le coefficient de corrélation des

molécules de test ($r^2test$), exprimé dans l'équation 8. Une valeur supérieure à 0.6 du coefficient de corrélation, indique un pouvoir prédictif élevé du modèle généré.

$$\boldsymbol{R^2test = 1 - \frac{\sum_i^n \left( \hat{A}\boldsymbol{i}\,(\boldsymbol{test}) - \boldsymbol{Ai}\,(\boldsymbol{test}) \right)^2}{\sum_i^n \left( \boldsymbol{Ai}\,(\boldsymbol{test}) - \bar{A}\,(\boldsymbol{entrainement}) \right)^2}} . \qquad Equation\ 8$$

Tel que : Âi et Ai représentent respectivement, l'activité biologique estimée et observée de la molécule « i » de l'ensemble de test ; et Ā représente la moyenne des activités biologiques des molécules de l'ensemble d'entrainement.

### 2.8.6. Domaine d'applicabilité

Pour s'assurer que le modèle prédictif des relations quantitatives structure-activité (QSAR) examiné à la fois par la validation externe et interne, est applicable sur une base largement suffisante de données expérimentales, il est nécessaire de mettre en œuvre le domaine d'applicabilité en tant qu'une approche critique visant à généraliser le modèle validé sur les ensembles d'entrainement et de test, tout en identifiant la région désirée dans laquelle la totalité des molécules étudiées peuvent être prédites en toute confiance, conformément aux exigences et priorités absolues de l'organisation de coopération et de développement économiques (OCDE) [24–26].

En principe, le domaine ou la zone d'applicabilité du modèle QSAR est généralement basée sur deux critères statistiques ; les résidus de prédiction standardisés qui doivent être comprises dans l'intervalle de confiance [-3, +3], et levier de prédiction qui ne doit pas dépasser le seuil du levier critique (h*), exprimé en équation 9. Ces deux critères sont examinés par le diagramme de Williams, de sorte que toute molécule située dehors de cette région est considérée comme une observation aberrante, susceptible à être supprimée [27–29].

$$\boldsymbol{h^* = [3 \times (p+1) / n].} \qquad Equation\ 9$$

Tel que : le nombre 3 indique la limite en valeur absolue de l'intervalle des résidus standardisés ; p est le nombre des descripteurs moléculaires ; et n est le nombre des molécules de l'ensemble d'entrainement.

### 2.8.7. Critères statistiques de Golbreikh et Tropsha

L'étude statistique de "Alexander Golbreikh et Alexander Tropsha" a démontré que la fiabilité et la qualité du modèle QSAR peuvent être testées sur la base d'un certain nombre de critères statistiques, présentés dans la Table 1, tout en vérifiant les seuils de tolérance standards associés aux paramètres de validation les plus pertinents [30]. Une fois que le modèle mathématique développé satisfait aux conditions énoncées dans l'approche théorique de Golbreikh et Tropsha, nous pouvons automatiquement conclure qu'il est fiable, robuste et offre

un haut niveau d'exactitude pour prédire l'activité biologique de nouvelles molécules, et que les descripteurs moléculaires sélectionnés peuvent forcément influencer l'activité biologique étudiée.

**Table 1. Critères statistiques de Golbreikh et Tropsha**

| Noms des Paramètres statistiques | Symboles statistiques | Formules mathématiques | Seuils critiques |
|---|---|---|---|
| Coefficient de corrélation du modèle établi sur le test d'entrainement | $R^2$ | $R^2 = 1 - \frac{\sum_i^n \left(\hat{A}i(entr) - Ai(entr)\right)^2}{\sum_i^n \left(Ai(entr) - \bar{A}(entr)\right)^2}$ | > 0.6 |
| Coefficient de corrélation d'ajustement | $R^2_{ajust}$ | $R^2 adj = \frac{(N-1) \times R^2 - p}{N - p - 1}$ | > 0.6 |
| Coefficient de corrélation de la validation croisée | $r^2cv$ | $r^2cv = 1 - \frac{\sum_i^n \left(\hat{A}i - Ai\right)^2}{\sum_i^n \left(Ai - \bar{A}\right)^2}$ | > 0.5 |
| Coefficient de corrélation de validation externe | $R^2test$ | $R^2test = 1 - \frac{\sum_i^n \left(\hat{A}i(test) - Ai\ (test)\right)^2}{\sum_i^n \left(Ai\ (test) - \bar{A}\ (ent)\right)^2}$ | > 0.6 |
| Coefficient de corrélation à n randomisations | $R^2_{random}$ | Moyenne des coefficients de corrélation pour n randomisations | < $R^2$ |
| Coefficient de corrélation de validation croisée à n randomisations | $r^2$cv random | Moyenne des coefficients de corrélation de validation croisée en n randomisations | < $r^2cv$ |
| Constante de randomisation à n randomisations | ${}^cR^2p_{random}$ | $cR^2p_{random} = R \times \sqrt[2]{R^2 - R^2_{random}}$ | > 0.5 |

### 3. Prédiction des propriétés physicochimiques et pharmacocinétiques ADME-Toxicité

Toute évaluation clinique nécessite au préalable une estimation des propriétés physicochimiques et pharmacocinétiques de l'absorption, de la distribution, du métabolisme, de l'excrétion et de la toxicité (ADMET) des candidats-médicaments dans le corps humain, sur la base d'un ensemble des conditions principales à travers lesquelles nous souhaitons vérifier la pharmacocinétique, la ressemblance aux médicaments et la compatibilité des petites molécules avec la chimie médicinale.

Tout d'abord, on cherche à vérifier que les cinq règles de Lipinski sont satisfaites pour les composés chimiques étudiés, dont le poids moléculaire est inférieur ou égale à 500 g/mol, l'indice de la réfractivité molaire (RM) est compris dans l'intervalle fermée [40, 130], coefficient de partage en solvant octanol/eau (Log P) inférieur à 5, pas plus de dix accepteurs de liaisons hydrogènes, et pas plus de cinq donneurs de liaisons hydrogènes [31–33]. On vérifie également que les nombres de violations de Veber [34], Egan [35,36], Muegge [37], et Ghose [38] sont respectés. Par la suite, on se base sur le modèle prédictif d'œuf à la coque d'Egan, comme un modèle précis largement appliqué dans la découverte des médicaments, en se basant

sur la prédiction d'inhibiteurs puissants du système nerveux central (SNC) [39], de sorte que toute molécule située dans la partie jaune d'œuf d'Egan devrait traverser passivement la barrière hémato-encéphalique (noté en anglais par (Blood-Brain Barrier "BBB"). Cependant, les molécules qui font partie de la zone blanche d'œuf d'Egan sont prédites d'être absorbés passivement par le tractus gastro-intestinal. En prenant en considération que les observations colorées en points bleus sont prédites comme des agents libérés du système nerveux central par la P-glycoprotéine (P-GP), et celles colorées en rouge sont prédites de ne pas être efflués du SNC par P-GP.

En dernière phase, on fait des prédictions in-silico des propriétés pharmacocinétiques ADMET d'inhibiteurs étudiés, de sorte que l'absorption des molécules dans l'organisme humain contrôlée par l'absorption intestinale humaine (AIH) soit importante, une distribution en corps humain et en particulier dans la circulation sanguine et les tissus, soit définie par des critères acceptables de perméabilités au système nerveux central (SNC) et au barrière hémato-encéphalique (BBB), un métabolisme significatif par les protéines et définitivement les cytochromes suivants : CYP1A2, CYP2C19, CYP2C9, CYP2D6, et CYP3A4, puis une élimination ou bien un excrétion par les urines, et enfin une intensité nulle de toxicité, tout en contrôlant les effets indésirables d'hépatotoxicité et d'allergie cutanée, ainsi que le potentiel d'un xénobiotique à engendrer la mutation des gènes (test biologique de AMES) [40]. Par conséquent, les propriétés physicochimiques et pharmacocinétiques jouent un rôle déterminant dans le succès des phases cliniques lors de la conception de nouveaux médicaments [41].

## 4. Docking moléculaire

### 4.1. Généralités

Le Docking moléculaire a progressivement évolué, passant d'un nouvel outil de visualisation pour les cristallographes de protéines et les modélisateurs moléculaires, à un véritable élément de la découverte de médicaments [42]. Il est destiné pour prédire la position relative de la petite molécule appelée ligand, dans une macromolécule qui est le récepteur protéique tout en déterminant l'énergie de liaison ($\Delta G_{liaison}$) en Kcal/mol du complexe moléculaire produit, qui fait intervenir la contribution en termes énergétiques des différentes types des interactions intermoléculaires comme présentés dans l'équation 10 [43].

$$\Delta G_{liaison} = \Delta G_{(interaction)} + \Delta G_{(solvent)} + \Delta G_{(conformation)} + \Delta G_{(mouvement)}. \quad Equation\ 10$$

Tel que : $\Delta G_{(interaction)}$ représente l'énergie correspondante aux interactions spécifiques entre le ligand et la protéine ; $\Delta G_{(solvent)}$ est l'énergie d'interaction du ligand et protéine vis-à-vis le solvant ; $\Delta G_{(conformation)}$ représente le terme énergétique associée aux changements

conformationnels du ligand et du récepteur ; et $\Delta G_{(mouvement)}$ c'est l'énergie liée aux mouvements de la protéine et du ligand pendant la formation du complexe.

### 4.2. Principe

En chimie computationnelle, le Docking moléculaire repose sur la détection des interactions intermoléculaires typiques produites entre les ligands actifs et les cibles protéiques responsables d'une activité biologique donnée, en se basant sur des complexes les plus stables en ordre énergétique (énergies de liaison $\Delta G_{liaison}$ plus basses). Par conséquent, Le complexe stable choisi parmi les dizaines de conformations possibles, fera l'objet d'une discussion approfondie portant sur les mécanismes réactionnels qui pourraient justifier l'activité biologique inhibitrice des candidats-médicaments.

### 4.3. Protocole de validation

La performance de la simulation du Docking moléculaire est dite validée, lorsque deux facteurs du protocole de validation du Docking moléculaire (ou Re-Docking) sont vérifiés ; le premier implique que le ligand examiné a été effectivement docké dans les sites actifs de la cible protéique, qui se déterminent expérimentalement vis-à-vis le ligand co-cristallisé lié à la cible protéique à l'aide de la méthode de diffraction des rayons X. Le second facteur tend à vérifier que l'écart quadratique moyen (noté en anglais par Root Mean Square Deviation : RMSD) de la position de superposition de ligand natif (ou co-cristallisé) et ligand docké, ne dépasse pas le seuil critique déterminé en 2 Å [44–46].

### 4.4. Objectif

À la lumière de ce qu'on a présenté, on comprend que l'approche du Docking moléculaire a été conçue spécifiquement pour découvrir un candidat médicament à condition qu'il présente une activité biologique importante, et pouvant réagir particulièrement dans les sites actifs d'une cible protéique avec un minimum d'énergie de liaison pour que le complexe produit soit le plus stable parmi les autres complexes possibles, en faisant intervenir des interactions intermoléculaires comme des liaisons hydrogènes connues par leur pouvoir stabilisant des complexes moléculaires.

### 4.5. Cibles protéiques

Les simulations du Docking moléculaire, réalisées dans nos travaux scientifiques publiés, ont été concentrées sur deux récepteurs protéiques ; la première est responsable de l'activité analgésique contre la douleur neuropathique, et la deuxième est responsable de l'activité inhibitrice du transporteur de glycine de type 1. Toutes les deux macromolécules ont été extraites de la banque des données de protéines (notée en anglais par Protein Data Bank : PDB) par codes 3QEL.pdb et 4M48.pdb, respectivement.

Avant qu'elles soient dockées vis-à-vis les ligands candidats, elles ont été préparées soigneusement sur une chaine précise dans laquelle se cristallise le ligand natif (ligand co-cristallisé). La procédure de préparation de la cible protéique se fait tout d'abord par sa désolvatation en supprimant toutes les molécules d'eau ($H_2O$) liées à la protéine. Ensuite, on supprime toutes autres espèces chimiques en suspension (molécules ou ions). Puis, on ajoute les charges de Gasteiger, ainsi que les atomes d'Hydrogène polaires pour que les liaisons hydrogènes soient détectables dans le complexe (protéine-ligand).

Les sites actifs de chaque protéine sont préalablement déterminés vis-à-vis les ligands co-cristallisés, dont les résidus d'acides aminés Phe176, Gln110, et Glu236 sont les sites actifs de la protéine codée 3QEL.pdb en complexe avec Ifenprodil cristallisé en chaine D, comme exposé dans la Figure 5. Et les résidus d'acides aminés Tyr124, Phe43, et Phe325 sont les sites actifs de la protéine codée 4M48.pdb en complexe avec Nortriptylène cristallisé en chaine A de la cible protéique, comme présenté dans la Figure 6.

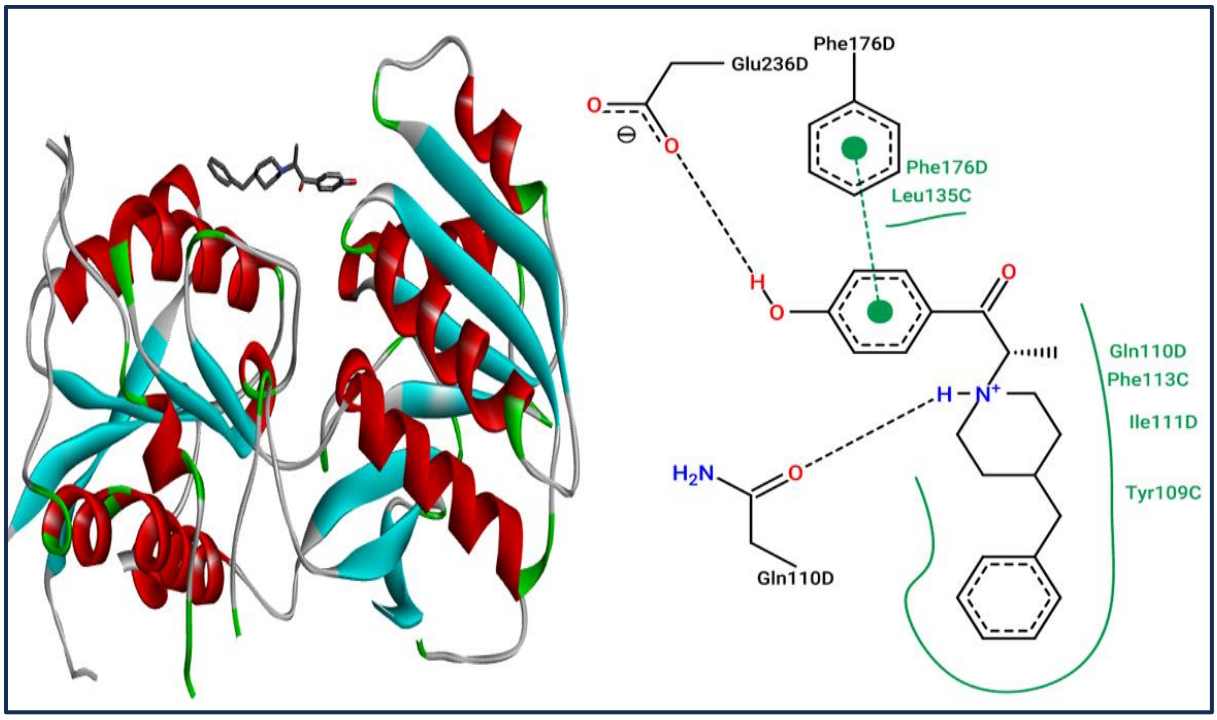



**Figure 5. Structure cristalline de la protéine préparée codée 3QEL.pdb, en complexe avec ligand co-cristallisé (partie gauche), et les sites actifs produits en chaine D de la cible protéique (partie droite)**

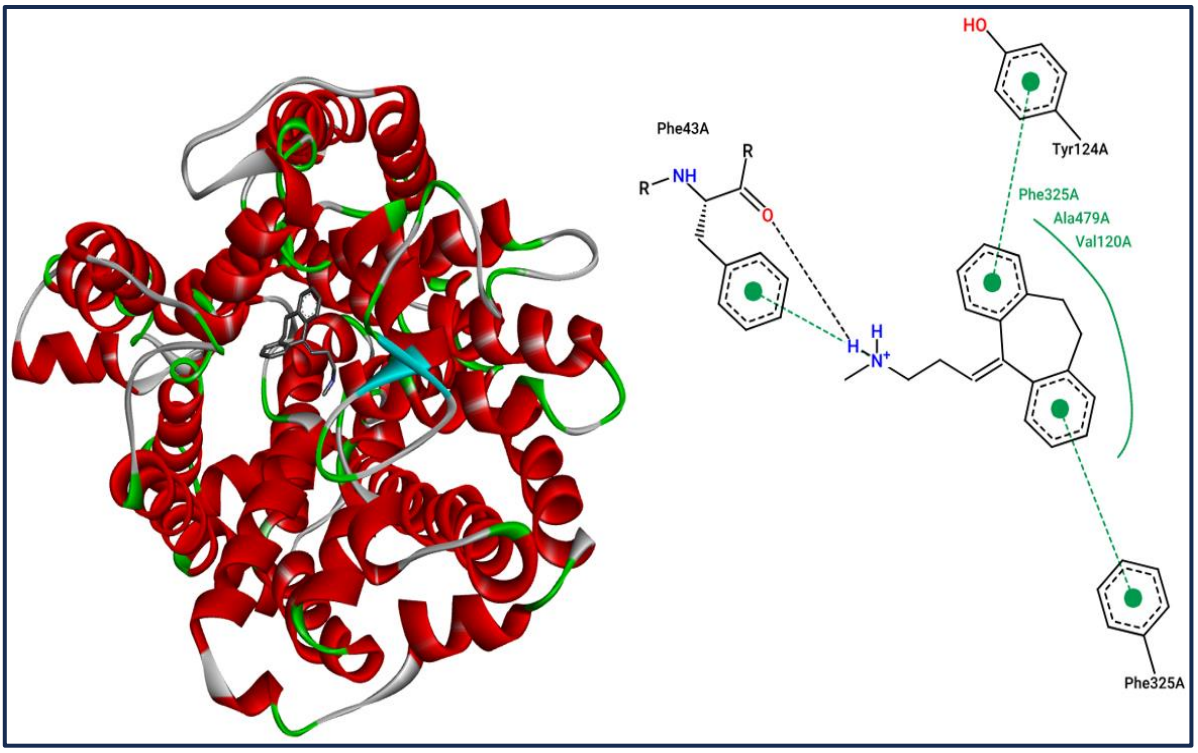



**Figure 6. Structure cristalline de la protéine préparée codée 4M48.pdb, en complexe avec ligand co-cristallisé (partie gauche), et les sites actifs produits en chaine A de la cible protéique (partie droite)**

## 5. Dynamique moléculaire

### 5.1. Définition

La dynamique moléculaire est une technique récemment développée pour simuler la trajectoire temporelle (ou dynamique) d'un système électronique (atomique ou moléculaire). Depuis sa conception, la technique de la dynamique moléculaire visait à examiner la stabilité des interactions intermoléculaires produites dans le complexe (ligand-protéine). Cependant, elle ne fait que répondre à la problématique posée par le Docking moléculaire ; dont la stabilité structurelle du complexe étudié ne sera pas absolument satisfaisante en raison des basses énergies de liaison en Kcal/mol. Pour cela, il est désormais nécessaire d'appliquer la technique de la dynamique moléculaire, pour mieux comprendre le mécanisme d'inhibition, ainsi que le comportement des petites molécules vis-à-vis les macromolécules (les cibles protéiques) en échelle atomique précise et une résolution temporelle bien détaillée [47–49], et surtout le suivi temporel des énergies de liaison hydrogènes, énergies de Van der Waals, et énergies de liaisons covalentes, grâce à l'approche de la mécanique moléculaire en borne généralisée et zone de surface, notée en anglais par : « Molecular Mechanics with Generalized Born and Surface Area (MMGBSA) » [50,51].

### 5.2. Principe

La simulation de la dynamique moléculaire repose sur l'utilisation de la loi fondamentale de la dynamique, connue en mécanique classique par la deuxième loi de Newton, comme elle est énoncée en équation 11 [52]. Contrairement à la première loi du mouvement de Newton, la seconde loi stipule que toutes les forces appliquées sur un corps sont en état de déséquilibre. Cela veut dire que la force nette est non nulle, elle est équivalente au taux de variation de la quantité de mouvement, comme le produit de la masse par accélération, ce dernier étant défini comme la dérivée de la vitesse par rapport au temps. Cela signifie que la force agissant sur le corps est proportionnelle à l'accélération générée (ai), et inversement proportionnelle à sa masse (m) [53].

$$\sum_{i=1}^{n} \overrightarrow{F}\, i(t) = m \times \overrightarrow{a}\, i(t). \qquad Equation\ 11$$

### 5.3. Objectif

La dynamique moléculaire est destinée au suivi du comportement dynamique d'un système électronique donné grâce au changements thermique produit lors de sa réaction avec une autre molécule, tout en détectant les changements conformationnels générés avec le temps de simulation estimé en picosecondes (ps) ou nanosecondes (ns).

## 6. Serveurs, plates-formes, et logiciels utilisés

Nos investigations computationnelles basées sur les simulations moléculaires, ont été réalisées grâce à une variété des serveurs, plates-formes, et logiciels gratuits et en libre accès. On cite sous-titre d'exemple : *ChemDraw Professional 16.0,* a été utilisé pour dessiner les composés chimiques (molécules étudiées) [53]; *ACD/ChemSketch,* a été utilisé pour calculer certains types des descripteurs moléculaires bidimensionnels [54,55]; *ChemBio3D,* a été utilisé pour calculer les descripteurs moléculaires liés aux structures 3D des molécules [56]; *Gaussien 09W,* a été utilisé pour calculer les descripteurs moléculaires géométriques à l'aide de théorie de la fonctionnelle de densité (notée par DFT) [57]; *XLSTAT 2014,* a été utilisé dans l'étude QSAR pour réduire la dimension des descripteurs calculés par l'analyse en composantes principales (ACP), pour générer les modèles mathématiques prédictifs de la régression linéaire multiple (RLM) et régression non linéaire multiple (RNLM), et les tester par la validation externe et interne (validation croisée LOO) [58]; *IBM SPSS Statistics 22,* a été utilisé pour effectuer le domaine d'applicabilité de williams [59] ; *java Platform SE binary,* a été utilisé pour réaliser le test d'hasardisation en Y [60]; *SYBYL-X 2.0,* a servi à l'élaboration de l'étude

3D-QSAR qui se repose sur les modèles de CoMFA et CoMSIA ; *pkCSM et SwissADME,* sont des serveurs en ligne pour mettre en œuvre des prédictions in-silico de l'absorption, distribution, métabolisme, excrétion, et de la toxicité des molécules étudiées [61–64] ; *Discovery Studio 2021 (BIOVIA),* a été utilisé pour préparer les complexes (ligand-protéine), et également pour visualiser les interactions intermoléculaires produites en 2D et 3D [65]; *AutoDock 4.2,* a été utilisé pour faire le Docking moléculaire entre ligands et protéines [67,68]; *Desmond, a package of Schrodinger LLC,* a été utilisé pour vérifier la stabilité moléculaire des complexes à l'aide de la technique de dynamique moléculaire [69].

**Références**

[1] S. Chtita, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments, (n.d.).

[2] D. Semrouni, Energétique et spectroscopie de peptides attachés à un cation métallique, champ de force de seconde génération et chimie quantique, (n.d.).

[3] N. Piekuś-Słomka, M. Zapadka, B. Kupcewicz, Methoxy and methylthio-substituted trans-stilbene derivatives as CYP1B1 inhibitors – QSAR study with detailed interpretation of molecular descriptors, Arabian Journal of Chemistry 15 (2022) 104204. https://doi.org/10.1016/j.arabjc.2022.104204.

[4] R. Hmamouchi, M. Bouachrine, Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété, (2016).

[5] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References, John Wiley & Sons, 2009.

[6] A. Goulon-Sigwalt-Abram, Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments, (n.d.).

[7] M. Karelson, Molecular descriptors in QSAR/QSPR, Wiley-Interscience, New York, 2000.

[8] A.T. Balaban, Highly discriminating distance-based topological index, Chemical Physics Letters 89 (1982) 399–404. https://doi.org/10.1016/0009-2614(82)80009-2.

[9] H. Wiener, Structural Determination of Paraffin Boiling Points, J. Am. Chem. Soc. 69 (1947) 17–20. https://doi.org/10.1021/ja01193a005.

[10] K. Chandrabose, Rationalization of Physicochemical Properties of Alkanoic Acid Derivatives towards Histone Deacetylase Inhibition, Internet Electronic Journal of … (2006). https://www.academia.edu/1976308/Rationalization_of_Physicochemical_Properties_of_Alkanoic_Acid_Derivatives_towards_Histone_Deacetylase_Inhibition (accessed July 23, 2023).

[11] R.A. Saunders, J.A. Platts, Scaled polar surface area descriptors: development and application to three sets of partition coefficientsElectronic supplementary information (ESI) available: tables giving the coefficients and t-ratios of the models, the statistics for the test sets and the full data set and descriptors employed. See http://www.rsc.org/suppdata/nj/b3/b307023a/, New J. Chem. 28 (2004) 166. https://doi.org/10.1039/b307023a.

[12] D. Jung, J. Floyd, T.M. Gund, A comparative molecular field analysis (CoMFA) study using semiempirical, density functional,ab initio methods and pharmacophore derivation using DISCOtech on sigma 1 ligands, J. Comput. Chem. 25 (2004) 1385–1399. https://doi.org/10.1002/jcc.10410.

[13] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967. https://doi.org/10.1021/ja00226a005.

[14] G. Klebe, U. Abraham, T. Mietzner, Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity, J. Med. Chem. 37 (1994) 4130–4146. https://doi.org/10.1021/jm00050a010.

[15] modlin.pdf, (n.d.). https://www.math.univ-toulouse.fr/~besse/pub/modlin.pdf (accessed July 28, 2023).

[16] leilclic168.pdf, (n.d.). https://www.mcours.net/cours/pdf/leilclic1/leilclic168.pdf (accessed July 28, 2023).

[17] G. Saporta, Probabilités, analyse des données et statistique, 2e éd. révisée et augmentée, Ed. Technip, Paris, 2006.
[18] A. Tropsha, P. Gramatica, V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci. 22 (2003) 69–77. https://doi.org/10.1002/qsar.200390007.
[19] P. Refaeilzadeh, L. Tang, H. Liu, Cross-Validation, in: L. LIU, M.T. ÖZSU (Eds.), Encyclopedia of Database Systems, Springer US, Boston, MA, 2009: pp. 532–538. https://doi.org/10.1007/978-0-387-39940-9_565.
[20] M. Clark, R.D. Cramer, The Probability of Chance Correlation Using Partial Least Squares (PLS), Quantitative Structure-Activity Relationships 12 (1993) 137. https://www.academia.edu/5722246/The_Probability_of_Chance_Correlation_Using_Partial_Least_Squares_PLS (accessed July 29, 2023).
[21] S. Wold, L. Eriksson, S. Clementi, Statistical Validation of QSAR Results, in: H. van de Waterbeemd (Ed.), Methods and Principles in Medicinal Chemistry, 1st ed., Wiley, 1995: pp. 309–338. https://doi.org/10.1002/9783527615452.ch5.
[22] H. van der Voet, Comparing the predictive accuracy of models using a simple randomization test, Chemometrics and Intelligent Laboratory Systems 25 (1994) 313–323. https://doi.org/10.1016/0169-7439(94)85050-X.
[23] S. Ekins, G. Bravi, S. Binkley, J.S. Gillespie, B.J. Ring, J.H. Wikel, S.A. Wrighton, Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors, Drug Metab Dispos 28 (2000) 994–1002.
[24] 9789264085442-en.pdf, (n.d.). https://www.oecd-ilibrary.org/docserver/9789264085442-en.pdf?expires=1690720167&id=id&accname=guest&checksum=485BB2D03424F82001B4624EB1068814 (accessed July 30, 2023).
[25] M. Balls, A.P. Worth, R.D. Combes, The Validation of Alternative Test Methods, in: The History of Alternative Test Methods in Toxicology, Elsevier, 2019: pp. 307–314. https://doi.org/10.1016/B978-0-12-813697-3.00033-0.
[26] J.C. Dearden, The History and Development of Quantitative Structure-Activity Relationships (QSARs):, International Journal of Quantitative Structure-Property Relationships 1 (2016) 1–44. https://doi.org/10.4018/IJQSPR.2016010101.
[27] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 52 , Altern Lab Anim 33 (2005) 155–173. https://doi.org/10.1177/026119290503300209.
[28] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs., Environ Health Perspect 111 (2003) 1361–1375. https://doi.org/10.1289/ehp.5758.
[29] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701. https://doi.org/10.1002/qsar.200610151.

[30] A. Golbraikh, A. Tropsha, Beware of q2!, Journal of Molecular Graphics and Modelling 20 (2002) 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1.
[31] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[32] C.M. Chagas, S. Moss, L. Alisaraie, Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski's Rule of Five, International Journal of Pharmaceutics 549 (2018) 133–149. https://doi.org/10.1016/j.ijpharm.2018.07.046.
[33] R. Barret, Lipinski's Rule of Five, in: Therapeutical Chemistry, Elsevier, 2018: pp. 97–100. https://doi.org/10.1016/B978-1-78548-288-5.50006-8.
[34] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[35] W.J. Egan, K.M. Merz,, J.J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics, J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[36] W.J. Egan, G. Lauri, Prediction of intestinal permeability, Advanced Drug Delivery Reviews 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.
[37] I. Muegge, S.L. Heald, D. Brittelli, Simple Selection Criteria for Drug-like Chemical Matter, J. Med. Chem. 44 (2001) 1841–1846. https://doi.org/10.1021/jm015507e.
[38] S. Yalcin, Molecular Docking, Drug Likeness, and ADMET Analyses of Passiflora Compounds as P-Glycoprotein (P-gp) Inhibitor for the Treatment of Cancer, Curr Pharmacol Rep 6 (2020) 429–440. https://doi.org/10.1007/s40495-020-00241-6.
[39] A. Daina, V. Zoete, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules, ChemMedChem 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.
[40] B. Canault, Développement d'une plateforme de prédiction in silico des propriétés ADME-Tox, phdthesis, Université d'Orléans, 2018. https://theses.hal.science/tel-02296720 (accessed August 1, 2023).
[41] L.L.G. Ferreira, A.D. Andricopulo, ADMET modeling approaches in drug discovery, Drug Discovery Today 24 (2019) 1157–1165. https://doi.org/10.1016/j.drudis.2019.03.015.
[42] J. Cole, E. Davis, G. Jones, C.R. Sage, 3.12 - Molecular Docking—A Solved Problem?, in: S. Chackalamannil, D. Rotella, S.E. Ward (Eds.), Comprehensive Medicinal Chemistry III, Elsevier, Oxford, 2017: pp. 297–318. https://doi.org/10.1016/B978-0-12-409547-2.12352-2.
[43] P.F.W. Stouten, R.T. Kroemer, Docking and Scoring, in: Comprehensive Medicinal Chemistry II, Elsevier, 2007: pp. 255–281. https://doi.org/10.1016/B0-08-045044-X/00253-4.
[44] S. C, D.K. S., V. Ragunathan, P. Tiwari, S. A., B.D. P, Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease, Journal of Biomolecular Structure and Dynamics 40 (2022) 585–611. https://doi.org/10.1080/07391102.2020.1815584.
[45] D. Gentile, V. Patamia, A. Scala, M.T. Sciortino, A. Piperno, A. Rescifina, Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study, Marine Drugs 18 (2020) 225. https://doi.org/10.3390/md18040225.

[46] N. Diniyah, M.B. Alam, A. Javed, F.H. Alshammari, H.-J. Choi, S.-H. Lee, In silico and docking studies on the binding activities of Keap1 of antioxidant compounds in non-oilseed legumes, Arabian Journal of Chemistry 16 (2023) 104414. https://doi.org/10.1016/j.arabjc.2022.104414.
[47] S.A. Hollingsworth, R.O. Dror, Molecular Dynamics Simulation for All, Neuron 99 (2018) 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.
[48] X. Hu, Z. Zeng, J. Zhang, D. Wu, H. Li, F. Geng, Molecular dynamics simulation of the interaction of food proteins with small molecules, Food Chemistry 405 (2023) 134824. https://doi.org/10.1016/j.foodchem.2022.134824.
[49] Y. Wang, T. Liu, J. Xie, M. Cheng, L. Sun, S. Zhang, J. Xin, N. Zhang, A review on application of molecular simulation technology in food molecules interaction, Current Research in Food Science 5 (2022) 1873–1881. https://doi.org/10.1016/j.crfs.2022.10.012.
[50] N. Forouzesh, N. Mishra, An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor, Molecules 26 (2021) 2383. https://doi.org/10.3390/molecules26082383.
[51] M. Ylilauri, O.T. Pentikäinen, MMGBSA As a Tool To Understand the Binding Affinities of Filamin–Peptide Interactions, J. Chem. Inf. Model. 53 (2013) 2626–2633. https://doi.org/10.1021/ci4002475.
[52] Les équations du mouvement [La dynamique moléculaire], (n.d.). https://uel.unisciel.fr/chimie/modelisation/modelisation_ch05/co/11_calcul_d_une_trajectoire.html (accessed August 20, 2023).
[53] Newton's Second Law Of Motion - Derivation, Applications, Solved Examples and FAQs, BYJUS (n.d.). https://byjus.com/physics/newtons-second-law-of-motion-and-momentum/ (accessed August 20, 2023).
[54] Télécharger chemdraw pro 16.0 gratuitement (Windows), (n.d.). https://en.freedownloadmanager.org/users-choice/Chemdraw_Pro_16.0.html (accessed July 31, 2023).
[55] T. Österberg, U. Norinder, Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors, European Journal of Pharmaceutical Sciences 12 (2001) 327–337. https://doi.org/10.1016/S0928-0987(00)00189-5.
[56] Télécharger ChemSketch (gratuit) - Clubic, Clubic.com (n.d.). https://www.clubic.com/telecharger-fiche431205-chemsketch.html (accessed July 31, 2023).
[57] G.W.A. Milne, Software Review of ChemBioDraw 12.0, J. Chem. Inf. Model. 50 (2010) 2053–2053. https://doi.org/10.1021/ci100385n.
[58] Citation gaussienne 09 | Gaussian.com, (n.d.). https://gaussian.com/g09citation/ (accessed July 31, 2023).
[59] Introducing the XLSTAT Free edition, XLSTAT, Your Data Analysis Solution (2016). https://www.xlstat.com/en/news/introducing-the-xlstat-free-edition (accessed July 31, 2023).
[60] SPSS Statistics 22.0 Available for Download, (2020). https://www.ibm.com/support/pages/spss-statistics-220-available-download (accessed July 31, 2023).

[61] Télécharger Java SE Platform (gratuit) - Clubic, Clubic.com (n.d.). https://www.clubic.com/telecharger-fiche10901-java-se-platform.html (accessed July 31, 2023).
[62] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci Rep 7 (2017) 42717. https://doi.org/10.1038/srep42717.
[63] D.E.V. Pires, T.L. Blundell, D.B. Ascher, pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures, J. Med. Chem. 58 (2015) 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104.
[64] pkCSM, (n.d.). https://biosig.lab.uq.edu.au/pkcsm/prediction (accessed July 31, 2023).
[65] SwissADME, (n.d.). http://www.swissadme.ch/ (accessed January 7, 2023).
[66] BIOVIA Discovery Studio 2021 Client. Get the software safely and easily., Software Informer (2023). https://biovia-discovery-studio-2021-client.software.informer.com/ (accessed July 31, 2023).
[67] AutoDock 4.2.6 / AutoDockTools 1.5.6 – Suite of Automated Docking Tools – My Biosoftware – Bioinformatics Softwares Blog, (2019). https://mybiosoftware.com/autodock-4-2-3-autodocktools-1-5-6-suite-automated-docking-tools.html (accessed July 31, 2023).
[68] A.P. Norgan, P.K. Coffman, J.-P.A. Kocher, D.J. Katzmann, C.P. Sosa, Multilevel Parallelization of AutoDock 4.2, J Cheminform 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[69] Desmond | Schrödinger, (n.d.). https://www.schrodinger.com/products/desmond (accessed July 31, 2023).

# Chapitre III. Modélisation moléculaire d'inhibiteurs sélectifs du récepteur NMDA, exerçant un effet analgésique contre la douleur neuropathique

## Résumé

Une production scientifique, publiée en 2022, portant sur une classe structurelle de trente-deux antagonistes sélectifs de la sous-unité 2B du récepteur N-Méthyl-D-Aspartate (NMDA), qui ont été systématiquement examinés grâce aux certaines techniques de modélisation moléculaire, afin d'identifier les meilleurs inhibiteurs analgésiques contre les douleurs neuropathiques.

Tout d'abord, des études tridimensionnelles des relations quantitatives entre la structure moléculaire et l'activité biologique (3D-QSAR) basées sur l'analyse comparative des champs moléculaires et l'analyse comparative de l'indice de similarité moléculaire (noté respectivement par des modèles classiques : CoMFA, et CoMSIA), ont montré que les champs stériques, électrostatiques et accepteurs de liaisons hydrogène jouent un rôle crucial dans l'activité analgésique contre la douleur neuropathique.

En plus, les prédictions in-silico des propriétés pharmacocinétiques connues par l'abréviation ADMET, ont révélé que les ligands L1, L2 et L3 ont été prédits comme des inhibiteurs non toxiques des cytochromes 1A2, 2C19 et 2C9, et pouvant traverser passivement la barrière hémato-encéphalique (BHE) et ont la plus forte probabilité de pénétrer le système nerveux central (SNC).

Par ailleurs, les résultats du Docking moléculaire ont indiqué également que ces ligands actifs (L1, L2 et L3) interagissent spécifiquement avec les résidus d'acides aminés Phe176, Glu235, Glu236, Gln110, Asp136 et Glu178 de la cible protéique codée par 3QEL.pdb.

En résumé, les ligands symbolisés par L1, L2, et L3 ont été prédits par un bon profil d'ADMET, et considérés comme des agents puissants du système nerveux central, en réagissant avec les sites actifs du récepteur NMDA comme il a été produit par la simulation du Docking moléculaire. Par conséquent, ils pourraient être utilisés comme des médicaments analgésiques pour le traitement de la douleur neuropathique.


**Mots clés :** 3D-QSAR; ADMET; douleur neuropathique; activité analgésique; Docking moléculaire; NMDA.

# Chapitre III: 3D-QSAR, ADME-Tox in Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists

Mohamed El fadili [1,*], Mohammed Er-rajy [1], Hamada Imtara [2,*], Mohammed Kara [3], Sara Zarougui [1], Najla Altwaijry [4], Omkulthom M. Al kamaly [4], Aisha Al Sfouk [4], and Menana Elhallaoui [1]

[1] Engineering Materials, Modeling and Environmental Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, BP 1796 Atlas, Fez 30000, Morocco;
[2] Faculty of Arts and Sciences, Arab American University Palestine, P.O. Box 240, Jenin 44862, Palestine;
[3] Laboratory of Biotechnology, Conservation and Valorisation of Naturals Resources, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, BP 1796 Atlas, Fez 30000, Morocco;
[4] Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint AbdulrahmanUniversity, P.O. Box 84428, Riyadh 11671, Saudi Arabia;

Corresponding author: mohamed.elfadili@usmba.ac.ma

**Abstract:** A new class of selective antagonists of the N-Methyl-D-Aspartate (NMDA) receptor subunit 2B have been developed using molecular modeling techniques. The three-dimensional quantitative structure–activity relationship (3D-QSAR) study, based on comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) models, indicate that steric, electrostatic and hydrogen bond acceptor fields have a key function in the analgesic activity against neuropathic pain. The predictive accuracy of the developed CoMFA model ($Q^2$ = 0.540, $R^2$ = 0.980, $R^2$ pred = 0.613) and the best CoMSIA model ($Q^2$ = 0.665, $R^2$ = 0.916, $R^2$ *pred* = 0.701) has been successfully examined through external and internal validation. Based on ADMET in silico properties, L1, L2 and L3 ligands are non-toxic inhibitors of 1A2, 2C19 and 2C9 cytochromes, predicted to passively cross the blood–brain barrier (BBB) and have the highest probability to penetrate the central nervous system (CNS). Molecular docking results indicate that the active ligands (L1, L2 and L3) interact specifically with Phe176, Glu235, Glu236, Gln110, Asp136 and Glu178 amino acids of the transport protein encoded as 3QEL. Therefore, they could be used as analgesic drugs for the treatment of neuropathic pain.

**Keywords:** 3D-QSAR; ADMET; neuropathic pain; analgesic activity; molecular docking; NMDA.

## 1. Introduction

Chronic pain is one of the most frequent reasons for medical consultations [1], and can be considered as a critical cause of morbidity and disability [2]. For this reason, a variety of analgesics have been designed to treat chronic pain, such as opioids, antidepressants, anticonvulsants, and nonsteroidal anti-inflammatory drugs (NSAIDs) [3]. However, opioid analgesic (OA) treatment of chronic pain remains controversial, its efficacy is unclear, and it has been associated with undesirable secondary effects [4], similar to those of morphine, such as constipation, nausea, vomiting, pruritus, drowsiness, cognitive impairment, respiratory depression, tolerance, physical dependence and addiction [3,5]. Consequently, the search for new potent non-opioid analgesics (NOA), more effective on the central nervous system (CNS), is an absolute priority [3] because they are highly desired in medical practice [6]. In this regard, the pain control is a world health problem, indicating an ever-increasing need for the discovery of new molecules with better analgesic activity and reduced secondary effects [7]. For this purpose, novel isoxazole, piperidine linker moieties, and phenylpiperazine derivatives, illustrated in Figure 7, show potential as analgesic drugs, and have been introduced to treat these types of acute and chronic diseases, reacting similar to selective antagonists of the N-Methyl-D-Aspartate (NMDA) receptor subunit 2B that inhibit the binding of [3H]-ifenprodil to brain membranes and do not cause side effects associated with non-selective N-Methyl-D-Aspartate receptor antagonists [8]. NMDA receptors are ligand-gated cationic channels expressed in brain tissue and naturally activated by the simultaneous action of glutamate and glycine [9]; in addition, they are required for both brain development and many higher order functions [10]. At the same time, they can cause the death of neurons as a result of receptor hyperactivation [9]. Consequently, NMDA receptors are considered as high profile therapeutic targets in the treatment of pain and neurodegenerative diseases such as Alzheimer's, Huntington's and Parkinson's disease [11]. The present study is conducted using three-dimensional quantitative structure activity relationships (3D-QSARs) a commonly used technology to discover new drugs, identifying high affinity ligands for a targeted protein [12,13]. The robustness, predictability and reliability of the established CoMFA and CoMSIA models have been examined with the help of a cross validation technique for the training set (26 molecules), and external validation for the test set (6 molecules), as a crucial and decisive step to judge the prediction accuracy of the constructed model for a novel database [14,15]. During the second part, we predicted ADMET in silico properties [16] of 32 molecules compared to a co-crystallized ligand (ifenprodil) as an anti-hypertensive agent, with neuroprotective activity through its effects on N-Methyl-D-Aspartate (NMDA) receptors [17],

based on the BOILED-Egg predictive model and Lipinski, Ghose, Muegge, Veber and Egan rules. Lastly, the predicted inhibitors (L1, L2 and L3) were chosen for molecular docking simulations to study their intermolecular interactions towards the active sites of the protein target encoded as 3QEL [18], and this approach has been validated using docking validation protocol [19].

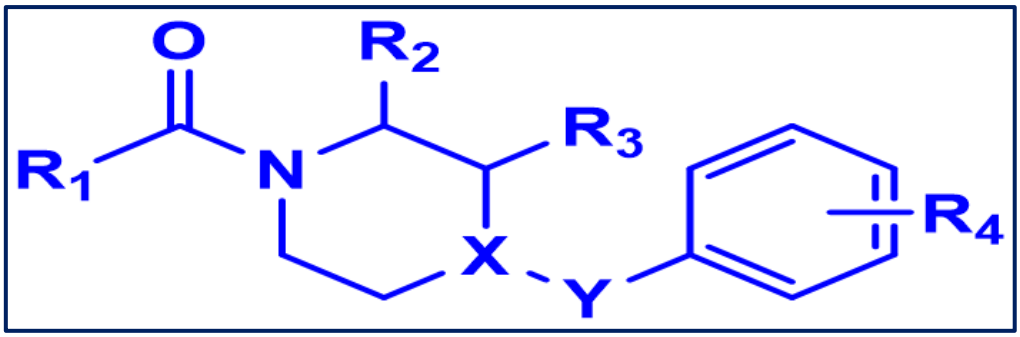


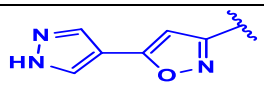
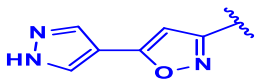
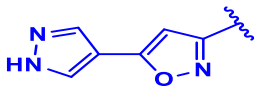
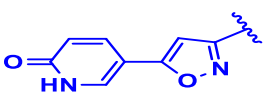
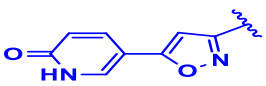

**Figure 7. Structural formula of isoxazole (1–6), piperidine linker moieties (6–14) and phenylpiperazine (12, 15–32) derivatives**

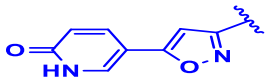
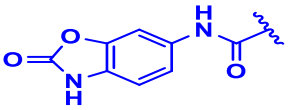
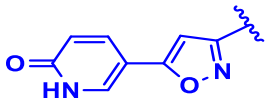
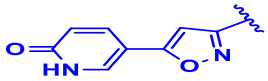
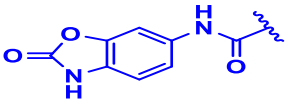
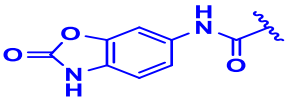
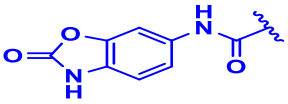
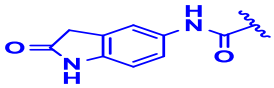
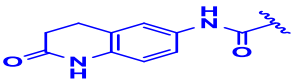
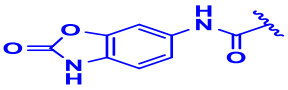
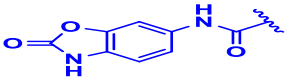
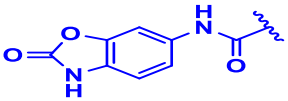
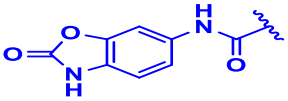
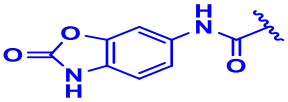
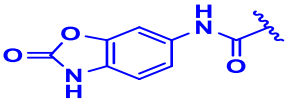
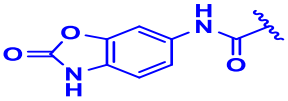
## 2. Materials and Methods

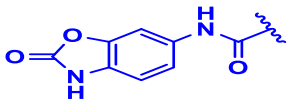
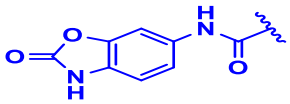
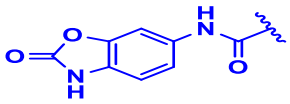
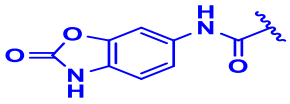
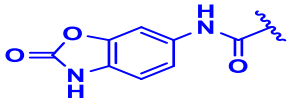
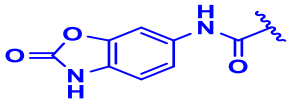
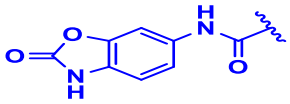
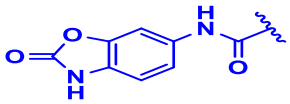
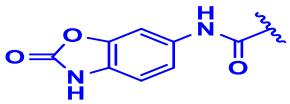
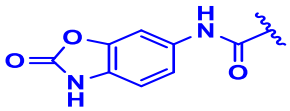
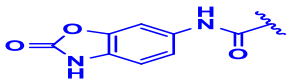
### 2.1.Experimental Database

To establish the 3D-QSAR models, we selected a set of 32 compounds from the recently published work of Anan, K et al., including NR2B-selective antagonists that inhibit the binding of [3H]-ifenprodil to rat brain membranes [8], as summarized in Figure 7. Then, we randomly divided this complete set into two subsets: the first one (training set) included 26 compounds that were used to build the model, and the second one (test set) included 6 compounds to validate the established model, as marked in Table 2.

**Table 2 . Studied compounds and their observed activities of $pIC_{50}$ order**

| Compounds | $R_1$ | $R_2$ | $R_3$ | $R_4$ | X | Y | $pIC_{50}$ (bind) |
|---|---|---|---|---|---|---|---|
| 1 |  | H | H | H | CH | $CH_2$ | 7.1 |
| 2 |  | H | H | 4-F | CH | $CH_2$ | 7.1 |
| 3 |  | H | H | 4-Cl | CH | $CH_2$ | 7.2 |
| 4 |  | H | H | H | CH | $CH_2$ | 7.2 |
| 5 |  | H | H | 4-F | CH | $CH_2$ | 7.3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 6 |  | H | H | 4-Cl | CH | $CH_2$ | 7.6 |
| 7 |  | H | H | 4-F | CH | $CH_2$ | 7.5 |
| 8 |  | H | H | 4-Cl | N | $CH_2$ | 6 |
| 9 * |  | H | H | 4-Cl | N | None | 6 |
| 10 |  | H | H | 4-Cl | CH | $CH_2$ | 7.5 |
| 11 |  | H | H | 4-Cl | N | $CH_2$ | 6.3 |
| 12 * |  | H | H | 4-Cl | N | None | 7.8 |
| 13 |  | H | H | 4-Cl | N | None | 6 |
| 14 |  | H | H | 4-Cl | N | None | 6 |
| 15 |  | H | H | H | N | None | 6 |
| 16 * |  | H | H | 4-F | N | None | 6.9 |
| 17 * |  | H | H | 4-$CF_3$ | N | None | 7.6 |
| 18 |  | H | H | 4-$OCH_3$ | N | None | 6.6 |
| 19 |  | H | H | 3.4-2F | N | None | 7.7 |
| 20 |  | H | H | 3.5-2F | N | None | 6.8 |
| 21 |  | H | H | 3-Cl.4-F | N | None | 7.5 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 22 |  | H | H | 3-F.4-Cl | N | None | 7.9 |
| 23 |  | H | H | 3-F.4-$CF_3$ | N | None | 7.8 |
| 24 * |  | H | H | 3-F.4-$OCH_3$ | N | None | 7.8 |
| 25 |  | $CH_3$ backward | H | 4-Cl | N | None | 7.5 |
| 26 |  | $CH_3$ forward | H | 4-Cl | N | None | 6 |
| 27 |  | H | $CH_3$ backward | 4-Cl | N | None | 7.2 |
| 28 |  | H | $CH_3$ forward | 4-Cl | N | None | 6.2 |
| 29 * |  | $CH_3$ backward | H | 3.4-2F | N | None | 7.3 |
| 30 |  | H | $CH_3$ backward | 3.4-2F | N | None | 7.2 |
| 31 |  | $CH_3$ backward | H | 3-F.4-Cl | N | None | 7.9 |
| 32 |  | H | $CH_3$ backward | 3-F.4-Cl | N | None | 7.3 |

* Indicates test set molecules.

### 2.2.Optimisation and Alignment

With the assistance of SYBYL-X 2.0 software (Tripos, Inc., United States of America USA), we first constructed the thirty-two molecules, which were optimized using the Tripos force fields and Gasteiger–Huckel atomic partial charges. Then, the convergence parameter of the Powell gradient algorithm was specified to be 0.001 kcal/(mol.A), with a maximum of 10,000 iterations to guarantee the conformational stability for each molecular system by making the energies in Kcal/mol completely minimized [20–22]. In the second stage, we aligned the set of molecules as one of the most significant steps to ensure the robustness of the CoMFA and CoMSIA models [23,24].

### 2.3.Development of 3D-QSAR Models

3D-QSAR models were generated via CoMFA and CoMSIA studies, using the partial least squares (PLS) analysis [25,26], where the steric and electrostatic fields of the CoMFA model were produced with a default energy cutoff of 30 kcal/mol. The attenuation factor and column filtering were defined as default values of 0.3 and 2.0 kcal/mol, respectively. Moreover, many additional physico-chemical descriptors, such as the hydrogen bond acceptor (HBA), hydrogen bond donor (HBD) and hydrophobic fields, were additionally calculated for the CoMSIA model in the identical conditions. A total of sixteen various combinations of fields were considered to develop the best CoMSIA model, as shown in Table 3.

**Table 3. Statistical results of CoMFA and CoMSIA models in various molecular field combinations using partial least squares analysis**

| | $Q^2$ | $R^2$ | *SEE* | *F* | *ONC* | $R^2$ *pr* | Fractions | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Steric | Electrostatic | Hydrophobic | Donor | Acceptor |
| CoMFA | 0.540 | 0.980 | 0.110 | 124.198 | 7 | 0.613 | 0.857 | 0.143 | - | - | - |
| CoMSIA/SEA | 0.667 | 0.903 | 0.228 | 37.421 | 5 | 0.420 | 0.304 | 0.524 | - | - | 0.172 |
| CoMSIA/SEH | 0.592 | 0.821 | 0.296 | 33.700 | 3 | 0.110 | 0.141 | 0.221 | 0.638 | - | - |
| CoMSIA/SED | 0.704 | 0.908 | 0.218 | 51.754 | 4 | 0.465 | 0.399 | 0.601 | - | 0 | - |
| CoMSIA/SHD | 0.537 | 0.755 | 0.347 | 22.618 | 3 | 0.131 | 0.204 | - | 0.796 | 0 | - |
| CoMSIA/SHA | 0.568 | 0.763 | 0.341 | 23.641 | 3 | 0.015 | 0.161 | - | 0.678 | - | 0.161 |
| CoMSIA/SDA | 0.320 | 0.564 | 0.462 | 9.505 | 3 | 0.522 | 0.5 | - | - | 0 | 0.5 |
| CoMSIA/EHA | 0.617 | 0.817 | 0.300 | 32.718 | 3 | 0.018 | - | 0.209 | 0.634 | - | 0.157 |
| CoMSIA/EHD | 0.579 | 0.756 | 0.338 | 35.624 | 2 | 0.010 | - | 0.211 | 0.789 | 0 | - |
| CoMSIA/EDA | 0.665 | 0.916 | 0.238 | 19.365 | 9 | 0.701 | - | 0.627 | - | 0 | 0.373 |
| CoMSIA/HDA | 0.546 | 0.762 | 0.342 | 23.449 | 3 | 0.025 | - | - | 0.801 | 0 | 0.199 |
| CoMSIA/EHDA | 0.617 | 0.817 | 0.300 | 32.718 | 3 | 0.018 | - | 0.209 | 0.634 | 0 | 0.157 |
| CoMSIA/SHDA | 0.568 | 0.763 | 0.341 | 23.641 | 3 | 0.015 | 0.161 | - | 0.678 | 0 | 0.161 |
| CoMSIA/SEDA | 0.667 | 0.903 | 0.228 | 37.421 | 5 | 0.420 | 0.304 | 0.524 | - | 0 | 0.172 |
| CoMSIA/SEHA | 0.620 | 0.808 | 0.307 | 30.844 | 3 | 0.007 | 0.122 | 0.172 | 0.568 | - | 0.138 |
| CoMSIA/SEHD | 0.592 | 0.821 | 0.296 | 33.700 | 3 | 0.110 | 0.141 | 0.221 | 0.638 | 0 | - |
| CoMSIA/SEHDA | 0.620 | 0.808 | 0.307 | 30.844 | 3 | 0.007 | 0.122 | 0.172 | 0.568 | 0 | 0.138 |

*Abbreviations:* ***$Q^2$***: the cross-validation determination coefficient, ***$R^2$***: the non-cross-validation determination coefficient, ***SEE***: the standard estimation error, ***F***: the Fischer test value, ***ONC***: the optimum number of components, and ***$R^2$ pr***: the external validation determination coefficient.

### 2.4.Partial Least Squares (PLS) Analysis

To model the linear relationship between a set of structural predictors and the response variable (analgesic activity of $pIC_{50}$ order) with a good quality of adjustment and a good predictive power [27,28], we have applied the partial least squares regression (PLSR) technology, with the help of SYBYL-X 2.0 software. The cross-validated determination coefficient (Q2) for an optimum number of principal components (*ONC*) was calculated using

the leave-one-out (LOO) procedure, and the non-cross-validated determination coefficient ($R^2$), *F*-test value and the standard estimation error (SEE) were calculated using the non-cross validation procedure. Where the best 3D-QSAR model was based on the optimum values of $Q^2$, $R^2$ must be greater than 0.5 and 0.6 [29,30] for the optimum number of components (*ONC*) and lowest standard error of estimation (SEE), respectively. To examine the reliability of the generated model, which was developed on 26 molecules (training set), we performed the external validation technique, using the 'predict' function included in the SYBYL package. Then, based on the previously developed model which was saved in sln format, we predicted the analgesic activities of $pIC_{50}$ order for 6 novel molecules (test set), as illustrated in Table 4. The required condition: $R^2ext > 0.6$ was verified afterwards using Excelstat software [31].

**Table 4. Results of external validation corresponding to CoMFA and CoMSIA/EDA models.**

| Compounds Number | Observed $pIC_{50}$ | Predicted $pIC_{50}$ of CoMFA | Predicted $pIC_{50}$ of CoMSIA/EDA |
|---|---|---|---|
| 9 * | 6 | 8.2 | 6.7 |
| 12 * | 7.8 | 9.6 | 7.3 |
| 16 * | 6.9 | 9.7 | 7.4 |
| 17 * | 7.6 | 10.0 | 7.4 |
| 24 * | 7.8 | 10.2 | 7.4 |
| 29 * | 7.3 | 10.6 | 7.5 |

* Indicates the test set molecules.

### 2.5.ADMET in Silico Pharmacocinetics

To identify the new drug with a high success level and reduced experimental study duration, it is mandatory to examine the absorption, distribution, metabolism, excretion and toxicity (ADMET) of these substances in the human organism before beginning the investigative process. [32,33], satisfying a few basic rules including Lipinski's [34], Veber's [35], Ghose's [36], Egan's [37,38] and Muegge's [39] regulations. This technology is equally applicable to remove the substances with potentially unfavorable physiological features, considering the pharmacokinetic qualities and toxicity [40]. For this purpose, we evaluated pharmacokinetic in silico characteristics of novel tested compounds as analgesic drugs against neuropathic pain, through the use of SwissADMET [41] and pkCSM [42] online servers, respectively.

### 2.6.Molecular Docking Modeling

Molecular docking is often used in computational chemistry to accelerate drug discovery at early stages. For this project, the molecular modeling technique was based on the cell key phenomenon, where the best position of the ligand or drug candidate (the agent) is the key that

can open the cell (or protein) to finally have a more stable complex by energetic order [43]. We started this study by extracting the protein responsible for the GluN2B subunit [44], coded as 3QEL.pdb from the protein data bank file [17], discovered by the X-ray diffraction method, in machine simulation with a resolution equal to 2.60 Å [45]. Then, we performed the dissolvation, removing all the water molecules bound to the protein; then we deleted the sodium atom and the suspended ligands, while adding the polar hydrogens using the discovery studio software, for the reason that the cavity method works best [46]. After the preparation of the protein and quoting its active sites, we launched the docking calculation in AutoDock 4.2 software [47]. With the help of algorithm AUTOGRID, we were able to centralize the grid box, putting the sizes 100, 100, and 100 in its three-dimensional structure with a spacing of 0.375 Å, and executing ten genetic algorithms, for a sum of 2,500,000 evals. Finally, we obtained the strongest complex out of the fifty obtained conformations [48], and we visualized 3D and 2D interactions of the protein–ligand with the use of discovery studio 2021 [49].

## 3. Results and Discussion

### 3.1.Molecular Alignement

To ensure the reliability and accuracy of CoMFA and CoMSIA models, the three-dimensional molecular structures of test and training sets were completely minimized and then aligned on the most active compound (molecule 22), which was selected as the template. *N*,*N*-diethylacetamide is the common core resulting from the superposition of the molecules on the template, as displayed in the Figure 8.

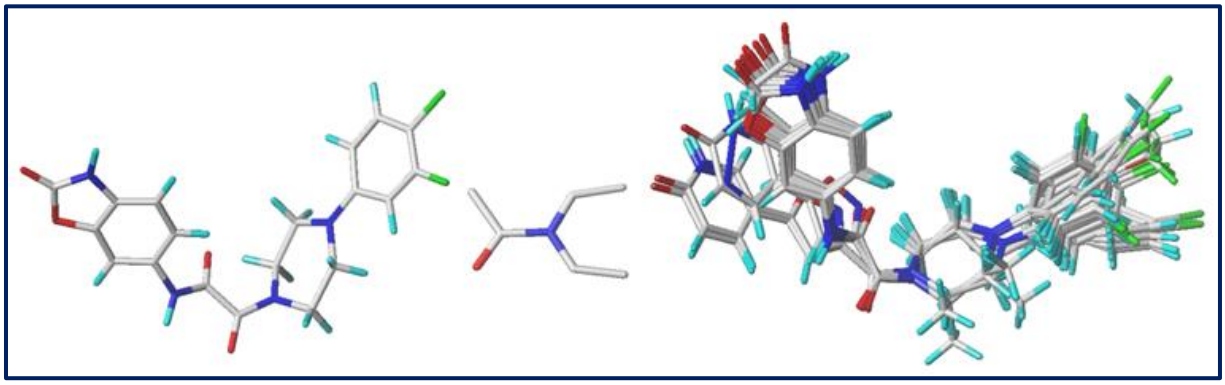


**Figure 8. 3D structure of the molecule 22 as template (left), 32 aligned and superposed molecules on the template (right) and the common core (middle).**

### 3.2.3D-QSAR Analysis

The present 3D-QSAR study was performed using comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) as two classical methods used to investigate the quantitative correlations between 3D molecular descriptors and

inhibitory activity [50]. The statistical results presented in Table 3 indicate that the CoMFA model obtained for an optimum number of principal components (*ONC* = 7) using the partial least squares analysis is given by an excellent non-cross-validated determination coefficient $R^2$ = 0.98 (much higher than 0.6), a good cross-validated determination coefficient $Q^2$ *cv* of 0.540, a high predictive value for the external validation of test set $R^2$ *pred* of 0.613, with a minimal standard estimation error SEE of 0.110 and *F*-test value of 124.198. According to the Golbraikh and Tropsha study, the established model is efficient and robust, because it has been successfully validated using the external validation performed for the test set ($R^2$ *pred* superior than 0.6), and the internal validation performed for the training test with the help of cross-validation technique (*Q2 cv* superior than 0.5) [30]. Therefore, the CoMFA analysis indicates that the biological activity of the $pIC_{50}$ order is affected by steric and electrostatic fields with a contribution rate of 85.7% and 14.3%, respectively. Moreover, we developed 16 various models from the comparative molecular similarity index analysis (CoMSIA), which included five three-dimensional descriptors, such as steric (S), electrostatic (E), hydrogen bond acceptor (HBA or A), hydrogen bond donor (HBD or D) and hydrophobic (H) fields; however, only the CoMSIA/EDA model was validated using internal and external validation. This mathematical model was produced for an optimum number of components (*ONC* = 9) using the partial least squares analysis, and given by a very good non-cross-validated determination coefficient $R^2$ = 0.916 (much higher than 0.6), a good cross-validated determination coefficient *Q2 cv* of 0.665, and a high predictive value for the external validation of test set $R^2$ *pred* of 0.701, with a minimal standard estimation error SEE of 0.238 and an *F*-test value of 19.365. Although the Golbraikh and Tropsha statistical criteria were verified, the non-cross-validated determination coefficient ($R^2$), the cross-validated determination coefficient ($Q^2$ *cv*) and the external validation of test set ($R^2$ *pred*) were greater than 0.6, 0.5 and 0.6 respectively; as such, the CoMSIA/EDA model was successfully validated, indicating that electrostatic and hydrogen bond acceptor fields have a key function in the analgesic activity against chronic pain, with a contribution rate of 62.7% and 37.3%, respectively.

### 3.3.A Graphical Explanation of CoMSIA/EDA and CoMFA Models

The structural features influencing the development of biological activity were visualized using the contour maps of the 3D-QSAR models [51], as shown in Figure 9, where the compound 22 structure was chosen as a reference. The green contours displayed in Figure 9A, contributing 80%, indicate the regions where bulky clusters can enhance the biological activity, while the yellow contours, contributing 20%, indicate the regions where bulky clusters decrease

the activity. Moreover, 80% of favorable contribution of the electrostatic field is expressed by the blue contours, and 20% of unfavorable contribution of the electrostatic field is represented by the red contours, as illustrated in the Figure 9B. In addition, the magenta-colored contour maps with 80% contribution indicate the hydrogen bond acceptor field that reinforces the activity, while the red contours of 20% contribution indicate the disadvantaged region, as shown in Figure 9C. We have noted the absence of the hydrogen bond donor (HBD) field, as shown in Figure 9D, since the results presented in Table 3 confirm no impact of HBD in the COMSIA models; therefore, steric (A), electrostatic (B) and hydrogen bond acceptor (C) fields are the main factors that influenced the biological activity. Additionally, we observed a green contour in meta and para positions, especially for the isoxazole derivatives, indicating that the selection of the voluminous substitution group in this area is necessary to make the ligand more active. Moreover, we detected a blue contour in the same disubstituted aromatic cycle in the meta and para positions, which makes the molecule most active. We also discovered, in the reverse part of the template molecule, a very important magenta contour in the [Oxazol-2(3H)-one] group, explaining the high activity of the inhibitor.

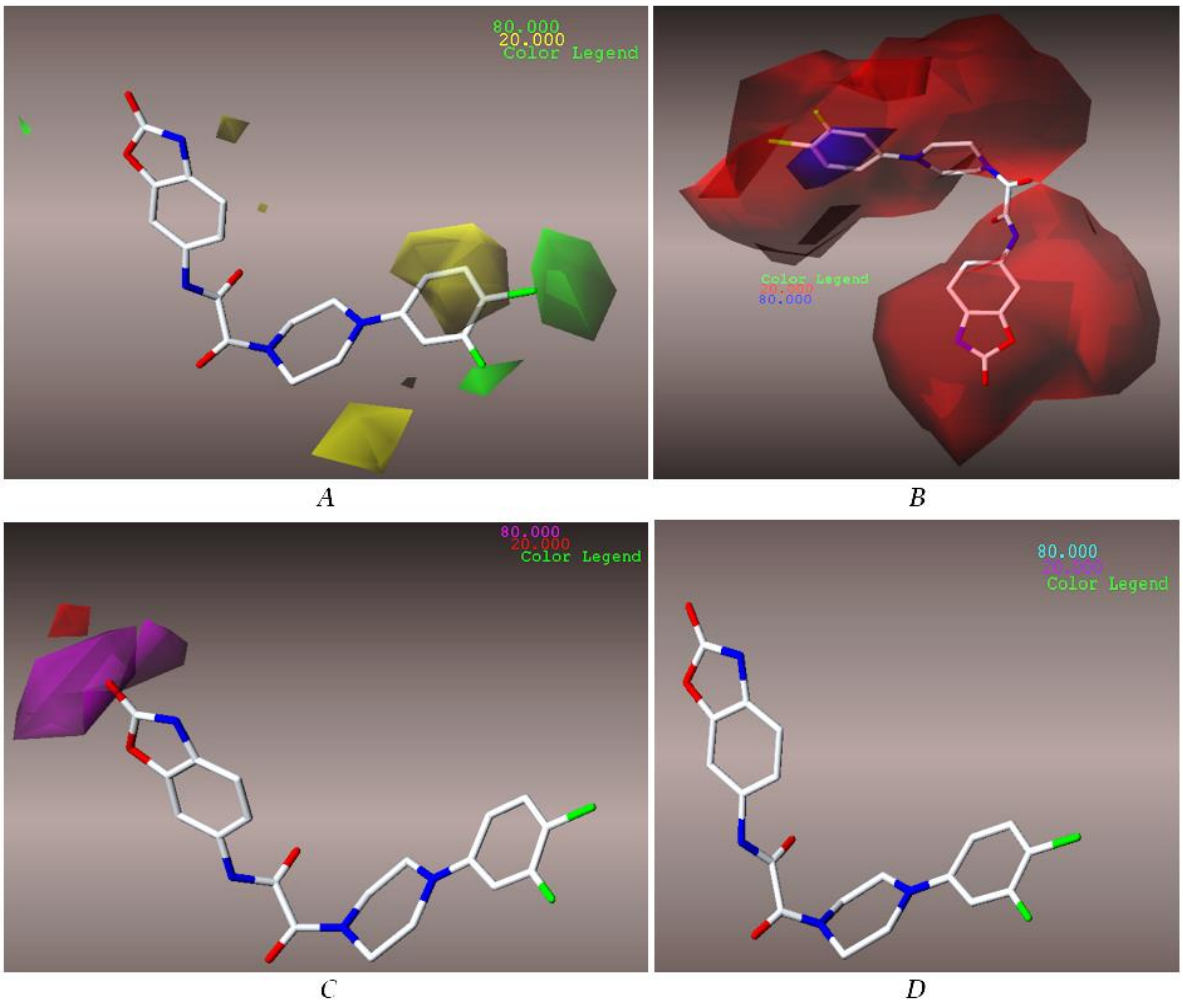



**Figure 9. Std*coeff. contour maps corresponding to steric (A), electrostatic (B), hydrogen bond acceptor (C) and hydrogen bond donor (D) fields explained for the compound 22 as a template**

### 3.4.ADMET in Silico Pharmacocinetics

Based on ifenprodil as an anti-hypertensive agent, with neuroprotective activity through its effects on N-Methyl-D-Aspartate (NMDA) receptors [17], we have tested the thirty-two molecules using BOILED-Egg as an accurate and reliable predictive model. This model is strongly applicable in the context of medicinal chemistry and drug discovery, focused on the evolution of lipophilicity, defined by the logarithm of the partition coefficient between n-octanol and water solvents (LogP O/W) in the Y-axis, as a function of polarity, given by the topological polar surface area (TPSA) in the X-axis [52]. Figure 10 presents only the L1, L2 and L3 ligands that are part of the yellow Egan Egg, due to the fact that they have a TPSA inferior to 80 $Å^2$; as such, they are the inhibitors with the greatest ability to cross the blood–brain barrier (BBB). In contrast, the other 29 molecules, which are part of the white egg zone, are most likely to be absorbed from the gastrointestinal tract [53]. Consequently L1, L2 and L3 inhibitors, as well as ifenprodil as a co-crystallized ligand, were further examined using

adsorption, distribution, metabolism, excretion and toxicity (ADMET) pharmacokinetic parameters on the basis of Lipinski, Veber, Egan, Muegge and Ghose rules [33,54]. Compared to ifenprodil, the results successively presented in Table 5 and Table 6 clearly illustrate that the three inhibitors respect the Lipinski thresholds, and satisfy the favorable violation numbers of Veber, Egan, Muegge and Ghose without exception. Moreover, they have efficient absorption at the level of human intestine (IAH over 89%), and a significant distribution, due to the fact that their volumes of distribution in humans are intended to be superior to −0.44 Log L/kg. At the blood–brain barrier (BBB), their permeabilities are significantly higher than −1 Log BB, except for the ligand L3. At the level of the central nervous system (CNS), they have permeabilities outside the (−2 to −3 Log PS) range, so they all access the central nervous system (CNS). In addition, they are all predicted to be inhibitors of the 1A2, 2C19 and 2C9 cytochromes; however, ifenprodil has been predicted to be an inhibitor of 1A2 and 2D6 cytochromes. Consequently, L1, L2 and L3 ligands are engineered as non-poisonous operators for the central nervous system (CNS), because of their high potential to cross the blood–brain barrier (BBB).

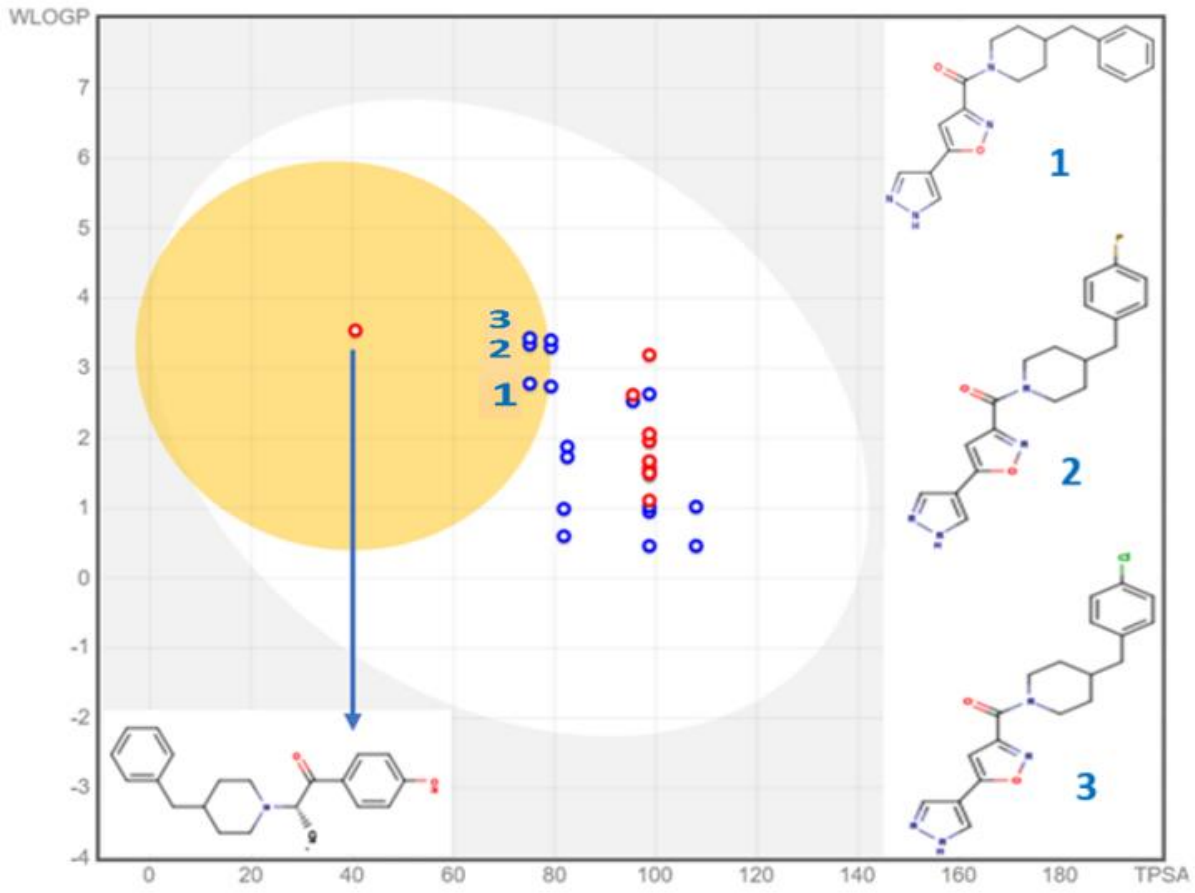



**Figure 10. BOILED-Egg predictive model of the co-crystallized ligand and the thirty-two inhibitors**

**Table 5. Physicochemical parameters prediction of ifenprodil, L1, L2 and L3 ligands on the basis of Lipinski, Veber, Egan, Muegge and Ghose regulations**

| Physical and Chemical Properties | | | | | | | Lipinski Violations | Veber Violations | Egan Violations | Ghose Violations | Muegge Violations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ligands number | Molecular weight (g/mol) | Molar refractive index | Rotatable bonds | Log P (Octanol /Water) | H bond acceptors | H-bond donors | | | | | |
| Threshold | ≤ 500 | 40 ≤ MR ≤ 130 | < 10 | < 5 | ≤ 10 | < 5 | Yes/No | Yes/No | Yes/No | Yes/No | Yes/No |
| L1 | 336.39 | 97.50 | 5 | 2.26 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| L2 | 354.38 | 97.46 | 5 | 2.41 | 5 | 1 | Yes | Yes | Yes | Yes | Yes |
| L3 | 370.83 | 102.51 | 5 | 2.53 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| Ifenprodil | 323.43 | 101.49 | 5 | 3.47 | 3 | 1 | Yes | Yes | Yes | Yes | Yes |

**Table 6. ADMET properties prediction of ifenprodil, L1, L2 and L3 ligands**

| | Absorption | | Distribution | | Metabolism | | | | | | | Excretion | Toxicity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ligands number | Intestinal Absorption (human) | VDss (human) | BBB permeability | CNS permeability | Substrate | Inhibitor | | | | | | Total Clearance | AMES toxicity |
| | | | | | CYP | | | | | | | | |
| | | | | | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | | |
| | Numeric (% Absorbed) | Numeric (Log L/kg) | Numeric (Log BB) | Numeric (Log PS) | Categorical (Yes/No) | | | | | | | Numeric (Log ml/min/kg) | Categorical (Yes/No) |
| L1 | 94.525 | 0.191 | −0.965 | −2.361 | Yes | Yes | Yes | Yes | Yes | No | Yes | 0.264 | Not toxic |
| L2 | 89.981 | 0.346 | −0.176 | −2.53 | No | Yes | Yes | Yes | Yes | No | No | 0.197 | Not toxic |
| L3 | 89.079 | 0.475 | −1.138 | −2.376 | No | Yes | Yes | Yes | Yes | No | No | 0.156 | Not toxic |
| Ifenprodil | 92.417 | 1.23 | 0.046 | −1.079 | Yes | Yes | Yes | No | No | Yes | No | 0.993 | Not toxic |

### 3.5.Molecular Docking

The results of molecular docking were concentrated on the transport protein of 3QEL.pdb connected to ifenprodil as a novel N-Methyl-D-Aspartate (NMDA) receptor antagonist that selectively inhibits receptors, including the NR2B subunit [55]. The crystal structure of this receptor was extracted from the protein data base (PDB) through the use of the X-ray diffraction process at a reasonable resolution equal to 2.6 Å [17,56]. Thus, the results of the intermolecular interactions established between L1, L2 and L3 ligands and the 3QEL encoded protein were compared to experimentally produced intermolecular interactions among the targeted protein and co-crystallized ligand **(**ifenprodil**)** thanks to the ProteinsPlus online server [57], which revealed that Phe176, Glu236 and Gln110 amino acids are the active locations of the targeted protein in its D-chain, as pictured in Figure 11.

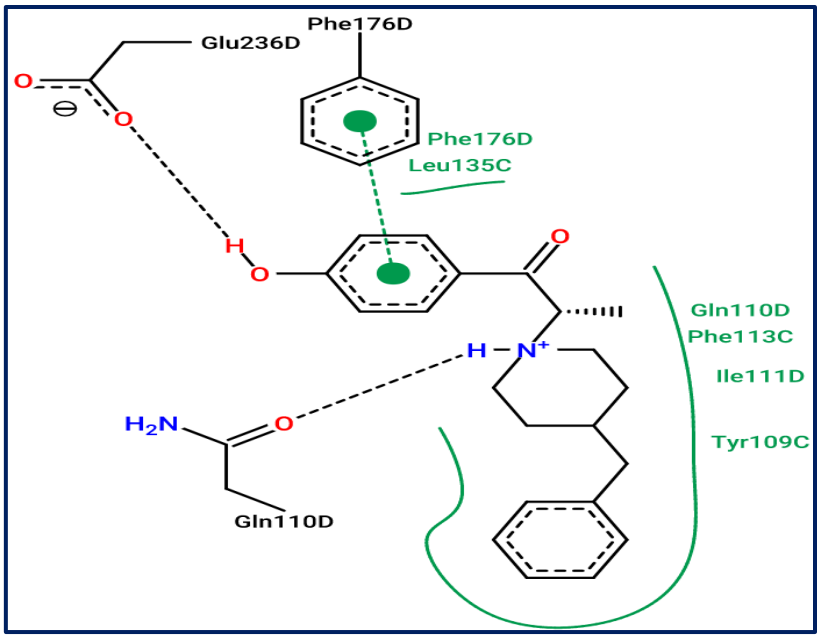



**Figure 11. Experimental pose view of ifenprodil with the protein's active sites**

### 3.6.Docking Validation Protocol

To validate the accuracy of the docking technique, we tested the efficiency of its algorithms, predicting the conformation of the ifenprodil as a co-crystallized ligand by means of the re-docking process, which is mainly based on the overlay of the docked ligand on the ligand attached to the protein in its D chain, as illustrated in Figure 12. Here, 3D (a) and 2D (b) visualizations of chemical interactions established between the docked ifenprodil (binding energy of −5.45 kcal/mol) and the responsible protein prove that Phe176, Glu236 and Gln110 amino acids are similar to those experimentally produced for co-crystallized ifenprodil. Additionally, the superposition result illustrated in Figure 12(c) reveals a root mean square deviation (RMSD) of 0.395 Å, smaller than 2, which provides an accurate pose prediction. Therefore, the molecular docking protocol is successfully validated [19].

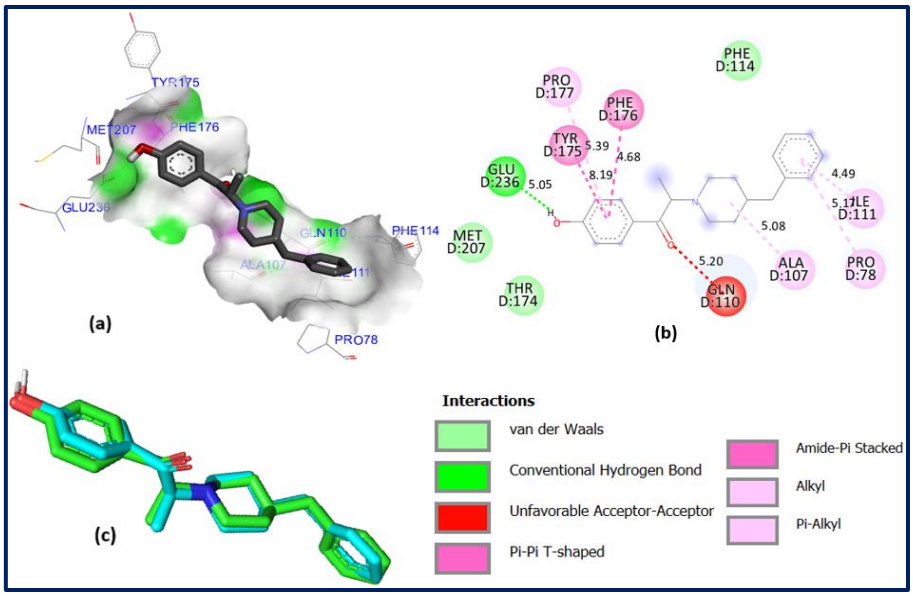
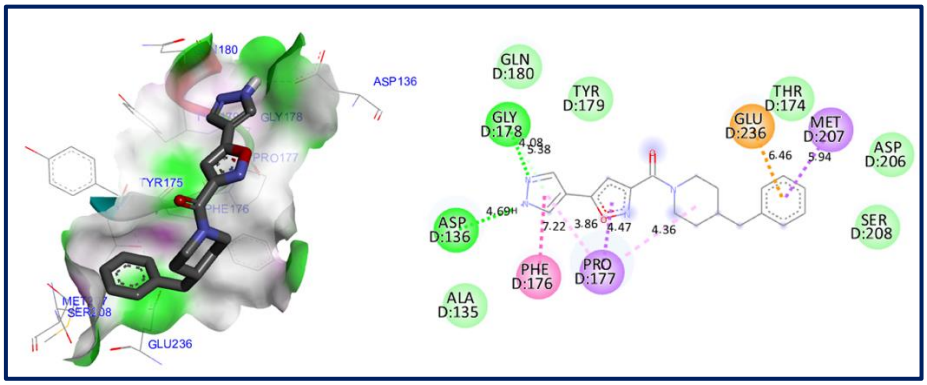



**Figure 12. 3D (a) and 2D (b) intermolecular interactions established between the targeted protein and docked ifenprodil with a binding energy equal to −5.45 kcal/mol, and redocking pose (c) with a RMSD value of 0.395 Å (original ifenprodil in cyan, and docked ifenprodil in green)**

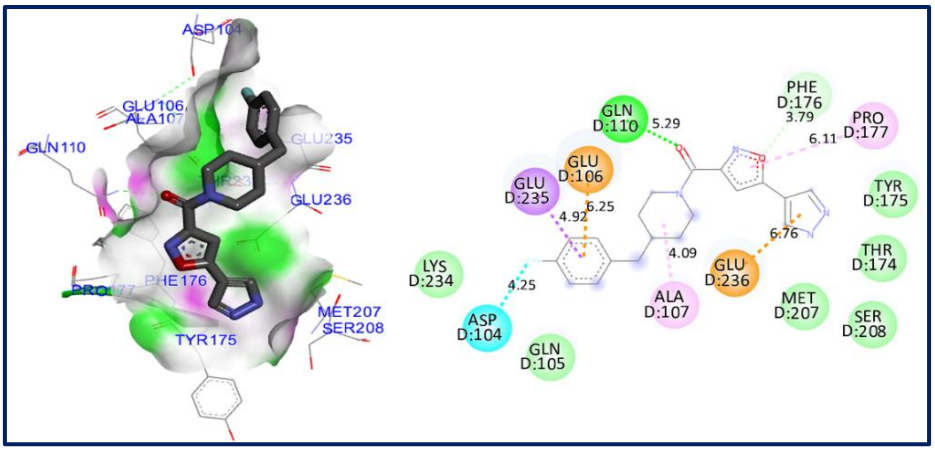
### 3.7.Molecular Docking Modeling for L1, L2 and L3 Ligands

Although the docking protocol was successfully validated, we docked L1, L2 and L3 inhibitors penetrating the blood–brain barrier (BBB) in the same active sites of the protein encoded as 3QEL, and we arrived at the molecular docking results presented in Figure 13. Firstly, the ligand L1 forms two hydrogen bonds, established between Asp136 and Glu178 amino acids and nitrogen atoms of the pyrazole group, with a nuclear distance of 4.69 Å and 4.08 Å, respectively. Moreover, it forms with the same group a chemical bond of amide-pi stacked-type with Phe176 amino acid at a distance of 7.22 Å, in addition to a pi-anion chemical bond with Glu236 amino acid at a nuclear distance of 6.46 Å. Secondly, the ligand L2 forms a hydrogen bond established between the oxygen atom of the acetamide cluster and the Gln110 amino acid, at a nuclear distance of 5.29 Å, as well as a pi-sigma chemical bond with Glu235 amino acid at a distance of 4.92 Å, in addition to a pi-anion chemical bond with Glu236 amino acid at a distance of 6.76 Å, and carbon hydrogen bond linked to the Phe176 amino acid. Thirdly, the ligand L3 forms only two chemical bonds: the first one is a pi-sigma chemical bond created with Glu235 amino acid at a distance of 4.40 Å, and the second one is a pi-anion chemical bond established with Glu236 amino acid, with a nuclear distance of 5.59 Å. Consequently, the results of molecular docking prove that L1, L2 and L3 ligands share a pi-anion chemical bond as a common molecular interaction with Glu236 amino acid. L2 and L3 ligands form a common pi-sigma chemical bond with the Glu235 amino acid. Moreover, both L1 and L2 ligands react with the Phe176 amino acid. As such, compared to the active sites of

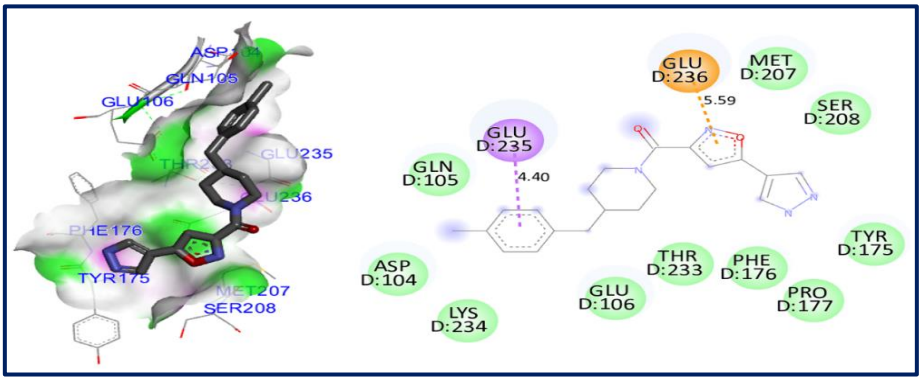
ifenprodil as a co-crystallized ligand, we conclude that Glu236, Glu235, Gln110, Phe176, Asp136 and Glu178 amino acids are the active sites in which L1, L2 and L3 ligands can inhibit the NMDA receptor subunit 2B, providing analgesic activity against neuropathic pain.




(L1-3QEL.pdb) complex




(L2-3QEL.pdb) complex




(L3-3QEL.pdb) complex

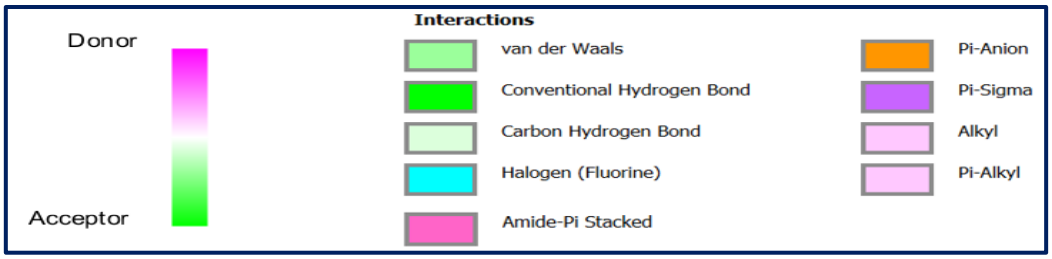



**Figure 13. 3D (left) and 2D (right) visualization of intermolecular interactions between L1, L2 and L3 ligands and the transport protein encoded as 3QEL, corresponding to the following binding energies in kcal/mol: −6.45, −5.57 and −6.32, respectively**

## 4. Conclusion

A systematic in silico study was conducted on a set of thirty-two selective antagonists of N-Methyl-D-Aspartate receptor subunit 2B to identify successful analgesic medications to treat neuropathic pain. Firstly, CoMFA and CoMSIA/EDA models were developed using the 3D-QSAR technique, and were evaluated using internal and external validation, indicating a significant effect of steric, electrostatic and hydrogen bond acceptor fields on the analgesic activity, which is explained through the presence of the [Oxazol-2(3H)-one] group and the halogen-like atomic groups in the meta and para positions of the aromatic cycle of the isoxazole derivatives. Afterwards, in silico ADMET pharmacokinetics prediction demonstrated a desirable profile of L1, L2 and L3 ligands, which were predicted as non-toxic inhibitors for 1A2, 2C19 and 2C9 cytochromes, respecting Lipinski, Veber, Egan, Ghose and Muegge rules, with a high level of absorption that exceeds 89%, and the highest chance of crossing over into the central nervous system (CNS). Finally, the obtained results were further qualified and reinforced using molecular docking simulation, which affirmed that L1, L2 and L3 isoxazoles react specifically with Phe176, Glu235, Glu236, Gln110, Asp136 and Glu178 amino acids of the transport protein encoded as 3QEL. Consequently, they can be applied as therapeutics in the field of medicine to prevent neuropathic diseases. Nevertheless, they need to be exposed to in vitro and in vivo investigations to assess their safety and effectiveness as analgesic medication.

**References**

[1] C. Bertin, N. Delage, B. Rolland, L. Pennel, M. Fatseas, A.-P. Trouvin, J. Delorme, C. Chenaf, N. Authier, Analgesic opioid use disorders in patients with chronic non-cancer pain: A holistic approach for tailored management, Neuroscience & Bi-obehavioral Reviews. 121 (2021) 160–174. https://doi.org/10.1016/j.neubiorev.2020.12.015.

[2] M.A. Sheikholeslami, S. Ghafghazi, S. Parvardeh, S. Koohsari, S.H. Aghajani, R. Pouriran, L.A. Vaezi, Analgesic effects of cuminic alcohol (4-isopropylbenzyl alcohol), a monocyclic terpenoid, in animal models of nociceptive and neuropathic pain: Role of opioid receptors, L-arginine/NO/cGMP pathway, and inflammatory cytokines, European Journal of Pharma-cology. 900 (2021) 174075. https://doi.org/10.1016/j.ejphar.2021.174075.

[3] T. Noguchi, Y. Umezaki, R. Takano, T. Fujimoto, Y. Domon, K. Kubota, K. Ueda, Y. Kawase, K. Yabe, M. Yokoyama, K. Shimada, A novel [5.2.1]bicyclic amine is a potent analgesic without μ opioid activity, Bioorganic & Medicinal Chemistry Letters. 36 (2021) 127790. https://doi.org/10.1016/j.bmcl.2021.127790.

[4] Y.R. Villarreal, A.L. Stotts, S.M. Paniagua, K. Rosen, M. Eckmann, R. Suchting, J.S. Potter, Mindfulness predicts current risk of opioid analgesic misuse in chronic low back pain patients receiving opioid therapy, Journal of Contextual Behavior-al Science. 18 (2020) 111–116. https://doi.org/10.1016/j.jcbs.2020.08.011.

[5] K. Wtorek, J. Piekielna-Ciesielska, T. Janecki, A. Janecka, The search for opioid analgesics with limited tolerance liability, Peptides. 130 (2020) 170331. https://doi.org/10.1016/j.peptides.2020.170331.

[6] T. Arita, M. Asano, K. Kubota, Y. Domon, N. Machinaga, K. Shimada, Discovery of DS34942424: An orally potent analge-sic without mu opioid receptor agonist activity, Bioorganic & Medicinal Chemistry. 28 (2020) 115714. https://doi.org/10.1016/j.bmc.2020.115714.

[7] M.F. Rossato, S.M. Oliveira, G. Trevisan, M. Rotta, P. Machado, M.A.P. Martins, J. Ferreira, Structural improvement of compounds with analgesic activity: AC-MPF4, a compound with mixed anti-inflammatory and antinociceptive activity via opioid receptor, Pharmacology Biochemistry and Behavior. 129 (2015) 72–78. https://doi.org/10.1016/j.pbb.2014.12.001.

[8] K. Anan, M. Masui, A. Tazawa, M. Tomida, Y. Haga, M. Kume, S. Yamamoto, S. Shinohara, H. Tsuji, S. Shimada, S. Yagi, N. Hasebe, H. Kai, Discovery of NR2B-selective antagonists via scaffold hopping and pharmacokinetic profile optimization, Bioorganic & Medicinal Chemistry Letters. 29 (2019) 1143–1147. https://doi.org/10.1016/j.bmcl.2019.02.017.

[9] N.S. Temnyakova, D.A. Vasilenko, O.I. Barygin, M.Y. Dron, E.B. Averina, Y.K. Grishin, V.V. Grigoriev, V.A. Palyulin, M.V. Fedorov, D.S. Karlov, Ifenprodil-like NMDA receptor modulator based on the biphenyl scaffold, Mendeleev Communica-tions. 30 (2020) 342–343. https://doi.org/10.1016/j.mencom.2020.05.027.

[10] L.P. Wollmuth, K. Chan, L. Groc, The diverse and complex modes of action of anti-NMDA receptor autoantibodies, Neu-ropharmacology. 194 (2021) 108624. https://doi.org/10.1016/j.neuropharm.2021.108624.

[11] V.P. Zambre, R.B. Patil, J.N. Sangshetti, S.D. Sawant, Comprehensive QSAR studies reveal structural insights into the NR2B subtype selective benzazepine derivatives as N-Methyl-

-Aspartate receptor antagonists, Journal of Molecular Struc-ture. 1197 (2019) 617–627. https://doi.org/10.1016/j.molstruc.2019.07.068.
[12] M.A. Lill, Multi-dimensional QSAR in drug discovery, Drug Discovery Today. 12 (2007) 1013–1017. https://doi.org/10.1016/j.drudis.2007.08.004.
[13] S. Vilar, G. Cozza, S. Moro, Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery, Current Topics in Medicinal Chemistry. 8 (2008) 1555–1572. https://doi.org/10.2174/156802608786786624.
[14] A.K. Halder, T. Jha, Validated predictive QSAR modeling of N-aryl-oxazolidinone-5-carboxamides for anti-HIV protease activity, Bioorganic & Medicinal Chemistry Letters. 20 (2010) 6082–6087. https://doi.org/10.1016/j.bmcl.2010.08.050.
[15] K. Roy, S. Kar, R.N. Das, Validation of QSAR Models, in: Understanding the Basics of QSAR for Applications in Pharma-ceutical Sciences and Risk Assessment, Elsevier, 2015: pp. 231–289. https://doi.org/10.1016/B978-0-12-801505-6.00007-7.
[16] L. El Mchichi, A. El Aissouq, R. Kasmi, A. Belhassan, R. El-Mernissi, A. Ouammou, T. Lakhlifi, M. Bouachrine, In silico design of novel Pyrazole derivatives containing thiourea skeleton as anti-cancer agents using: 3D QSAR, Drug-Likeness studies, ADMET prediction and molecular docking, Materials Today: Proceedings. 45 (2021) 7661–7674. https://doi.org/10.1016/j.matpr.2021.03.152.
[17] R.P.D. Bank, RCSB PDB - 3QEL: Crystal structure of amino terminal domains of the NMDA receptor subunit GluN1 and GluN2B in complex with ifenprodil, (n.d.). https://www.rcsb.org/structure/3QEL (accessed May 11, 2022).
[18] M. Mazigh, C. El'mbarki, H. Hadni, M. Elhallaoui, QSAR studies combined with DFT-calculations and Molecular docking of polyamine-sensitive inhibitors of the NMDA receptor, Mediterr.J.Chem.,. 9 (2019) 164–174. https://doi.org/10.13171/mjc93190924930hh.
[19] M. Abdullahi, G.A. Shallangwa, A. Uzairu, In silico QSAR and molecular docking simulation of some novel aryl sulfona-mide derivatives as inhibitors of H5N1 influenza A virus subtype, Beni-Suef Univ J Basic Appl Sci. 9 (2020) 2. https://doi.org/10.1186/s43088-019-0023-y.
[20] W.P. Purcell, J.A. Singer, A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method, J. Chem. Eng. Data. 12 (1967) 235–246. https://doi.org/10.1021/je60033a020.
[21] A. Ghaleb, A. Aouidate, M. Bouachrine, T. Lakhlifi, A. Sbai, In Silico Exploration of Aryl Halides Analogues as Check-pointKinase 1 Inhibitors by Using 3D QSAR, Molecular Docking Study,and ADMET Screening, Adv Pharm Bull. 9 (2019) 84–92. https://doi.org/10.15171/apb.2019.011.
[22] M. Clark, R.D. Cramer, N. Van Opdenbosch, Validation of the general purpose tripos 5.2 force field, J. Comput. Chem. 10 (1989) 982–1012. https://doi.org/10.1002/jcc.540100804.
[23] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967. https://doi.org/10.1021/ja00226a005.
[24] G. Klebe, U. Abraham, T. Mietzner, Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity, J. Med. Chem. 37 (1994) 4130–4146. https://doi.org/10.1021/jm00050a010.

[25] L. Ståhle, S. Wold, 6 Multivariate Data Analysis and Experimental Design in Biomedical Research, in: Progress in Medici-nal Chemistry, Elsevier, 1988: pp. 291–338. https://doi.org/10.1016/S0079-6468(08)70281-9.
[26] B.L. Bush, R.B. Nachbar, Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA, J Computer-Aided Mol Des. 7 (1993) 587–619. https://doi.org/10.1007/BF00124364.
[27] R. Rosipal, N. Krämer, Overview and Recent Advances in Partial Least Squares, in: C. Saunders, M. Grobelnik, S. Gunn, J. Shawe-Taylor (Eds.), Subspace, Latent Structure and Feature Selection, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006: pp. 34–51. https://doi.org/10.1007/11752790_2.
[28] E. Polat, S. Gunay, A New Robust Partial Least Squares Regression Method Based on Multivariate MM-estimators, Inter-national Journal of Mathematics and StatisticsTM. 18 (2017) 82–99.
[29] A. Tropsha, P. Gramatica, V. Gombar, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, QSAR Comb. Sci. 22 (2003) 69–77. https://doi.org/10.1002/qsar.200390007.
[30] A. Golbraikh, A. Tropsha, Beware of q2!, Journal of Molecular Graphics and Modelling. 20 (2002) 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1.
[31] M. Baroni, S. Clementi, G. Cruciani, G. Costantino, D. Riganelli, E. Oberrauch, Predictive ability of regression models. Part II: Selection of the best predictive PLS model, J. Chemometrics. 6 (1992) 347–356. https://doi.org/10.1002/cem.1180060605.
[32] O. Daoui, S. Elkhattabi, S. Chtita, R. Elkhalabi, H. Zgou, A.T. Benjelloun, QSAR, molecular docking and ADMET proper-ties 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 (2021) e07463. https://doi.org/10.1016/j.heliyon.2021.e07463.
[33] O. Ebenezer, N. Damoyi, M.A. Jordaan, M. Shapi, Unveiling of Pyrimidindinones as Potential Anti-Norovirus Agents—A Pharmacoinformatic-Based Approach, Molecules. 27 (2022) 380. https://doi.org/10.3390/molecules27020380.
[34] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews. 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[35] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[36] S. Yalcin, Molecular Docking, Drug Likeness, and ADMET Analyses of Passiflora Compounds as P-Glycoprotein (P-gp) Inhibitor for the Treatment of Cancer, Current Pharmacology Reports. 6 (2020) 429–440. https://doi.org/10.1007/s40495-020-00241-6.
[37] W.J. Egan, Merz Kenneth M., J.J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics, J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[38] W.J. Egan, G. Lauri, Prediction of intestinal permeability, Advanced Drug Delivery Reviews. 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.

[39] I. Muegge, S.L. Heald, D. Brittelli, Simple Selection Criteria for Drug-like Chemical Matter, J. Med. Chem. 44 (2001) 1841–1846. https://doi.org/10.1021/jm015507e.
[40] D. Gomes, S. Silvestre, A.P. Duarte, A. Venuti, C.P. Soares, L. Passarinha, Â. Sousa, In Silico Approaches: A Way to Unveil Novel Therapeutic Drugs for Cervical Cancer Management, Pharmaceuticals. 14 (2021) 741. https://doi.org/10.3390/ph14080741.
[41] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.
[42] pkCSM, (n.d.). http://biosig.unimelb.edu.au/pkcsm/run_example? (accessed March 8, 2022).
[43] A. Serrano, B. Imbernón, H. Pérez-Sánchez, J.M. Cecilia, A. Bueno-Crespo, J.L. Abellán, QN-Docking: An innovative mo-lecular docking methodology based on Q-Networks, Applied Soft Computing. 96 (2020) 106678. https://doi.org/10.1016/j.asoc.2020.106678.
[44] M.L. Berger, D. Maciejewska, J.J. Vanden Eynde, M. Mottamal, J. Żabiński, P. Kaźmierczak, M. Rezler, I. Jarak, I. Piantani-da, G. Karminski-Zamola, A. Mayence, P. Rebernik, A. Kumar, M.A. Ismail, D.W. Boykin, T.L. Huang, Pentamidine ana-logs as inhibitors of [3H]MK-801 and [3H]ifenprodil binding to rat brain NMDA receptors, Bioorganic & Medicinal Chem-istry. 23 (2015) 4489–4500. https://doi.org/10.1016/j.bmc.2015.06.012.
[45] A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P.E. Bourne, H.M. Berman, The RCSB PDB information portal for structural genomics, Nucleic Acids Research. 34 (2006) D302–D305. https://doi.org/10.1093/nar/gkj120.
[46] Signals Lead Discovery - Chemical Analytics - PerkinElmer Informatics, PerkinElmer. (n.d.). https://perkinelmerinformatics.com/products/research/signals-lead-discovery/ (accessed September 19, 2021).
[47] A.P. Norgan, P.K. Coffman, J.-P.A. Kocher, D.J. Katzmann, C.P. Sosa, Multilevel Parallelization of AutoDock 4.2, J Cheminform. 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[48] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDock-Tools4: Automated docking with selective receptor flexibility, Journal of Computational Chemistry. 30 (2009) 2785–2791. https://doi.org/10.1002/jcc.21256.
[49] BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes®, (n.d.). https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/ (accessed November 17, 2021).
[50] K. El Khatabi, R. El-mernissi, I. Aanouz, M.A. Ajana, T. Lakhlifi, M. Shahinozzaman, M. Bouachrine, Benzimidazole De-rivatives in Identifying Novel Acetylcholinesterase Inhibitors: A Combination of 3D-QSAR, Docking and Molecular Dy-namics Simulation, Physical Chemistry Research. 10 (2022) 237–249. https://doi.org/10.22036/pcr.2021.277306.1895.
[51] H. Hadni, M. Elhallaoui, 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents, Heliyon. 6 (2020) e03580. https://doi.org/10.1016/j.heliyon.2020.e03580.
[52] M. El fadili, M. Er-Rajy, M. Kara, A. Assouguem, A. Belhassan, A. Alotaibi, N.N. Mrabti, H. Fidan, R. Ullah, S. Ercisli, S. Zarougui, M. Elhallaoui, QSAR, ADMET In Silico

Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia, Pharmaceuti-cals. 15 (2022) 670. https://doi.org/10.3390/ph15060670.
[53] A. Daina, V. Zoete, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules, ChemMedChem. 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.
[54] A. Daina, O. Michielin, V. Zoete, iLOGP: A Simple, Robust, and Efficient Description of n -Octanol/Water Partition Coeffi-cient for Drug Design Using the GB/SA Approach, J. Chem. Inf. Model. 54 (2014) 3284–3301. https://doi.org/10.1021/ci500467k.
[55] K. Williams, Ifenprodil, a novel NMDA receptor antagonist: site and mechanism of action, Curr Drug Targets. 2 (2001) 285–298. https://doi.org/10.2174/1389450013348489.
[56] A. Adamski, D. Kruszka, Z. Dutkiewicz, M. Kubicki, A. Gorczyński, V. Patroniak, Novel family of fused tricyclic [1,4]diazepines: Design, synthesis, crystal structures and molecular docking studies, Tetrahedron. 73 (2017) 3377–3386. https://doi.org/10.1016/j.tet.2017.05.015.
[57] Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server, (n.d.). https://proteins.plus/ (accessed March 8, 2022).

# Chapitre IV. Étude in-silico de nouveaux antagonistes potentiels du récepteur NMDA pénétrant dans le cerveau en tant qu'agents thérapeutiques contre les accidents vasculaires cérébraux (AVC)

## Résumé

L'activation des récepteurs NMDA induite par les sous-unités GluN2B revêt une grande importance dans les troubles du système nerveux central (SNC), notamment la maladie de Parkinson, la maladie d'Alzheimer et les accidents vasculaires cérébraux, en raison de sa forte contribution au phénomène d'excitotoxicité, ce qui fait des antagonistes sélectifs des récepteurs NMDA l'un des agents thérapeutiques potentiels pour le traitement des maladies neurodégénératives, en particulier les accidents vasculaires cérébraux (AVC).

La présente étude vise à examiner une famille structurelle de trente antagonistes des sous-unités GluN2B du récepteur N-méthyl-D-aspartate (NMDA) pouvant pénétrer au sein du cerveau, en utilisant la technique de la conception de médicaments assistée par ordinateur, afin de découvrir des médicaments hautement candidats pour les accidents vasculaires cérébraux de type ischémiques. Au début, l'analyse des propriétés physicochimiques et pharmacocinétiques ADMET a révélé que les composés C13 et C22 ont été prédits en tant qu'inhibiteurs non toxiques des cytochromes CYP2D6 et CYP3A4, avec une absorption intestinale chez l'homme supérieure à 90 %, mais également en tant qu'agents efficaces pour le système nerveux central (SNC) du fait de leurs plus grandes chances de franchir la barrière hémato-encéphalique (BHE). Comparés à l'ifenprodil, un ligand co-cristallisé complexé avec la protéine de transport codée sous 3QEL.pdb, nous avons noté que les composés chimiques C13 et C22 étaient définis par des profils ADME-Toxicité satisfaisants, répondant aux règles de Lipinski, Veber, Egan, Ghose, et Muegge.

De plus, les résultats du docking moléculaire indiquent que les ligands C22 et C13 réagissent spécifiquement avec les résidus d'acides aminés des sous-unités GluN1 et GluN2B du récepteur NMDA. Ces interactions intermoléculaires produites entre les candidats médicaments et la protéine ciblée dans la chaîne B restent stables sur 200 nanosecondes de temps de simulation de dynamique moléculaire. En conclusion, les ligands C22 et C13 sont hautement recommandés comme médicaments thérapeutiques contre l'AVC en raison de leur sécurité et de leur stabilité moléculaire vis-à-vis des récepteurs NMDA.


**Mots-clés** : NMDA ; SNC ; AVC ; ADMET ; Docking moléculaire ; Dynamique moléculaire.

# Chapitre IV: An in-silico Investigation Based on Molecular Simulations of Novel and Potential Brain-penetrant GluN2B NMDA Receptor Antagonists as Anti-stroke Therapeutic Agents

Mohamed El fadili [1,*], Mohammed Er-rajy [1], Wafa Ali Eltayb [2], Mohammed Kara [3], Hamada Imtara [4], Sara Zarougui [1], Nawal Al-Hoshani [5], Abdullah Hamadi [6], and Menana Elhallaoui [1]

[1] LIMAS Laboratory, Faculty of Sciences Dhar El Mehraz, Sidi Mohammed Ben Abdellah University, Fez, Morocco;
[2] Biotechnology Department, Faculty of Sciences and Technology, Shendi University, Shendi, Sudan;
[3] Laboratory of Biotechnology, Conservation and Valorisation of Naturals Resources, Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco;
[4] Faculty of Arts and Sciences, Arab American University Palestine, Jenin, Palestine;
[5] Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia;
[6] Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, University of Tabuk, Tabuk, Saudi Arabia;

**Abstract:** GluN2B-induced activation of NMDA receptors plays a key function in central nervous system (CNS) disorders, including Parkinson, Alzheimer, and stroke, as it is strongly involved in excitotoxicity, which makes selective NMDA receptor antagonists one of the potential therapeutic agents for the treatment of neurodegenerative diseases, especially stroke. The present study aims to examine a structural family of thirty brain-penetrating GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists, using virtual computer-assisted drug design (CADD) to discover highly candidate drugs for ischemic strokes. Initially, the physicochemical and ADMET pharmacokinetic properties confirmed that C13 and C22 compounds were predicted as non-toxic inhibitors of CYP2D6 and CYP3A4 cytochromes, with human intestinal absorption (HIA) exceeding 90 %, and designed to be as efficient central nervous system (CNS) agents due to the highest probability to cross the blood-brain barrier (BBB). Compared to ifenprodil, a co-crystallized ligand complexed with the transport protein encoded as 3QEL.pdb, we have noticed that C13 and C22 chemical compounds were defined by good ADME-Toxicity profiles, meeting Lipinski, Veber, Egan, Ghose, and Muegge rules. The molecular docking results indicated that C22 and C13 ligands react specifically with the amino acid residues of the NMDA receptor subunit GluN1 and GluN2B. These intermolecular interactions produced between the candidate drugs and the targeted protein in the B chain remain stable over 200 nanoseconds of molecular dynamics simulation time. In conclusion, C22 and C13 ligands are highly recommended as anti-stroke therapeutic drugs due to their safety and molecular stability towards NMDA receptors.

**Keywords:** NMDA; CNS; CADD; ADMET; BBB; Molecular Docking; Molecular Dynamics

## 1. Introduction

According to the 2015 Global Burden of Disease (GBD) report, strokes were classified as the second most common cause of death and the third major cause of disability in the world [1]. Ischemic strokes make up about 87% of all strokes and include oxidative stress, energy failure, apoptosis, brain edema, high-frequency amino acid neurotoxins, and intracellular calcium overload [2,3]. The present study was performed using a virtual screening of a set of thirty brain-penetrant GluN2B NMDAR antagonists, which were previously designed, and biologically evaluated as anti-stroke therapeutic agents, with the help of pharmacophore-merging strategy, through the combination of popular GluN2B ligands and 3-n-Butylphthalide (NBP) structures [4]. Nowadays, computer-assisted drug design (CADD) based on *in-silico* techniques has gained great importance and has been largely reported in the literature [5–9]. In this regard, we have examined GluN2B NMDAR antagonists, through the use of computer-assisted drug design (CADD) technology, to detect and identify non-toxic, highly effective CNS agents, intended to be the best anti-stroke therapeutic drugs. During the first part of this work, we examined all thirty NMDA receptor antagonists, using the BOILED-Egg as an accurate predictive model, highly practical in medicinal chemistry to discover the candidate drugs [10]. Thereafter, we predicted the physicochemical and pharmacokinetic properties of thirty compounds compared to ifenprodil as a co-crystallized ligand complexed to amino-terminal domains of the NMDA receptor subunit GluN1 and GluN2B, encoded as 3QEL.pdb [11] (https://www.rcsb.org/structure/3QEL), and we assessed their similarity to drugs with the assistance of bioavailability radars, taking into account the evaluation of six important parameters for the design of orally administered drugs, like lipophilicity, polarity, size, solubility, saturation, and flexibility [12,13]. In the second part, the molecular docking simulation was performed to explore the intermolecular interactions produced between the transport protein of 3QEL.pdb code, and the most active compounds scored as C13, C22, and C23 which were experimentally defined by the highest level of neuroprotective activity (survival rate of 81.91%, 89.80%, and 88.35%, respectively), meeting the five rules of Lipinski, and the bioavailability conditions of Veber, Egan, Muegge, and Ghose, with a good ADME-Tox profile [14]. Thereafter, the molecular mechanics with generalized born and surface area (MMGBSA) solvation was conducted to calculate the binding free energies of three studied complexes [15,16]. Then, the conformational changes of each complex were examined during 200 nanoseconds of simulation time, with the help of the molecular dynamics technique, a successful technology widely applied to examine the stability of chemical interactions obtained

by molecular docking, studying the behavior of proteins towards small molecules with atomic precision and detailed temporal resolution [17,18]. In the last step, the stability and reactivity of candidate drugs was also examined using the density functional theory (DFT) to discover the most powerful drug candidate for the treatment of ischemic stroke [19,20].

## 2. Materials and Methods

### 2.1.Experimental database

The current study was carried out on thirty brain-penetrating GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists, which were experimentally synthesized through the pharmacophore-merging strategy. This experimental study which was performed by Qinlong Xu et al in China concluded that about half of 3-ethylisobenzofuran-1(3H)-one derivatives (Figure 14) presented a superior neuroprotective activity to 3-n-butylphthalide (NBP) against NMDA-induced neurotoxicity in hippocampal neurons at 10 μM, as displayed in Table 7, and also concluded that the chemical compound C22 merits further functional evaluation as novel anti-stroke therapeutic agent [4]. For this reason, we have examined this experimental database using machine learning techniques, based on ADME and toxicity predictions, drug kinetics assessments, molecular docking, molecular dynamics simulations, and DFT studies to verify the toxicity, ADME pharmacokinetics properties, and stability of intermolecular interactions produced towards the targeted protein.

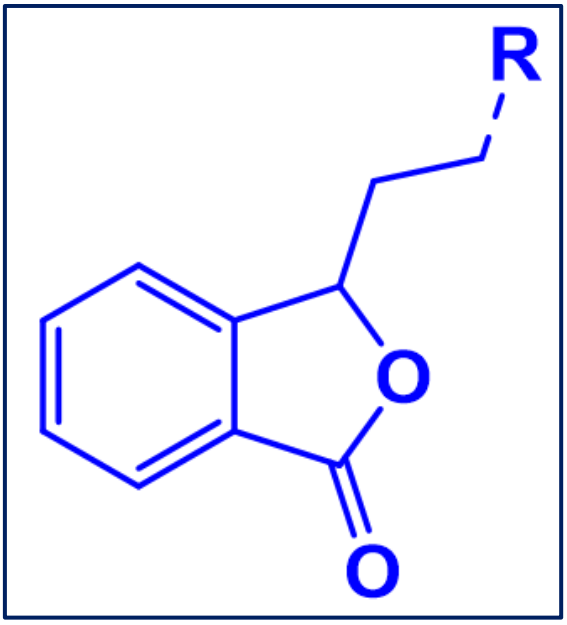
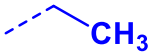
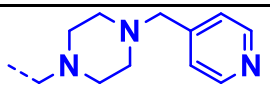
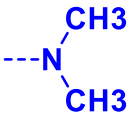
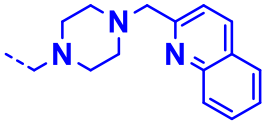
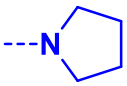
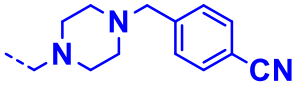
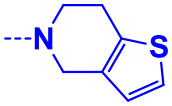
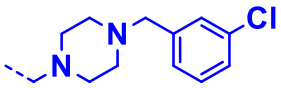
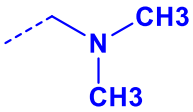
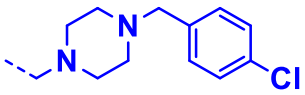
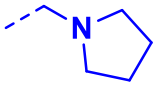
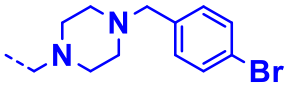
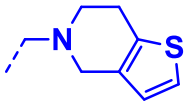
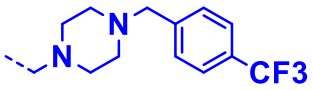
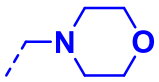
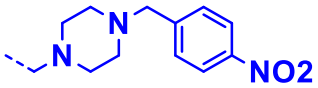
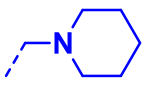
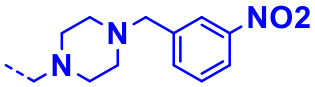
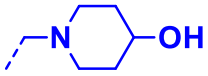
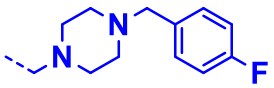
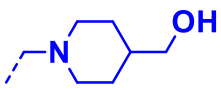
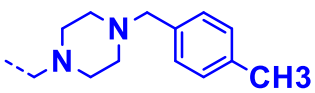
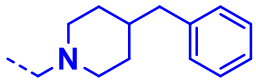
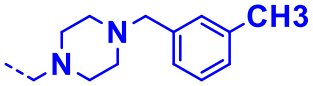
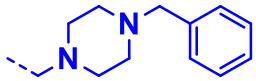
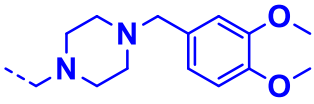
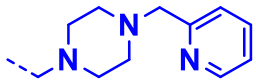
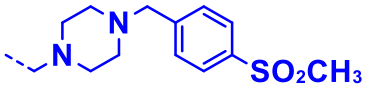
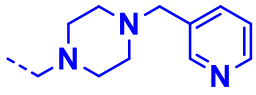
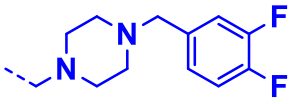



**Figure 14. Structural formula of 3-ethylisobenzofuran-1(3H)-one derivatives.**

**Table 7. Neuroprotection of thirty chemical compounds against NMDA-induced neurotoxicity in hippocampal neurons at 10 µM.**

| Compounds number | R | Survival rate (%) | Compounds number | R | Survival rate (%) |
|---|---|---|---|---|---|
| C1 |  | 73.45 | C16 |  | 75.54 |
| C2 |  | 73.56 | C17 |  | 62.96 |
| C3 |  | 70.66 | C18 |  | 60.33 |
| C4 |  | 66.05 | C19 |  | 62.96 |
| C5 |  | 78.18 | C20 |  | 60.33 |
| C6 |  | 73.16 | C21 |  | 46.13 |
| C7 |  | 66.52 | C22 |  | **89.80** |
| C8 |  | 76.22 | C23 |  | **88.35** |
| C9 |  | 77.85 | C24 |  | 73.28 |
| C10 |  | 77.13 | C25 |  | 72.2 |
| C11 |  | 78.14 | C26 |  | 68.38 |

| | | | | | |
|---|---|---|---|---|---|
| C12 |  | 79.12 | C27 |  | 67.43 |
| C13 |  | **81.91** | C28 |  | 48.44 |
| C14 |  | 74.65 | C29 |  | 61.02 |
| C15 |  | 63.33 | C30 |  | 75.45 |

### 2.2.In-silico pharmacokinetics, Drug-Likeliness, and ADME-Toxicity Prediction

A certain number of conditions must be satisfied before a drug can be approved for clinical trials. Initially, the BOILED-Egg model based on the calculation of lipophilicity given by the logarithm of the partition coefficient between n-octanol and water (Log $P_{O/W}$) and polarity signaled by the topological polar surface area (TPSA) of small molecules was applied to identify the potent central nervous system (CNS) agents, with the highest probability to cross the blood-brain barrier (BBB) [10,21]. Secondly, the bioavailability radars were also evaluated on this set of molecules to assess their similarity to drugs meeting the oral bioavailability, considering six molecular descriptors as: lipophilicity, polarity, size, solubility, saturation, and flexibility [22–24]. Moreover, the physicochemical properties of each compound were studied in the basis of Lipinski [25], Egan [26,27], Veber [28], Muegge [29], and Ghose [30] rules. Finally, the pharmacokinetic properties of absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) were predicted and compared to the ADME-Tox profile of ifenprodil as a selective NMDA receptor antagonist, with the help of the online SwissADMET (http://www.swissadme.ch/) and pkCSM (https://biosig.lab.uq.edu.au/pkcsm/prediction) servers [13,31].

### 2.3.Molecular Docking

Molecular docking is an efficient and successful tool for drug discovery [32]. It's largely applied in the literature to explore the binding affinity and the type of intermolecular interactions produced between the candidate drug and targeted protein [5,7,33–35]. For this purpose, we extracted the experimentally-determined 3D structure of the N-methyl-D-aspartate

(NMDA) receptor from the protein data bank (PDB) archive. The crystal structure of this transport protein encoded as 3QEL.pdb was discovered using the X-RAY diffraction method with a resolution of 2.60 Å [11] (https://www.rcsb.org/structure/3QEL). Then, we prepared our protein using the Discovery Studio 2021 (BIOVIA) package software [36], deleting the water molecules, ion sodium, ifenprodil, and all other co-crystallized ligands linked to protein. Moreover, the Gasteiger charges were computed and the polar hydrogens were added to improve the cavity performance. So, the prepared protein was docked to C13, C22, and C23 compounds as three most active ligands, using AutoDock 4.2.6 / AutoDockTools 1.5.6 software [37,38]. With the help of AUTOGRID algorithm, we centralized the grid box on (81.43 Å, 15.25 Å, -21.157 Å), putting the sizes: (100, 100, 100) in the 3D structure of (candidate ligand – 3QEL.pdb protein) complex, and executing 10 genetic algorithms of a maximum number of evals equal to 25000000 for each complex. Lastly, the intermolecular interactions resulting from the strongest complexes with the lowest binding energies in Kcal/mol were visualized using Discovery Studio 2021 software [7].

### 2.4.Molecular dynamics

To assess the level of molecular recognition of the candidate drug, we evaluated its molecular behavior towards the protein target, using the molecular dynamics technique performed by the use of Desmond software, a package from Schrödinger LLC [39]. For this goal, the output files of molecular docking were selected as input files of molecular dynamics, to verify the stability of the (C13, C22, C23 ligands – 3QEL.pdb protein) complexes, during 200 nanoseconds of molecular dynamics (MD) simulation time. Initially, the studied complexes were prepared using Protein Preparation Maestro and further optimized in the System Builder tool utilizing of the transferable interaction potential 3-point (TIP3P) solvent model, with an OPLS-type force field [40]. Subsequently, the adopted models were neutralized by the addition of the counter ions and water molecules to reproduce the physiological environment of 0.15 M salt (Na+, Cl-), with a pressure of 1 atm and a temperature of 300 K. Finally, the conformational changes of root mean square fluctuation (RMSF), root mean square deviation (RMSD), polar surface area (PSA), solvent accessible surface area (SASA), radius of gyration (Rg), and molecular surface area (MolSA) were all registered at 200 nanoseconds of molecular dynamics simulation time [41,42].

## 3. Results and discussion

### 3.1.Drug Kinetics Assessment, Drug-likeliness, and ADME-Toxicity Profile Analysis

The BOILED-Egg model has been applied on thirty antagonists of GluN2B N-methyl-D-aspartate receptor (NMDAR), which have been compared to ifenprodil, as a noncompetitive antagonist of this receptors type, known earlier thanks to its neuroprotective activity against hypertension [43]. The results presented in Figure 15, indicate that all the studied compounds are part of the yellow Egan eggs, except compound 23 which is part of the white eggs. Thus, compound 23 belonging to the white region of the egg, has the highest probability of being absorbed from the gastrointestinal tract, and all other compounds have the highest probability to cross the blood-brain barrier (BBB). The chemical compounds C1, C2, C5, C28, and ifenprodil ligand are marked in red, showing that they are predicted not to be effluated from the central nervous system (CNS) by the P-glycoprotein. Inversely, all other compounds colored in bleu, are predicted to be successfully evacuated from the CNS by the P-glycoprotein except molecule 23, because it is most likely to be passively absorbed by the gastrointestinal tract [10].

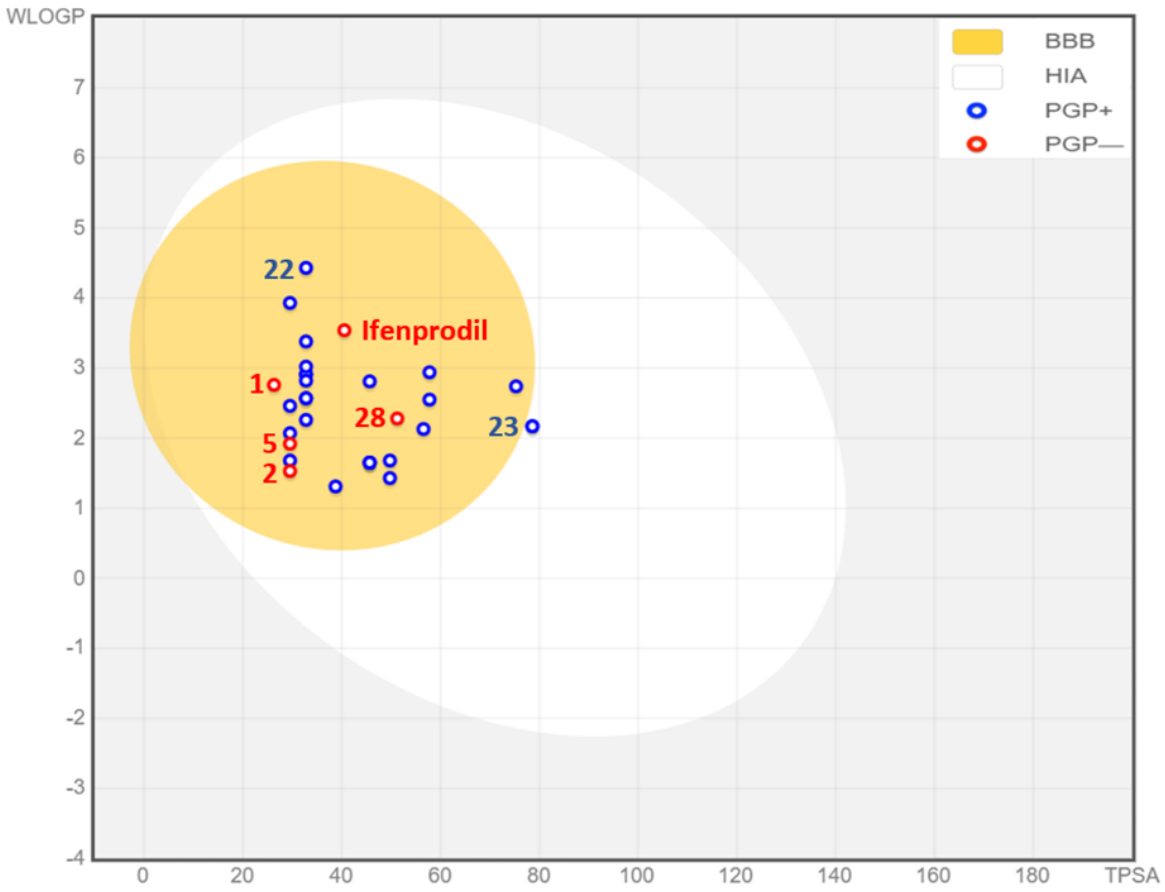



**Figure 15. BOILED-Egg model of thirty inhibitors and ifenprodil as co-crystallized ligand.**

Moreover, the predicted physicochemical properties of thirty GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists, were compared to ifenprodil, based on five rules of Lipinski, and the violations number of Veber, Egan, Ghose, and Muegge. The results presented in Table 8, demonstrate that GluN2B NMDA receptor antagonists satisfy all five rules of Lipinski, where: molecular weight (MW) ≤ 500, 40 ≤ molar refractivity (MR) ≤ 130, Log P (octanol/water) < 5, hydrogen bond acceptor (HBA) ≤ 10, and hydrogen bond donor (HBD) < 5. Then, the bioavailability number of Veber, Egan, Ghose, and Muegge is successfully verified, except for compound C1, which didn't satisfy the Muegge rule, because it's defined by a molecular weight inferior to 200 g/mol.

**Table 8. Physicochemical properties of ifenprodil and thirty GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists, based on Lipinski, Veber, Egan, and Ghose violations.**

| N° | Physico-chemical properties | | | | | | Lipinski violation | Veber violation | Egan violation | Ghose violation | Muegge violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Molecular weight (g/mol) | Molar refractive index | Rotatable bonds | Log P | H-BA | H-BD | Categorical (Yes/No) | | | | |
| Rule | ≤500 | 40≤ MR ≤130 | <10 | <5 | ≤10 | <5 | Yes/No | Yes/No | Yes/No | Yes/No | Yes/No |
| C1 | 190.24 | 54.99 | 3 | 2.53 | 2 | 0 | Yes | Yes | Yes | Yes | No |
| C2 | 205.25 | 57.89 | 3 | 2.39 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C3 | 231.29 | 69.31 | 3 | 2.66 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C4 | 299.39 | 87.02 | 3 | 2.95 | 3 | 3 | Yes | Yes | Yes | Yes | Yes |
| C5 | 219.28 | 62.70 | 4 | 2.64 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C6 | 245.32 | 74.11 | 4 | 2.82 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C7 | 313.41 | 91.83 | 4 | 3.13 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C8 | 261.32 | 75.20 | 4 | 2.72 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C9 | 259.34 | 78.92 | 4 | 3.03 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C10 | 275.34 | 80.08 | 4 | 2.64 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C11 | 289.37 | 84.89 | 5 | 2.82 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C12 | 349.47 | 108.21 | 6 | 3.80 | 3 | 0 | Yes | Yes | Yes | Yes | Yes |
| C13 | 350.45 | 110.22 | 6 | 3.67 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C14 | 351.44 | 108.01 | 6 | 3.42 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C15 | 351.44 | 108.01 | 6 | 3.31 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C16 | 351.44 | 108.01 | 6 | 3.36 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C17 | 401.50 | 125.52 | 6 | 3.98 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C18 | 375.46 | 114.93 | 6 | 3.66 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C19 | 384.90 | 115.23 | 6 | 3.98 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C20 | 384.90 | 115.23 | 6 | 4.05 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C21 | 429.35 | 117.92 | 6 | 4.09 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C22 | 418.45 | 115.22 | 7 | 4.03 | 7 | 0 | Yes | Yes | Yes | Yes | Yes |
| C23 | 395.45 | 119.04 | 7 | 3.45 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| C24 | 395.45 | 119.04 | 7 | 3.46 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| C25 | 368.44 | 110.18 | 6 | 3.81 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| C26 | 364.48 | 115.18 | 6 | 3.66 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C27 | 364.48 | 115.18 | 6 | 3.65 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| C28 | 410.51 | 123.20 | 8 | 4.10 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| C29 | 428.54 | 123.31 | 7 | 3.35 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| C30 | 386.43 | 110.13 | 6 | 3.89 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| Ifenprodil | 323.43 | 101.49 | 5 | 3.47 | 3 | 1 | Yes | Yes | Yes | Yes | Yes |

The pharmacokinetic properties of absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) of thirty molecules, were also compared to the ADMET profile of Ifenprodil, as resulted in Table 9. We have recorded that all thirty GluN2B NMDA receptor antagonists and a noncompetitive antagonist of NMDA receptors (ifenprodil) are characterized by very good human intestinal absorption (HIA superior to 90 %), and defined by blood-brain barrier (BBB) permeabilities superior to -1 Log BB, and central nervous system (CNS) permeabilities included between -1 and -3 Log PS. In addition, they are predicted as inhibitors of CYP1A2, CYP2D6, and CYP3A4 cytochromes, but they have any effect on CYP2C9 and CYP2C19 cytochromes. All chemical compounds are predicted by similar excretion, which is defined by a total clearance that does not decrease by 0.8 log ml/min/kg, except for C4 compound (total clearance of -1.208 log ml/min/kg). The AMES toxicity test indicates that all GluN2B NMDA receptor antagonists are predicted as not toxic inhibitors, except C4, C23, and C24 compounds. Most of these compounds do not cause skin sensitization, but they are predicted to have significant hepatotoxicity in the human organism, as an adverse side effect like any other medications.

**Table 9. ADMET in-silico pharmacokinetic properties of ifenprodil and thirty GluN2B N-methyl-D-aspartate (NMDA) receptor antagonists.**

| | Absorption | Distribution | | Metabolism | | | | | | | Excretion | Toxicity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intestinal Absorption (Human) | BBB permeability | CNS Permeability | Substrate | | Inhibitor | | | | | Total Clearance | AMES toxicity | Hepatotoxicity | Skin Sensitization |
| | | | | Cytochromes | | | | | | | | | | |
| N° | | | | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | | | | |
| | Numeric (% Absorbed) | Numeric (Log BB) | Numeric (LogPS) | Categorical (Yes/No) | | | | | | | Numeric (Log ml/min/kg) | Categorical (Yes/No) | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 95.198 | 0.237 | -1.911 | No | No | Yes | No | No | No | No | 0.808 | No | Yes | Yes |
| C2 | 95.246 | 0.226 | -1.93 | No | No | No | No | No | No | No | 0.992 | No | No | Yes |
| C3 | 94.194 | 0.241 | -2.16 | No | Yes | No | No | No | Yes | No | 1.202 | No | Yes | Yes |
| C4 | 91.909 | 0.298 | -1.208 | No | Yes | No | No | No | Yes | No | -1.208 | Yes | No | No |
| C5 | 94.646 | 0.3 | -1.886 | No | Yes | Yes | No | No | Yes | No | 0.996 | No | Yes | Yes |
| C6 | 93.66 | 0.27 | -2.106 | No | Yes | No | No | No | Yes | No | 1.25 | No | Yes | Yes |
| C7 | 91.335 | 0.327 | -1.268 | No | Yes | No | No | No | Yes | No | 1.461 | No | No | No |
| C8 | 94.975 | 0.13 | -2.201 | No | Yes | No | No | No | No | No | 1.248 | No | Yes | Yes |
| C9 | 93.271 | 0.257 | -2.152 | No | Yes | No | No | No | Yes | No | 1.205 | No | Yes | Yes |
| C10 | 92.782 | 0.088 | -2.499 | No | Yes | No | No | No | No | No | 1.188 | No | No | No |
| C11 | 92.312 | 0.056 | -2.53 | No | Yes | No | No | No | Yes | No | 1.082 | No | No | No |
| C12 | 93.056 | 0.464 | -1.195 | Yes | Yes | No | No | No | Yes | No | 1.133 | No | Yes | No |
| C13 | 92.838 | 0.266 | -1.145 | Yes | Yes | No | No | No | Yes | No | 1.142 | No | Yes | No |
| C14 | 94.803 | 0.127 | -1.995 | No | Yes | No | No | No | Yes | Yes | 1.159 | No | Yes | No |
| C15 | 95.02 | 0.135 | -1.832 | No | Yes | No | No | No | Yes | No | 1.129 | No | Yes | No |
| C16 | 95.048 | 0.135 | -1.832 | No | Yes | No | No | No | Yes | No | 1.162 | No | Yes | No |
| C17 | 92.643 | 0.25 | -1.419 | Yes | Yes | Yes | No | No | Yes | No | 1.162 | No | Yes | No |
| C18 | 93.001 | 0.013 | -2.164 | Yes | Yes | No | No | No | Yes | No | 1.16 | No | Yes | No |
| C19 | 91.984 | 0.268 | -1.309 | Yes | Yes | No | No | No | Yes | No | 1.022 | No | Yes | No |
| C20 | 91.492 | 0.316 | -1.392 | Yes | Yes | No | No | No | Yes | Yes | 1.065 | No | Yes | No |
| C21 | 91.425 | 0.314 | -1.392 | Yes | Yes | No | No | No | Yes | Yes | 1.045 | No | Yes | No |
| C22 | 90.338 | 0.283 | -1.36 | Yes | Yes | No | No | No | Yes | Yes | 0.917 | No | Yes | No |
| C23 | 92.104 | -0.345 | -2.383 | Yes | Yes | No | No | No | Yes | No | 0.83 | Yes | Yes | No |
| C24 | 92.137 | -0.293 | -2.378 | Yes | Yes | No | No | No | No | Yes | 0.824 | Yes | Yes | No |
| C25 | 92.394 | 0.309 | -1.418 | Yes | Yes | No | No | No | Yes | Yes | 1.026 | No | Yes | No |
| C26 | 92.95 | 0.329 | -1.392 | Yes | Yes | No | No | No | Yes | Yes | 1.078 | No | Yes | No |
| C27 | 93.442 | 0.281 | -1.309 | Yes | Yes | No | No | No | Yes | No | 1.051 | No | Yes | No |
| C28 | 93.179 | 0.223 | -2.52 | Yes | Yes | No | No | No | Yes | Yes | 1.117 | No | Yes | No |
| C29 | 94.23 | 0.021 | -2.624 | No | Yes | No | No | No | No | No | 0.944 | No | Yes | No |
| C30 | 91.635 | 0.301 | -1.43 | Yes | Yes | No | No | No | Yes | Yes | 1.002 | No | Yes | No |
| ifenprodil | 92.417 | 0.046 | -1.079 | Yes | Yes | Yes | No | No | Yes | No | 0.993 | No | No | No |

As a result, C4, C23, and C24 compounds were predicted as toxic products, according to the AMES toxicity test. C23 compound was predicted to be absorbed by the gastrointestinal tract, because it's part of white Egan Egg. C1, C2, C5, and C28 ligands are predicted not to be evacuated from the central nervous system (CNS) by the P-glycoprotein. While the majority of the remaining compounds do not cause skin sensitization, but they present marked hepatotoxicity in the human organism. C22 ligand was successfully examined with a good ADMET profile, respecting Lipinski, Veber, Egan, Muegge, and Rhose rules.

Recently, increasing attention has been directed to the necessity of assessing the bioavailability problems of potential drugs. This need was particularly marked for drug

discovery projects [22]. For this reason, we have examined the bioavailability of ifenprodil and GluN2B N-methyl-D-aspartate receptor antagonists, based on bioavailability radars technology, considering six physicochemical properties ideally adapted for oral bioavailability, namely lipophilicity, polarity, size, solubility, saturation, and flexibility. Where, lipophilicity (LIPO): -0.7 < XLOGP3 < +5, SIZE: 150 < MV < 500 g/mol, polarity (POLAR): 20 Å2 < TPSA < 130 Å2, insolubility (INSOLU): -6 < LOG S < 0, insaturation (INSATU): 0.25 < Fraction Csp3 < 1, and flexibility (FLEX): 0 < Number of rotatable bonds < 9. The pink-colored area is the suitable physicochemical space for oral bioavailability, in which the graph of each molecule must be fully adjusted to be declared as drug-like [23]. The present study shows that all compounds adhere to the suitable space for oral bioavailability because they are part of the pink zone of bioavailability radars without exception, as displayed in Figure 16.




LIPO
SIZE
POLAR
INSOLU
INSATU
FLEX
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20
C21
C22
C23
C24

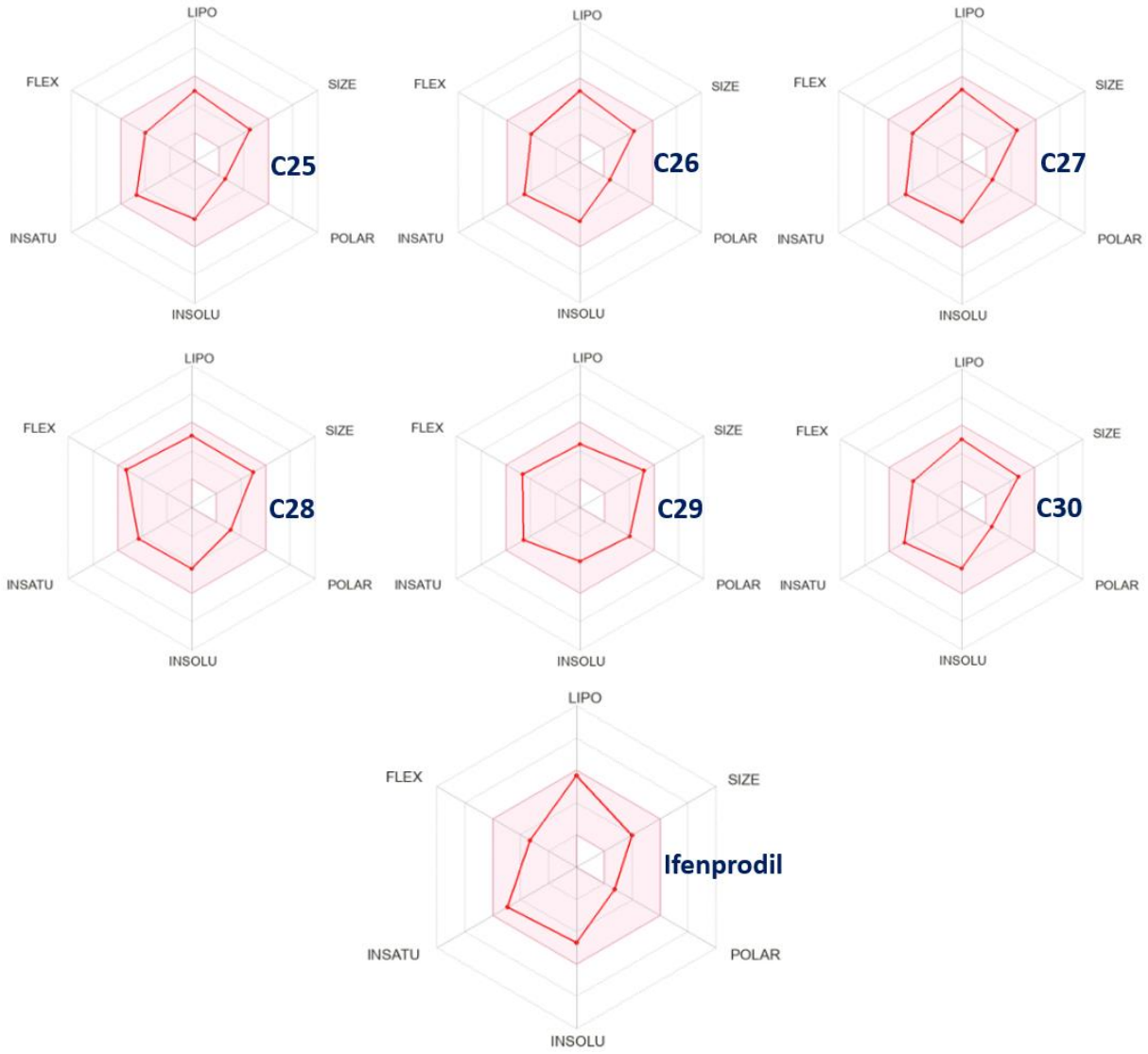
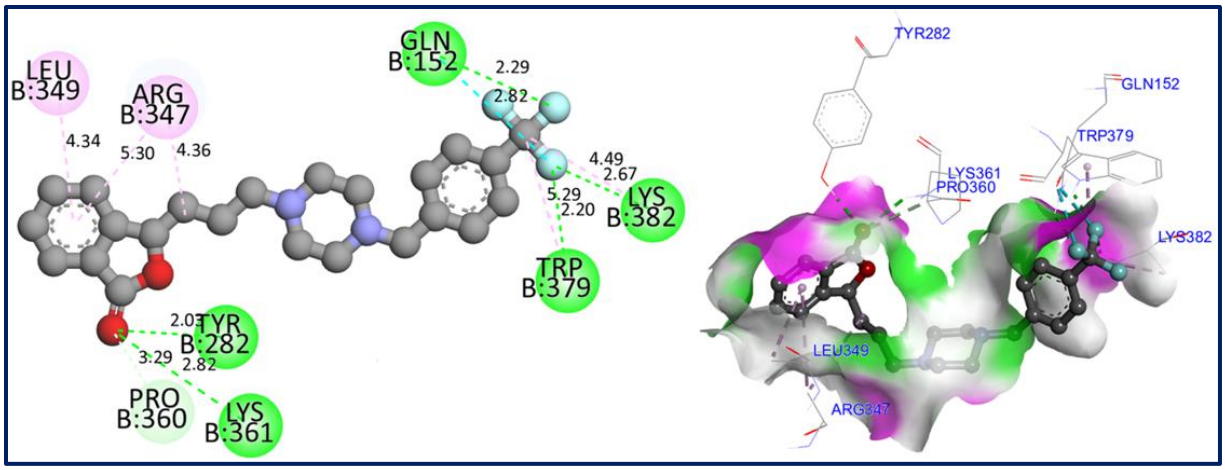



**Figure 16. Bioavailability radars of ifenprodil and thirty compounds, taking into account six physicochemical properties ideal for oral bioavailability, namely lipophilicity (LIPO), polarity (POLAR), size (SIZE), solubility (INSOLU), saturation (INSATU), and flexibility (FLEX).**

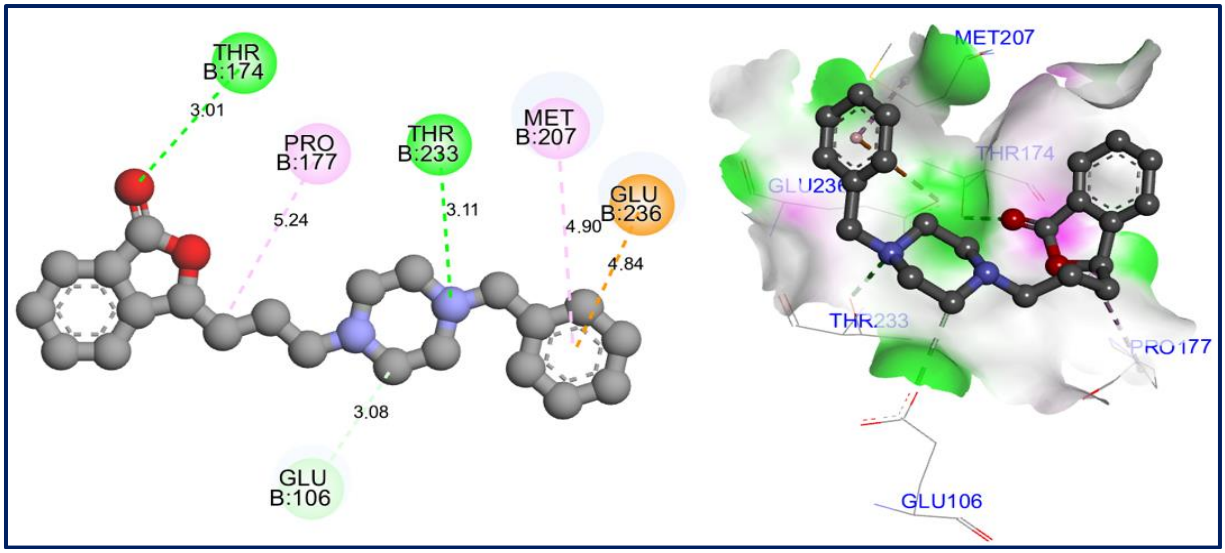
### 3.2.Molecular Docking Simulation

2D and 3D visualizations of molecular docking resulted in Figure 17, show that the C22 compound reacts specifically with Gln 152, Lys 382, Trp 379, Tyr 282, and Lys 361 amino acid residues of 3QEL.pdb protein with a binding energy of -7.27 kcal/mol, creating five conventional hydrogen bonds, with nuclear distances of 2.29 Å, 2.67 Å, 2.20 Å, 2.03 Å, and 2.82 Å, respectively. Three first conventionnel hydrogen bonds were produced with fluorine atoms, and last two were created with oxygen atom of ketonic function. This type of intermolecular interaction makes the C22 ligand so stable toward the targeted protein. In addition, one carbon hydrogen bond was produced with Pro 360 amino acid residue, and two

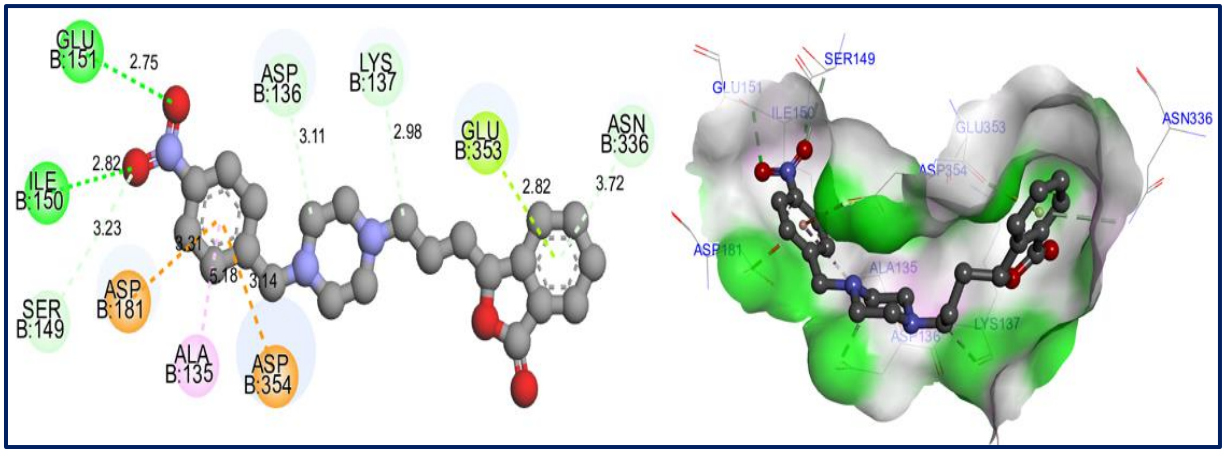
Pi-alkyl bonds were formed with Arg 347, and Leu 349 amino acid residues. While, the C13 compound was docked to the targeted protein with a binding energy of -7.89 kcal/mol, producing two hydrogen bonds towards Thr 174 and Thr 233 amino acid residues, with nuclear distances of 3.01 Å, and 3.11 Å, respectively. Then, one Alkyl bond was formed with Pro 177 amino acid residue at 5.24 Å, one Pi-Alkyl bond was created with Met 207 amino acid residue at 4.90 Å, and one other Pi-Anion bond was created to Glu 236 amino acid residue at 4.84 Å, and Carbon Hydrogen bond was formed with Glu 106 amino acid residue at 3.08 Å. Concerning the C23 compound, it was docked to the protein target with a binding energy of -9.17 kcal/mol, forming two Hydrogen bonds towards Glu 151, and Ile 150 amino acid residues, with nuclear distances of 2.75 Å, and 2.82 Å, respectively. Moreover, two Pi-Anion bonds were created with Asp 181 and Asp 354 at 3.31 Å and 3.14 Å, respectively. One Pi-Alkyl bond was produced with Ala 135 amino acid residue at 5.18 Å, and one Pi-Lone Pair bond was formed Glu 353 amino acid residue at 2.82 Å. In addition, four Carbon Hydrogen bonds were produced with Ser 149, Asp 136, Lys 137, and Asn 336 amino acid residues, at 3.23 Å, 3.11 Å, 2.98 Å, and 3.72 Å, respectively. These chemical bonds include intermolecular interactions similar to those detected towards the Glu 236, Pro 177, Met 207, and Glu 106 amino acid residues, which have been found by candidate selective antagonists of the N-Methyl-D-Aspartate (NMDA) receptor subunit 2B, which were designed to treat neuropathic pain, in particular Parkinson's, Huntington's, and Alzheimer's diseases [7]. The key amino acid residues, namely Glu 236, and Pro 177 from GluN2B subtypes of NMDA receptor were also detected for the most active indoles in complex with the same targeted protein encoded as 3QEL.pdb. As well as, the key residue Glu 236, was equally detected in the same way as the Gln 110 residue for ifenprodil, and Ro 25-6981 as selective and competitive antagonists of NMDA receptors subunit 2B, providing pharmacological therapies for chronic neurodegenerative diseases, in particular Alzheimer, Parkinson, depression, schizophrenia, and neuronal loss after stroke [44,45]. Therefore, we conclude that the most active ligands labelled C13, C22, and C23 reacted similarly to N-Methyl-D-Aspartate receptors (NMDARs) subunit 2B (3QEL.pdb), producing almost the same intermolecular interactions towards the active sites of the targeted protein. As a result, they could be potential treatments for neurodegenerative disorders, acting in the same way as well-known drugs such as ifenprodil, Ro 25-6981, Tacrine, memantine, and MK-801 [46–50].




(C22-3QEL.pdb) complex




(C13-3QEL.pdb) complex




(C23-3QEL.pdb) complex




**Figure 17. 2D and 3D visualizations of intermolecular interactions established between 3QEL.pdb protein and C22, C13, and C23 ligands, with binding energy of -7.27 Kcal/mol, -7.89 Kcal/mol, and -9.17 Kcal/mol, respectively.**

### 3.3.Molecular Dynamics Simulation

To verify the stability of intermolecular interactions produced between the transport protein encoded as 3QEL.pdb and C22, C13, and C23 compounds, we have performed the molecular dynamics (MD) technique on (targeted protein-candidate drug) complexes during 200 nanoseconds of MD simulation time. The Molecular Mechanics with Generalized Born and Surface Area (MMGBSA) solvation was conducted to calculate the binding free energies of the studied complexes, as resulted in Table 10 [15,16]. The MMGBSA approach shows that the binding energies are negatively very large, indicating that the candidate ligands have reacted with the protein target with low energies (minimal ΔG Bind scores), which makes the complexes stable with optimal energies during the molecular dynamic's simulation time.

**Table 10. Binding energies in Kcal/mol of the studied complexes using MMGBSA approach.**

| Studied complex | MMGBSA ΔG Bind | MMGBSA ΔG Bind Coulomb | MMGBSA ΔG Bind Covalent | MMGBSA ΔG Bind Hbond | MMGBSA ΔG Bind Lipo | MMGBSA ΔG Bind packing | MMGBSA ΔG Bind Solv GB | MMGBSA ΔG Bind vdw |
|---|---|---|---|---|---|---|---|---|
| C22 - protein | **-68.4791617** | **-89.4000894** | **3.587672457** | **-0.80935759** | **-24.58127916** | **-2.529968808** | **97.4565515** | **-52.2026907** |
| C13 - protein | **-52.2744679** | **-83.1266545** | **3.849614219** | **-0.54666909** | **-21.03976226** | **-2.218290244** | **92.2753436** | **-41.4680496** |
| C23 - protein | **-47.7334229** | **-113.000573** | **2.859691276** | **-1.27631036** | **-14.30883244** | **-0.869315669** | **132.261756** | **132.261756** |

Additionally, the results of conformational changes presented in Figure 18, indicate that root mean square deviation (RMSD) values of C22 compound as pictured in red, were predicted with a good level of molecular stability. Furthermore, the candidate drug does not diffuse further from the targeted protein, as marked in blue, because the ligand still linked to the responsible protein throughout 200 ns of MD simulation time, which mean that the previous interactions produced by molecular docking, render the complex so stable. Additionally, the protein root mean square fluctuation (RMSF) variation shows only two remarkable fluctuations for some protein residue indexes, so all other fluctuations were estimated to be negligible as they oscillate around the mean fluctuation. Furthermore, the conformational changes corresponding to the C13 and C23 ligands appear almost identical, as the RMSD values oscillated parallel to the protein target during the first fifty nanoseconds, but deviate slightly during the last 50 nanoseconds until the end of the simulation time. Moreover, the RMSF values of the protein target in blue were detected in a small range from 1 Å to 9 Å with negligible fluctuations after being bound to C13 and C23 ligands, similar to C22 compound.

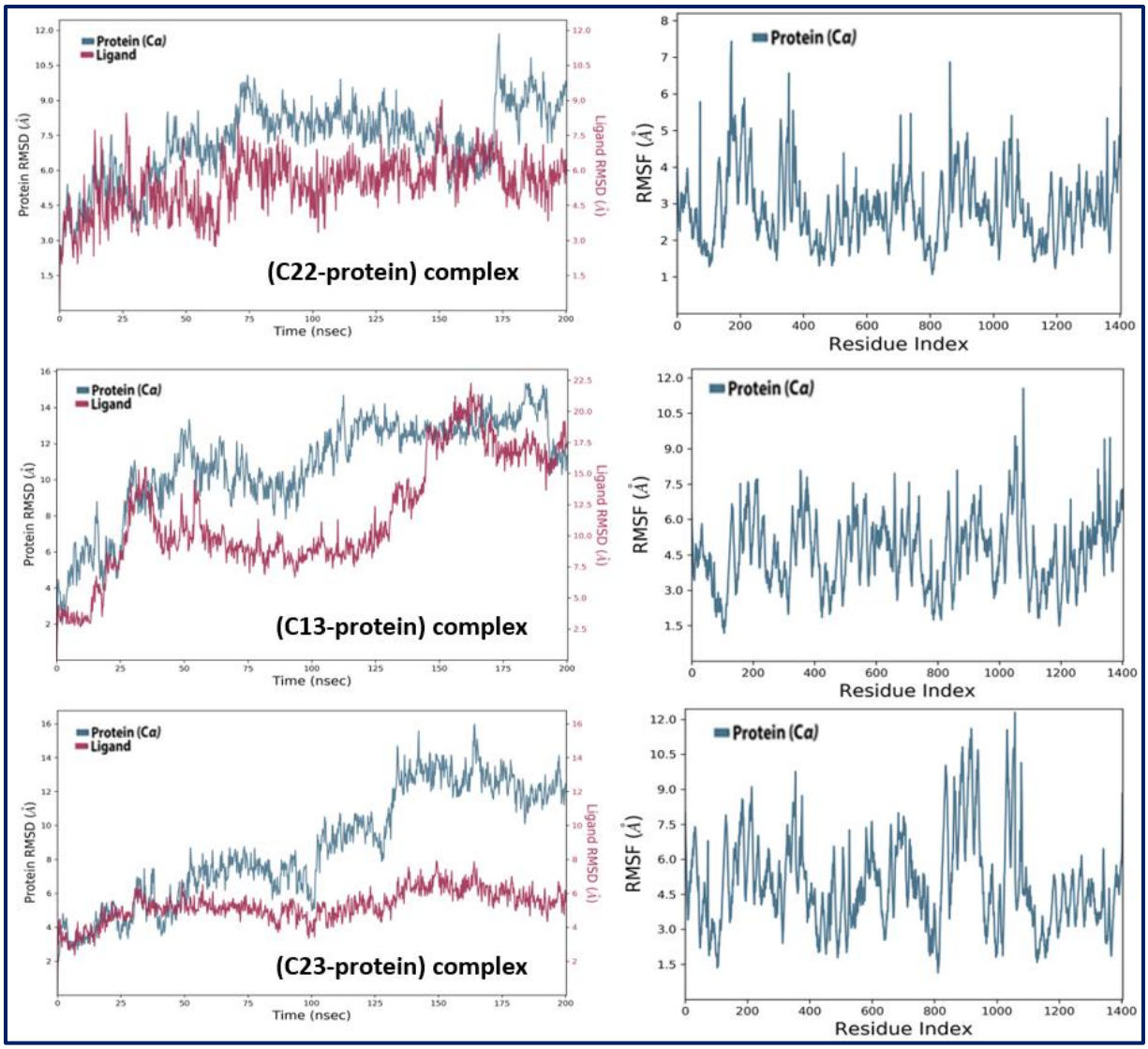



**Figure 18. Changes in root mean square deviation (RMSD) and root mean square fluctuation (RMSF) during 200 nanoseconds of molecular dynamics (MD) simulation time for C22, C13, and C23 ligands complexed to 3QEL.pdb protein.**

The same observations were reported regarding the conformational changes in the ligands properties as the radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA), and polar surface area (PSA), where they seem unstable during some nanoseconds, while they become more stable during the remaining time of the simulation time, as displayed in Figure 19, which demonstrates that C22, C13, and C23 ligands were very flexible after being bound to 3QEL.pdb protein, with few negligible changes in ligand compactness, due to the contribution of polar atoms such as nitrogen, oxygen, and fluorine elements, which confirms the highest level of molecular stability attained by the (ligand-protein) complexes. Consequently, the stability of intermolecular interactions produced

between the candidate drugs and protein target was successfully examined by molecular dynamics technique during 200 nanoseconds of MD simulation time.

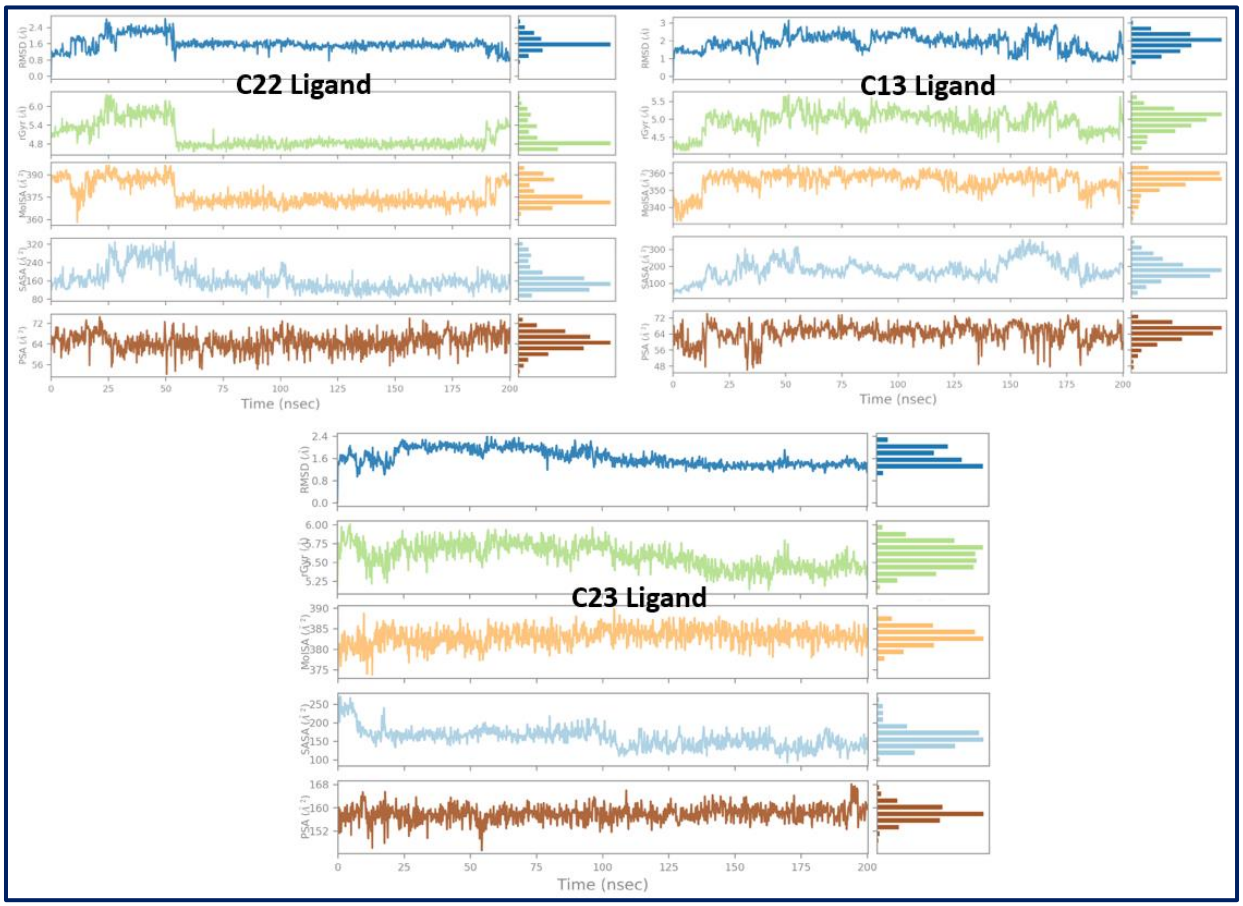



**Figure 19. Variations of root mean square deviation (RMSD), radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA) and polar surface area (PSA) during 200 ns of MD simulation time for C22, C13, and C23 ligands complexed to 3QEL.pdb protein.**

### 3.4. Chemical reactivity using DFT calculations

To examine the stability and chemical reactivity of the most active compounds (C22, C13, and C23), the calculations of Density Functional Theory (DFT) were applied on B3LYP, 6–31 + G (d, p) basis. The results of DFT calculations presented in Table 11, summarize all energetic values obtained for the electrostatic descriptors involved in the chemical reactivity of the candidate drugs, as the difference in binding energy [$\Delta E = E._{HOMO} - E._{LUMO}$], where the chemical compound defined by a lower energy gap corresponds to a higher stability and probably exhibits a high level of reactivity towards the protein target. Moreover, the negative values of Pauling electronegativity ($\chi$), indicate favorable interaction behavior towards the protein target, and positive values of chemical electron potential ($\mu$) display spontaneous chemical reactivity for the candidate molecules. Similarly, the positive values for chemical

softness (S) and chemical hardness (η) again confirm the molecular interactions of the candidate drugs. In addition, one other molecular property like electrophilicity index (ω) which explains the ability to accept an electron from the outside environment, especially when expressed positively.

The transfer of electrons that largely contributes to the binding of the candidate compounds with the protein target was presented by the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). For this reason, we focused on the Frontier Molecular Orbitals (FMO) analysis, which was done to identify the most likely reactive sites of the C22, C13, and C23 compounds, as resulted in Figure 20. Where, the red areas represent the attractive potential, while the blue areas represent the repulsive potential, which play an essential role in the chemical reactivity. In addition, the negative values of frontier orbitals confirm the stability of the compound, and the minimal value of gap energies (ΔE) values equal to 0.159603816, 0.149697224, and 0.120297517 for C22, C13, and C23 respectively, which confirm again that most active molecules are so flexible to react with targeted protein, and the reactive sites are in good agreement with the intermolecular interactions resulted in molecular docking simulation [19,20]. Therefore, all three molecules could exhibit a stable behavior during their interactions with the targeted protein.

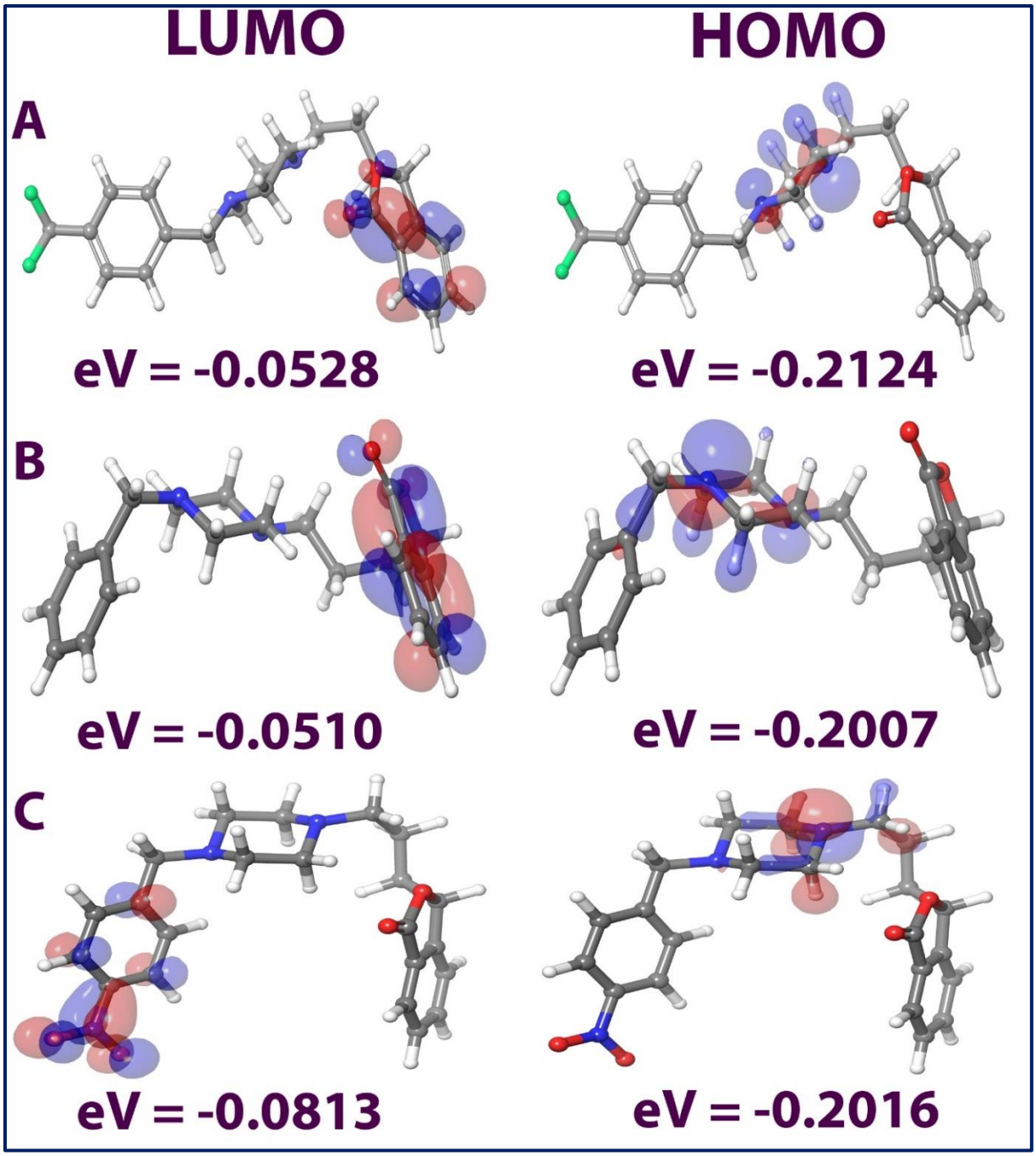



**Figure 20. Frontier molecular orbitals (FMO) of the most active compounds C22 (A), C13 (B), and C23 (C)**

**Table 11. DFT Calculations for the candidate drugs**

| | E. LUMO (eV) | E. HOMO (eV) | ΔE (eV) | χ Pauling | η (eV) | μ (eV) | S($eV^{-1}$) | ω (eV) |
|---|---|---|---|---|---|---|---|---|
| C22 | -0.052845041 | -0.2124489 | 0.159603816 | -0.13265 | 0.079802 | 0.132647 | 6.265514 | 0.110243 |
| C13 | -0.051013771 | -0.200711 | 0.149697224 | -0.12586 | 0.074849 | 0.125862 | 6.680151 | 0.105823 |
| C23 | -0.081295983 | -0.2015935 | 0.120297517 | -0.14144 | 0.060149 | 0.141445 | 8.312724 | 0.166309 |

## 4. Conclusion

The virtual screening adopted in the current study succeeded to predict the toxicity of three molecules marked as C4, C23, and C24 according to the AMES toxicity test, that we do not recommend as therapeutic agents for stroke. In contrast, most of the remaining compounds do not cause skin sensitization, but they present significant hepatotoxicity for the human body. In addition, the C22 and C13 compounds were successfully tested as a non-toxic inhibitor of CYP2D6 and CYP3A4 cytochromes, with a good ADMET profile, and designed to be an efficient central nervous system (CNS) agent due to the highest probability to cross the blood-brain barrier (BBB), meeting the violations number of Egan, Veber, Rhose, and Muegge, and complies with all five Lipinski rules, forming stable intermolecular interactions with the amino acid residues of the transport protein coded as 3QEL.pdb over 200 ns of molecular dynamic (MD) simulation time. Therefore, we strongly recommend the chemical compounds called C22 and C13 as powerful anti-stroke drugs.

## References

[1] M. Katan, A. Luft, Global Burden of Stroke, Semin Neurol. 38 (2018) 208–211. https://doi.org/10.1055/s-0038-1649503.

[2] M.J. Reeves, C.D. Bushnell, G. Howard, J.W. Gargano, P.W. Duncan, G. Lynch, A. Khatiwoda, L. Lisabeth, Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes, The Lancet Neurology. 7 (2008) 915–926. https://doi.org/10.1016/S1474-4422(08)70193-5.

[3] D. Barthels, H. Das, Current advances in ischemic stroke research and therapies, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 1866 (2020) 165260. https://doi.org/10.1016/j.bbadis.2018.09.012.

[4] Q. Xu, M. Hu, J. Li, X. Ma, Z. Chu, Q. Zhu, Y. Zhang, P. Zhu, Y. Huang, G. He, Discovery of novel brain-penetrant GluN2B NMDAR antagonists via pharmacophore-merging strategy as anti-stroke therapeutic agents, European Journal of Medicinal Chemistry. 227 (2022) 113876. https://doi.org/10.1016/j.ejmech.2021.113876.

[5] M. El fadili, M. Er-Rajy, M. Kara, A. Assouguem, A. Belhassan, A. Alotaibi, N.N. Mrabti, H. Fidan, R. Ullah, S. Ercisli, S. Zarougui, M. Elhallaoui, QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia, Pharmaceuticals. 15 (2022) 670. https://doi.org/10.3390/ph15060670.

[6] M. Radan, T. Djikic, D. Obradovic, K. Nikolic, Application of in vitro PAMPA technique and in silico computational methods for blood-brain barrier permeability prediction of novel CNS drug candidates, European Journal of Pharmaceutical Sciences. 168 (2022) 106056. https://doi.org/10.1016/j.ejps.2021.106056.

[7] M. El fadili, M. Er-rajy, H. Imtara, M. Kara, S. Zarougui, N. Altwaijry, O. Al kamaly, A. Al Sfouk, M. Elhallaoui, 3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists, Processes. 10 (2022) 1462. https://doi.org/10.3390/pr10081462.

[8] M. Er-rajy, M.E. Fadili, S. Mujwar, F.Z. Lenda, S. Zarougui, M. Elhallaoui, QSAR, molecular docking, and molecular dynamics simulation–based design of novel anti-cancer drugs targeting thioredoxin reductase enzyme, Struct Chem. (2023). https://doi.org/10.1007/s11224-022-02111-x.

[9] M. Er-rajy, M. El fadili, H. Imtara, A. Saeed, A. Ur Rehman, S. Zarougui, S.A. Abdullah, A. Alahdab, M.K. Parvez, M. Elhallaoui, 3D-QSAR Studies, Molecular Docking, Molecular Dynamic Simulation, and ADMET Proprieties of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer, Life. 13 (2023) 127. https://doi.org/10.3390/life13010127.

[10] A. Daina, V. Zoete, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules, ChemMedChem. 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.

[11] R.P.D. Bank, RCSB PDB - 3QEL: Crystal structure of amino terminal domains of the NMDA receptor subunit GluN1 and GluN2B in complex with ifenprodil, (n.d.). https://www.rcsb.org/structure/3QEL (accessed June 13, 2023).

[12] T.J. Ritchie, P. Ertl, R. Lewis, The graphical representation of ADME-related molecule properties for medicinal chemists, Drug Discovery Today. 16 (2011) 65–72. https://doi.org/10.1016/j.drudis.2010.11.002.
[13] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.
[14] M. El fadili, M. Er-rajy, H. Imtara, O.M. Noman, R.A. Mothana, S. Abdullah, S. Zerougui, M. Elhallaoui, QSAR, ADME-Tox, molecular docking and molecular dynamics simulations of novel selective glycine transporter type 1 inhibitors with memory enhancing properties, Heliyon. (2023) e13706. https://doi.org/10.1016/j.heliyon.2023.e13706.
[15] S. Faisal, S.L. Badshah, B. Kubra, M. Sharaf, A.-H. Emwas, M. Jaremko, M. Abdalla, Identification and Inhibition of the Druggable Allosteric Site of SARS-CoV-2 NSP10/NSP16 Methyltransferase through Computational Approaches, Molecules. 27 (2022) 5241. https://doi.org/10.3390/molecules27165241.
[16] C.J. Ononamadu, M. Abdalla, G.O. Ihegboro, J. Li, T.A. Owolarafe, T.D. John, Q. Tian, In silico identification and study of potential anti-mosquito juvenile hormone binding protein (MJHBP) compounds as candidates for dengue virus - Vector insecticides, Biochemistry and Biophysics Reports. 28 (2021) 101178. https://doi.org/10.1016/j.bbrep.2021.101178.
[17] S.A. Hollingsworth, R.O. Dror, Molecular Dynamics Simulation for All, Neuron. 99 (2018) 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.
[18] X. Hu, Z. Zeng, J. Zhang, D. Wu, H. Li, F. Geng, Molecular dynamics simulation of the interaction of food proteins with small molecules, Food Chemistry. 405 (2023) 134824. https://doi.org/10.1016/j.foodchem.2022.134824.
[19] J.H. Zothantluanga, M. Abdalla, M. Rudrapal, Q. Tian, D. Chetia, J. Li, Computational Investigations for Identification of Bioactive Molecules from *Baccaurea ramiflora* and *Bergenia ciliata* as Inhibitors of SARS-CoV-2 $M^{pro}$, Polycyclic Aromatic Compounds. (2022) 1–29. https://doi.org/10.1080/10406638.2022.2046613.
[20] A.A. Alameen, M. Abdalla, H.M. Alshibl, M.R. AlOthman, M.M. Alkhulaifi, T.O. Mirgany, R. Elsayim, In-silico studies of glutathione peroxidase4 activators as candidate for multiple sclerosis management, Journal of Saudi Chemical Society. 26 (2022) 101554. https://doi.org/10.1016/j.jscs.2022.101554.
[21] A. Daina, O. Michielin, V. Zoete, iLOGP: A Simple, Robust, and Efficient Description of *n* -Octanol/Water Partition Coefficient for Drug Design Using the GB/SA Approach, J. Chem. Inf. Model. 54 (2014) 3284–3301. https://doi.org/10.1021/ci500467k.
[22] Y.C. Martin, A Bioavailability Score, J. Med. Chem. 48 (2005) 3164–3170. https://doi.org/10.1021/jm0492002.
[23] F. Kandsi, A. Elbouzidi, F.Z. Lafdil, N. Meskali, A. Azghar, M. Addi, C. Hano, A. Maleb, N. Gseyra, Antibacterial and Antioxidant Activity of Dysphania ambrosioides (L.) Mosyakin and Clemants Essential Oils: Experimental and Computational Approaches, Antibiotics. 11 (2022) 482. https://doi.org/10.3390/antibiotics11040482.
[24] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.

[25] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews. 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[26] W.J. Egan, Merz Kenneth M., J.J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics, J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[27] W.J. Egan, G. Lauri, Prediction of intestinal permeability, Advanced Drug Delivery Reviews. 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.
[28] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[29] I. Muegge, Selection criteria for drug-like compounds, Med. Res. Rev. 23 (2003) 302–321. https://doi.org/10.1002/med.10041.
[30] S. Yalcin, Molecular Docking, Drug Likeness, and ADMET Analyses of Passiflora Compounds as P-Glycoprotein (P-gp) Inhibitor for the Treatment of Cancer, Current Pharmacology Reports. 6 (2020) 429–440. https://doi.org/10.1007/s40495-020-00241-6.
[31] pkCSM, (n.d.). https://biosig.lab.uq.edu.au/pkcsm/prediction (accessed November 19, 2022).
[32] D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, S. Moro, Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro, Pharmaceuticals. 15 (2022) 180. https://doi.org/10.3390/ph15020180.
[33] M. Er-rajy, M. El Fadili, H. Hadni, N.N. Mrabti, S. Zarougui, M. Elhallaoui, 2D-QSAR modeling, drug-likeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives, Struct Chem. 33 (2022) 973–986. https://doi.org/10.1007/s11224-022-01913-3.
[34] M. Er-Rajy, M. El Fadili, N.N. Mrabti, S. Zarougui, M. Elhallaoui, QSAR, molecular docking, ADMET properties in silico studies for a series of 7-propanamide benzoxaboroles as potent anti-cancer agents, Chinese Journal of Analytical Chemistry. 50 (2022) 100163. https://doi.org/10.1016/j.cjac.2022.100163.
[35] M. Er-rajy, M. El fadili, S. Mujwar, S. Zarougui, M. Elhallaoui, Design of novel anti-cancer drugs targeting TRKs inhibitors based 3D QSAR, molecular docking and molecular dynamics simulation, Journal of Biomolecular Structure and Dynamics. 0 (2023) 1–14. https://doi.org/10.1080/07391102.2023.2170471.
[36] BIOVIA Discovery Studio 2021 Client. Get the software safely and easily., Software Informer. (n.d.). https://biovia-discovery-studio-2021-client.software.informer.com/ (accessed November 19, 2022).
[37] A.P. Norgan, P.K. Coffman, J.-P.A. Kocher, D.J. Katzmann, C.P. Sosa, Multilevel Parallelization of AutoDock 4.2, J Cheminform. 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[38] AutoDock 4.2.6 / AutoDockTools 1.5.6 – Suite of Automated Docking Tools – My Biosoftware – Bioinformatics Softwares Blog, (n.d.). https://mybiosoftware.com/autodock-4-2-3-autodocktools-1-5-6-suite-automated-docking-tools.html (accessed November 19, 2022).

[39] Drug Discovery | Schrödinger, (n.d.). https://www.schrodinger.com/platform/drug-discovery (accessed November 19, 2022).
[40] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B. 105 (2001) 6474–6487. https://doi.org/10.1021/jp003919d.
[41] M. Abdalla, A.M. Rabie, Dual computational and biological assessment of some promising nucleoside analogs against the COVID-19-Omicron variant, Computational Biology and Chemistry. 104 (2023) 107768. https://doi.org/10.1016/j.compbiolchem.2022.107768.
[42] W.A. Eltayb, M. Abdalla, A.M. Rabie, Novel Investigational Anti-SARS-CoV-2 Agent Ensitrelvir "S-217622": A Very Promising Potential Universal Broad-Spectrum Antiviral at the Therapeutic Frontline of Coronavirus Species, ACS Omega. 8 (2023) 5234–5246. https://doi.org/10.1021/acsomega.2c03881.
[43] E. Karakas, N. Simorowski, H. Furukawa, Subunit arrangement and phenylethanolamine binding in GluN1/GluN2B NMDA receptors, Nature. 475 (2011) 249–253. https://doi.org/10.1038/nature10180.
[44] M.R. Buemi, L. De Luca, S. Ferro, E. Russo, G. De Sarro, R. Gitto, Structure-guided design of new indoles as negative allosteric modulators (NAMs) of N-methyl-d-aspartate receptor (NMDAR) containing GluN2B subunit, Bioorganic & Medicinal Chemistry. 24 (2016) 1513–1519. https://doi.org/10.1016/j.bmc.2016.02.021.
[45] L. Temme, B. Frehland, D. Schepmann, D. Robaa, W. Sippl, B. Wünsch, Hydroxymethyl bioisosteres of phenolic GluN2B-selective NMDA receptor antagonists: Design, synthesis and pharmacological evaluation, European Journal of Medicinal Chemistry. 144 (2018) 672–681. https://doi.org/10.1016/j.ejmech.2017.12.054.
[46] F.A. Alhumaydhi, M.A. Aljasir, A.S.M. Aljohani, S.A. Alsagaby, A.S.S. Alwashmi, M. Shahwan, Md.I. Hassan, A. Islam, A. Shamsi, Probing the interaction of memantine, an important Alzheimer's drug, with human serum albumin: In silico and in vitro approach, Journal of Molecular Liquids. 340 (2021) 116888. https://doi.org/10.1016/j.molliq.2021.116888.
[47] M.L. Crismon, Tacrine: First Drug Approved for Alzheimer's Disease, Ann Pharmacother. 28 (1994) 744–751. https://doi.org/10.1177/106002809402800612.
[48] Z. Liu, X. Qiu, S. Mak, B. Guo, S. Hu, J. Wang, F. Luo, D. Xu, Y. Sun, G. Zhang, G. Cui, Y. Wang, Z. Zhang, Y. Han, Multifunctional memantine nitrate significantly protects against glutamate-induced excitotoxicity via inhibiting calcium influx and attenuating PI3K/Akt/GSK3beta pathway, Chemico-Biological Interactions. 325 (2020) 109020. https://doi.org/10.1016/j.cbi.2020.109020.
[49] S.K. Sonkusare, C.L. Kaul, P. Ramarao, Dementia of Alzheimer's disease and other neurodegenerative disorders—memantine, a new hope, Pharmacological Research. 51 (2005) 1–17. https://doi.org/10.1016/j.phrs.2004.05.005.
[50] S. Thakral, A. Yadav, V. Singh, M. Kumar, P. Kumar, R. Narang, K. Sudhakar, A. Verma, H. Khalilullah, M. Jaremko, A.-H. Emwas, Alzheimer's disease: Molecular aspects and treatment opportunities using herbal drugs, Ageing Research Reviews. 88 (2023) 101960. https://doi.org/10.1016/j.arr.2023.101960.

# Chapitre V. Modélisation moléculaire de nouveaux bicyclo (aryl methyl) benzamides en tant qu'inhibiteurs puissants de GlyT1 destinés au traitement de la schizophrénie

## Résumé

Une classe structurelle de quarante-quatre bicyclo ((aryl) méthyl) benzamides, agissant en tant qu'inhibiteurs du transporteur de glycine type 1 (GlyT1), assurant la régulation de la fonction des récepteurs N-méthyl-D-Aspartate (NMDA) en modulant la concentration de glycine au niveau des synapses glutamatergiques, a été examinée à l'aide des techniques In-silico basées sur la modélisation moléculaire.

Tout d'abord, les modèles prédictifs QSAR générés par les techniques des régressions linéaires et non linéaires multiples affirment que l'activité biologique inhibitrice contre la schizophrénie est fortement et significativement corrélée aux descripteurs physicochimiques, géométriques et topologiques, en particulier : donneur de liaison hydrogène, polarisabilité, tension de surface, énergies d'étirement et de torsion, ainsi que le diamètre topologique.

Également, les propriétés pharmacocinétique d'ADMET, ont révélé que les ligands les plus actifs (L6, L9, L30, L31 et L37) sont les molécules qui ont la plus forte probabilité de pénétrer dans le système nerveux central (SNC), par contre la molécule L32 a la plus forte probabilité d'être absorbée par le tractus gastro-intestinal.

Encore, les résultats du Docking moléculaire indiquent que les acides aminés Tyr124, Phe43, Phe325, Asp46, Phe319 et Val120 sont les sites actifs de la protéine membranaire du transporteur de dopamine (DAT), dans laquelle les ligands les plus actifs peuvent inhiber le transporteur de glycine de type 1 (GlyT1).

Enfin, les résultats des simulations de la dynamique moléculaire ont montré que les cinq inhibiteurs sont restés stables dans les sites actifs de la protéine DAT pendant 100 ns, ce qui démontre leur rôle prometteur en tant que médicaments candidats pour le traitement de la schizophrénie.


**Mots-clés:** GlyT1; QSAR; Schizophrénie; ADMET; Docking moléculaire; DAT; Dynamique moléculaire.

# Chapitre V: QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia

Mohamed El fadili [1,*] , Mohammed Er-Rajy [1] , Mohammed Kara [2,*] , Amine Assouguem [3] , Assia Belhassan [4] , Amal Alotaibi [5] , Nidal Naceiri Mrabti [1] , Hafize Fidan [6] , Riaz Ullah [7] , Sezai Ercisli [8] , Sara Zarougui [1] and Menana Elhallaoui [1]

[1] Engineering Materials, Modeling and Environmental Laboratory, Faculty of Sciences Dhar El Mehraz, Sidi Mohammed Ben Abdellah University, Fez BP 1796, Morocco;
[2] Laboratory of Biotechnology, Conservation and Valorisation of Naturals Resources, Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco;
[3] Laboratory of Functional Ecology and Environment, Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Imouzzer Street, Fez P.O. Box 2202, Morocco;
[4] Molecular Chemistry and Natural Substances Laboratory, Department of Chemistry, Faculty of Sciences, University Moulay Ismail, Meknes, Morocco;
[5] Department of Basic Science, College of Medicine, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia;
[6] Department of Tourism and Culinary Management, Faculty of Economics, University of Food Technologies, 4000 Plovdiv, Bulgaria;
[7] Department of Pharmacognosy (MAPPRC), College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia;
[8] Department of Horticulture, Agricultural Faculty, Ataturk University, Erzurum TR-25240, Turkey;
***** Corresponding author: mohamed.elfadili@usmba.ac.ma

**Abstract:** Forty-four bicyclo ((aryl) methyl) benzamides, acting as glycine transporter type 1 (GlyT1) inhibitors, are developed using molecular modeling techniques. QSAR models generated by multiple linear and non-linear regressions affirm that the biological inhibitory activity against the schizophrenia disease is strongly and significantly correlated with physicochemical, geometrical and topological descriptors, in particular: Hydrogen bond donor, polarizability, surface tension, stretch and torsion energies and topological diameter. According to in silico ADMET properties, the most active ligands (L6, L9, L30, L31 and L37) are the molecules having the highest probability of penetrating the central nervous system (CNS), but the molecule 32 has the highest probability of being absorbed by the gastrointestinal tract. Molecular docking results indicate that Tyr124, Phe43, Phe325, Asp46, Phe319 and Val120 amino acids are the active sites of the dopamine transporter (DAT) membrane protein, in which the most active ligands can inhibit the glycine transporter type 1 (GlyT1). The results of molecular dynamics (MD) simulation revealed that all five inhibitors remained stable in the active sites of the DAT protein during 100 ns, demonstrating their promising role as candidate drugs for the treatment of schizophrenia.

**Keywords:** GlyT1; QSAR; schizophrenia; ADMET; molecular docking; DAT; MD

## 1. Introduction

About 1% of the worldwide population is affected by schizophrenia as a serious neuropsychiatric disease [1]. Despite the current regimens with favorable levels of efficacy and the great advancement in the treatment of schizophrenia, no antipsychotic medication can completely treat the cognitive dysfunction associated with this disorder, because its present treatments are accompanied by undesirable secondary effects. Therefore, the discovery of more clinically effective antipsychotic drugs are still necessary [2]. For this goal, the glycine transporter type 1 (GlyT1) inhibitors approved by the Food and Drug Administration (FDA) are a key therapeutic development strategy to treat a variety of central nervous system (CNS) disorders, in particular schizophrenia and cognitive disorders [3,4]. In this regard, type 1 glycine transporters regulate N-methyl-D-Aspartate (NMDA) receptor function via modulation of glycine concentration at the glutamatergic synapses, but their deficiency may affect the higher central nervous system functions [5,6]. In this paper, a systematic in silico study was performed on 44 GlyT1 inhibitors, which were tested in a locomotor activity assay (LMA) of the MK801 mouse to model the treatment of positive and negative symptoms of schizophrenia [4], by means of the following molecular modeling techniques: first of all, the quantitative structure activity relationships (QSAR) as a technology widely used in drug discovery, indicating ligands with a high affinity for a given macromolecular target and optimizing the quantitative linear and non-linear relationship established between structure and inhibitory activity [7,8]; secondly, in silico ADMET prediction of newly engineered drugs [9]; and third, the molecular docking study as an approach designed in computational chemistry to accelerate drug discovery at the early stages through the detection of typical intermolecular interactions, established between the potent ligands and the responsible protein target [10]. The last step concerns the molecular dynamics (MD) simulation as an efficient technique to investigate the dynamic conformational changes of the selected complexes (active ligands-protein target) [11,12]. In this context, we started our study with a molecular descriptors calculation for each GlyT1 inhibitor, using a quantum chemistry computation with the assistance of the molecular modeling method of MM2 type and the density functional theory (DFT) based on B3LYP/6-31 + G(d,p) level, in order to optimize the molecular configurations of all inhibitors [13]. Then, we reduced the dimension of the molecular descriptors using a principal component analysis (PCA) based on the correlation matrix. Next, two QSAR models were developed using multiple linear regression (MLR) and multiple nonlinear regression (MNLR). The robustness and reliability of the established QSAR models were examined using the external validation technique, followed by Y-randomization test, an applicability domain and a cross-validation technique with the Leave-One-Out process,

as one of the decisive steps to assess the confidence of the developed model's predictions for a new data set [14,15]. Moreover, we predicted the molecules having the highest inhibitory activity, based on their adsorption, distribution, metabolism, excretion and toxicity (ADMET) properties and the conditions mentioned in the rules of Lipinski, Ghose, Veber and Egan [16]. Additionally, we studied the intermolecular interactions established between the more active ligands and the dopamine transporter (DAT) membrane protein, encoded 4M48 as a crucial target for schizophrenia, withthe assistance of the molecular docking approach [2,17], which was validated using docking validation protocol [18]. Lastly, we performed the molecular dynamics simulations to analyze and elaborate the details of interaction and stability of the potent ligands in the protein targets [19].

## 2. Results and Discussion

### 2.1.Pricipal Component Analysis

Principal component analysis (PCA) is one of the most widely applied multivariate techniques. It is used to reduce the size of the variables into a limited number of principal components (linear combinations of the original variables) [20]. In this paper, we calculated 40 different descriptors, which were later reduced to 27 descriptors based on the correlation matrix, since descriptors that are strongly correlated with each other (*r pearson* > 0.9) were removed. From this reduced number of variables, we were able to visualize the projection of the new database on the first two principal components (factorial axes), as shown in Figure 21, which clearly indicates that molecules 1 and 32 are poorly explained. Consequently, they are considered as outliers.

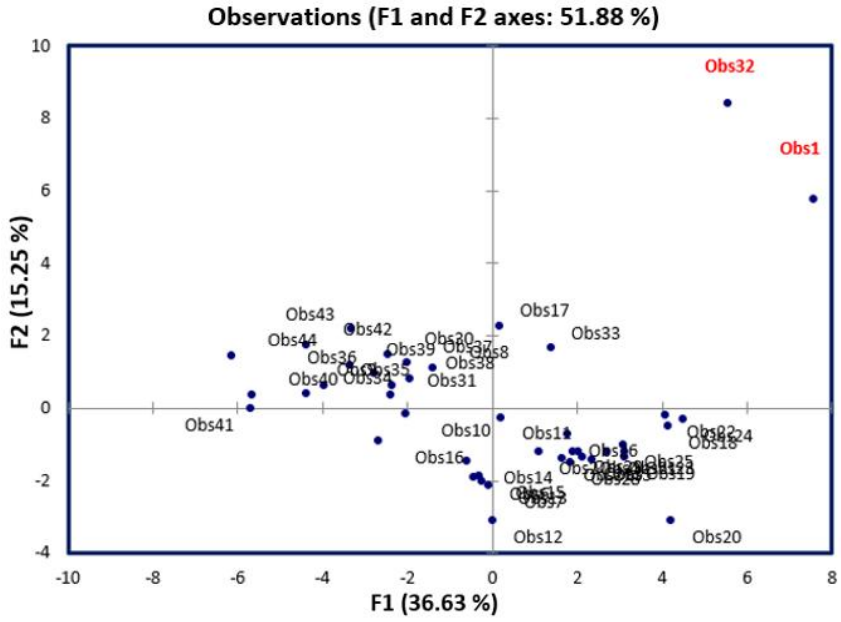



**Figure 21. Data visualization on the first two principal components.**

### 2.2.Statistical Database

Although observations 1 and 32 are considered as outliers, the new database will be represented by a matrix of 27 descriptors and 42 molecules. Using the k-means method, we randomly divided the database into training and test sets. The first one includes 80% of the total data (35 molecules) and was taken to develop the QSAR models, while the second one contains 20% of the total data (7 molecules) and was used to assess the validity of the developed models [21].

### 2.3.Multiple Linear Regression

The Quantitative structure-activity relationships (QSAR) have the potential to reduce the time and effort of molecular screening using mathematical predictive models [22]. One of these models is obtained by the multiple linear regression (MLR) technique, as a statistical tool for estimating the linear relationship between more than two variables which have cause-effect relations [23]. Thus, the first QSAR model was applied using the MLR technique with stepwise selection, on a training set of thirty-five molecules ($N$ = 35), where the process was repeated more than a thousand times based on statistical criteria: in particular, the determination and correlation coefficients, provided that they will be validated in the next stage. Accordingly, the best QSAR model is given by the following equation:

$$\mathbf{Log_{10}IC_{50} = -10{,}407 - 0.279 \times \alpha e + 0{,}069 \times \gamma + 0.156 \times TE + 1.83 \times HBD + 1.716 \times SE + 1.029 \times TD.} \quad \mathbf{(1)}$$

This constructed model shows that the biological activity at the log scale is a quantitative variable affected by the following six descriptors: polarizability (αe), surface tension (ɣ), torsion energy (TE), Hydrogen bond donor (HBD), stretch energy (SE) and topological diameter (TD), which have been calculated and presented in Table 12. Moreover, the significance test demonstrates that the slope of each variable has a probability inferior to 5% as shown in Table 13, and so the selected descriptors have a significant weight on the biological inhibitory activity at a 95% confidence interval. Except the polarizability, all five molecular descriptors affect positively the biological activity as shown in Figure 22, where a molecule can be more active if it is less polar and has higher values of surface tension, torsion and stretch energies, hydrogen bond donor and topological diameter.

**Table 12. Values of selected descriptors for 44 molecules.**

| N° | ae | ɣ | TE | HBD | SE | TD | $Log_{10}IC_{50}$ |
|---|---|---|---|---|---|---|---|
| 2 | 41.05 | 40.3 | 12.9136 | 1 | 2.609 | 11 | 0.47712126 |
| 3 | 43.9 | 43.9 | 12.833 | 1 | 3.43 | 11 | 1.61278386 |
| 4 | 43.9 | 43.9 | 12.2837 | 1 | 3.7251 | 11 | 0.77815125 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5 | 41.96 | 44.6 | 10.5041 | 1 | 3.3398 | 11 | 1.96378783 |
| 6 | 43.87 | 50.9 | 5.8562 | 1 | 3.5148 | 10 | 0 |
| 7 | 43.11 | 54.3 | 10.9678 | 1 | 3.5012 | 10 | 0.77815125 |
| 8 | 44.29 | 57.7 | 10.6827 | 1 | 3.31 | 10 | 0.47712126 |
| 10 | 43.05 | 49.2 | 13.6092 | 1 | 3.581 | 10 | 2.69372695 |
| 11 | 43.82 | 51.6 | 16.8366 | 1 | 3.7574 | 10 | 2.67577834 |
| 12 | 42.28 | 53.7 | 12.0135 | 1 | 3.2349 | 10 | 1.87506126 |
| 13 | 43.11 | 54.3 | 15.9864 | 1 | 3.5181 | 10 | 2.90794852 |
| 14 | 43.87 | 50.9 | 12.0765 | 1 | 3.6081 | 10 | 1.36172784 |
| 15 | 43.11 | 54.3 | 15.211 | 1 | 3.5393 | 10 | 1.65321251 |
| 17 | 45.28 | 46.3 | 11.4242 | 1 | 4.1245 | 10 | 0.77815125 |
| 18 | 46.55 | 43.3 | 11.1295 | 1 | 4.5556 | 11 | 2.53147892 |
| 19 | 45.81 | 43.2 | 11.5098 | 1 | 4.5198 | 11 | 2.71096312 |
| 20 | 46.95 | 45.7 | 12.3354 | 1 | 3.7527 | 11 | 0.95424251 |
| 21 | 45.81 | 43.2 | 11.5767 | 1 | 3.7965 | 11 | 1.462398 |
| 22 | 46.55 | 43.3 | 12.3649 | 1 | 4.0183 | 12 | 2.24797327 |
| 23 | 45.81 | 43.2 | 11.5647 | 1 | 3.7752 | 12 | 3.23121465 |
| 24 | 46.55 | 43.3 | 12.3272 | 1 | 3.9781 | 13 | 3.24526584 |
| 25 | 41.5 | 38.4 | 18.5378 | 1 | 3.5883 | 11 | 3.52659771 |
| 26 | 40.85 | 43.8 | 14.1068 | 1 | 3.3294 | 11 | 1.8920946 |
| 29 | 43.14 | 46.6 | 17.5352 | 1 | 3.5318 | 11 | 2.95616843 |
| 30 | 42.67 | 45.3 | 11.6496 | 1 | 3.2701 | 10 | 0 |
| 31 | 43.05 | 49.2 | 15.9695 | 1 | 3.5677 | 10 | 0 |
| 33 | 52.92 | 45 | 13.4879 | 1 | 3.9397 | 14 | 2.97589114 |
| 34 | 41.86 | 48.5 | 7.3285 | 2 | 2.3774 | 10 | 0.47712126 |
| 37 | 45.11 | 46 | 10.9824 | 1 | 3.3677 | 10 | 0 |
| 38 | 44.35 | 48.9 | 16.6401 | 1 | 3.3293 | 10 | 1 |
| 40 | 39.53 | 51.1 | 7.6774 | 2 | 2.2337 | 10 | 1.76342799 |
| 41 | 38.77 | 54.5 | 13.3491 | 2 | 2.2124 | 10 | 1.65321251 |
| 42 | 40.15 | 49.3 | 8.3209 | 1 | 3.3316 | 10 | 0.84509804 |
| 43 | 41.98 | 47.7 | 10.8208 | 1 | 3.622 | 10 | 0.47712126 |
| 44 | 38.2 | 52.2 | 5.2114 | 2 | 2.8062 | 10 | 2.22530928 |
| 1 * | 45.25 | 41.3 | 39.2347 | 1 | 2.6809 | 13 | 1.56820172 |
| 9 * | 43.81 | 46.2 | 10.3084 | 1 | 3.6309 | 10 | 0 |
| 16 * | 43.6 | 55.2 | 8.802 | 2 | 3.3159 | 10 | 0.90308999 |
| 27 * | 43.14 | 46.6 | 18.0072 | 1 | 3.6639 | 11 | 1.25527251 |
| 28 * | 43.14 | 46.6 | 18.0105 | 1 | 3.682 | 11 | 2.32428246 |
| 32 * | 53.84 | 50.1 | 58.0893 | 1 | 5.1759 | 14 | 0 |
| 35 * | 41.48 | 47.3 | 9.0526 | 1 | 2.9116 | 10 | 0.47712126 |
| 36 * | 43.31 | 45.9 | 9.4978 | 1 | 3.0768 | 10 | 0.60205999 |
| 39 * | 44.32 | 54.8 | 9.9783 | 1 | 2.9 | 10 | 0.30103 |

***** indicates test set molecules.

**Table 13. Significance test of the slopes.**

| Source | Value | Standard Deviation | t | Pr > \|t\| | Lower Terminal (95%) | Higher Terminal (95%) |
|---|---|---|---|---|---|---|
| Constante | −10.408 | 3.172 | −3.281 | 0.003 | −16.905 | −3.910 |
| *αe* | −0.279 | 0.079 | −3.550 | 0.001 | −0.441 | −0.118 |
| ɣ | 0.070 | 0.033 | 2.101 | 0.045 | 0.002 | 0.138 |
| TE | 0.156 | 0.042 | 3.731 | 0.001 | 0.071 | 0.242 |
| HBD | 1.830 | 0.571 | 3.208 | 0.003 | 0.661 | 2.999 |
| SE | 1.716 | 0.387 | 4.429 | 0.000 | 0.922 | 2.510 |
| TD | 1.030 | 0.208 | 4.956 | <0.0001 | 0.604 | 1.455 |

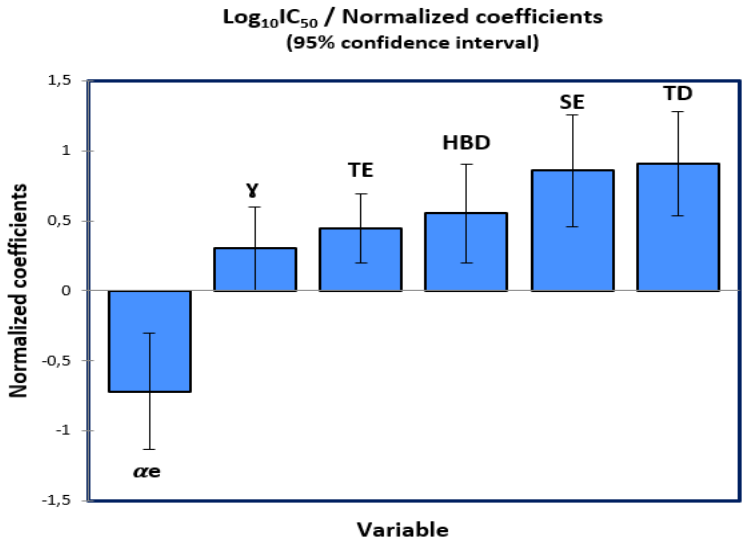



**Figure 22. Influence degree of the descriptors on the biological activity.**

Additionally, the null hypothesis (H0) postulated by the Fisher statistical test is rejected, because the calculated Fisher value ($F$ = 10.325) is so much higher than its critical value: [$F$ (35,6) = 2.37, $p$ < 0.0001], as presented in the one Anova test (Table 14). Therefore, the variance between the response ($Log_{10}IC_{50}$) and the six predictor variables is homogeneous. Moreover, the correlation and determination coefficients of $R$ = 0.83 and $R^2$ = 0.69, respectively, confirm that there is a strong relationship between the descriptors and the inhibitory activity. Thus, the first QSAR model generated via MLR technique has a good predictive performance, with a low standard error (*RMSE* = 0.66).

**Table 14. Variance analysis.**

| Source | DDL | Total Square | Mean Square | *F* | *Pr > F* |
|---|---|---|---|---|---|
| Model | 6 | 26.753 | 4.459 | 10.325 | < 0.0001 |
| Error | 28 | 12.092 | 0.432 | | |
| Adjusted total | 34 | 38.846 | | | |

### 2.4.Multiple Non-Linear Regression

The multiple non-linear regression (MNLR) technique is applied using a set of adapted algorithms to generate the quantitative predictive models [24]. In the present study, we relied on the programmed function of the type:

$$Y = a0 + \sum_{i=1}^{n} (ai \times Xi + bi \times Xi^2) \quad (2)$$

As:

Y: is the predicted biological activity ($Log_{10}IC_{50}$)

Xi: is the explicative variable

a0: is the constant of the QSAR model

ai and bi: are the slopes of each descriptor to one and two degrees, respectively.

Finally, we arrived at the second QSAR model given by the following equation:

$$\mathbf{Log_{10}IC_{50} = -19.699 - 0.009 \times \alpha e - 0.056 \times \gamma - 0.161 \times TE + 1.466 \times HBD + 0.5 \times SE + 2.693 \times TD - 0.002 \times \alpha e^{\wedge 2} + 0.001 \times \gamma^{\wedge 2} + 0.013 \times TE^{\wedge 2} + 0.148 \times SE^{\wedge 2} - 0.07 \times TD^{\wedge 2}.} \quad \mathbf{(3)}$$

This mathematical model has a good predictive capacity, justified by a strong non-linear relationship between the biological activity and the six descriptors, as it is defined by a good correlation coefficient ($R = 0.84$) and a good coefficient of determination ($R^2 = 0.71$), in addition to its minimal mean square error ($RMSE = 0.72$).

### 2.5.QSAR Model Validation

#### 2.5.1. Applicability Domain

The applicability domain (AD) of a quantitative structure-activity relationship (QSAR) model is necessary to verify its reliability on new compounds (test set) that were not considered during its development [25]. This technique has been evaluated by an analysis expressed as a Williams diagram (Figure 23), which confirms that the molecules (1 and 32) belonging to the test set are really outliers, because they exceed the warning leverage ($h^* = 0.6$), where: $h^* = 3 \times K/n$ *and* $K = p + 1$, *(*$p = 6$, $K = 7$, $n = 35$*)* as, *n*: is the number of training set, and *p*: is the number of predictor descriptors [26,27]. Next, we noted that the compound 33 from training test is not an outlier because it does not exceed the critical leverage *(h*)*. Therefore, except for molecules (1 and 32), all the others are well explained because they have in addition a normalized residual included in the 3 times standard deviation interval. Consequently, the 42 remaining molecules are tested in the applicability domain and the QSAR model was predicted correctly.

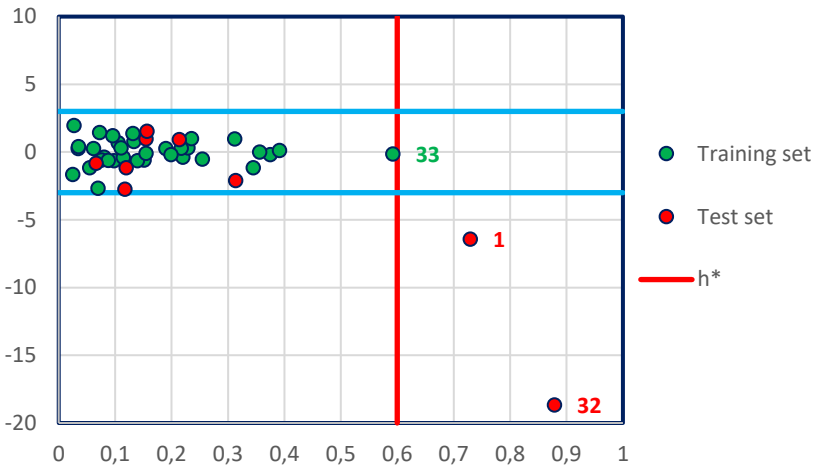



**Figure 23. William's diagram of the MLR model established by Equation (1). 1 and 32 are outliers in the test set and 33 is a non-aberrant molecule in the training test.**

### 2.5.2. External Validation

To assess the accuracy of the QSAR predictive model and guarantee its generalizability, it is absolutely needed to validate it on new molecules included in the test set, before its application in clinical practice [28]. Based on a training test (35 molecules), we tested the seven new molecules from the test set and got the results presented in Table 15.

Table 15. External validation results of the MLR and MNLR models.

| Molecule Number | Observed $Log_{10}IC_{50}$ | Predicted $Log_{10}IC_{50}$ (MLR) | Predicted $Log_{10}IC_{50}$ (MNLR) |
|---|---|---|---|
| 9 * | 0.000 | 0.545 | 0.380 |
| 16 * | 0.903 | 2.285 | 2.134 |
| 27 * | 1.255 | 3.051 | 3.342 |
| 28 * | 2.324 | 3.083 | 3.372 |
| 35 * | 0.477 | −0.158 | −0.164 |
| 36 * | 0.602 | −0.414 | −0.429 |
| 39 * | 0.301 | −0.304 | −0.046 |

* Indicates test set molecules.

The results mentioned in Figure 24 indicate that the MLR QSAR model is given by an external validation correlation coefficient ($R^2ext$ = 0.63), and the results noted in Figure 25 indicate that the MNLR QSAR model is characterized by an external validation correlation coefficient of $R^2ext$ = 0.68. According to the Alexander Golbraikh and Alexander Tropsha theory, a QSAR model is externally validated if the correlation coefficient of its external validation is greater than 0.6. Therefore, the mathematical models developed with the help of MLR and MNLR techniques are externally validated.

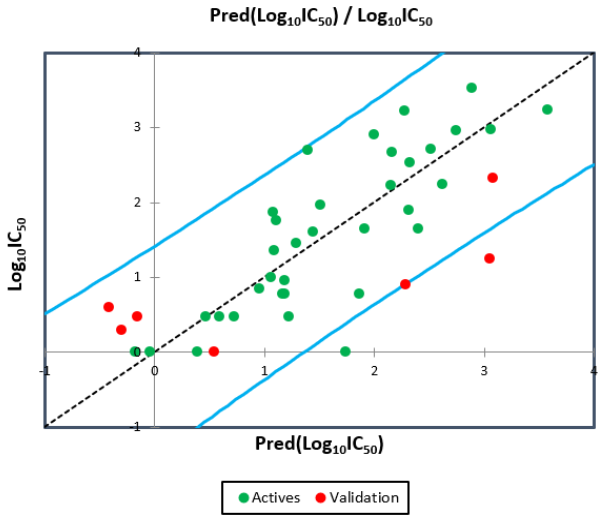



**Figure 24. Correlation between the observed and predicted activities using MLR technique.**

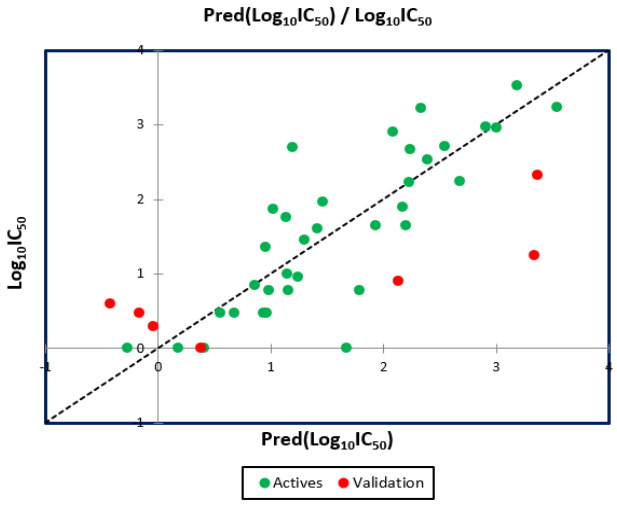



**Figure 25. Correlation between the observed and predicted activities using MNLR technique.**

### 2.5.3. Internal Validation

To validate internally the QSAR model, we applied the cross-validation technique with the leave-one-out procedure (CVLOO), so that each observation is tested exactly once, by executing a new model each run on thirty-four compounds (*N*-1 = 34) and predicting the biological activity of the removed sample, as shown in Table 16. This technique is based on the calculation of the quadratic coefficient of cross validation ($Q^2cv$), which is expressed in the following equation [29,30]: $Q^2cv = 1 - \frac{\sum_i^n (Ypred-Yobs)^2}{\sum_i^n (Yobs-Ymean)^2}$ (4) AS: *Ypred*: is the predicted activity value, *Yobs*: is the observed activity value, *Ymean*: is the mean of the observed activity

values. A high value of $Q^2cv$ = 0.57 (superior than 0.5) signifies that the established model is reliable, robust and has better internal predictivity.

**Table 16. Observed and predicted activity values from the QSAR models.**

| Molecules Number | Observed $Log_{10}IC_{50}$ | Predicted $Log_{10}IC_{50}$ (MLR) | Predicted $Log_{10}IC_{50}$ (MNLR) | Predicted $Log_{10}IC_{50}$ (CV) |
|---|---|---|---|---|
| 2 | 0.47712126 | 0.588 | 0.679 | 0.664 |
| 3 | 1.61278386 | 1.439 | 1.415 | 1.428 |
| 4 | 0.77815125 | 1.860 | 1.790 | 1.922 |
| 5 | 1.96378783 | 1.511 | 1.464 | 1.441 |
| 6 | 0 | −0.040 | 0.405 | −0.053 |
| 7 | 0.77815125 | 1.186 | 1.157 | 1.245 |
| 8 | 0.47712126 | 0.721 | 0.933 | 0.801 |
| 10 | 2.69372695 | 1.397 | 1.195 | 1.321 |
| 11 | 2.67577834 | 2.157 | 2.240 | 2.057 |
| 12 | 1.87506126 | 1.083 | 1.018 | 0.970 |
| 13 | 2.90794852 | 2.000 | 2.081 | 1.827 |
| 14 | 1.36172784 | 1.093 | 0.954 | 1.075 |
| 15 | 1.65321251 | 1.915 | 1.933 | 1.959 |
| 17 | 0.77815125 | 1.163 | 0.982 | 1.247 |
| 18 | 2.53147892 | 2.322 | 2.392 | 2.250 |
| 19 | 2.71096312 | 2.520 | 2.548 | 2.458 |
| 20 | 0.95424251 | 1.189 | 1.245 | 1.217 |
| 21 | 1.462398 | 1.289 | 1.302 | 1.272 |
| 22 | 2.24797327 | 2.623 | 2.680 | 2.667 |
| 23 | 3.23121465 | 2.281 | 2.338 | 2.174 |
| 24 | 3.24526584 | 3.578 | 3.538 | 3.710 |
| 25 | 3.52659771 | 2.891 | 3.183 | 2.562 |
| 26 | 1.8920946 | 2.311 | 2.171 | 2.396 |
| 29 | 2.95616843 | 2.751 | 3.000 | 2.717 |
| 30 | 0 | 0.391 | 0.176 | 0.443 |
| 31 | 0 | 1.744 | 1.675 | 1.934 |
| 33 | 2.97589114 | 3.062 | 2.907 | 3.204 |
| 34 | 0.47712126 | 0.463 | 0.555 | 0.454 |
| 37 | 0 | −0.178 | −0.276 | −0.228 |
| 38 | 1 | 1.056 | 1.151 | 1.068 |
| 40 | 1.76342799 | 1.104 | 1.136 | 0.867 |
| 41 | 1.65321251 | 2.404 | 2.199 | 2.852 |
| 42 | 0.84509804 | 0.959 | 0.855 | 0.993 |
| 43 | 0.47712126 | 1.226 | 0.961 | 1.294 |

### 2.5.4. Validation Using Y-Randomisation Test

The statistical study of Alexander Golbraikh and Alexander Tropsha confirms that the cross-validation technique is necessary but not sufficient, as the internal predictive accuracy of the cross-validation procedure tends to be overestimated and the high value of the quadratic coefficient may be the result of chance correlation. For this reason, the Y-randomisation test is necessary [31]. Using java Platform SE binary, we tested the QSAR model quality by running one hundred randomizations, as presented in Table 17. The results of the Y-randomisation test demonstrate that the ($cR^2p$ = 0.602) criteria is superior than 0.5; moreover, the $R$, $R^2$ and $R^2cv$ values of the original model are much better than the values obtained by 100 randomizations. Consequently, the biological activity values predicted by the original model are not due to chance.

**Table 17. Y-randomization test results.**

| Model | *R* | *R^2* | *Q^2* | Model | *R* | *R^2* | *Q^2* |
|---|---|---|---|---|---|---|---|
| **Original** | **0.829884** | **0.688707** | **0.572045** | Random 51 | 0.206983 | 0.042842 | −0.46272 |
| Random 1 | 0.252331 | 0.063671 | −0.50878 | Random 52 | 0.537396 | 0.288794 | −0.15592 |
| Random 2 | 0.457615 | 0.209411 | −0.17702 | Random 53 | 0.379861 | 0.144294 | −0.43774 |
| Random 3 | 0.47795 | 0.228436 | −0.41001 | Random 54 | 0.367538 | 0.135084 | −0.29876 |
| Random 4 | 0.375518 | 0.141014 | −0.34708 | Random 55 | 0.179251 | 0.032131 | −0.54121 |
| Random 5 | 0.422447 | 0.178462 | −0.39625 | Random 56 | 0.663141 | 0.439756 | 0.029755 |
| Random 6 | 0.480602 | 0.230979 | −0.17775 | Random 57 | 0.36146 | 0.130653 | −0.41471 |
| Random 7 | 0.306791 | 0.09412 | −0.47744 | Random 58 | 0.445943 | 0.198865 | −0.22915 |
| Random 8 | 0.354955 | 0.125993 | −0.40713 | Random 59 | 0.417956 | 0.174687 | −0.19669 |
| Random 9 | 0.209847 | 0.044036 | −0.71484 | Random 60 | 0.204369 | 0.041767 | −0.88175 |
| Random 10 | 0.395267 | 0.156236 | −0.36218 | Random 61 | 0.557804 | 0.311145 | 0.016035 |
| Random 11 | 0.520928 | 0.271366 | −0.1806 | Random 62 | 0.50639 | 0.256431 | −0.3063 |
| Random 12 | 0.510412 | 0.260521 | −0.21009 | Random 63 | 0.37293 | 0.139077 | −0.46899 |
| Random 13 | 0.427634 | 0.182871 | −0.23082 | Random 64 | 0.383643 | 0.147182 | −0.41262 |
| Random 14 | 0.445148 | 0.198156 | −0.41414 | Random 65 | 0.414428 | 0.171751 | −0.30301 |
| Random 15 | 0.21278 | 0.045275 | −0.4451 | Random 66 | 0.292763 | 0.08571 | −0.36258 |
| Random 16 | 0.516892 | 0.267178 | −0.45198 | Random 67 | 0.526141 | 0.276824 | −0.1287 |
| Random 17 | 0.37686 | 0.142024 | −0.55449 | Random 68 | 0.284657 | 0.08103 | −0.54548 |
| Random 18 | 0.154692 | 0.023929 | −0.85659 | Random 69 | 0.456042 | 0.207974 | −0.2171 |
| Random 19 | 0.491084 | 0.241163 | −0.24676 | Random 70 | 0.451139 | 0.203526 | −0.15451 |
| Random 20 | 0.424795 | 0.180451 | −0.30099 | Random 71 | 0.402163 | 0.161735 | −0.15347 |
| Random 21 | 0.513699 | 0.263886 | −0.1961 | Random 72 | 0.480122 | 0.230517 | −0.17729 |
| Random 22 | 0.316251 | 0.100015 | −0.30938 | Random 73 | 0.426294 | 0.181727 | −0.22948 |
| Random 23 | 0.301949 | 0.091173 | −0.63655 | Random 74 | 0.475859 | 0.226442 | −0.23411 |
| Random 24 | 0.332628 | 0.110641 | −0.8224 | Random 75 | 0.462608 | 0.214006 | −0.12839 |
| Random 25 | 0.633727 | 0.401609 | 0.166923 | Random 76 | 0.53816 | 0.289616 | −0.33075 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Random 26 | 0.328704 | 0.108046 | −0.48201 | Random 77 | 0.383709 | 0.147233 | −0.30145 |
| Random 27 | 0.46585 | 0.217016 | −0.16011 | Random 78 | 0.38822 | 0.150715 | −0.41903 |
| Random 28 | 0.441731 | 0.195126 | −0.25279 | Random 79 | 0.528782 | 0.279611 | −0.29561 |
| Random 29 | 0.355019 | 0.126039 | −0.31878 | Random 80 | 0.330001 | 0.1089 | −0.41611 |
| Random 30 | 0.329982 | 0.108888 | −0.42698 | Random 81 | 0.413654 | 0.171109 | −0.22613 |
| Random 31 | 0.378435 | 0.143213 | −0.25482 | Random 82 | 0.493491 | 0.243533 | −0.12853 |
| Random 32 | 0.462326 | 0.213746 | −0.13151 | Random 83 | 0.381202 | 0.145315 | −0.49761 |
| Random 33 | 0.343488 | 0.117984 | −0.53921 | Random 84 | 0.323593 | 0.104712 | −0.30559 |
| Random 34 | 0.462673 | 0.214066 | −0.27221 | Random 85 | 0.32106 | 0.103079 | −0.33856 |
| Random 35 | 0.35063 | 0.122941 | −0.3394 | Random 86 | 0.30071 | 0.090427 | −0.55488 |
| Random 36 | 0.522964 | 0.273491 | −0.09258 | Random 87 | 0.518334 | 0.26867 | −0.1494 |
| Random 37 | 0.222631 | 0.049564 | −0.75169 | Random 88 | 0.387695 | 0.150307 | −0.45639 |
| Random 38 | 0.241784 | 0.058459 | −0.47485 | Random 89 | 0.36652 | 0.134337 | −0.30196 |
| Random 39 | 0.339537 | 0.115286 | −0.4132 | Random 90 | 0.279562 | 0.078155 | −0.47573 |
| Random 40 | 0.448316 | 0.200987 | −0.47037 | Random 91 | 0.575806 | 0.331552 | −0.03852 |
| Random 41 | 0.487561 | 0.237716 | −0.34662 | Random 92 | 0.5706 | 0.325585 | 0.021398 |
| Random 42 | 0.369003 | 0.136164 | −0.33599 | Random 93 | 0.381837 | 0.1458 | −0.44739 |
| Random 43 | 0.400756 | 0.160605 | −0.30621 | Random 94 | 0.385236 | 0.148406 | −0.6547 |
| Random 44 | 0.343595 | 0.118058 | −0.42487 | Random 95 | 0.251773 | 0.06339 | −0.49809 |
| Random 45 | 0.390289 | 0.152325 | −0.27962 | Random 96 | 0.446548 | 0.199405 | −0.47359 |
| Random 46 | 0.350185 | 0.12263 | −0.22911 | Random 97 | 0.316743 | 0.100326 | −0.95316 |
| Random 47 | 0.463947 | 0.215247 | −0.27397 | Random 98 | 0.367366 | 0.134958 | −0.3352 |
| Random 48 | 0.37435 | 0.140138 | −0.27999 | Random 99 | 0.631342 | 0.398592 | 0.090509 |
| Random 49 | 0.452168 | 0.204456 | −0.40118 | Random 100 | 0.56162 | 0.315417 | 0.016006 |
| Random 50 | 0.266881 | 0.071225 | −0.47945 | | | | |

### 2.5.5. Golbreikh and Tropsha Criteria

The quantitative structure-activity relationship (QSAR) model, defined by the first Equation (1), satisfies the threshold criteria postulated by Golbraikh and Tropsha theory, as shown in Table 18.

**Table 18. Golbraikh and Tropsha statistical criteria to validate the designed QSAR model.**

| Parameter | Equation | Model Score | Threshold | Comment |
|---|---|---|---|---|
| $R^2$ | $R^2 = 1 - \frac{\sum (\ Yobs - Y\ cal\ )^{\ 2}}{\sum (Yobs\ - \overline{Yobs})^{\ 2}}$ | 0.69 | >0.6 | Accepted |
| $R^2adj$ | $R^2adj = \frac{(N-1)R^{\ 2} - p}{N - p - 1}$ | 0.62 | >0.6 | Accepted |
| $R^2test$ | $R^2test = 1 - \frac{\sum (\ Ycal(test) - Y\ obs(test)\ )^{\ 2}}{\sum (Yobs(test)\ - \overline{Yobs}(train))^{\ 2}}$ | 0.63 | >0.6 | Accepted |
| $Q^2cv$ | $Q^2cv = 1 - \frac{\sum (\ Ycal - Y\ obs\ )^{\ 2}}{\sum (Yobs\ - \overline{Yobs})^{\ 2}}$ | 0.57 | >0.5 | Accepted |
| $R^2$rand | Average of the 100 $R^2$rand (i) | 0.17 | $<R^2$ | Accepted |

| | | | | |
|---|---|---|---|---|
| $Q^2cv$ 'LOO' rand | Average of the 100 $Q^2cv$ 'LOO' rand (i) | −0.34 | $<Q^2cv$ | Accepted |
| $cR^2p$ | $cR^2p = R*\sqrt{(R^2 - (Average\ R_{rand})^2)}$ | 0.60 | >0.5 | Accepted |

*Yobs* and *Ycalc*: refer to the observed and calculated/predicted response values. $\overline{Yobs}$ and $\overline{Ycal}$: refer to the mean of the observed and calculated/predicted response values. *N* and *p* refer to the number of data points (compounds) and descriptors.

### 2.6.In Silico Pharmacokinetics ADMET Prediction

The most active ligands (L6, L9, L30, L31, L32 and L37), acting as inhibitors of type 1 glycine transporters, were tested based on the rules of Lipinski, Veber, Egan, and Ghose, and the pharmacokinetic properties (ADMET) [32], which were compared to the obtained results for nortriptyline as a co-crystallized ligand bound to the dopamine transporter (DAT) membrane protein encoded 4M48. The results presented in Table 19 indicate that all molecules respect the rules of Lipinski, Veber, Egan and Ghose except the ligand 32, because its molar refractivity index exceeds 130 and its Ghose violation number is equal to 2 (exceed 1). Additionally, the exact predictive model (BOILED-Egg), highly practical in the context of drug discovery and medicinal chemistry, and based on the calculation of lipophilicity given by the logarithm of the partition coefficient between n-octanol and water (Log $P_{O/W}$) and polarity signaled by the topological polar surface area (TPSA) of small molecules, clearly shows that the molecule (L32) is the only one that does not belong to the yellow Egan-egg, as presented in Figure 26. Therefore, the five ligands (L6, L9, L30, L31 and L37) are the molecules having the highest probability to penetrate the brain. In comparison, the molecule 32 belonging to the white region of the egg has the highest probability of being absorbed by the gastrointestinal tract [33], which is why it was an outlier in the previous QSAR study.

The pharmacokinetic parameters of adsorption, distribution, metabolism, excretion and toxicity (ADMET) of the most active ligands as presented in Table 20 indicate that the ligands have a good absorption in the human intestine (IAH so higher than 70%), and a good distribution, since their human distribution volumes are estimated to be greater than −0.44 Log L/kg. Their permeability to the blood-brain barrier (BBB) is greater than −1 Log BB, and their permeability to the central nervous system (CNS) outside the interval (of −2 to −3) Log PS. Thus, they all penetrate the central nervous system (CNS) with the exception of ligands L32 and L30. In addition, the molecules are all predicted as inhibitors of cytochrome 2D6 except ligand 32. Consequently, the ligands (L6, L9, L30, L31 and L37) are designed to be agents of the central nervous system due to the highest probability of penetrating the blood-brain barrier (BBB).

**Table 19. Prediction of the physicochemical properties of nortriptyline and more active ligands, based on Lipinski, Veber, Egan and Ghose violations.**

| Ligands Number | Physico-Chemical Propities | | | | | | Lipinski Violations | Veber Violations | Egan Violations | Ghose Violations | Synthetic Accessibility |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Molecular Weight (g/mol) | Molar Refractive Index | Rotatable Bonds | Log *p* (Octanol/Water) | H-BA | H-BD | | | | | |
| Rule | ≤500 | 40 ≤ MR ≤ 130 | <10 | <5 | ≤10 | <5 | ≤1 | Yes/No | Yes/No | Yes/No | 0 < S.A < 10 |
| (1) L6 | 403.34 | 115.02 | 5 | 3.60 | 2 | 1 | 1 | Yes | Yes | Yes | 4.24 |
| (2) L9 | 362.51 | 114.93 | 5 | 3.58 | 2 | 1 | 0 | Yes | Yes | Yes | 4.21 |
| (3) L30 | 366.50 | 112.17 | 5 | 3.75 | 3 | 1 | 0 | Yes | Yes | Yes | 4.97 |
| (4) L31 | 363.50 | 112.73 | 5 | 3.15 | 3 | 1 | 0 | Yes | Yes | Yes | 4.28 |
| (5) L32 | 480.66 | 140.34 | 8 | 3.61 | 4 | 1 | 0 | Yes | Yes | No | 5.19 |
| (6) L37 | 377.52 | 121.65 | 6 | 3.23 | 3 | 2 | 0 | Yes | Yes | Yes | 4.47 |
| (7) nortriptyline | 265,39 | 85.74 | 4 | 3.24 | 1 | 1 | 1 | Yes | Yes | Yes | 3.28 |

**Table 20. Prediction of ADMET pharmacokinetic properties of nortriptyline and more active ligands.**

| Ligands Number | Absorption | Distribution | | | Metabolism | | | | | | | Excretion | Toxicity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intestinal Absorption (Human) | VDss (Human) | BBB Permeability | CNS Permeability | Substrate | Inhibitor | | | | | | Total Clearance | AMES Toxicity |
| | | | | | CYP | | | | | | | | |
| | | | | | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | | |
| | Numeric (% Absorbed) | Numeric (Log L/kg) | Numeric (Log BB) | Numeric (Log PS) | Categorical (Yes/No) | | | | | | | Numeric (Log ml/min/kg) | Categorical (Yes/No) |
| (1) L6 | 91.194 | 1.431 | 0.199 | −1.06 | Yes | Yes | Yes | Yes | No | Yes | Yes | 1.058 | Not toxic |
| (2) L9 | 93.373 | 1.477 | 0.223 | −1.072 | Yes | Yes | Yes | No | No | Yes | No | 0.978 | Not toxic |
| (3) L30 | 93.344 | 1.242 | 0.176 | −2.055 | Yes | Yes | No | No | No | Yes | No | 0.883 | Not toxic |
| (4) L31 | 95.105 | 1.187 | 0.048 | −1.976 | Yes | Yes | No | No | No | Yes | Yes | 0.948 | Not toxic |
| (5) L32 | 94.331 | 1.038 | −0.378 | −2.005 | No | Yes | Yes | No | No | No | Yes | 0.85 | Not toxic |
| (6) L37 | 92.765 | 1.814 | 0.044 | −0.657 | Yes | Yes | No | No | No | Yes | No | 0.905 | Not toxic |
| (7) nortriptyline | 98.519 | 1.688 | 0,854 | −1.287 | No | Yes | Yes | No | No | Yes | No | 1.077 | Not toxic |

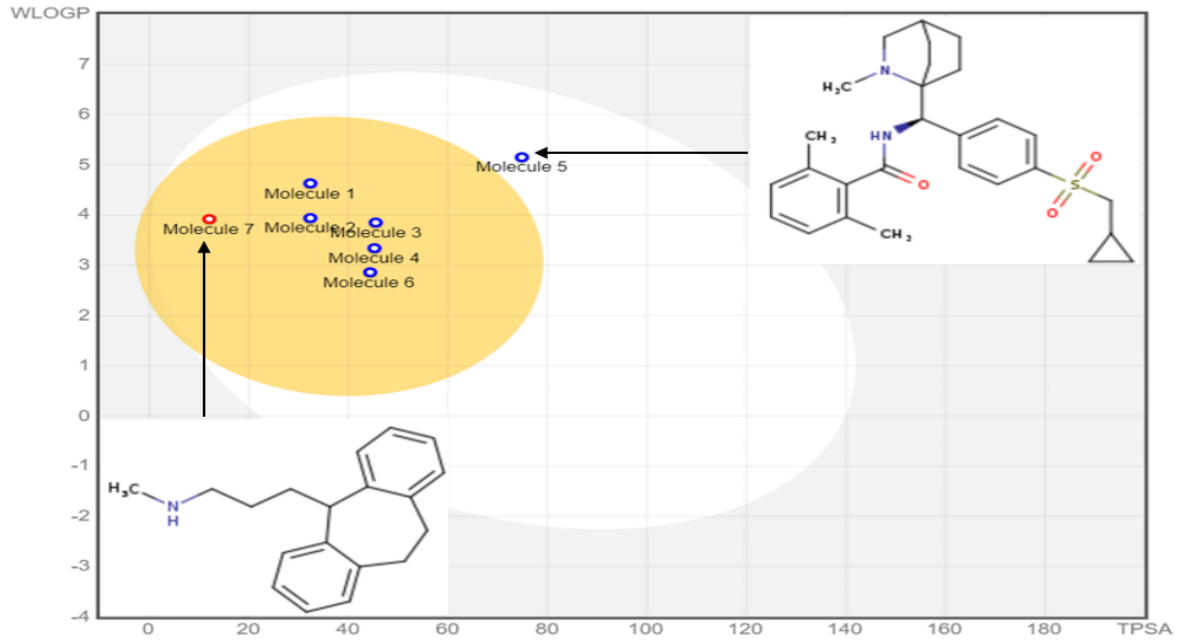



**Figure 26. BOILED-egg predictive model of the most active ligands.**

## 2.7.Molecular Docking

Molecular docking results are focused on the dopamine transporter (DAT) bound to the tricyclic antidepressant nortriptyline, as a transmembrane protein that removes the neurotransmitter dopamine from the synaptic cleft and transports it into the cytosol of surrounding cells. The crystal structure of this receptor is extracted using the X-ray diffraction method at a resolution of 2.96 Å taken from the protein data base (PDB) [34–36]. In this part of the research, the molecular docking process is started for the following most active molecules (L6, L9, L30, L31 and L37) to predict the type of Intermolecular interactions established with the protein encoded 4M48, compared to the established interactions with the co-crystallized ligand (nortriptyline) pictured in Figure 27, which indicate that Phe43A, Phe325A and Tyr124A amino acids, are the active sites of the target protein, as sourced using the ProteinsPlus online server [37].

The results of molecular docking applied on the more active ligands, presented in Figure 28, show that the ligands L6 and L9 share common molecular interactions as the chemical bonds of type pi-sigma and Pi-Pi T-shaped established between the benzenic cycle and (Val120 and Tyr124) amino acids respectively, in addition to two bonds of alkyl type with Phe325 and Phe43 amino acids. L30 and L31 ligands also form common bonds, like the hydrogen bond linked to the nitrogen atom, with the amino acid Asp46 at the same nuclear distance (5.5 Å), in addition to the alkyl bond with bicyclo group and Tyr124 amino acid. The same type of bond was established between the methyl group and the amino acid Phe325 at a nuclear distance of 5.5

Å, more than Pi-Pi bonds with Phe43 and Phe319 amino acids. Even the ligand L37 formed an alkyl bond with Tyr124 amino acid, and two Pi-Pi chemical bonds with Phe325 and Phe319 amino acids. Therefore, we can conclude that Tyr124, Phe43, Phe325, Asp46, Phe319 and Val120 amino acids are the active sites of the dopamine transporter (DAT) membrane protein, in which the most active ligands can inhibit the glycine transporter type 1 (GlyT1).

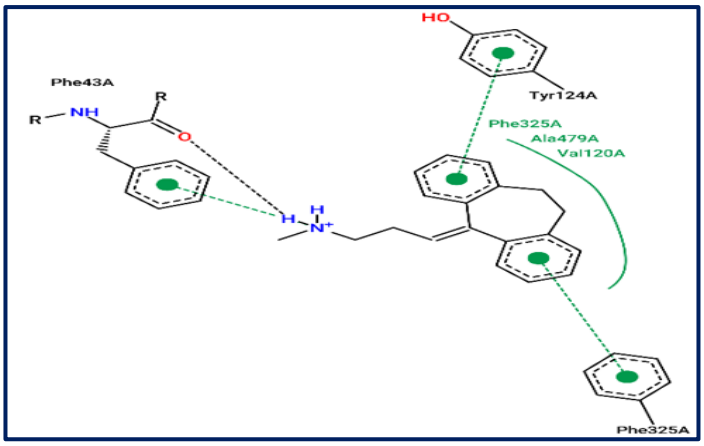
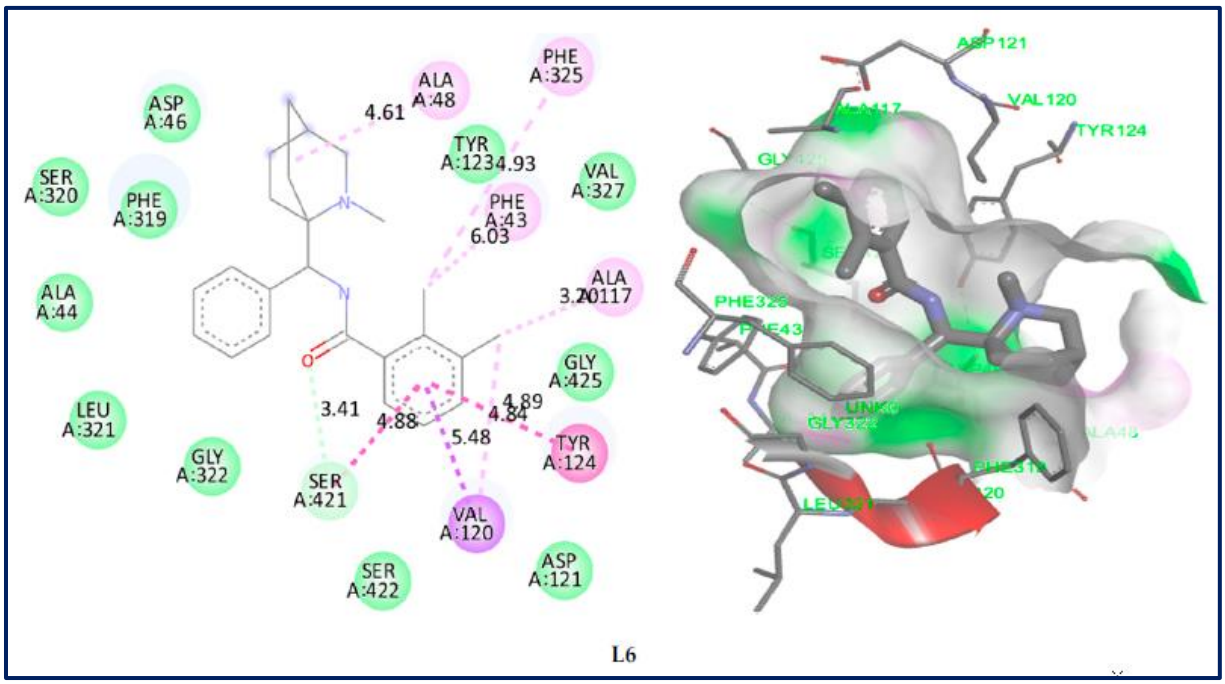
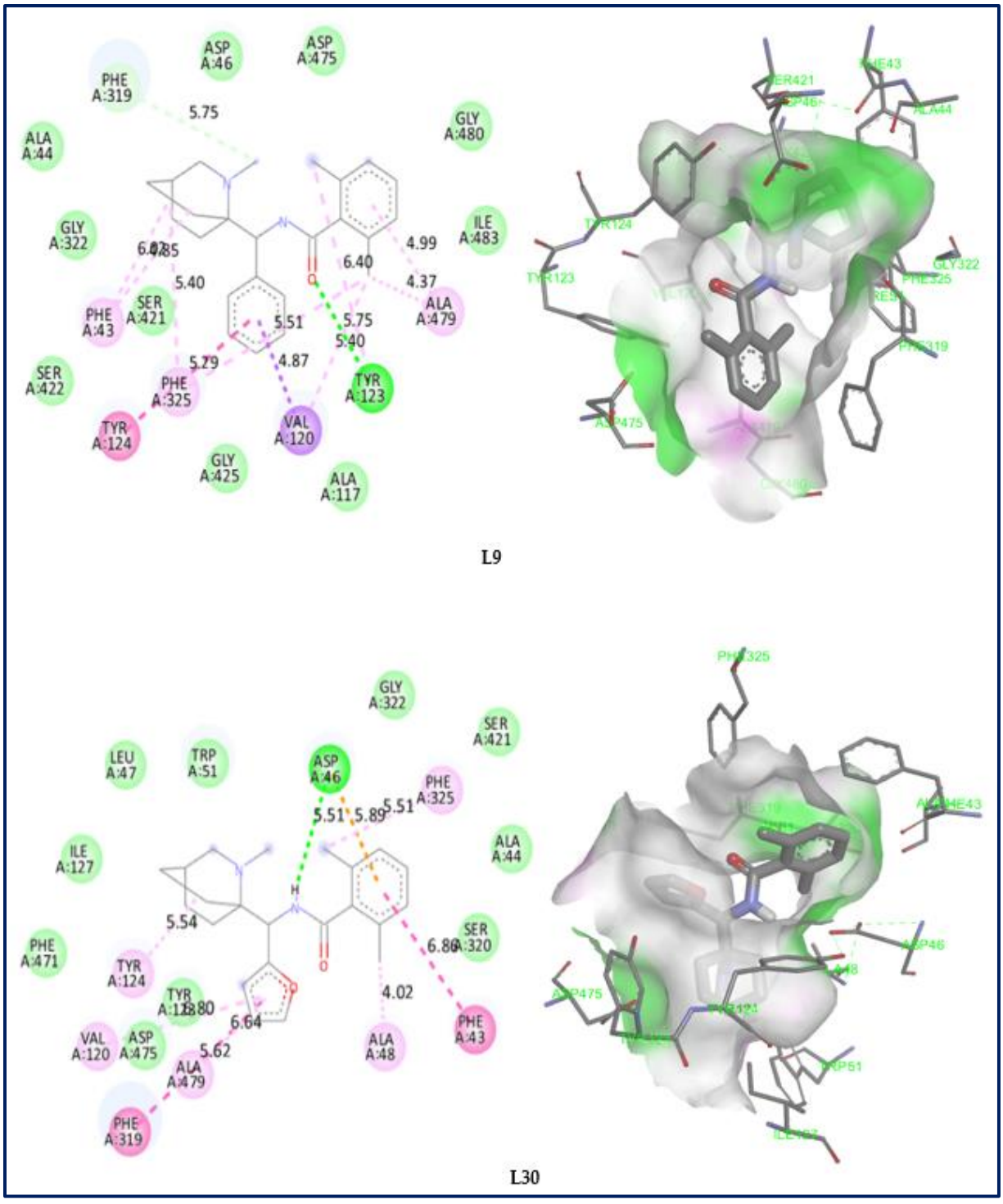



**Figure 27. Experimental pose view of Nortriptyline with the protein's active sites**




ASP A:46
ASP A:475
PHE A:319
5.75
GLY A:480
ALA A:44
GLY A:322
ILE A:483
6.40
4.99
4.37
5.40
PHE A:43
SER A:421
ALA A:479
5.51
5.75
5.40
5.29
4.87
SER A:422
PHE A:325
TYR A:123
TYR A:124
VAL A:120
GLY A:425
ALA A:117
L9
GLY A:322
SER A:421
LEU A:47
TRP A:51
ASP A:46
PHE A:325
5.51
5.89
5.51
ILE A:127
ALA A:44
5.54
SER A:320
6.86
PHE A:471
TYR A:124
4.02
6.64
5.62
VAL A:120
ASP A:475
ALA A:48
PHE A:43
ALA A:479
PHE A:319
L30

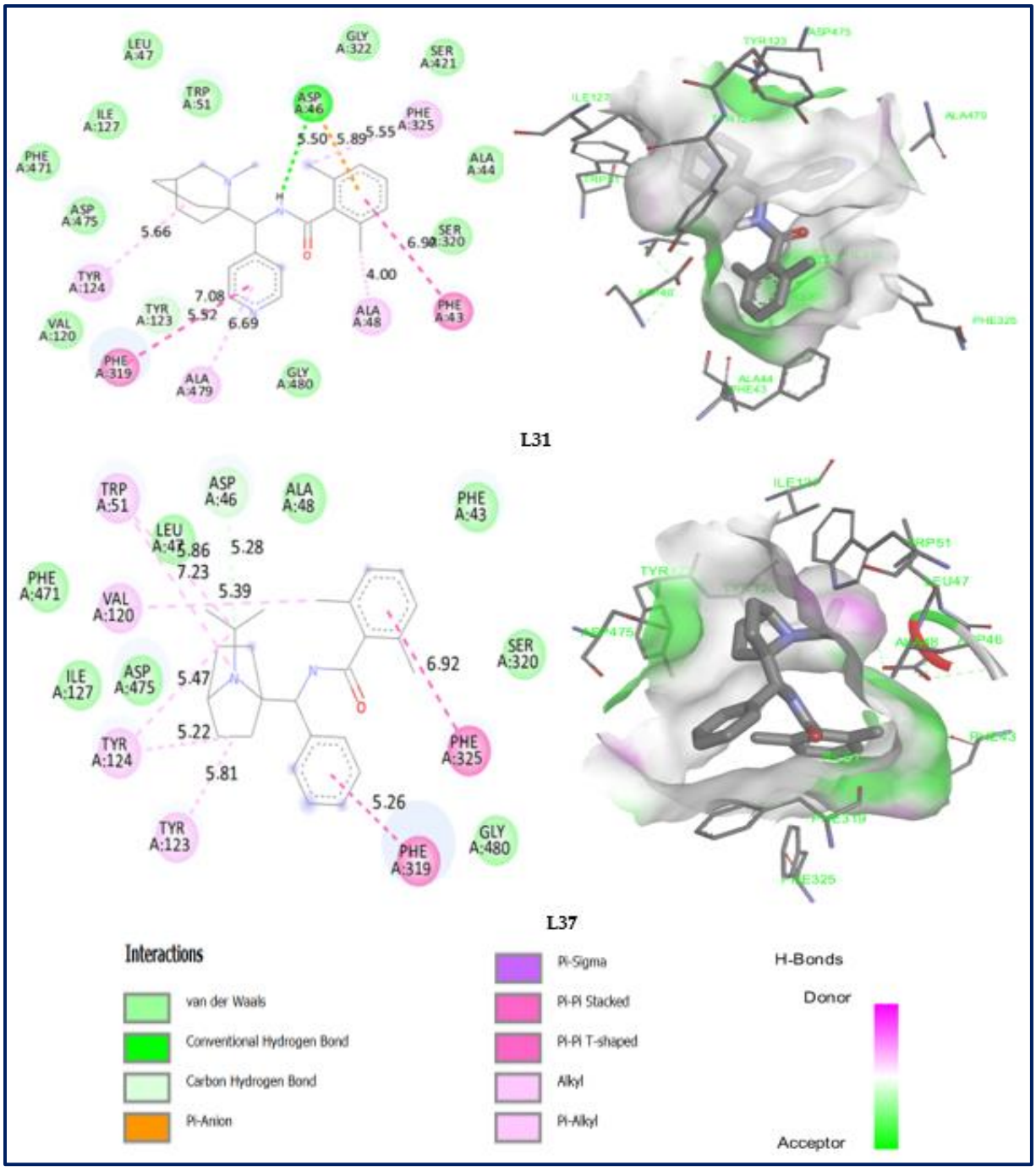



**Figure 28. 2D and 3D visualization of intermolecular interactions between DAT (PDB code: 4M48) and the more active ligands (L6, L9, L30, L31 and L37), with binding energies of −10.74 kcal/mol, −10.22 kcal/mol, −8.46 kcal/mol, −8.78 kcal/mol and −8.59 kcal/mol, respectively.**

## 2.8.Docking Validation Protocol

The efficiency of the molecular docking algorithms was tested using the re-docking methodology, which is based on the superposition of the docked ligand on the protein-bound ligand, as shown in Figure 29. The superposition result indicates a root mean square deviation

smaller than 2 (*RMSD* = 0.022 Å), which explains an exact pose prediction. Additionally, 2D and 3D visualization (Figure 30) of the intermolecular interaction between the docked nortriptyline and the protein target indicates that the chemical bonds formed with Phe43A and Tyr124A amino acids are the same as those observed experimentally. Thus, the molecular docking protocol is successfully validated [18].

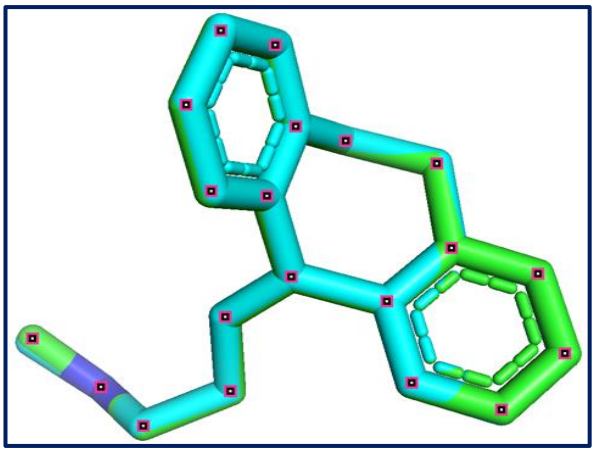


**Figure 29. Re-docking pose with RMSD equal to 0.022 Å (original nortriptyline of cyan color, and docked nortriptyline of green color).**

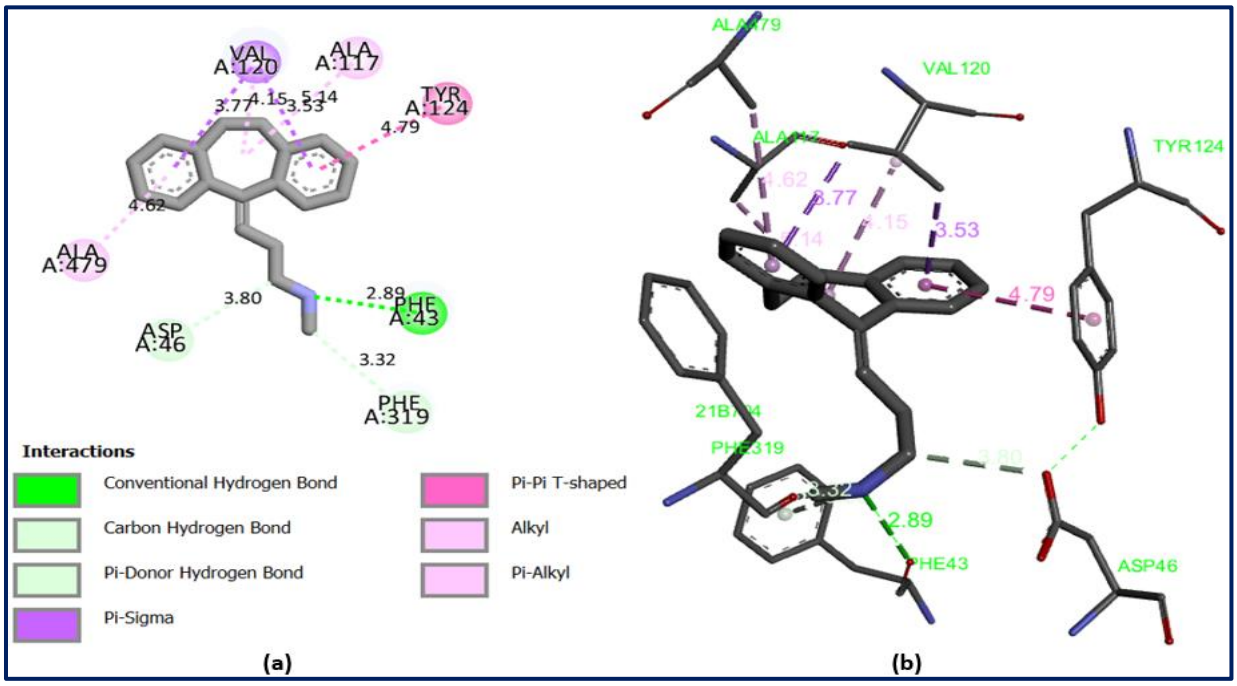



**Figure 30. 2D (a) and 3D (b) visualization of intermolecular interaction between the docked nortriptyline and the protein target (binding energy of −9.11 kcal/mol).**

### 2.9.Molecular Dynamics Simulations

The most active ligands (L6, L9, L30, L31 and L37) were chosen for the molecular dynamic's simulation during 100 ns, to examine their stability toward DAT protein, where the

conformational changes of one of these ligands are presented in Figure 31, and the others were presented in Figure S1.

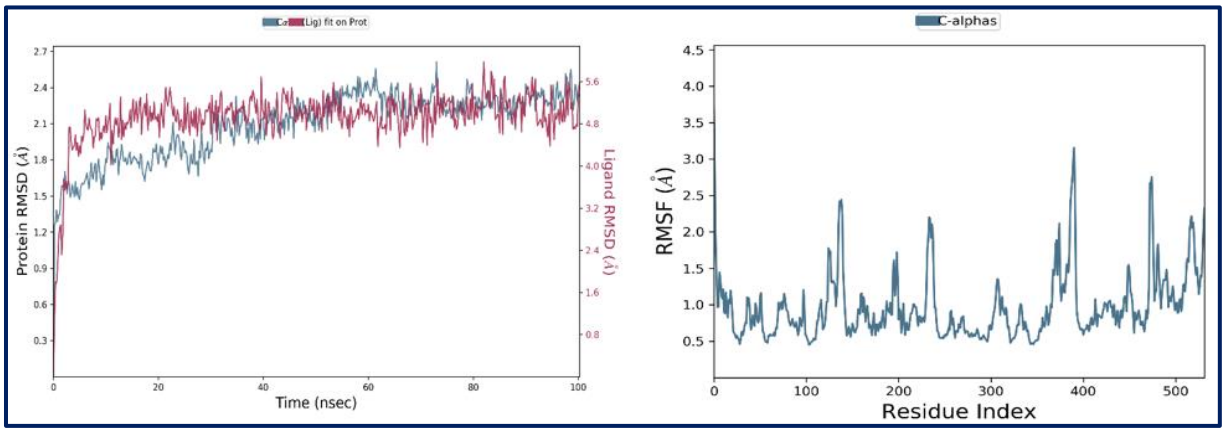



**Figure 31. RMSD and RMSF graphs for the ligand 9 complexed with the dopamine transporter membrane protein during 100 ns.**

The dynamic changes of conformation for (L9-protein) complex shown in Figure 31, indicate that the simulation is well-equilibrated, as the fluctuations of the root mean square deviation (RMSD) of the protein (left *Y*-axis) are around the thermal mean structure throughout the simulation time (100 ns), because the changes of the order of 1–3 Å are perfectly acceptable for small globular proteins. Moreover, the RMSD evolution of the heavy atoms of the ligand (right *Y*-axis) shows its stability with respect to the protein, when the protein-ligand complex is first aligned on the protein backbone of the reference, because the observed values are significantly smaller than the RMSD of the protein, and so the ligand did not diffuse away from its initial binding site.

The root mean square fluctuation (RMSF) values are also computed to examine the impact of the ligand binding on the internal dynamics of the target protein during 100 ns, where the tails (N- and C-terminal) fluctuate more than any other part of the protein and the secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein; for this reason, they fluctuate less than the loop regions. Except for a single fluctuation of 3.2 Å, detected in the loop region of residue 390, all fluctuations were less than 3 Å, indicating the binding strength between the ligand 9 and the DAT protein, and no significant change in the protein conformation resulting from ligand binding.

Additionally, the radius of gyration (r Gyr) values fluctuated in a small interval from 3.45 to 3.76 Å until the end of the simulation, as shown in Figure 32, indicating that there are just some changes in the compactness of the ligand; thus, the protein has a good flexibility after its binding with the ligand 9. Moreover, the solvent accessibility of the protein-ligand 9 complex

was evaluated by the solvent accessible surface area (SASA) analysis, which fluctuated between 0 and 15 $Å^2$ for 100 ns; this graph revealed that the structure of compound 9 was relatively stable during the simulation time. The polar surface area (PSA) is a solvent accessible surface area in a molecule contributed only by oxygen and nitrogen atoms; this parameter varies between 8 and 32 $Å^2$, accompanied by some maximal and minimal fluctuations during the simulation time. The contributions of this type of atoms make the ligand relatively unstable. However, the molecular surface area (MolSA) illustrates the molecular surface calculation with a probe radius of 1.4 Å, equivalent to a Van der Waals surface, showing only minimal fluctuations.

Lastly, the graph of the total energy presented in Figure 32 shows a minimal variation about the average −53,8682 $kcal.mol^{-1}$, which means that the energy of the L9- DAT protein complex remained in equilibrium throughout the MD simulation.

The dynamic changes of conformation for other complexes are available in the supplementary material (Figure S1). The protein-ligands interactions fluctuate with a root mean square deviation (RMSD) of 1 to 3 Å along the simulation time (100 ns), except for the L31-protein complex, which oscillates for the first 20 nanoseconds from 1 to 3 Å, then destabilizes until 50 ns, and stabilizes again with a deviation 4 Å of about until the end of the simulation time.

Minimal fluctuations have been observed during the established interactions between the protein and the ligand 6, such as the observed root mean square fluctuations (RMSF) of 3.4 Å, 4.4 Å and 3.9 Å detected in the loop region of 10, 240 and 480 residues, respectively, in addition to a maximal fluctuation of 5.5 Å recorded in the loop area of residue 380. No fluctuation was noticed greater than 3 Å for L30-protein and L37-protein complexes. Three fluctuations have been recorded for L31-protein complex of 3.6 Å, 3.5 Å and 7 Å reported in the loop zone of 10, 480 and 520 residues, respectively.

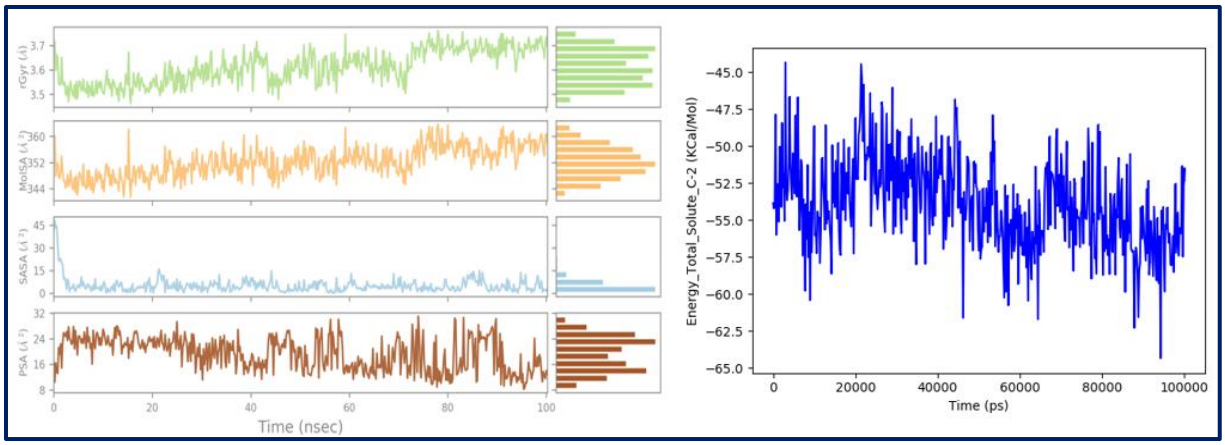



**Figure 32. Rg, MolSA, SASA and PSA during 100 ns of MD simulation, and the variation of total free energy for the ligand 9 complexed with DAT protein.**

Overall, we note that the protein has a good flexibility of binding with L6, L30, L31 and L37 inhibitors, as there are just a few changes in the compactness of the ligand, since the r Gyr, SASA, MolSA and PSA parameters are varied with minimal fluctuations about the mean along 100 ns.

Finally, the total energy plots presented in Figure S2 show a minimal variance around the average energy of the total system, which was in Kcal/mol of −47.7003, −48.8821 −49.5236 and −62.5749 for L30, L31, L37 and L6 inhibitors, respectively, indicating that the energies of the ligands-protein complexes have remained in equilibrium over the course of the MD simulation.

We conclude that the molecular dynamics simulations reinforce the previous results obtained by QSAR and docking studies, since the ligands L6, L9, L30, L31 and L37 are the most active inhibitors, forming typical static interactions with some amino acids of the target protein. These interactions form dynamically stable complexes during the 100 ns of the simulation time, as there is no change in their properties, except for the minimal fluctuations that were observed.

## 3. Materials and Methods

### 3.1.Database

The present study is performed on 44 bicyclo ((aryl) methyl) benzamides as glycine transporter type 1 (GlyT1) inhibitors, whose biological activities are expressed on a logarithmic decimal scale ($\log_{10}IC_{50}$), as illustrated in Table S1.

### 3.2.Molecular Descriptors Calculation

To build the quantitative structure-activity relationship (QSAR) models that provide information on the correlation between activities and structure-based molecular descriptors, we calculated various types of molecular descriptors [38], as shown in Table S2. Initially, the

constitutional descriptors were calculated using the ACD/chemsketch software [39]. Subsequently, the thermodynamic and physicochemical descriptors were extracted using the MM2 technique via the ChemBio3D software [40]. Lastly, the quantum descriptors are calculated through Gaussian 09 software [41], using the density function theory (DFT)/B3LYP [42], combined with the 6-31 + G(d,p) basis set, in order to ensure the molecules stability and optimize their three-dimensional geometries.

### 3.3.Statistical Methods

The Quantitative structure-activity relationships (QSARs) are developed with the help of XLSTAT 2014 software [43], using different statistical methods such as: the principal component analysis (PCA), multiple linear regression (MLR) and multiple non-linear regression (MNLR). The principal component analysis method is a very important step that serves to minimize the molecular descriptor dimension so as to identify the most predictive variables [44]. This limited number of descriptors is mathematically modeled by the multiple linear regression (MLR) and multiple non-linear regression (MNLR) techniques. Therefore, the two obtained QSAR models were generated to predict the linear and non-linear relationships established between the biological activity of glycine transporter type 1 (GlyT1) inhibitors and their relevant descriptors. For their applicability, these two models have been evaluated by external and internal validation, as well as the molecules, which have been tested in the applicability domain [45]. Additionally, the Golbreikh and Tropsha criteria and the Y-randomization test were used to verify the robustness and predictive potential of the established QSAR model [31,46].

### 3.4.Drug Likeness and In Silico Pharmacokinetics ADMET Prediction

To make the drugs applicable in clinical trials, it is necessary to study their absorption, distribution, metabolism, excretion and toxicity (ADMET) in the human body before starting the investigation protocols [21,47], respecting some important rules such as those of Lipinski [48], Veber [49], Ghose [50] and Egan [51,52]. This technique is also applied to eliminate the compounds with potentially undesirable physiological qualities, taking into account toxicity and pharmacokinetic properties [53]. For this task, we estimated the drug similarity and in silico pharmacokinetic properties of the newly selected molecules as GluT1 inhibitory agents, using the online SwissADMET [54] and pkCSM [55] servers, respectively.

### 3.5.Molecular Docking Modeling

The computational technique of molecular docking is an efficient, fast and powerful tool for drug discovery [56]. For this project, we uploaded the three-dimensional coordinates of the target protein from the protein data bank (pdb) using the Discovery Studio 2021 (BIOVIA)

software package [57]. To improve the performance of the cavity method, water molecules and suspended ligands bound to the protein were removed and polar hydrogens were added. Accordingly, the prepared protein was docked with the most active ligands, previously optimized by the density functional theory (DFT), with the assistance of AutoDock 4.2 [58]. Moreover, the grid box was centralized on (−42.562 Å, −0.46 Å, −55.066 Å) with the help of AUTOGRID algorithm, by putting the sizes (80, 80, 80) in their three-dimensional structure, and running 10 genetic algorithms with a total of 25 million trials. Finally, the molecular interactions of the protein-ligand were visualized using discovery studio 2021 [59].

### 3.6.Molecular Dynamics

Based on QSAR and molecular docking results, the five best-docked ligands, having the highest activity, were chosen for the molecular dynamics simulations in order to identify the molecular recognition between the ligand and the dopamine transporter (DAT) membrane protein. The MD simulations were performed for 100 nanoseconds using Desmond software, a package of Schrödinger LLC [60]. The first stage in the molecular dynamic's simulation of protein-ligand complexes was obtained by docking studies, and preprocessed using Protein Preparation Maestro, which performs optimization and minimization of complexes that have been prepared by the System Builder tool using a solvent model with an orthorhombic box that was chosen as TIP3P (transferable intermolecular interaction potential 3 points), using the OPLS force field [61]. At 300 K temperature and 1 atm pressure, the models were made neutral with the addition of water molecules, and counter ions, such as 0.15 M salt ($Na^+$,$Cl^-$) were added to mimic the physiological conditions. Finally, the trajectories were saved after every 10 ps for analysis, and the stability of simulations was evaluated by calculating the root mean square deviation (RMSD) of the protein and ligand over time. In addition, the root-mean-square fluctuation (RMSF), gyration radius (Rg), solvent accessible surface area (SASA), molecular surface area (MolSA) and polar surface area (PSA) were recorded for 100 ns, and the free energies of the inhibitors-protein interactions were evaluated using MM-GBSA approach [62].

## 4. Conclusion

A systematic in silico study was applied on 44 bicyclo((aryl)methyl)benzamide derivatives as glycine transporter type 1 (GlyT1) inhibitors to discover effective antipsychotic candidates for the treatment of schizophrenia. Initially, two QSAR models were developed using MLR and MNLR techniques and were examined through external and internal validation, applicability domain (AD), Y-randomization test and Golbreikh and tropsha criteria, indicating a significant effect of hydrogen bond donor, polarizability, surface tension, stretch and torsion energies and topological diameter on the locomotor activity (LMA). Subsequently, ADMET in silico

pharmacokinetics prediction revealed a favorable profile of the most active ligands, where L6, L9, L30, L31 and L37 were predicted as non-toxic inhibitors for 2D6 cytochrome, which respect the rules of Lipinski, Veber, Egan and Ghose, with an excellent absorption exceeded 91% and highest probability to penetrate the central nervous system (CNS). In contrast, the ligand L32 as an outlier in the QSAR study has an unfavorable ADMET profile with the highest probability of being absorbed by the gastrointestinal tract. Lastly, the obtained results were further strengthened and qualified using molecular docking and molecular dynamics studies, which confirm that L6, L9, L30, L31 and L37 react specifically with Tyr124, Phe43, Phe325, Asp46, Phe319 and Val120 amino acids of the dopamine transporter (DAT) membrane protein in a way that blocks glycine transporter type 1 (GlyT1) forming dynamically stable complexes during 100 ns of MD simulation time. Therefore, they could be used as therapeutics in medicine to treat schizophrenia. However, they must be subjected to in vitro and in vivo investigations to evaluate their efficacy and safety as anti-schizophrenia drugs.

**References**

[1] X. Zheng, C. Wang, N. Zhai, X. Luo, G. Liu, X. Ju, In Silico Screening of Novel α1-GABAA Receptor PAMs towards Schizophrenia Based on Combined Modeling Studies of Imidazo [1,2-a]-Pyridines, Int. J. Mol. Sci. 22 (2021) 9645. https://doi.org/10.3390/ijms22179645.

[2] L. Gao, C. Hao, J. Chen, R. Ma, L. Zheng, Q. Wu, X. Liu, B.-F. Liu, G. Zhang, Y. Chen, J. Jin, Discovery of a new class of multi-target heterocycle piperidine derivatives as potential antipsychotics with pro-cognitive effect, Bioorg. Med. Chem. Lett. 40 (2021) 127909. https://doi.org/10.1016/j.bmcl.2021.127909.

[3] A. Shahsavar, P. Stohler, G. Bourenkov, I. Zimmermann, M. Siegrist, W. Guba, E. Pinard, M.A. Seeger, T.R. Schneider, R.J.P. Dawson, P. Nissen, Structural insights into glycine reuptake inhibition, Neuroscience, 2020. https://doi.org/10.1101/2020.12.20.110478.

[4] J.G. Varnes, H. Xiong, J.M. Forst, C.R. Holmquist, G.E. Ernst, W. Frietze, B. Dembofsky, D.W. Andisik, W.E. Palmer, L. Hinkley, G.B. Steelman, D.E. Wilkins, G. Tian, G. Jonak, W.M. Potts, X. Wang, T.A. Brugel, C. Alhambra, M.W. Wood, C.A. Veale, J.S. Albert, Bicyclo((aryl)methyl)benzamides as inhibitors of GlyT1, Bioorg. Med. Chem. Lett. 28 (2018) 1043–1049. https://doi.org/10.1016/j.bmcl.2018.02.029.

[5] B. Cubelos, C. Giménez, F. Zafra, The glycine transporter GLYT1 interacts with Sec3, a component of the exocyst complex, Neuropharmacology. 49 (2005) 935–944. https://doi.org/10.1016/j.neuropharm.2005.07.021.

[6] B.L. Marques, O.C. Oliveira-Lima, G.A. Carvalho, R. de Almeida Chiarelli, R.I. Ribeiro, R.C. Parreira, E.M. da Madeira Freitas, R.R. Resende, F. Klempin, H. Ulrich, R.S. Gomez, M.C.X. Pinto, Neurobiology of glycine transporters: From molecules to behavior, Neurosci. Biobehav. Rev. 118 (2020) 97–110. https://doi.org/10.1016/j.neubiorev.2020.07.025.

[7] M.A. Lill, Multi-dimensional QSAR in drug discovery, Drug Discov. Today. 12 (2007) 1013–1017. https://doi.org/10.1016/j.drudis.2007.08.004.

[8] S. Vilar, G. Cozza, S. Moro, Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery, Curr. Top. Med. Chem. 8 (2008) 1555–1572. https://doi.org/10.2174/156802608786786624.

[9] S. Tian, J. Wang, Y. Li, D. Li, L. Xu, T. Hou, The application of in silico drug-likeness predictions in pharmaceutical research, Adv. Drug Deliv. Rev. 86 (2015) 2–10. https://doi.org/10.1016/j.addr.2015.01.009.

[10] A. Serrano, B. Imbernón, H. Pérez-Sánchez, J.M. Cecilia, A. Bueno-Crespo, J.L. Abellán, QN-Docking: An innovative molecular docking methodology based on Q-Networks, Appl. Soft Comput. 96 (2020) 106678. https://doi.org/10.1016/j.asoc.2020.106678.

[11] K. El Khatabi, R. El-mernissi, I. Aanouz, M.A. Ajana, T. Lakhlifi, M. Shahinozzaman, M. Bouachrine, Benzimidazole Derivatives in Identifying Novel Acetylcholinesterase Inhibitors: A Combination of 3D-QSAR, Docking and Molecular Dynamics Simulation, Phys. Chem. Res. 10 (2022) 237–249. https://doi.org/10.22036/pcr.2021.277306.1895.

[12] T.A. de Oliveira, L.R. Medaglia, E.H.B. Maia, L.C. Assis, P.B. de Carvalho, A.M. da Silva, A.G. Taranto, Evaluation of Docking Machine Learning and Molecular Dynamics Methodologies for DNA-Ligand Systems, Pharmaceuticals. 15 (2022) 132. https://doi.org/10.3390/ph15020132.

[13] H. Wang, Y. Jia, G. Jing, X. Wu, A novel toxicity prediction model for hydrazine compounds based on 1D–3D molecular descriptors, Comput. Toxicol. 18 (2021) 100169. https://doi.org/10.1016/j.comtox.2021.100169.
[14] A.K. Halder, T. Jha, Validated predictive QSAR modeling of N-aryl-oxazolidinone-5-carboxamides for anti-HIV protease activity, Bioorg. Med. Chem. Lett. 20 (2010) 6082–6087. https://doi.org/10.1016/j.bmcl.2010.08.050.
[15] K. Roy, S. Kar, R.N. Das, Validation of QSAR Models, in: Underst. Basics QSAR Appl. Pharm. Sci. Risk Assess., Elsevier, 2015: pp. 231–289. https://doi.org/10.1016/B978-0-12-801505-6.00007-7.
[16] L. El Mchichi, A. El Aissouq, R. Kasmi, A. Belhassan, R. El-Mernissi, A. Ouammou, T. Lakhlifi, M. Bouachrine, In silico design of novel Pyrazole derivatives containing thiourea skeleton as anti-cancer agents using: 3D QSAR, Drug-Likeness studies, ADMET prediction and molecular docking, Mater. Today Proc. 45 (2021) 7661–7674. https://doi.org/10.1016/j.matpr.2021.03.152.
[17] M. Mazigh, C. El'mbarki, H. Hadni, M. Elhallaoui, QSAR studies combined with DFT-calculations and Molecular docking of polyamine-sensitive inhibitors of the NMDA receptor, Mediterr. J. Chem. 9 (2019) 164–174. https://doi.org/10.13171/mjc93190924930hh.
[18] M. Abdullahi, G.A. Shallangwa, A. Uzairu, In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of H5N1 influenza A virus subtype, Beni-Suef Univ. J. Basic Appl. Sci. 9 (2020) 2. https://doi.org/10.1186/s43088-019-0023-y.
[19] E. Krieger, T. Darden, S.B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: Force-field parameterization in crystal space, Proteins Struct. Funct. Bioinforma. 57 (2004) 678–683. https://doi.org/10.1002/prot.20251.
[20] A. Bastianoni, E. Guastaldi, A. Barbagli, S. Bernardinetti, A. Zirulia, M. Brancale, T. Colonna, Multivariate Analysis Applied to Aquifer Hydrogeochemical Evaluation: A Case Study in the Coastal Significant Subterranean Water Body between "Cecina River and San Vincenzo", Tuscany (Italy), Appl. Sci. 11 (2021) 7595. https://doi.org/10.3390/app11167595.
[21] O. Daoui, S. Elkhattabi, S. Chtita, R. Elkhalabi, H. Zgou, A.T. Benjelloun, 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 (2021) e07463. https://doi.org/10.1016/j.heliyon.2021.e07463.
[22] Y. Matsuzaka, Y. Uesawa, A Deep Learning-Based Quantitative Structure–Activity Relationship System Construct Prediction Model of Agonist and Antagonist with High Performance, Int. J. Mol. Sci. 23 (2022) 2141. https://doi.org/10.3390/ijms23042141.
[23] G.K. Uyanık, N. Güler, A Study on Multiple Linear Regression Analysis, Procedia - Soc. Behav. Sci. 106 (2013) 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027.
[24] N. Kravić, J. Savosina, M. Agafonova-Moroz, V. Babain, A. Legin, D. Kirsanov, Nonlinear Multivariate Regression Algorithms for Improving Precision of Multisensor Potentiometry in Analysis of Spent Nuclear Fuel Reprocessing Solutions, Chemosensors. 10 (2022) 90. https://doi.org/10.3390/chemosensors10030090.

[25] I.L. Ruiz, M.Á. Gómez-Nieto, Study of the Applicability Domain of the QSAR Classification Models by Means of the Rivality and Modelability Indexes, Molecules. 23 (2018) 2756. https://doi.org/10.3390/molecules23112756.
[26] E. Pourbasheer, S. Riahi, M.R. Ganjali, P. Norouzi, Quantitative structure–activity relationship (QSAR) study of interleukin-1 receptor associated kinase 4 (IRAK-4) inhibitor activity by the genetic algorithm and multiple linear regression (GA-MLR) method, J. Enzyme Inhib. Med. Chem. 25 (2010) 844–853. https://doi.org/10.3109/14756361003757893.
[27] S. Chtita, A. Belhassan, M. Bakhouch, A.I. Taourati, A. Aouidate, S. Belaidi, M. Moutaabbid, S. Belaaouad, M. Bouachrine, T. Lakhlifi, QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods, Chemom. Intell. Lab. Syst. 210 (2021) 104266. https://doi.org/10.1016/j.chemolab.2021.104266.
[28] K.B. de C. Rezende, A.J.L.A. da Cunha, J. Amim Junior, R.G. Bornia, External validation of the Fetal Medicine Foundation algorithm for the prediction of preeclampsia in a Brazilian population, Pregnancy Hypertens. 17 (2019) 64–68. https://doi.org/10.1016/j.preghy.2019.05.006.
[29] M. Rafało, Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis, ICT Express. (2021) S2405959521000552. https://doi.org/10.1016/j.icte.2021.05.001.
[30] S. Chtita, M. Ghamali, A. Ousaa, A. Aouidate, A. Belhassan, A.I. Taourati, V.H. Masand, M. Bouachrine, T. Lakhlifi, QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors, Heliyon. 5 (2019) e01304. https://doi.org/10.1016/j.heliyon.2019.e01304.
[31] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1.
[32] A. Daina, O. Michielin, V. Zoete, iLOGP: A Simple, Robust, and Efficient Description of *n* -Octanol/Water Partition Coefficient for Drug Design Using the GB/SA Approach, J. Chem. Inf. Model. 54 (2014) 3284–3301. https://doi.org/10.1021/ci500467k.
[33] A. Daina, V. Zoete, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules, ChemMedChem. 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.
[34] A. Belhassan, H. Zaki, M. Benlyas, T. Lakhlifi, M. Bouachrine, Study of novel triazolo-benzodiazepine analogues as antidepressants targeting by molecular docking and ADMET properties prediction, Heliyon. 5 (2019) e02446. https://doi.org/10.1016/j.heliyon.2019.e02446.
[35] A. Adamski, D. Kruszka, Z. Dutkiewicz, M. Kubicki, A. Gorczyński, V. Patroniak, Novel family of fused tricyclic [1,4]diazepines: Design, synthesis, crystal structures and molecular docking studies, Tetrahedron. 73 (2017) 3377–3386. https://doi.org/10.1016/j.tet.2017.05.015.
[36] A. Penmatsa, K.H. Wang, E. Gouaux, X-ray structure of dopamine transporter elucidates antidepressant mechanism, Nature. 503 (2013) 85–90. https://doi.org/10.1038/nature12533.
[37] Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server, (n.d.). https://proteins.plus/ (accessed March 8, 2022).
[38] J. Spiegel, H. Senderowitz, Evaluation of QSAR Equations for Virtual Screening, Int. J. Mol. Sci. 21 (2020) 7828. https://doi.org/10.3390/ijms21217828.

[39] T. Österberg, U. Norinder, Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors, Eur. J. Pharm. Sci. 12 (2001) 327–337. https://doi.org/10.1016/S0928-0987(00)00189-5.
[40] G.W.A. Milne, Software Review of ChemBioDraw 12.0, J. Chem. Inf. Model. 50 (2010) 2053–2053. https://doi.org/10.1021/ci100385n.
[41] Citation gaussienne 09 | Gaussian.com, (n.d.). https://gaussian.com/g09citation/ (accessed March 8, 2022).
[42] R.G. Parr, Y. Weitao, Density-Functional Theory of Atoms and Molecules, Oxford University Press, New York, 1995. https://doi.org/10.1093/oso/9780195092769.001.0001.
[43] Introducing the XLSTAT Free edition, XLSTAT Your Data Anal. Solut. (n.d.). https://www.xlstat.com/en/news/introducing-the-xlstat-free-edition (accessed March 8, 2022).
[44] C. Guerra Tort, V. Aguiar Pulido, V. Suárez Ulloa, F. Docampo Boedo, J.M. López Gestal, J. Pereira Loureiro, Electronic Health Records Exploitation Using Artificial Intelligence Techniques, Proceedings. 54 (2020) 60. https://doi.org/10.3390/proceedings2020054060.
[45] H. Hadni, M. Elhallaoui, 3D-QSAR, docking and ADMET properties of aurone analogues as antimalarial agents, Heliyon. 6 (2020) e03580. https://doi.org/10.1016/j.heliyon.2020.e03580.
[46] H. van der Voet, Comparing the predictive accuracy of models using a simple randomization test, Chemom. Intell. Lab. Syst. 25 (1994) 313–323. https://doi.org/10.1016/0169-7439(94)85050-X.
[47] O. Ebenezer, N. Damoyi, M.A. Jordaan, M. Shapi, Unveiling of Pyrimidindinones as Potential Anti-Norovirus Agents—A Pharmacoinformatic-Based Approach, Molecules. 27 (2022) 380. https://doi.org/10.3390/molecules27020380.
[48] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[49] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[50] S. Yalcin, Molecular Docking, Drug Likeness, and ADMET Analyses of Passiflora Compounds as P-Glycoprotein (P-gp) Inhibitor for the Treatment of Cancer, Curr. Pharmacol. Rep. 6 (2020) 429–440. https://doi.org/10.1007/s40495-020-00241-6.
[51] W.J. Egan, Merz Kenneth M., J.J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics, J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[52] W.J. Egan, G. Lauri, Prediction of intestinal permeability, Adv. Drug Deliv. Rev. 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.
[53] D. Gomes, S. Silvestre, A.P. Duarte, A. Venuti, C.P. Soares, L. Passarinha, Â. Sousa, In Silico Approaches: A Way to Unveil Novel Therapeutic Drugs for Cervical Cancer Management, Pharmaceuticals. 14 (2021) 741. https://doi.org/10.3390/ph14080741.
[54] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.

[55] pkCSM, (n.d.). http://biosig.unimelb.edu.au/pkcsm/run_example? (accessed March 8, 2022).
[56] D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, S. Moro, Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro, Pharmaceuticals. 15 (2022) 180. https://doi.org/10.3390/ph15020180.
[57] D. Iqbal, M.T. Rehman, A. Bin Dukhyil, S.M.D. Rizvi, M.F. Al Ajmi, B.M. Alshehri, S. Banawas, M.S. Khan, W. Alturaiki, M. Alsaweed, High-Throughput Screening and Molecular Dynamics Simulation of Natural Product-like Compounds against Alzheimer's Disease through Multitarget Approach, Pharmaceuticals. 14 (2021) 937. https://doi.org/10.3390/ph14090937.
[58] A.P. Norgan, P.K. Coffman, J.-P.A. Kocher, D.J. Katzmann, C.P. Sosa, Multilevel Parallelization of AutoDock 4.2, J. Cheminformatics. 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[59] BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes®, (n.d.). https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/ (accessed March 8, 2022).
[60] Desmond | Schrodinger, (n.d.). https://www.schrodinger.com/products/desmond (accessed April 28, 2022).
[61] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B. 105 (2001) 6474–6487. https://doi.org/10.1021/jp003919d.
[62] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J.Z.H. Zhang, T. Hou, End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design, Chem. Rev. 119 (2019) 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055.

# Chapitre VI. Modélisation moléculaire d'inhibiteurs sélectifs du transporteur de glycine type 1, destinés pour le traitement des perturbations de la mémoire

## Résumé

Une production scientifique, publiée en 2023, portant sur une famille structurelle de quarante inhibiteurs du transporteur de glycine de type 1 (GlyT1), qui ont été étudié à l'aide des techniques de modélisation moléculaire. A savoir, la modélisation moléculaire linéaire et non linéaire basée sur les relations quantitatives structure-activité (QSAR), a indiqué que l'activité humaine du GlyT1 est fortement et significativement affectée par les descripteurs constitutionnels, géométriques, physicochimiques et topologiques.

En outre, les prédictions en In-silico des propriétés pharmacocinétiques d'ADME-Toxicité, ont révélé que les ligands L28 et L30 étaient prédits comme des inhibiteurs non toxiques définis par un bon profil ADME et haute probabilité de pénétrer dans le système nerveux central (SNC).

D'autre part, les résultats du Docking moléculaire ont indiqué que les inhibiteurs prédits bloquent GlyT1, en réagissant spécifiquement avec les acides aminés Phe319, Phe325, Tyr123, Tyr 124, Arg52, Asp475, Ala117, Ala479, Ile116 et Ile483 de la protéine membranaire du transporteur de dopamine (DAT).

Ces résultats ont été nuancés et renforcés par une étude de dynamique moléculaire (MD), qui a affirmé que les interactions intermoléculaires établies pour les deux complexes (L28, L30-protéine DAT) restent parfaitement stables pendant 50 ns du temps de simulation de la dynamique moléculaire. Par conséquent, ils pourraient être fortement recommandés en médecine comme thérapeutiques pour améliorer les performances de la mémoire.


**Mots clés** : QSAR; GlyT1; ADME-Tox; SNC; DAT; Docking moléculaire ; dynamique moléculaire.

# Chapitre VI: QSAR, ADME-Tox, Molecular Docking and Molecular Dynamics Simulations of Novel Selective Glycine Transporter Type 1 Inhibitors with Memory Enhancing Properties

Mohamed El fadili[a,*], Mohammed Er-rajy[a] , Hamada Imtara[b] , Omar M. Noman[c] , Ramzi A. Mothana[c] , Sheaf Abdullah[d] , Sara Zerougui[a] , and Menana Elhallaoui[a]

[a] LIMAS Laboratory, Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, BP 1796 Atlas, Fez 30000, Morocco;

[b] Faculty of Arts and Sciences, Arab American University Palestine, Jenin BP Box 240, Palestine;

[c] Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia;

[d] Department of Hand Surgery and Microsurgery, University Medicine Greifswald, Greifswald, Germany;

* Corresponding author: mohamed.elfadili@usmba.ac.ma

**Abstract:** A structural class of forty glycine transporter type 1 (GlyT1) inhibitors, was examined using molecular modeling techniques. The quantitative structure-activity relationships (QSAR) technology confirmed that human GlyT1 activity is strongly and significantly affected by constitutional, geometrical, physicochemical and topological descriptors. ADME-Tox in-silico pharmacokinetics revealed that L28 and L30 ligands were predicted as non-toxic inhibitors with a good ADME profile and the highest probability to penetrate the central nervous system (CNS). Molecular docking results indicated that the predicted inhibitors block GlyT1, reacting specifically with Phe319, Phe325, Tyr123, Tyr 124, Arg52, Asp475, Ala117, Ala479, Ile116 and Ile483 amino acids of the dopamine transporter (DAT) membrane protein. These results were qualified and strengthened using molecular dynamics (MD) study, which affirmed that the established intermolecular interactions for (L28, L30–DAT protein) complexes remain perfectly stable along 50 ns of MD simulation time. Therefore, they could be strongly recommended as therapeutics in medicine to improve memory performance.

**Keywords:** QSAR; GlyT1; ADME-Tox; CNS; DAT; Molecular docking; MD simulations.

## 1. Introduction

About 25% of the worldwide population suffers from neurological disorders, according to the World Health Organization (WHO) report, published in 2001 [1]. Many of these neurological diseases are the result of memory dysfunctions [2]. So, the rationale for this virtual screening is the discovery of successful candidate drugs that can improve memory performance. For this goal, we have processed with the assistance of molecular modeling techniques, a set of forty pyrrolo[3,4-c]pyrazole and azetidine-based analogs, given in the Table S3, as selective glycine transporter type 1 (GlyT1) inhibitors, having the potential therapies for memory disorders, considering that GluT1 is one of the most essential neurotransmitters that control learning and memory functions [3,4].

Over the last decade, the glycine transporter type 1 (GlyT1) has received enormous interest as it is one of the key constituents of glycine metabolism in the central nervous system (CNS). Whereas, the amino acid glycine is an excitatory and inhibitory neurotransmitter in the vertebrate central nervous system. In the cortex, hippocampus, and thalamus areas, glycine regulates excitatory neurotransmission as a mandatory co-agonist at N-methyl-D-aspartate (NMDA) receptors (glycine B sites). Inversely, it acts as an inhibitory neurotransmitter at strychnine-sensitive glycine receptor ion channels (glycine A sites) mainly present in the brainstem, cerebellum, and spinal cord [3,5]. Hence, GlyT1 inhibitors have been introduced into the clinic as potent treatments for obsessive-compulsive disorder (OCD), negative and cognitive symptoms associated with schizophrenia, addiction, depression, and Parkinson's disease [3,5,6]. Nowadays, computer-assisted drug design (CADD) based on in silico techniques has gained enormous importance and has been widely reported in the literature. In this regard, we have examined all 40 GlyT1 inhibitors, through the use of computer-assisted drug design (CADD) technology, to detect and identify a non-toxic, highly effective CNS drug candidate intended to treat memory impairments [7].

In the first stage, we started this task, using quantitative structure-activity relationships (QSAR) as a technique largely applied in the field of drug discovery, predicting the human GlyT1 activity of 40 pyrrolo[3,4-c]pyrazole and azetidine derivatives, based on two mathematical QSAR models, which were established thanks to the use of multiple linear regression (MLR) and multiple nonlinear regression (MNLR) respectively, to model the linear and nonlinear relationships between human GlyT1 activity and different types of molecular descriptors [6,8].

In the second part, we have predicted adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties for forty selective inhibitors of glycine transporter type 1, with

the aim of discovering qualified drugs with a favorable ADMET profile [9,10]. Subsequently, we have performed the molecular docking study for the predicted ligands that traverse the central nervous system (CNS), to identify the established intermolecular interactions with the amino acids from the membrane protein of dopamine transporter (DAT) coded as 4M48.pdb [11,12].

Lastly, this work has been concluded using molecular dynamics simulations during 50 nanoseconds, to verify the stability of (predicted ligands-protein target) complexes, and monitoring the total free energy variation in Kcal/mol of each complex based on the MM-GBSA approach [6,13,14].

## 2. Materials and methods

### 2.1.Experimental Database

The current study includes forty pyrrolo[3,4-c]pyrazole and azetidine-based derivatives (Figure S3) having human GlyT1 activities [3,5], that were subsequently expressed on a logarithmic decimal scale ($pIC_{50} = - \log_{10} IC_{50}$), as presented in the Table S3.

### 2.2.Molecular Descriptors Calculation

To model the quantitative structure-activity relationships (QSAR), various types of molecular descriptors have been calculated as presented in Table S4. With the help of Gaussian 09 software [15], we have calculated the quantum descriptors, to optimize the 3D geometries and ensure molecular stability with minimized energies using the density functional theory (DFT) coupled with B3LYP/6-31+G(d,p) basis [16,17]. Then, we executed the thermodynamic, topological, and physicochemical descriptors based on the MM2 technique using the ChemBio3D software [18]. Finally, the constitutional descriptors have been processed through the ACD/chemsketch software [19].

### 2.3.Statistical Methods

To minimize the size of the molecular descriptors into a reduced number of factorial axes (or principal components), we have applied the principal component analysis (PCA), as a very important step to identify the most predictive variables that can influence the biological activity of inhibitors [20]. Thereafter, two statistical methods were performed on the training set (N=32) to develop a predictive QSAR model using XLSTAT 2014 software [21]. The first one is the multiple linear regression (MLR) technique, which is widely used to investigate the linear relationship between the biological activity of $pIC_{50}$ order and the previously calculated molecular descriptors. The second, is the multiple nonlinear regression (MNLR) analysis, which is adapted to examine the nonlinear relationship established between the inhibitory activity and the calculated descriptors [22]. Next, the MNLR and MLR QSAR models were

checked with external validation for a test set (N=8) and further evaluated with internal validation using the leave-one-out (LOO) procedure of the cross-validation (CV) technique. Additionally, the Tropsha and Golbreikh criteria, applicability domain, and the Y-randomization test were also utilized to evaluate the predictive potential and robustness of the developed QSAR models [23,24].

**2.4.In silico pharmacokinetics, drug-likeliness, and ADMET depiction exploration**

To obtain a novel medication with high efficiency and less experimental investigation time, it's mandatory to evaluate their absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties in the human organism before beginning any clinical trials, verifying the five rules of Lipinski [25], and respecting the Veber, Ghose, Egan and Muegge violations number of drug candidates [26–30]. For this purpose, we have predicted in silico pharmacokinetics and ADMET properties of the candidate's molecules as GlyT1 inhibitors enhancing memory performance, using pkCSM and online SwissADME servers, respectively [31,32].

**2.5.Molecular Docking Simulation**

Molecular docking is an accurate, quick, and successful technology in the field of drug discovery [33]. It is largely applied to examine the intermolecular interactions established between the candidate drug and the targeted protein. For this goal, we have extracted the 3D coordinates of the drosophila melanogaster dopamine transporter encoded as 4M48 from the data bank of proteins (pdb) [34], with a resolution equal to 2.96 Å, using the X-ray diffraction method [35]. With the assistance of the Discovery Studio 2021 (BIOVIA) software package [36], we have deleted the water molecules ($H_2O$) and suspended ligands that were bound to the protein, such as nortriptyline ($C_{19}H_{21}N$), cholesterol ($C_{27}H_{46}O$), chloride and sodium ions (Na+, Cl-). Then, we added the polar hydrogens, to ensure that the performance of the cavity method is improved [37]. Thereafter, we launched the prepared protein, L28, and L30 ligands in AutoDock 4.2 software [38], going from PDB to PDBQT format. With the assistance of the AUTOGRID algorithm, the grid box dimensions were fixed on (−42.56 Å, −0.46 Å, −55.07 Å), choosing the 3D- structure sizes on (110,110,110) with a spacing of 0.375 Å, and running ten genetic algorithms for twenty-five million evals. Lastly, 3D and 2D intermolecular interactions of candidate drug-targeted protein complexes were visualized with the help of Discovery Studio 2021 (BIOVIA) software [36,37].

**2.6.Molecular Dynamic**

To ensure the intermolecular interactions stability of protein-ligand complexes, we have performed the molecular dynamics technique during 50 nanoseconds using Desmond software,

a package of Schrödinger LLC [39]. The output files of molecular docking were defined as input files of molecular dynamic simulations, which were prepared by Protein Preparation Maestro, which optimizes and minimizes the pre-treated complexes in the System Builder utilizing a solvent model having an orthorhombic box of transferable intermolecular interaction potential 3 points (TIP3P) type, and an OPLS-like force field [40]. In addition, the models were neutralized with the insertion of water molecules and counter ions, where the physiological conditions were reproduced in 0.15 M salts (Na+, Cl-), under a temperature and pressure of 300 K and 1 atm, respectively. After every 10 picoseconds of evaluation, we recorded the trajectories, and we examined the stability of the simulations, studying the root mean square deviation (RMSD) for each complex. Thereafter, we have supervised the root-mean-square fluctuation (RMSF), polar surface area (PSA), solvent accessible surface area (SASA), gyration radius (Rg), and molecular surface area (MolSA) variations [6]. Lastly, the free energies of the protein-candidate drug interactions were tested at 50 ns of simulations time, with the help of MM-GBSA technology [41].

## 3. Results and discussion

### 3.1.Principal Component Analysis

To minimize the dimension of molecular descriptors in a reduced number of principal components, we have focused on the correlation matrix given by the principal component analysis (PCA) technique, as one of the most applied multivariate analyses [42]. We have started this task with a database of 40 molecules and 40 different descriptors as illustrated in Table S4. Then, we reduced this uncontrollable number of descriptors into 25 uncorrelated descriptors, as the strongly correlated descriptors between them (Pearson correlation coefficient>0.9) were removed, so to facilitate the development of QSAR mathematical models.

### 3.2.Experimental Database

The present study is intended to develop a mathematical QSAR model between the structural descriptors of human GlyT1 inhibitors and their biological activities of $pIC_{50}$ order. For this reason, we have worked on an experimental database of forty compounds and twenty-five uncorrelated descriptors. Then, we based on thousands of randomizations with a ‘stepwise’ option, dividing our statistical database into training and test sets. The first set contains 80% of global data (32 molecules) taken to establish the QSAR model, and the second set includes 20% (8 molecules) selected to validate the established model [6,37,43].

### 3.3.Multiple Linear Regression

The first QSAR model was constructed for a training set of thirty-two molecules (N=32) using multiple linear regression (MLR) as a statistical method capable to evaluate the linear

relationship between the molecular descriptors and the inhibitory activity [44]. The process was based on the best values of the correlation, determination, and adjustment coefficients ($R$, $R^{2,}$ and $R^2adj$). So, the optimal MLR QSAR model was developed in Equation 1:

$$\mathbf{pIC_{50} = 3.82 + 0.01 \times MW - 0.32 \times NRB + 0.28 \times Log\ P - 0.22 \times TD + 0.045 \times PSA\ (1)}$$

The predictive QSAR model introduces the following selected descriptors: molecular weight (MW), partition coefficient (Log P), number of rotatable bonds (NRB), polar surface area (PSA), and topological diameter (TD), as included in the Table S5.

The slope of each selected descriptor has a probability inferior to 5%, as indicated by the significance test in Table 21. So, the selected variables have a significant impact on the activity at a confidence interval of 95%. Molecular weight, partition coefficient, and polar surface area have a positive effect on activity, but the number of rotatable bonds and topological diameter influence it negatively, as illustrated in Figure 33. Thus, the compound can be highly active if it's defined by a small value of TD and NRB and given by higher values of MW, Log P, and PSA. According to Fisher's statistical test, the null hypothesis (H0) is rejected because the analysis of variance (Anova) proves that the critical value: [$F$ (32, 5) = 2.51, $p < 0.0001$] is much smaller than its calculated value ($F$=13.832), as presented in Table 22. Thus, the dependent variable ($pIC_{50}$) and the independent variables (5 selected descriptors) are defined by a homogeneous variance. In addition, the correlation, determination, and adjustment coefficients are equal to $R$=0.85, $R^2$=0.73, and $R^2adj$=0.67, respectively. which confirms a solid relationship between the inhibitory activity and the molecular descriptors, and a strong predictive performance of the MLR QSAR model, with a minimal root mean square error (RMSE=0.297). Consequently, we conclude that human GlyT1 activity is strongly and significantly correlated with constitutional (MW), physicochemical (Log P), geometrical (NRB), and topological (PSA and TD) descriptors.

**Table 21. Significant weights of the selected descriptors**

| Source | Value | Standard deviation | $t$ | $Pr > \|t\|$ | Lower terminal (95%) | Higher terminal (95%) |
|---|---|---|---|---|---|---|
| Constante | 3.823 | 1.307 | 2.926 | **0.007** | 1.137 | 6.509 |
| MW | 0.010 | 0.002 | 5.608 | **< 0.0001** | 0.007 | 0.014 |
| NRB | -0.321 | 0.056 | -5.769 | **< 0.0001** | -0.436 | -0.207 |
| Log P | 0.280 | 0.095 | 2.950 | **0.007** | 0.085 | 0.474 |
| TD | -0.222 | 0.105 | -2.118 | **0.044** | -0.437 | -0.007 |
| PSA | 0.045 | 0.010 | 4.453 | **0.000** | 0.024 | 0.065 |

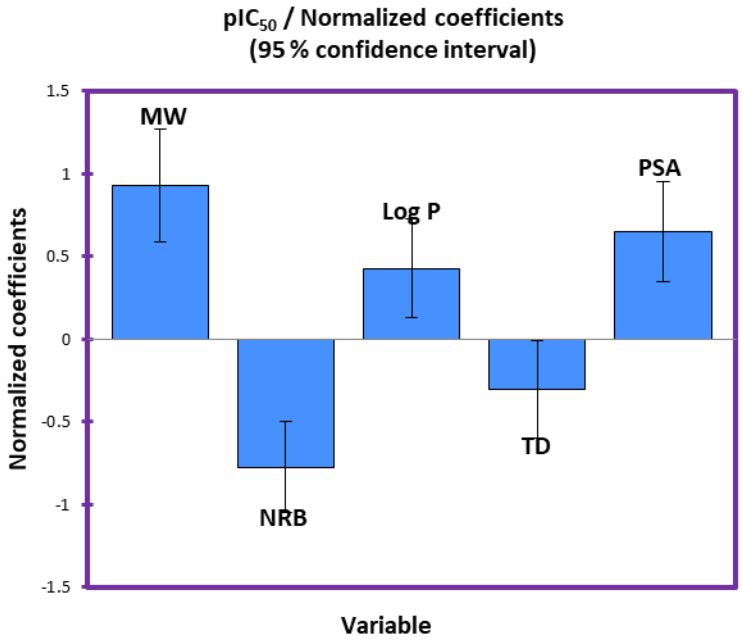



**Figure 33. Impact of molecular descriptors on the biological activity**

**Table 22. Analysis of variance using the fisher test**

| Source | DDL | Total square | Mean square | F | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 6.086 | 1.217 | **13.832** | < 0.0001 |
| Error | 26 | 2.288 | 0.088 | | |
| Adjusted total | 31 | 8.374 | | | |

### 3.4.Multiple Non-Linear Regression

To model the nonlinear relationship between the previously selected descriptors and human GlyT1 activity, we have applied the multiple non-linear regression (MNLR) technique [45], for a training set of thirty-two molecules (N=32) using the following programmed function (Equation 2):

$$Y = a0 + \sum_{i=1}^{n} (ai \times Xi + bi \times Xi^2) \qquad (2)$$

Where, Y: is the predicted activity with MNLR QSAR model, a0: is the constant of the mathematical model, Xi: is the selected descriptor, ai and bi: are the slopes of each molecular descriptor for 1 and 2 degrees, respectively [6]. Thus, we have obtained the second QSAR model as presented by Equation 3:

**$pIC_{50}$ = - 0.487 + 0.006×MW - 0.629×NRB + 0.577×Log P - 0.477×TD + 0.232×PSA + 3.887E-6* MW^2 + 0.024×NRB^2 - 0.055×Log P^2 + 0.010×TD^2 - 0.001×PSA^2 (3)**

This predictive model is qualified by a strong correlation coefficient of $R$=0.88, a good determination coefficient ($R^2$=0.78) and a low root mean square error ($RMSE$=0.296). So, it affirms that human GlyT1 activity is affected by these five selected descriptors.

### 3.5.Validation of predictive QSAR Models

#### 3.5.1. Applicability Domain

To examine the applicability of quantitative activity-structure relationship models, we have performed the applicability domain (AD) technology. using IBM SPSS Statistics 22 software. based on William's diagram shown in Figure 34. The warning leverage ($h^*$=0.56) was indicated using the following equation: $h^* = [3 \times (p + 1) / n]$, ($p$=5, $n$=32). Where, p: is the number of selected descriptors. n: is the number of molecules in the training set, and 3: is the absolute value limit of the standardized residuals interval on the Y-axis [46]. We note that all molecules in the training and test sets have a leverage under the critical leverage value, and they are all placed in the confidence interval of [-3, +3], except two molecules (19 and 29) of the test set, which are considered as outliers because they have a standardized residual of -4.601 and -3.785, respectively. So, the QSAR model was correctly predicted, where the outliers were removed and the 38 remaining molecules were successfully examined in the applicability domain [6,43,47].

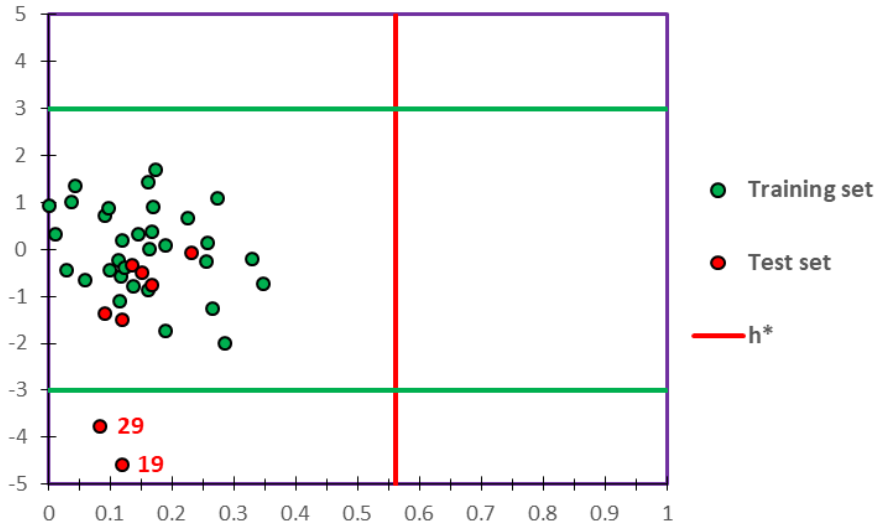



**Figure 34. William's diagram of the predictive QSAR models**

#### 3.5.2. External Validation

To evaluate the reliability of the QSAR predictive model and to ensure its application, it is strictly necessary to validate it on some novel compounds comprised in the test set using the external validation technique [48]. Based on a training test of thirty-two molecules, we have

tested six new compounds from the test set and we have obtained the results of predicted activities using the linear and nonlinear regression techniques, as displayed in Table 23.

**Table 23. Observed and predicted activities of $pIC_{50}$ order using MLR and MNLR QSAR models**

| Molecules number | Observed $pIC_{50}$ | Predicted $pIC_{50}$ (MLR) | Predicted $pIC_{50}$ (MNLR) |
|---|---|---|---|
| 4* | 7.638 | 8.044 | 7.897 |
| 10* | 7.796 | 7.825 | 7.908 |
| 14* | 8.222 | 8.323 | 8.410 |
| 17* | 7.569 | 8.017 | 7.954 |
| 18* | 7.921 | 8.069 | 8.012 |
| 31* | 6.848 | 7.076 | 7.200 |

* Indicates the test set molecules

The first QSAR model, which was developed using the multiple linear regression (MLR) technique, is given by an external validation correlation coefficient equal to $R^2ext = 0.867$, as presented in Figure 35, and the second QSAR model, developed by multiple nonlinear regression (MNLR) technique, is defined by an external validation correlation coefficient of $R^2ext = 0.946$, as illustrated in Figure 36. Thus, MLR and MNLR QSAR models are externally validated because they are characterized by an external validation correlation coefficient greater than 0.6.

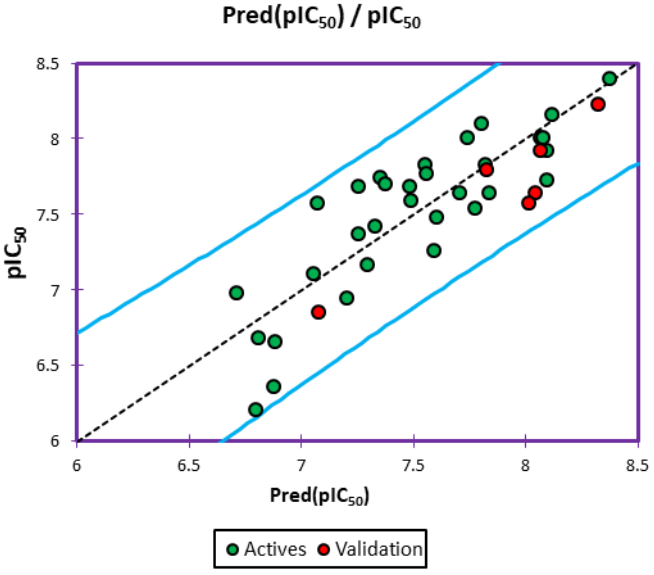



**Figure 35. Correlation among the predicted and observed activities using multiple linear regression (MLR)**

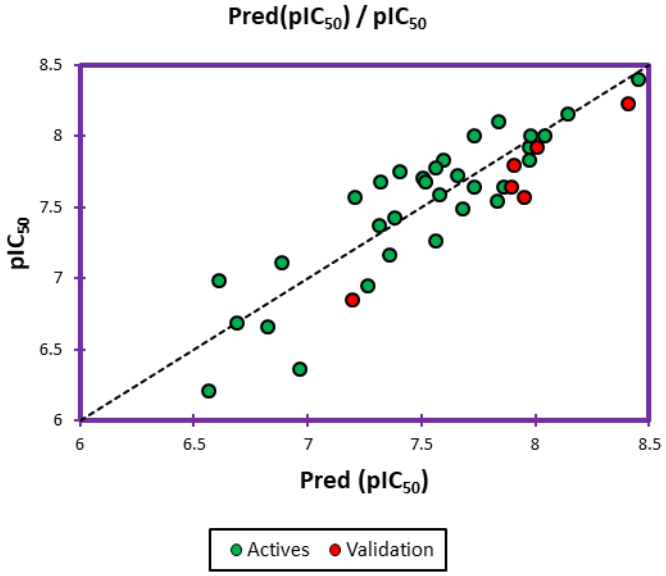



**Figure 36. Correlation among the predicted and observed activities using multiple nonlinear regression (MNLR)**

### 3.5.3. Internal Validation

The leave-one-out procedure (LOO) of the cross-validation technique (CV) was performed to validate internally the developed QSAR models. For this purpose, we tested each molecule once, with the development of a new mathematical model on thirty-one compounds (N-1=31) and prediction of the inhibitory activity of this tested molecule [49], as resulted in Table 24, this process was repeated for all training set molecules [50]. Thereafter, we calculated the quadratic coefficient ($Q^2cv$) of cross-validation using the following equation:

$$Q^2cv = 1 - \frac{\sum_i^n (Yobs - Ypred)^2}{\sum_i^n (Yobs - Ymean)^2} \qquad (4)$$

Where, Yobs: signifies the observed activity, Ypred: is the predicted activity, and Ymean: presents the average of the observed activities. Although the obtained value of quadratic coefficient ($Q^2cv$=0.55) is higher than 0.5, the established QSAR model is robust, reliable, and has a stronger internal predictivity.

**Table 24. Observed and predicted activities using MLR. MNLR and CVLOO techniques**

| Molecules number | Observed $pIC_{50}$ | Predicted$pIC_{50}$ (MLR) | Predicted $pIC_{50}$ (MNLR) | Predicted $pIC_{50}$ (CVLOO) |
|---|---|---|---|---|
| 1 | 7.824 | 7.551 | 7.601 | 7.541 |
| 2 | 8.097 | 7.802 | 7.844 | 7.780 |
| 3 | 7.638 | 7.836 | 7.866 | 7.856 |
| 5 | 7.921 | 8.097 | 7.976 | 8.128 |

| | | | | |
|---|---|---|---|---|
| 6 | 7.770 | 7.559 | 7.563 | 7.528 |
| 7 | 7.638 | 7.709 | 7.732 | 7.721 |
| 8 | 8.000 | 7.743 | 7.734 | 7.704 |
| 9 | 7.420 | 7.329 | 7.387 | 7.325 |
| 11 | 8.000 | 8.067 | 8.042 | 8.105 |
| 12 | 8.155 | 8.119 | 8.148 | 8.105 |
| 13 | 8.000 | 8.081 | 7.983 | 8.114 |
| 15 | 8.398 | 8.376 | 8.458 | 8.370 |
| 16 | 7.538 | 7.775 | 7.833 | 7.824 |
| 20 | 7.108 | 7.056 | 6.887 | 7.047 |
| 21 | 7.161 | 7.298 | 7.361 | 7.307 |
| 22 | 7.745 | 7.351 | 7.406 | 7.318 |
| 23 | 6.678 | 6.813 | 6.691 | 6.833 |
| 24 | 6.979 | 6.712 | 6.616 | 6.645 |
| 25 | 6.947 | 7.205 | 7.266 | 7.267 |
| 26 | 7.678 | 7.258 | 7.322 | 7.156 |
| 27 | 6.356 | 6.878 | 6.970 | 7.027 |
| 28 | 7.699 | 7.376 | 7.511 | 7.233 |
| 30 | 7.260 | 7.592 | 7.562 | 7.650 |
| 32 | 7.481 | 7.602 | 7.684 | 7.625 |
| 33 | 7.678 | 7.485 | 7.522 | 7.418 |
| 34 | 7.721 | 8.098 | 7.660 | 8.259 |
| 35 | 7.569 | 7.073 | 7.208 | 6.944 |
| 36 | 6.201 | 6.797 | 6.569 | 7.075 |
| 37 | 6.658 | 6.882 | 6.825 | 7.019 |
| 38 | 7.367 | 7.257 | 7.316 | 7.230 |
| 39 | 7.585 | 7.493 | 7.580 | 7.473 |
| 40 | 7.824 | 7.823 | 7.973 | 7.823 |

### 3.5.4. Validation Using Y-Randomization Test

Golbraikh and Tropsha's study argues that the internal predictive accuracy of the quantitative structure-activity relationships (QSAR) model, explained by a large value of the quadratic coefficient of leave one out cross-validation (CVLOO), has a tendency to be overestimated and may be the result of a lucky correlation. For this reason, the Y-Randomization test is necessary [6,51]. With the assistance of java Platform SE binary, we have tested the quality of the established QSAR model running 100 randomizations as resulted in Table S6. The results of the Y-Randomization test indicate that the value of the $cR^2p$ criteria is 0.65, which is greater than 0.5. Additionally, the values of correlation, determination, and cross-validation with the leave-one-out procedure coefficients of the original model are so much higher than the resulting values for all one hundred randomizations. As a consequence, the

human GlyT1 activities as predicted by the originally designed model are not due to random chance [6].

### 3.5.5. Statistical Criteria of Golbreikh and Tropsha's Study

The predictive MLR QSAR model given in Equation 1, successfully meets the postulated threshold conditions of Tropsha and Golbraikh's theory, as illustrated in the following Table 25.

**Table 25. Tropsha and Golbraikh conditions to examine the accuracy of generated MLR QSAR model**

| Parameter | Equation | Model score | Threshold | Comment |
|---|---|---|---|---|
| $R^2$ | $R^2 = 1 - \frac{\sum (\ Yobs - Y\ cal\ )^2}{\sum (Yobs\ - \overline{Yobs})^2}$ | 0.73 | > 0.6 | Accepted |
| $R^2adj$ | $R^2adj = \frac{(N-1)R^2 - p}{N - p - 1}$ | 0.67 | > 0.6 | Accepted |
| $R^2test$ | $R^2tes = 1 - \frac{\sum (\ Ycal(test) - Y\ obs(test)\ )^2}{\sum (Yobs(test)\ - \overline{Yobs}(train))^2}$ | 0.87 | > 0.6 | Accepted |
| $Q^2cv$ | $Q^2cv = 1 - \frac{\sum (\ Ycal - Y\ obs\ )^2}{\sum (Yobs\ - \overline{Yobs})^2}$ | 0.55 | > 0.5 | Accepted |
| $R^2$rand | Average of the 100 $R^2$rand (i) | 0.16 | < $R^2$ | Accepted |
| $Q^2cv$ 'LOO' rand | Average of the 100 $Q^2cv$ 'LOO' rand (i) | -0.30 | < $Q^2cv$ | Accepted |
| $cR^2p$ | $cR^2p = R * \sqrt[2]{(R^2 - (\ Average\ R_{rand}\ )^2)}$ | 0.65 | > 0.5 | Accepted |

### 3.6.In Silico Pharmacokinetics ADMET Prediction

In the beginning, we predicted the human GlyT1 inhibitors using BOILED-Egg as an accurate predictive model, very practical in medicinal chemistry to discover the candidate drugs as presented in Figure 37. We have noted that 28 and 30 molecules in blue are part of the yellow Egan-egg, so they are predicted as inhibitors with the greatest ability to cross the blood-brain barrier (BBB) and predicted to be effluated from the central nervous system (CNS) by the P-glycoprotein. Thus, 24 molecules and the nortriptyline as a co-crystallized ligand bound to the membrane protein of dopamine transporter (DAT), as colored in red, are also part of yellow Egan-egg, but they are predicted not to be effluated from the central nervous system (CNS) by the P-glycoprotein. In contrast, the compounds that are part of the white Egan-egg, are predicted as molecules passively absorbed by the gastrointestinal tract [52]. At the second stage, we have predicted the nortriptyline as a co-crystallized ligand compared to 28 and 30 inhibitors which are considered as candidate inhibitors able to penetrate the blood-brain barrier (BBB) and evacuated from the central nervous system (CNS) by the P-glycoprotein in the basis of Lipinski,

Egan, Veber, Ghose, and Muegge rules, as shown in Table 26. Thus, we have noted that these candidate drugs respect all Lipinski rules like as: MW≤500, 40≤ MR ≤130, NRB<10, Log $P_{(octanol/water)}$ <5, HBA≤10, and HBD<5 [43]. Additionally, they satisfy Veber, Egan, Ghose, and Muegge regulations [53]. In the last step, we studied the pharmacokinetic properties of adsorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) of the candidate molecules. which were compared to the obtained results for nortriptyline, as resulted in Table 27. We have noticed that L28 and L30 ligands have an excellent human intestinal absorption (HIA superior than 96%), good distribution because they are defined by the human distribution volumes higher than −0.44 Log L/kg, and characterized by a BBB and CNC permeability values greater than -1 Log BB and included in (-2 to -3) Log PS, respectively. So, they are well permeable to the blood brain barrier (BBB) and the central nervous system (CNS). Moreover, they are all predicted as potent inhibitors of 2C9, 2C19 and 3A4 cytochromes. Then, they present a total clearance of 0.563 and 0.711 Log ml/min/kg respectively, without any toxicity on the human body. Therefore, we have concluded that L28 and L30 ligands are successfully predicted as not toxic inhibitors of 2C9, 2C19, and 3A4 cytochromes, respecting Lipinski, Veber, Egan, Ghose, and Muegge rules, and they are estimated to penetrate the blood-brain barrier (BBB) with the highest probability. So, they are designed as powerful central nervous system (CNS) agents.

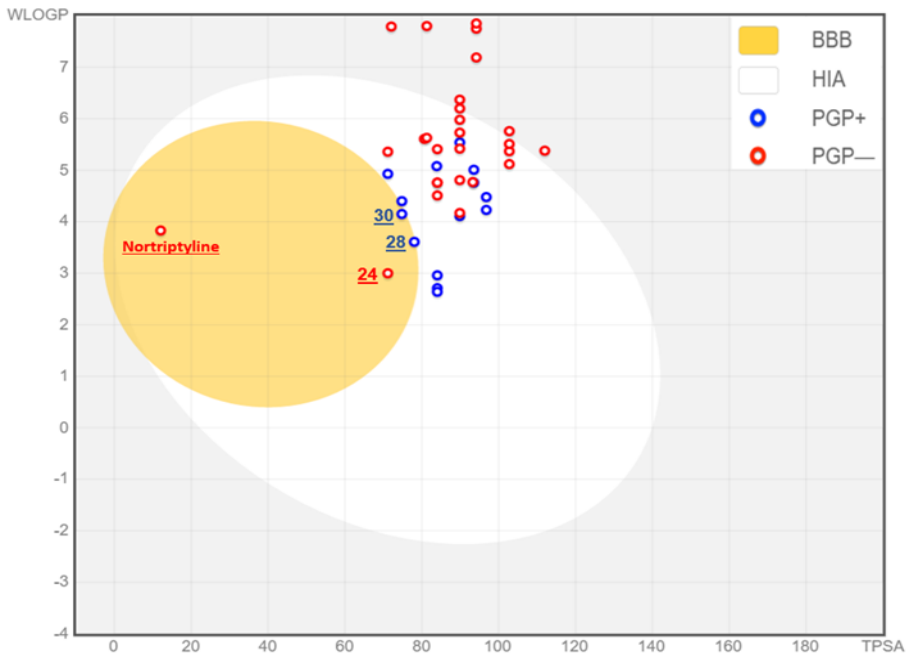



**Figure 37. BOILED-Egg model of nortriptyline and forty pyrrolo[3,4-c]pyrazole and azetidine derivative**

**Table 26. Prediction of physico-chemical parameters of L28, L30, and nortriptyline ligands based on Lipinski, Egan, Veber, Ghose, and Muegge regulations**

| | Physical and chemical properties | | | | | | Lipinski violation | Veber violation | Egan violation | Ghose violation | Muegge violation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ligands number | Molecular weight (g/mol) | Molar refractive index | Rotatable bonds Number | Log P (Octanol / Water) | Hydrogen bond acceptors | H-bond donors Number | | | | | |
| Threshold | MW≤500 | 40≤ MR ≤130 | <10 | <5 | ≤10 | <5 | Yes/No | Yes/No | Yes/No | Yes/No | Yes/No |
| L28 | 448.47 | 124.97 | 4 | 3.35 | 6 | 0 | Yes | Yes | Yes | Yes | Yes |
| L30 | 423.44 | 120.78 | 4 | 3.23 | 5 | 0 | Yes | Yes | Yes | Yes | Yes |
| Nortriptyline | 263.38 | 86.06 | 3 | 3.22 | 1 | 1 | Yes | Yes | Yes | Yes | No |

**Table 27. ADMET-Tox pharmacokinetic properties of L28, L30, and Nortriptyline ligands**

| | Absorption | Distribution | | | Metabolism | | | | | | | Excretion | Toxicity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ligands number | Intestinal Absorption (human) | VDss (human) | BBB permeability | CNS permeability | Substrate CYP | | | | Inhibitor | | | Total Clearance | AMES toxicity |
| | | | | | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | | |
| | Numeric (% Absorbed) | Numeric (Log L/kg) | Numeric (Log BB) | Numeric (Log PS) | Categorical (Yes/No) | | | | | | | Numeric (Log ml/min/kg) | Categorical (Yes/No) |
| L28 | 96.007 | 0.234 | -0.818 | -2.839 | No | Yes | No | Yes | Yes | No | Yes | 0.563 | Not toxic |
| L30 | 98.673 | 0.219 | -0.353 | -1.865 | No | Yes | Yes | Yes | Yes | No | Yes | 0.711 | Not toxic |
| Nor | 97.482 | 1.575 | 0.808 | -1.196 | No | Yes | Yes | No | No | Yes | No | 0.929 | Not toxic |

### 3.7.Molecular Docking

Although L28 and L30 ligands are engineered as potent inhibitors with a favorable ADMET profile, thus we have studied their intermolecular interactions towards the drosophila melanogaster dopamine transporter (DAT) of 4M48.pdb code, as a transmembrane protein able to extract the neurotransmitter dopamine out of the synaptic cleft and transfer it into the cytosol of neighboring cells [6,54–56]. The molecular docking simulations of (L28, L30-protein) complexes were done in the active sites of tricyclic antidepressant nortriptyline (as co-crystallized ligand linked to the targeted protein in his A-chain) which resulted from the ProteinsPlus online server [57], like Tyr124A, Phe325A, and Phe43A amino acids, as presented in Figure 38.

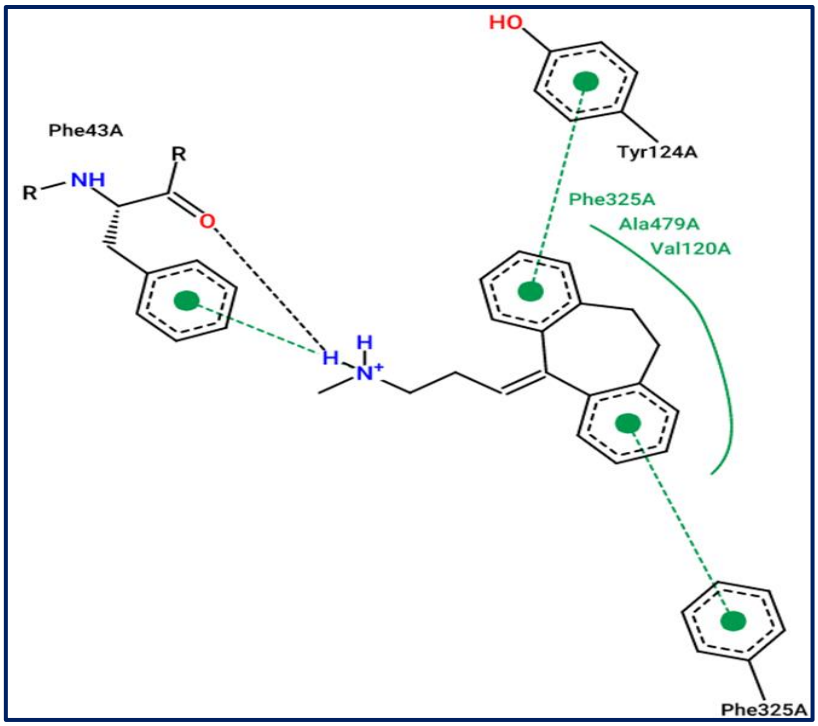
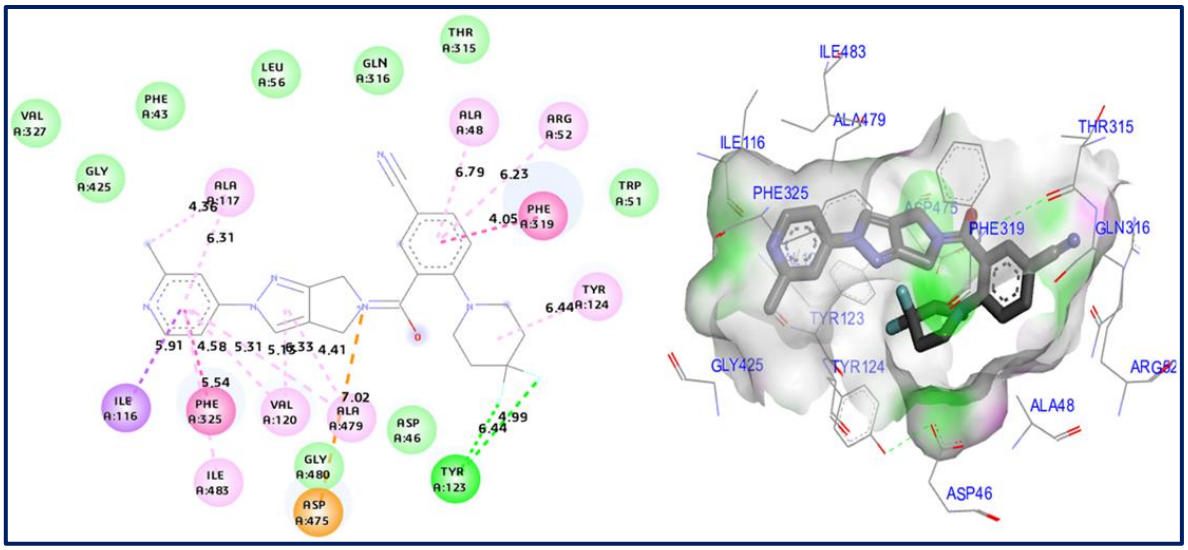
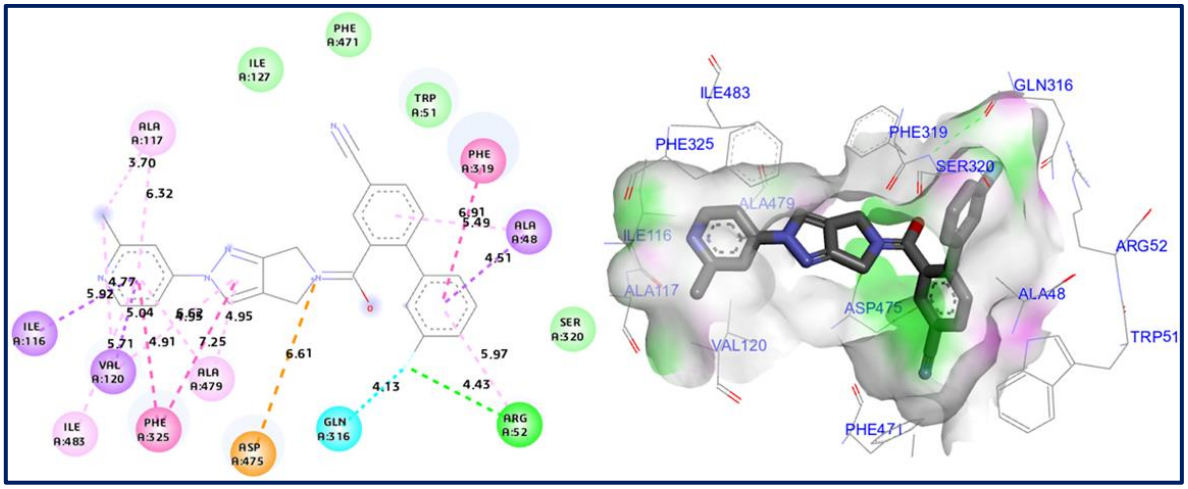



**Figure 38. Active sites of co-crystalized ligand bound to 4M48.pdb protein in A-chain**

The molecular docking results presented in Figure 39, indicate that L28 and L30 ligands have several common intermolecular interactions such as: Two stacked Pi-Pi chemical bonds, established between Phe325 and Phe319 amino acids of the targeted protein and the candidate drugs. A common Pi-Sigma chemical bond was formed between the candidate drugs and Ile116 amino acid with a nuclear distance of 5.91 Å. Three Pi-Alkyl chemical bonds were produced with Ala479, Ile483, and Ala117 amino acids of the responsible protein. A common bond of attractive charge with the nitrogen atom was created with Asp475 amino acid. Additionally, we note that the L28 ligand formed at 6.44 Å a bond of Alkyl type towards Tyr124 amino acid. and created for each one of its two fluorine atoms, a hydrogen bond with Tyr123 amino acid, rendering the L28-protein complex more stable. Also, the L30-protein complex was stabilized via the formation of a hydrogen bond between the candidate drug and the Arg52 amino acid of the targeted protein. Therefore, we conclude that Phe325, Phe319, Ile116, Ala479, Ile483, Ala117, Asp475, Tyr124, Tyr123, and Arg52 amino acids are the responsible sites of human glycine transporter type one (GLUT1) inhibitory activities to improve memory performance.




**(L28 ligand-DAT protein) complex**




**(L30 ligand-DAT protein) complex**

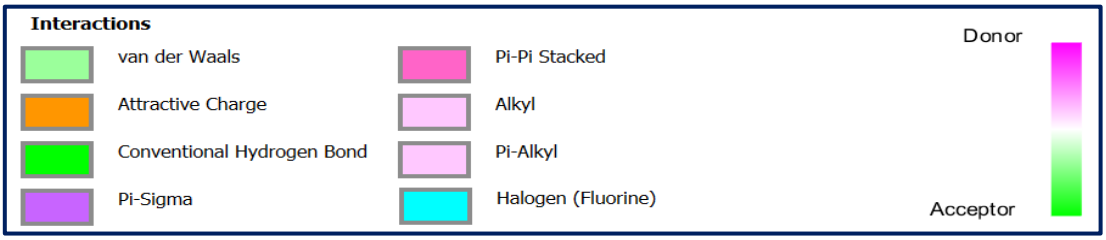



**Figure 39. 2D (left) and 3D (right) inter-molecular interactions, produced between L28 and L30 ligands and DAT protein of 4M48.pdb code with binding energies of −8.72 kcal/mol and −8.70 kcal/mol, respectively.**

### 3.8.Docking Validation Protocol

The efficiency and accuracy of the molecular docking algorithms were examined through the re-docking method, based on the superposition of the co-crystallized ligand on the docked nortriptyline ligand. The overlay results shown in Figure 40C, reflects a root mean square deviation (RMSD) of 0.552 Å, which is inferior than 2 Å, so the prediction pose is precise and

accurate. In addition, 2D and 3D intermolecular interactions, successively presented in Figure 40A and Figure 40B, reveal that docked nortriptyline ligand of cyan color, reacts to Tyr124 and Phe43 amino acids, which are the same sites as those experimentally produced. As a result, the docking validation protocol was successfully examined [6,37,58].

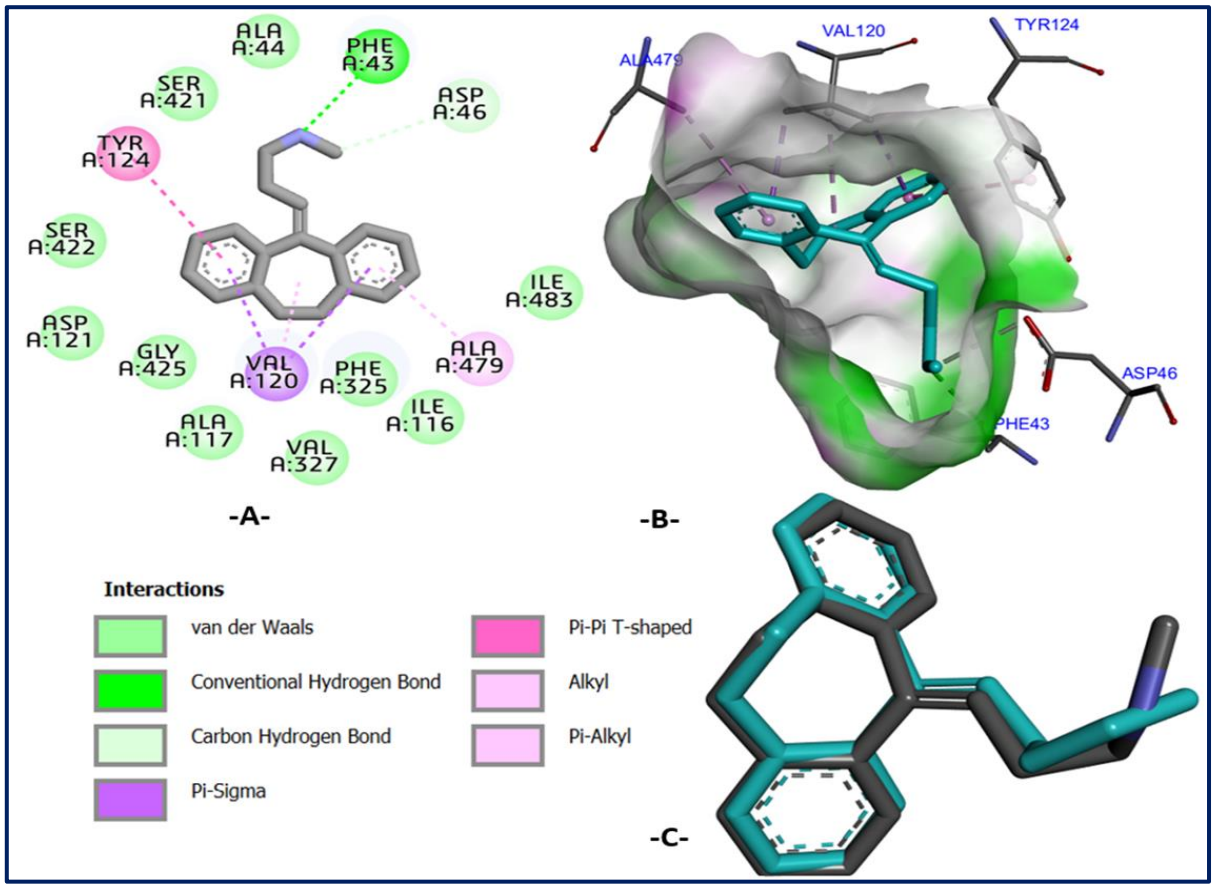



**Figure 40. 2D (A) and 3D (B) visualization of intermolecular interactions created between the docked nortriptyline and 4M48.pdb protein with a binding energy of −8.36 kcal/mol, and re-docking pose (c) defined by a RMSD value of 0.552 Å (docked nortriptyline in cyan and co-crystallized nortriptyline colored by the atomic elements)**

### 3.9.Molecular Dynamic simulation

To examine the stability of intermolecular interactions, produced between L28 and L30 candidate drugs and the targeted protein encoded as 4M48.pdb, we have performed the molecular dynamics simulations of these two complexes during 50 nanoseconds [6,14].

The results of dynamic conformational changes of (the L28-DAT protein) complex shown in Figure 41A, illustrate that protein root mean square deviation (RMSD) in the left Y-axis. fluctuates around the mean thermal structure during 50 ns of simulation time, because deviation changes in the range of 1-3 Å are acceptable for small globular proteins. Thus, the simulation was well-balanced. In addition, the evolution of the ligand RMSD in the right Y-axis through

the heavy atoms, reveals the ligand stability with respect to the protein targets, when the studied complex was first aligned with the reference protein backbone, as the observed values of ligand RMSDs are nearly the same as the protein RMSDs values. Thus, the ligand had not diffused away of its primary binding site. Moreover, the evolution of the root mean square fluctuation (RMSF) was also controlled to investigate the influence of ligand binding on the internal dynamics of DAT protein over 50 ns, in which C- and N-terminal tails tend to fluctuate more than any other protein part. Similarly, alpha helices and beta strands as secondary structure elements are generally more rigorously rigid than the non-structured protein part, so they fluctuate considerably less than the buckle regions. Just one oscillation of 3.7 Å, has been detected in the buckle region of the residue 390. In contrast, all other fluctuations were inferior to 3 Å, as noted in Figure 41B, which indicates the absence of a conformational change of the protein caused by its binding towards the ligand, and the stability of the produced complex. The values of gyration radiuses (r-Gyr) have oscillated in a relatively short range of 4.40 to 5.35 Å along the simulation time, as noticed in Figure 41C, which reveals a few conformational changes in the compactness of the L28 ligand, so the DAT protein has an acceptable flexibility after being bound to the candidate drug. Also, the evolution of solvent accessible surface area (SASA), was done to evaluate the solvent accessibility of (L28-DAT protein) complex over 50 ns of simulation time, as we have recorded a minimal oscillation, included between 20 and 90 $Å^2$, which demonstrates that the drug structure was approximately unchanged during the simulation process. Moreover, the molecular surface area (MolSA) associated with a Van Der Waals surface was calculated with a probe radius of 1.4 Å, has been fluctuated between 400 and 430 $Å^2$. Furthermore, we have noticed that the polar surface area (PSA) which is mainly influenced by nitrogen and oxygen atoms, changes between 75 and 145 Å2, with some very small maximum and minimum fluctuations. Eventually, the graph of total energy, obtained using the MM-GBSA approach, displays some minimal oscillations around the mean of -58.360388 kcal.mol-1, which are progressively stabilized from 25 ns until the end of the simulation as displayed in Figure 41D.

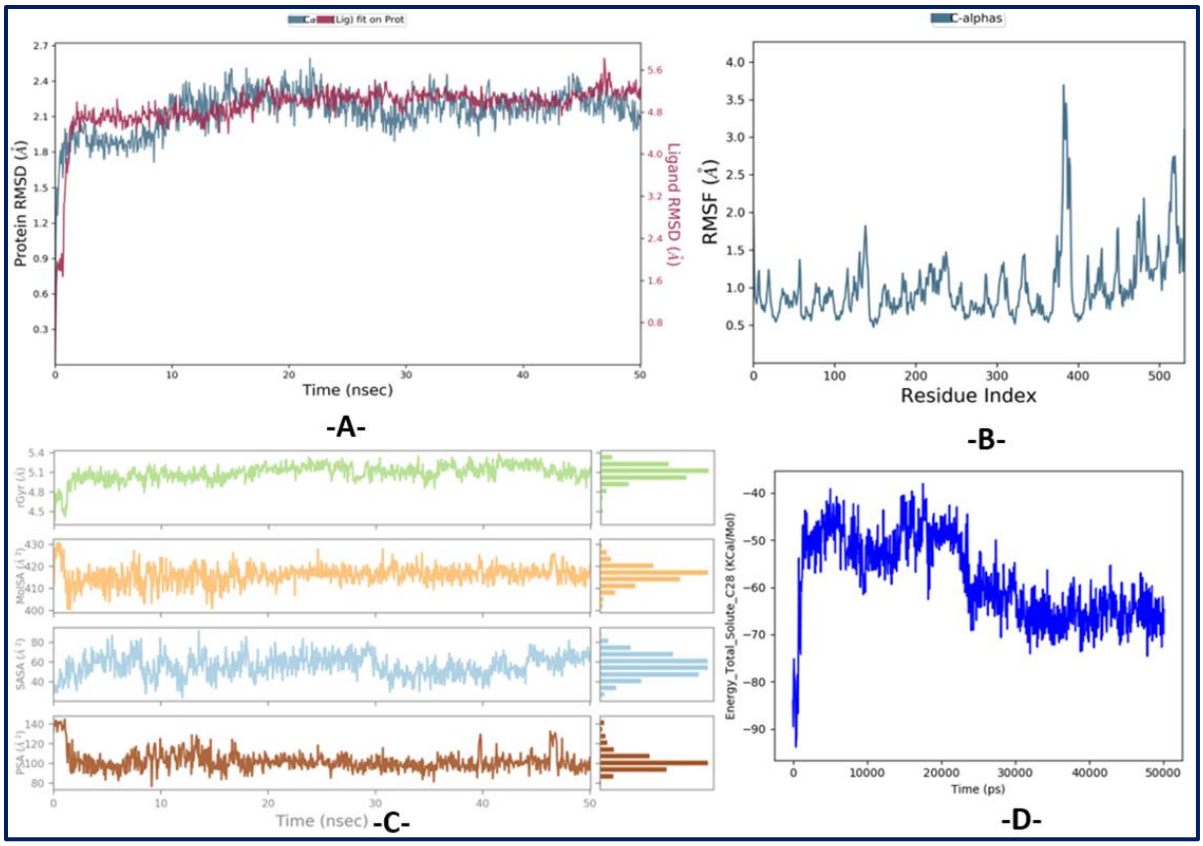



**Figure 41. Evolution of RMSD. RMSF. Rg. MolSA. SASA. PSA and total free energy during 50 ns. for (L28-DAT protein) complex**

Regarding the second complex, we noticed that the DAT protein was also characterized by a root mean square deviation (RMSD) included between 1 and 3 Å, where the observed values of the ligand RMSD oscillate in parallel with those of the protein RMSD, as presented in the Figure 42A. So, the simulation was well equilibrated for the (L30-DAT protein) complex, and the candidate drug had not diffused away from its primary link site. Additionally, the graph of root mean square fluctuations (RMSFs), indicates only two fluctuations greater than 3 Å. The first one was recorded at 4.6 Å for residue 230, and the second one was detected at 4.7 Å for residue 480, as shown in Figure 42B, indicating that the DAT protein binds to the drug with extremely small conformational changes. Subsequently, we have monitored the evolution of gyration radiuses (r-Gyr), solvent accessible surface area (SASA), molecular surface area (MolSA), and polar surface area (PSA) properties, which are generally subject to very small fluctuations as displayed in the Figure 42C. So, the drug structure remained approximately unchanged during 50 ns of simulation time, and the DAT protein has an agreeable flexibility during its interaction with the L30 candidate drug. Finally, the total energy graph of the (L30-DAT protein) complex, given with the assistance of the MM-GBSA approach, displays slight

fluctuations about the mean of -61.193294 kcal.mol-1, as shown in Figure 42D, which means that the complex remains very stable over the simulation time.

Consequently, we conclude that intermolecular interactions produced in the (DAT protein-L28, and L30 ligands) complexes are dynamically more stable because their studied properties remained in equilibrium throughout the MD simulation.

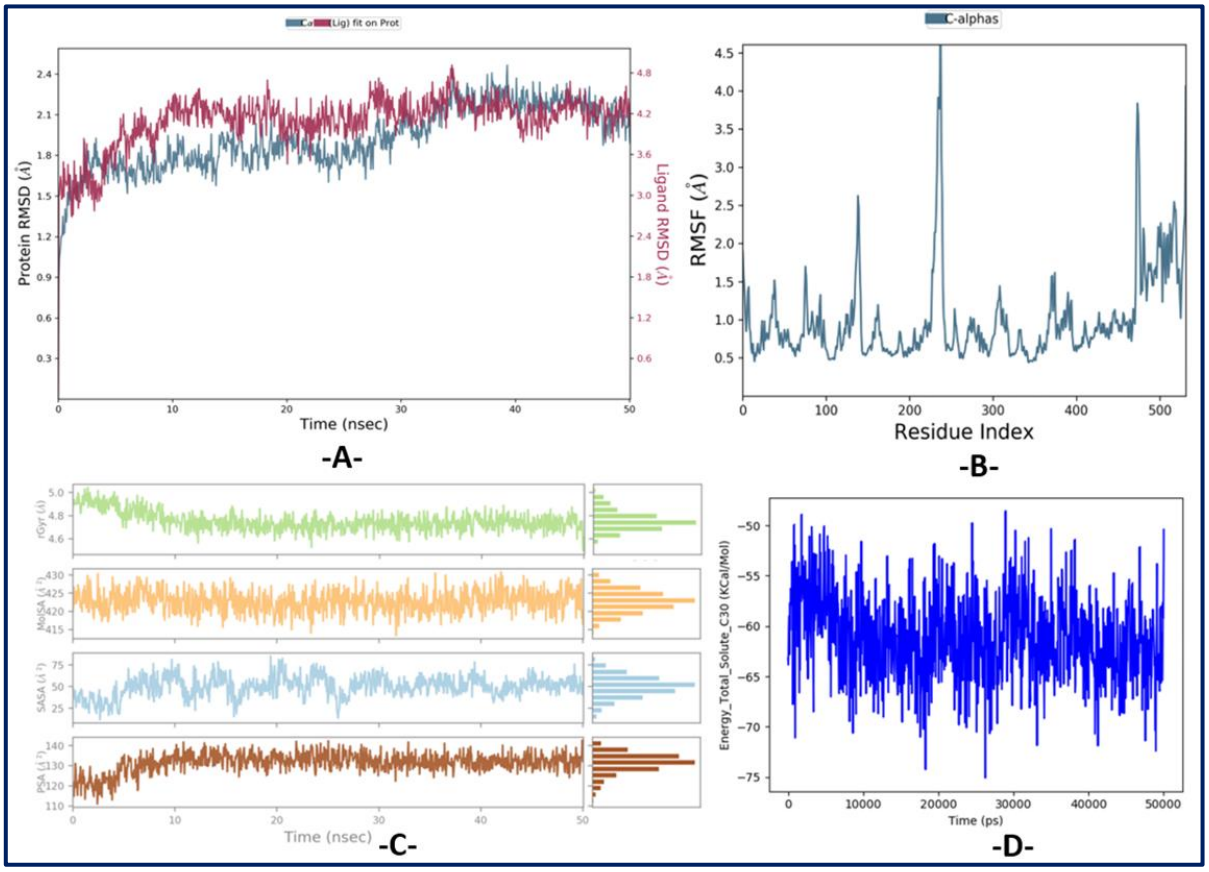



**Figure 42. Evolution of RMSD, RMSF, Rg, MolSA, SASA, PSA, and total free energy during 50 ns, for (L30-DAT protein) complex**

## 4. Conclusion

An in silico investigation was performed for forty pyrrolo[3,4-c]pyrazole and azetidine-based derivatives, to discover an effective drug for enhancing human memory performance. In the beginning, two predictive QSAR models were developed using MNLR and MLR techniques and were successfully validated by internal and external validation, Y-randomization test, applicability domain, Tropsha, and Golbreikh criteria, indicating that MW, Log P, NRB, PSA, and TD descriptors have a key function in the human glycine transporter type one (GlyT1) activity. Afterward, a pharmacoinformatic-based approach revealed that L28 and L30 ligands were predicted as non-toxic inhibitors of 2C9, 2C19, and 3A4 cytochromes, with a desirable ADMET profile, meeting the rules of Lipinski, Egan Veber, Muegge, and Ghose. So, they are

designed as candidate drugs, most likely to penetrate the central nervous system (CNS). For this reason, they were specially selected for molecular docking simulation, which revealed that they react with Phe319, Phe325, Tyr123, Tyr 124, Arg52, Asp475, Ala117, Ala479, Ile116, and Ile483 amino acids of a membrane protein from the dopamine transporter (DAT) of 4M48.pdb code. Finally, the stability of intermolecular interactions produced between the candidate drugs and the targeted protein was successfully examined through the molecular dynamics technique along 50 ns of simulation time. Therefore, L28 and L30 drugs could be safely and securely applicable as therapeutics in medicine to enhance human memory performance.

**References**

[1] N. Nazir. M. Nisar. M. Zahoor. F. Uddin. S. Ullah. R. Ullah. S.A. Ansari. H.M. Mahmood. A. Bari. A. Alobaid. Phytochemical Analysis. In Vitro Anticholinesterase. Antioxidant Activity and In Vivo Nootropic Effect of Ferula ammoniacum (Dorema ammoniacum) D. Don. in Scopolamine-Induced Memory Impairment in Mice. Brain Sciences. 11 (2021) 259. https://doi.org/10.3390/brainsci11020259.

[2] B. Couto. G. Rojas. C. Gelormini-Lezama. S. O'Neill. Memory. Executive Function and Social Cognition in Neurological Disorders. in: Encyclopedia of Behavioral Neuroscience. 2nd Edition. Elsevier. 2022: pp. 140–147. https://doi.org/10.1016/B978-0-12-819641-0.00032-3.

[3] V.J. Santora. T.A. Almos. R. Barido. J. Basinger. C.L. Bellows. B.C. Bookser. J.G. Breitenbucher. N.J. Broadbent. C. Cabebe. C.-K. Chai. M. Chen. S. Chow. D.M. Chung. L. Crickard. A.M. Danks. G.C. Freestone. D. Gitnick. V. Gupta. C. Hoffmaster. A.R. Hudson. A.P. Kaplan. M.R. Kennedy. D. Lee. J. Limberis. K. Ly. C.C. Mak. B. Masatsugu. A.C. Morse. J. Na. D. Neul. J. Nikpur. M. Peters. R.E. Petroski. J. Renick. K. Sebring. S. Sevidal. A. Tabatabaei. J. Wen. Y. Yan. Z.W. Yoder. D. Zook. Design and Synthesis of Novel and Selective Glycine Transporter-1 (GlyT1) Inhibitors with Memory Enhancing Properties. J. Med. Chem. 61 (2018) 6018–6033. https://doi.org/10.1021/acs.jmedchem.8b00372.

[4] V. Eulenburg. W. Armsen. H. Betz. J. Gomeza. Glycine transporters: essential regulators of neurotransmission. Trends in Biochemical Sciences. 30 (2005) 325–333. https://doi.org/10.1016/j.tibs.2005.04.004.

[5] A.R. Hudson. V.J. Santora. R.E. Petroski. T.A. Almos. G. Anderson. R. Barido. J. Basinger. C.L. Bellows. B.C. Bookser. N.J. Broadbent. C. Cabebe. C.-K. Chai. M. Chen. S. Chow. D.M. Chung. L. Heger. A.M. Danks. G.C. Freestone. D. Gitnick. V. Gupta. C. Hoffmaster. A.P. Kaplan. M.R. Kennedy. D. Lee. J. Limberis. K. Ly. C.C. Mak. B. Masatsugu. A.C. Morse. J. Na. D. Neul. J. Nikpur. J. Renick. K. Sebring. S. Sevidal. A. Tabatabaei. J. Wen. S. Xia. Y. Yan. Z.W. Yoder. D. Zook. M. Peters. J.G. Breitenbucher. Azetidine-based selective glycine transporter-1 (GlyT1) inhibitors with memory enhancing properties. Bioorganic & Medicinal Chemistry Letters. 30 (2020) 127214. https://doi.org/10.1016/j.bmcl.2020.127214.

[6] M. El fadili. M. Er-Rajy. M. Kara. A. Assouguem. A. Belhassan. A. Alotaibi. N.N. Mrabti. H. Fidan. R. Ullah. S. Ercisli. S. Zarougui. M. Elhallaoui. QSAR. ADMET In Silico Pharmacokinetics. Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia. Pharmaceuticals. 15 (2022) 670. https://doi.org/10.3390/ph15060670.

[7] M. Radan. T. Djikic. D. Obradovic. K. Nikolic. Application of in vitro PAMPA technique and in silico computational methods for blood-brain barrier permeability prediction of novel CNS drug candidates. European Journal of Pharmaceutical Sciences. 168 (2022) 106056. https://doi.org/10.1016/j.ejps.2021.106056.

[8] S. Vilar. G. Cozza. S. Moro. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. Current Topics in Medicinal Chemistry. 8 (2008) 1555–1572. https://doi.org/10.2174/156802608786786624.

[9] O. Ebenezer. N. Damoyi. M.A. Jordaan. M. Shapi. Unveiling of Pyrimidindinones as Potential Anti-Norovirus Agents—A Pharmacoinformatic-Based Approach. Molecules. 27 (2022) 380. https://doi.org/10.3390/molecules27020380.

[10] S. Tian. J. Wang. Y. Li. D. Li. L. Xu. T. Hou. The application of in silico drug-likeness predictions in pharmaceutical research. Advanced Drug Delivery Reviews. 86 (2015) 2–10. https://doi.org/10.1016/j.addr.2015.01.009.
[11] A. Serrano. B. Imbernón. H. Pérez-Sánchez. J.M. Cecilia. A. Bueno-Crespo. J.L. Abellán. QN-Docking: An innovative molecular docking methodology based on Q-Networks. Applied Soft Computing. 96 (2020) 106678. https://doi.org/10.1016/j.asoc.2020.106678.
[12] R.P.D. Bank. RCSB PDB - 4M48: X-ray structure of dopamine transporter elucidates antidepressant mechanism. (n.d.). https://www.rcsb.org/structure/4M48 (accessed July 4. 2022).
[13] E. Wang. H. Sun. J. Wang. Z. Wang. H. Liu. J.Z.H. Zhang. T. Hou. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem. Rev. 119 (2019) 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055.
[14] T.A. de Oliveira. L.R. Medaglia. E.H.B. Maia. L.C. Assis. P.B. de Carvalho. A.M. da Silva. A.G. Taranto. Evaluation of Docking Machine Learning and Molecular Dynamics Methodologies for DNA-Ligand Systems. Pharmaceuticals. 15 (2022) 132. https://doi.org/10.3390/ph15020132.
[15] Gaussian 09 Citation | Gaussian.com. (n.d.). https://gaussian.com/g09citation/ (accessed July 6. 2022).
[16] R.G. Parr. Density Functional Theory of Atoms and Molecules. in: K. Fukui. B. Pullman (Eds.). Horizons of Quantum Chemistry. Springer Netherlands. Dordrecht. 1980: pp. 5–15. https://doi.org/10.1007/978-94-009-9027-2_2.
[17] H. Wang. Y. Jia. G. Jing. X. Wu. A novel toxicity prediction model for hydrazine compounds based on 1D–3D molecular descriptors. Computational Toxicology. 18 (2021) 100169. https://doi.org/10.1016/j.comtox.2021.100169.
[18] G.W.A. Milne. Software Review of ChemBioDraw 12.0. J. Chem. Inf. Model. 50 (2010) 2053–2053. https://doi.org/10.1021/ci100385n.
[19] T. Österberg. U. Norinder. Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors. European Journal of Pharmaceutical Sciences. 12 (2001) 327–337. https://doi.org/10.1016/S0928-0987(00)00189-5.
[20] C. Guerra Tort. V. Aguiar Pulido. V. Suárez Ulloa. F. Docampo Boedo. J.M. López Gestal. J. Pereira Loureiro. Electronic Health Records Exploitation Using Artificial Intelligence Techniques. in: 3rd XoveTIC Conference. MDPI. 2020: p. 60. https://doi.org/10.3390/proceedings2020054060.
[21] Introducing the XLSTAT Free edition. XLSTAT. Your data analysis solution. (n.d.). https://www.xlstat.com/en/news/introducing-the-xlstat-free-edition (accessed March 8. 2022).
[22] N. Mostoufi. A. Constantinides. Chapter 8 - Linear and nonlinear regression analysis. in: N. Mostoufi. A. Constantinides (Eds.). Applied Numerical Methods for Chemical Engineers. Academic Press. 2023: pp. 403–476. https://doi.org/10.1016/B978-0-12-822961-3.00008-X.
[23] A.K. Halder. T. Jha. Validated predictive QSAR modeling of N-aryl-oxazolidinone-5-carboxamides for anti-HIV protease activity. Bioorganic & Medicinal Chemistry Letters. 20 (2010) 6082–6087. https://doi.org/10.1016/j.bmcl.2010.08.050.

[24] K. Roy. S. Kar. R.N. Das. Validation of QSAR Models. in: Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. Elsevier. 2015: pp. 231–289. https://doi.org/10.1016/B978-0-12-801505-6.00007-7.
[25] C.A. Lipinski. F. Lombardo. B.W. Dominy. P.J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[26] D.F. Veber. S.R. Johnson. H.-Y. Cheng. B.R. Smith. K.W. Ward. K.D. Kopple. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[27] W.J. Egan. Merz Kenneth M.. J.J. Baldwin. Prediction of Drug Absorption Using Multivariate Statistics. J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[28] W.J. Egan. G. Lauri. Prediction of intestinal permeability. Advanced Drug Delivery Reviews. 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.
[29] S. Yalcin. Molecular Docking. Drug Likeness. and ADMET Analyses of Passiflora Compounds as P-Glycoprotein (P-gp) Inhibitor for the Treatment of Cancer. Current Pharmacology Reports. 6 (2020) 429–440. https://doi.org/10.1007/s40495-020-00241-6.
[30] I. Muegge. S.L. Heald. D. Brittelli. Simple Selection Criteria for Drug-like Chemical Matter. J. Med. Chem. 44 (2001) 1841–1846. https://doi.org/10.1021/jm015507e.
[31] pkCSM. (n.d.). http://biosig.unimelb.edu.au/pkcsm/run_example? (accessed July 6. 2022).
[32] A. Daina. O. Michielin. V. Zoete. SwissADME: a free web tool to evaluate pharmacokinetics. drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.
[33] D. Bassani. M. Pavan. G. Bolcato. M. Sturlese. S. Moro. Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro. Pharmaceuticals. 15 (2022) 180. https://doi.org/10.3390/ph15020180.
[34] R.P.D. Bank. RCSB PDB - 4M48: X-ray structure of dopamine transporter elucidates antidepressant mechanism. (n.d.). https://www.rcsb.org/structure/4M48 (accessed August 23. 2022).
[35] A. Kouranov. The RCSB PDB information portal for structural genomics. Nucleic Acids Research. 34 (2006) D302–D305. https://doi.org/10.1093/nar/gkj120.
[36] BIOVIA Discovery Studio - BIOVIA - Dassault Systèmes®. (n.d.). https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/ (accessed August 23. 2022).
[37] M. El fadili. M. Er-rajy. H. Imtara. M. Kara. S. Zarougui. N. Altwaijry. O. Al kamaly. A. Al Sfouk. M. Elhallaoui. 3D-QSAR. ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists. Processes. 10 (2022) 1462. https://doi.org/10.3390/pr10081462.

[38] A.P. Norgan. P.K. Coffman. J.-P.A. Kocher. D.J. Katzmann. C.P. Sosa. Multilevel Parallelization of AutoDock 4.2. J Cheminform. 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[39] Desmond | Schrodinger. (n.d.). https://www.schrodinger.com/products/desmond (accessed April 28. 2022).
[40] G.A. Kaminski. R.A. Friesner. J. Tirado-Rives. W.L. Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B. 105 (2001) 6474–6487. https://doi.org/10.1021/jp003919d.
[41] E. Wang. H. Sun. J. Wang. Z. Wang. H. Liu. J.Z.H. Zhang. T. Hou. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem. Rev. 119 (2019) 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055.
[42] A. Bastianoni. E. Guastaldi. A. Barbagli. S. Bernardinetti. A. Zirulia. M. Brancale. T. Colonna. Multivariate Analysis Applied to Aquifer Hydrogeochemical Evaluation: A Case Study in the Coastal Significant Subterranean Water Body between “Cecina River and San Vincenzo”. Tuscany (Italy). Applied Sciences. 11 (2021) 7595. https://doi.org/10.3390/app11167595.
[43] M. Er-rajy. M. El Fadili. H. Hadni. N.N. Mrabti. S. Zarougui. M. Elhallaoui. 2D-QSAR modeling. drug-likeness studies. ADMET prediction. and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives. Struct Chem. 33 (2022) 973–986. https://doi.org/10.1007/s11224-022-01913-3.
[44] G.K. Uyanık. N. Güler. A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences. 106 (2013) 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027.
[45] N. Kravić. J. Savosina. M. Agafonova-Moroz. V. Babain. A. Legin. D. Kirsanov. Nonlinear Multivariate Regression Algorithms for Improving Precision of Multisensor Potentiometry in Analysis of Spent Nuclear Fuel Reprocessing Solutions. Chemosensors. 10 (2022) 90. https://doi.org/10.3390/chemosensors10030090.
[46] E. Pourbasheer. S. Riahi. M.R. Ganjali. P. Norouzi. Quantitative structure–activity relationship (QSAR) study of interleukin-1 receptor associated kinase 4 (IRAK-4) inhibitor activity by the genetic algorithm and multiple linear regression (GA-MLR) method. Journal of Enzyme Inhibition and Medicinal Chemistry. 25 (2010) 844–853. https://doi.org/10.3109/14756361003757893.
[47] M. Er-rajy. M. El fadili. N.N. Mrabti. S. Zarougui. M. Elhallaoui. QSAR. Molecular docking. ADMET properties in silico studies for a series of 7-propanamide benzoxaboroles as potent anti-cancer Agents. Chinese Journal of Analytical Chemistry. (2022) 100163. https://doi.org/10.1016/j.cjac.2022.100163.
[48] K.B. de C. Rezende. A.J.L.A. da Cunha. J. Amim Junior. R.G. Bornia. External validation of the Fetal Medicine Foundation algorithm for the prediction of preeclampsia in a Brazilian population. Pregnancy Hypertension. 17 (2019) 64–68. https://doi.org/10.1016/j.preghy.2019.05.006.
[49] M. Rafało. Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis. ICT Express. (2021) S2405959521000552. https://doi.org/10.1016/j.icte.2021.05.001.

[50] R. Kasmi. E. Hadaji. O. Chedadi. A. El Aissouq. M. Bouachrine. A. Ouammou. 2D-QSAR and docking study of a series of coumarin derivatives as inhibitors of CDK (anticancer activity) with an application of the molecular docking method. Heliyon. 6 (2020) e04514. https://doi.org/10.1016/j.heliyon.2020.e04514.
[51] A. Golbraikh. A. Tropsha. Beware of q2!. Journal of Molecular Graphics and Modelling. 20 (2002) 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1.
[52] A. Daina. V. Zoete. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. ChemMedChem. 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.
[53] C.-Y. Jia. J.-Y. Li. G.-F. Hao. G.-F. Yang. A drug-likeness toolbox facilitates ADMET study in drug discovery. Drug Discovery Today. 25 (2020) 248–258. https://doi.org/10.1016/j.drudis.2019.10.014.
[54] A. Adamski. D. Kruszka. Z. Dutkiewicz. M. Kubicki. A. Gorczyński. V. Patroniak. Novel family of fused tricyclic [1.4]diazepines: Design. synthesis. crystal structures and molecular docking studies. Tetrahedron. 73 (2017) 3377–3386. https://doi.org/10.1016/j.tet.2017.05.015.
[55] A. Penmatsa. K.H. Wang. E. Gouaux. X-ray structure of dopamine transporter elucidates antidepressant mechanism. Nature. 503 (2013) 85–90. https://doi.org/10.1038/nature12533.
[56] A. Belhassan. H. Zaki. M. Benlyas. T. Lakhlifi. M. Bouachrine. Study of novel triazolo-benzodiazepine analogues as antidepressants targeting by molecular docking and ADMET properties prediction. Heliyon. 5 (2019) e02446. https://doi.org/10.1016/j.heliyon.2019.e02446.
[57] Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server. (n.d.). https://proteins.plus/ (accessed August 29. 2022).
[58] M. Abdullahi. G.A. Shallangwa. A. Uzairu. In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of H5N1 influenza A virus subtype. Beni-Suef Univ J Basic Appl Sci. 9 (2020) 2. https://doi.org/10.1186/s43088-019-0023-y.

# Chapitre VII. Étude in-silico d'inhibiteurs sélectifs et compétitifs du transporteur de glycine type 1 destinés au traitement des perturbations associés à l'hypofonctionnement du récepteur NMDA

## Résumé

En dernière application, nous avons mené une étude systématique in-silico basée sur des techniques de modélisation moléculaire, à partir de laquelle nous avons examiné une série de trente dérivés de pyrrolidine sulfonamide 3,4-disubstitués afin d'identifier le médicament candidat pour traiter la schizophrénie et les déficiences associées à l'hypofonctionnement des récepteurs NMDA, via l'inhibition sélective et compétitive du transporteur de glycine de type 1 (GlyT1).

L'analyse QSAR a montré que les descripteurs géométriques et constitutionnels jouent un rôle clé dans l'activité humaine du transporteur de glycine type 1. L'étude in silico a conclu que le ligand le plus actif, marqué C19, devait être un inhibiteur non toxique, avec un profil ADME-toxicité souhaitable et une probabilité significative de pénétrer dans le système nerveux central (SNC).

Les simulations du Docking moléculaire ont confirmé que le composé C19 était docké aux sites actifs de la protéine du transporteur de dopamine de la drosophile melanogaster (DAT), créant une variété de liaisons chimiques vers les résidus d'acides aminés suivants : TYR 124, ASP 475, GLU 480, ALA 479, et VAL 120.

La technique de dynamique moléculaire (MD) combinée à l'approche MM-GBSA a confirmé que les interactions intermoléculaires produites pour le complexe (protéine DAT-ligand C19) restent stables pendant 100 nanosecondes du temps de simulation de la dynamique moléculaire. Par conséquent, le ligand plus actif C19 est fortement recommandé pour le traitement de la schizophrénie et d'autres handicaps liés à l'hypofonctionnement des récepteurs glutaminergiques NMDA.


**Mots clés :** NMDA; GlyT1; ADME-Toxicité; SNC; DAT; dynamique moléculaire.

# Chapitre VII: In-silico screening based on molecular simulations of 3,4-disubstituted pyrrolidine sulfonamides as selective and competitive GlyT1 inhibitors

Mohamed El fadili [a,*], Mohammed Er-rajy [a] , Wafa Ali Eltayb [b] , Mohammed Kara [c,*], Amine Assouguem [d] , Asmaa Saleh [e], Omkulthom Al Kamaly [e], Sara Zarougui [a] , Menana Elhallaoui [a]

[a] LIMAS Laboratory, Faculty of Sciences Dhar El Mehraz, Sidi Mohammed Ben Abdellah University, BP 1796 Atlas, Fez 30000, Morocco;
[b] Biotechnology Department, Faculty of Sciences and Technology, Shendi University, Shendi 11111, Sudan;
[c] Laboratory of Biotechnology, Conservation and Valorization of Naturals Resources, Faculty of Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, BP 1796 Atlas, Fez, Morocco;
[d] Laboratory of Functional Ecology and Environment, Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Imouzzer Street, Fez P.O. Box 2202, Morocco;
[e] Department of Pharmaceutical Sciences, College of Pharmacy, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia;
* Corresponding author: mohamed.elfadili@usmba.ac.ma

**Abstract:** A systematic in-silico study based on molecular modeling techniques was conducted on thirty 3,4-disubstituted pyrrolidine sulfonamides derivatives to identify the drug candidate for treating schizophrenia and impairments associated with NMDA receptor hypofunction, through selective and competitive inhibition of GlyT1. QSAR analysis demonstrates that geometric and constitutional descriptors have a key function in human GlyT1 activity. The in-silico study concluded that the most active ligand labeled C19 was predicted to be a non-toxic inhibitor, with a desired ADME-Toxicity profile and a significant probability to penetrate the central nervous system (CNS). Molecular docking simulations confirmed that the C19 compound was docked to the active sites of drosophila melanogaster dopamine transporter (DAT) protein, creating a variety of chemical bonds towards TYR 124, ASP 475, GLU 480, ALA 479, and VAL 120 amino acids residues. The molecular dynamic (MD) technique combined with the MMGBSA approach confirmed that produced intermolecular interactions for the (DAT protein–C19 ligand) complex remain so stable during 100 nanoseconds of MD simulation time. Consequently, the C19 ligand is highly recommended for the treatment of schizophrenia and other disabilities linked to the hypofunction of glutaminergic NMDA receptors.

**Keywords:** NMDA; GlyT1; ADME-Toxicity; CNS; MD.

## 1. Introduction

Over the last decade, the glycine transporter type 1 (GlyT1) localized at glutamatergic synapses has gained great attention as one of the major neurotransmitters that control memory and learning functions, and a key element of glycine metabolism in the central nervous system (CNS) [1–4]. Many pharmaceutical companies have focused on developing selective and competitive GluT1 inhibitors, to provide potential treatments for neurodegenerative disorders associated with glutamine N-methyl-D-aspartate (NMDA) receptor hypofunction, including schizophrenia [5–8]. A number of these GlyT1 inhibitors have been repeatedly proven to be effective and successful in animal models of schizophrenia and clinically assessed for several symptom domains [2,9,10]. A few years ago, Roche succeeded in developing a noncompetitive GlyT1 inhibitor as the most advanced compound, namely bitopertin (RG1678), which attained positive results to treat negative symptoms of schizophrenia patients. However, it did not successfully reduce these negative symptoms in the later stages of clinical trials, due to its very limited efficacy [9,11]. For this reason, the discovery of selective and competitive glycine transporter type 1 inhibitors continues to be of great interest to chemists in a wide range of literature [1,2,10,12–15]. Recently, a structural family of 3,4-disubstituted pyrrolidine sulfonamides, was synthesized and experimentally evaluated through in vivo and in vitro assays to selectively and competitively inhibit the glycine transporter type 1. In this regard, our in-silico investigations were carried out on thirty derivatives of 3,4-disubstituted pyrrolidine sulfonamides which were detected thanks to their inhibitory activity against the type 1 of glycine transporter [16]. In the beginning, we applied the quantitative structure-activity relationship (QSAR), which gained an advanced position in modern chemistry [17], by developing two validated QSAR models which will be able to predict human GluT1 activity for 3,4-disubstituted pyrrolidine sulfonamides family, using multiple linear regression (MLR) and multiple nonlinear regression (MNLR) techniques [18]. Thereafter, we predicted the physicochemical and pharmacokinetic properties of candidate inhibitors able to cross the blood-brain barrier (BBB), penetrating the central nervous system (CNS). Afterwards, the most active inhibitor having an excellent absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) profile and meeting all drug-likeliness conditions [19], was chosen for molecular docking simulation as a technology widely applied in the field of drug discovery [20], indicating the intermolecular interactions produced towards the dopamine transporter membrane protein encoded as 4M48.pdb https://www.rcsb.org/structure/4M48 [21]. Moreover, the process of molecular docking for the (candidate ligand-targeted protein) complex was successfully validated using the docking validation protocol [22]. The molecular stability of the studied

complex was examined with the help of the molecular dynamics (MD) technique during 100 nanoseconds of MD simulation time, studying the conformational changes behaviors of the targeted protein towards the candidate drug with detailed temporal resolution and highest atomic precision [23–26]. Finally, the molecular mechanics with generalized born and surface area (MMGBSA) solvation was equally conducted to calculate the binding free energies of the studied complex, to check again the molecular stability of the candidate ligand in complex with the protein target [27,28]. Thus, the present study aims to discover a highly effective drug among thirty selective and competitive GluT1 inhibitors, predicted with a good ADME-Toxicity profile, and defined with a good level of molecular stability towards the targeted protein, which we can strongly recommend for the treatment of schizophrenia and other impairments associated with NMDA receptor hypofunction.

## 2. Materials and methods

### 2.1.Experimental Database

The present work aims to study thirty of 3,4-disubstituted pyrrolidine sulfonamides shown in Table S7 [16], whose the human GlyT1 activities of Ki order were converted to pKi scale ($pKi = - \log_{10} Ki$), and the experimental database was divided on training and test sets which contain 80 % and 20 % of active molecules, respectively.

### 2.2.Molecular Descriptors Calculation

To develop the quantitative structure-activity relationships (QSAR), a number of molecular descriptors were calculated, as resulted in Table S8. The constitutional descriptors were executed using ACD/chemsketch software [29], while the geometric, physicochemical, thermodynamic, and topological descriptors were obtained using ChemBio3D software [30].

### 2.3.Statistical Methods

Initially, the principal component analysis (PCA) was applied as a multivariate statistical method, to reduce the dimension of calculated descriptors based on the correlation matrix, in which the non-correlated descriptors were kept well, while one of the two highly correlated descriptors ($r > 0.9$) was removed [31–33]. Secondly, the linear and nonlinear relationships between human GluT1 activity and uncorrelated descriptors were studied using multiple linear regression (MLR) and multiple nonlinear regression (MNLR) techniques, respectively [18]. So, MLR and MNLR QSAR models, which were developed for training set of 24 molecules, were externally validated for test set on 6 novel molecules. Moreover, the mathematical models were internally validated with the help of cross validation with the leave one out procedure (CV-LOO) [34].

**2.4.In-silico pharmacokinetics, drug-likeliness, and ADMET depiction exploration**

Before proceeding with any clinical investigation, it is necessary to carry out an assessment of absorption, distribution, metabolism, excretion and toxicity (ADME-Tox) properties in the human body, checking the five rules of Lipinski [35], and respecting the violations number of Veber, Ghose, Egan, and Muegge [36–39] for the drug candidates, which were predicted as central nervous system (CNS) agents according to Boild-Eggs model. To achieve this objective, we examined the physicochemical and in-silico pharmacokinetics properties of the candidate's compounds, with the help of pkCSM [40], and SwissADME [41] servers.

**2.5.Molecular Docking Simulation**

To discover the chemical bonds produced between a drug candidate and a target protein, we performed molecular docking, an effective drug discovery technique [42]. The dopamine transport protein, was uploaded from proteins data bank (PDB), which was discovered by the X-ray diffraction technique with a resolution of 2.96 Å [43,44]. The targeted protein coded as 4M48.pdb was prepared using Discovery Studio 2021 (BIOVIA) software package [45], removing the water molecules ($H_2O$), chloride and sodium ions (Na+, Cl-) and all suspended ligands like as cholesterol ($C_{27}H_{46}O$) and nortriptyline ($C_{19}H_{21}N$). The gasteiger charges and polar hydrogens were added. Then, the prepared protein was docked to the candidate drug using AutoDock 4.2 software [46], converting the complex to PDBQT format. The sizes of grid box were centralized on (−42.424 Å, −0.573 Å, 55.066 Å), putting the 3D- structure dimensions on (120, 120, 120) with a spacing of 0.375 Å. Finally, the results of intermolecular interactions produced for the most stable complex were visualized by mean of Discovery Studio 2021 (BIOVIA) software.

**2.6.Molecular Dynamic**

To investigate the stability of intermolecular interactions produced between the candidate drug and targeted protein, we performed the molecular dynamics for studied complex during 100 nanoseconds of simulation time, using Desmond software, a package of Schrödinger LLC [47]. The (candidate drug-targeted protein) complex was prepared with the assistance of Protein Preparation Maestro. Then, the pre-treated complex was optimized in the System Builder, working on a solvent model of TIP3P orthorhombic box, with application of OPLS force field [48]. Additionally, the selected model was neutralized by the reproduction of physiological conditions, adding water molecules ($H_2O$) and counter ions ([$Na^+$, $Cl^-$] equal to 0.15 M) with a temperature of 300 K and a pressure of 1 atm.

## 3. Results and discussion

### 3.1.Quantitative Structure-Activity Relationships (QSARs)

#### 3.1.1. Principal Components Analysis

The quantitative structure-activity relationship technique is intended to develop a predictive QSAR model between biological activity and uncorrelated descriptors. For this goal, we applied the principal component analysis (PCA) as a statistical method widely applied to minimize the size of molecular descriptors in a reduced number of factorial axes (or principal components) based on the correlation matrix [49], as resulted in Table S8, which indicates thirty-two strong correlations (correlation coefficients of Pearson superior than 0.9). So, the development of QSAR model was focused only on weakly correlated descriptors.

#### 3.1.2. Multiple Linear Regression

To study the linear relationship between the structural descriptors and human GluT1 activity, we applied the multiple linear regression (MLR) technique [50] on 24 molecules of training set, based on stepwise selection for hundreds of randomizations. Therefore, the best MLR QSAR model was obtained in equation 1, considering the best values of determination, adjustment, and correlation coefficients ($R^2$, $R^2$ *adjusted*, and $R$) and minimal root mean square deviation (RMSD).

**pKi = -60.528+43.061×D+3.569×%H+0.321×VDW E-0.087×T E-0.642×%N-0.539×%O**

**Equation 1**

The first predictive QSAR model developed in equation 1, demonstrates that human Glut1 activity of pKi order is strongly correlated with constitutional descriptors (in particular %H, %N, %O and D) and geometric descriptors (like as T E, and VDW E), with minimal root mean square deviation (RMSD of 0.546) and good values of determination, correlation, and adjustment coefficients ($R^2$= 0.746, $R^2$ *adjusted*= 0.657, and $R$= 0.864, respectively). The six selected descriptors (Table S8) have significant impacts on human GluT1 activity, because the slope of each one is defined by a probability less than 0.05 (5%) for 95% of confidence interval, as displayed in Table 28.

**Table 28. Significant weights of the selected descriptors**

| Source | Value | Standard deviation | t | Pr > \|t\| | Lower terminal (95%) | Higher terminal (95%) |
|---|---|---|---|---|---|---|
| Constante | -60.528 | 12.047 | -5.024 | 0.000 | -85.945 | -35.111 |
| D | 43.061 | 7.133 | 6.037 | < 0.0001 | 28.011 | 58.110 |
| %H | 3.569 | 0.710 | 5.025 | 0.000 | 2.071 | 5.068 |
| VDW E | 0.321 | 0.101 | 3.162 | 0.006 | 0.107 | 0.534 |

| | | | | | | |
|---|---|---|---|---|---|---|
| T E | -0.087 | 0.026 | -3.366 | 0.004 | -0.142 | -0.033 |
| %N | -0.642 | 0.129 | -4.969 | 0.000 | -0.915 | -0.369 |
| %O | -0.539 | 0.117 | -4.628 | 0.000 | -0.785 | -0.293 |

The developed MLR model shows that electronic density (D), mass percentage of hydrogen (%H), and van der Waals energy (VDW E) have a positive effect on human GluT1 activity, but mass percentages of nitrogen and oxygen (%N, %O), and torsion energy (T E) affect it negatively, as displayed in Figure 43.

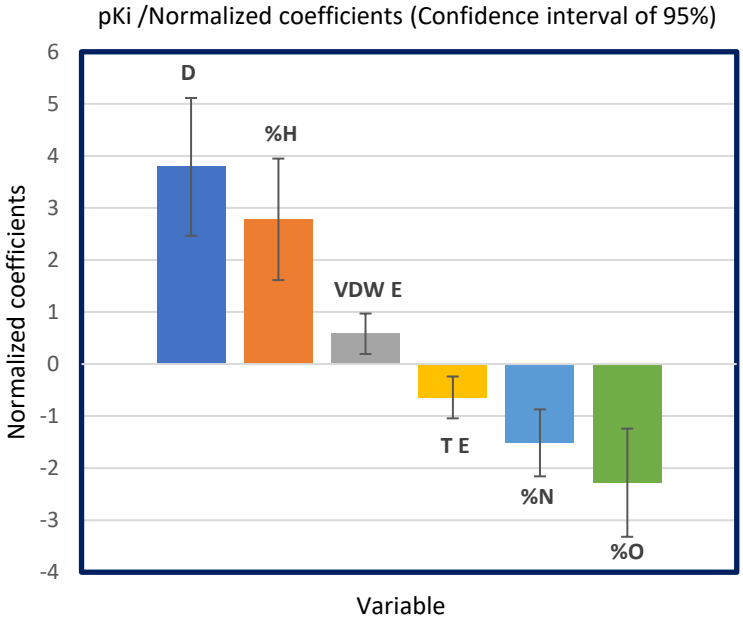



**Figure 43. Impact of selected descriptors on human GluT1 activity**

The analysis of variance (Anova test) shows that calculated value of Fisher (F= 8.332) is superior to its critical value (F (30, 6) = 2.42, p<5%), which indicates that null hypothesis (H0) was rejected, so there is an homogenous variance between the inhibitory activity and six selected descriptors as displayed in Table 29.

**Table 29. Analysis of variance using the fisher test**

| Source | DDL | Total square | Mean square | F | Pr > F |
|---|---|---|---|---|---|
| Model | 6 | 14.923 | 2.487 | 8.332 | 0.000 |
| Error | 17 | 5.075 | 0.299 | | |
| Adjusted total | 23 | 19.998 | | | |

### 3.2.Multiple Non-Linear Regression

To model the nonlinear regression between the six uncorrelated descriptors, signaled by MLR method and GlyT1 activity, we performed the multiple non-linear regression (MNLR) method [51], for twenty-four inhibitors (training set) based on following programmed function:

$$Y = a0 + \sum_{i=1}^{n} (ai \times Xi + bi \times Xi^2) \qquad [24]$$

Y: predicted activity by MNLR QSAR model. a0: constant of predictive model. Xi: selected descriptor. ai, bi: slopes of each descriptor for one and two degrees, respectively.

The second QSAR model, developed by the pre-programmed function of multiple non-linear regression confirms also that human GluT1 activity of pki order is affected by the six selected descriptors as expressed in equation 2. This mathematical model is done by a low deviation (RMSD equal to 0.636) and good values of determination and correlation coefficients ($R^2$= 0.777 and $R$= 0.882, respectively).

**pKi=87.623-193.402×D+11.223×%H+1.393×VDW E-0.048×T E-1.080×%N -0.479×%O+80.55646×$D^2$-0.866×%$H^2$-0.027×VDW $E^2$-0.00038×T $E^2$+0.014×%$N^2$ 0.004×%$O^2$ . Equation 2**

### 3.3.Validation of predictive QSAR Models

#### 3.3.1. External Validation

External validation is an essential and critical step to examine the reliability of MLR and MNLR QSAR models, ensuring their application to new molecules included in the test set [52]. Based on developed models in equations 1 and 2 for twenty-four molecules (training set), we tested six novel molecules (2, 10, 11, 13, 19, 24) which are part of test set, and we obtained an external correlation coefficient of $R^2ext$=0.648 and $R^2ext'$=0.872 for MLR and MNLR models, respectively. According to Golbreikh and Tropsha's study, the linear and non-linear models are externally validated because the external correlation coefficients are superior to 0.6. The Table 30, shows the predicted activities for six novel molecules by MLR and MNLR models, where the quantitative correlations are translated in Figure 44 and 45, successively.

**Table 30. Observed and predicted activities of pKi order using MLR and MNLR models**

| Compounds number | Molecules nomenclature | Observed pKi | Predicted pKi (MLR) | Predicted pKi (MNLR) |
|---|---|---|---|---|
| 2* | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N,4-diphenylpyrrolidin-3-amine. | 6.701 | 5.447 | 5.606 |
| 10* | N-((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)-5-(trifluoromethyl)pyridin-3-amine. | 5.936 | 7.218 | 7.088 |
| 11* | (3R,4S)-N-(3-chlorophenyl)-4-(3-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 7.377 | 8.286 | 8.257 |

| | | | | |
|---|---|---|---|---|
| 13* | N-((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)-2-(trifluoromethyl)pyridin-4-amine. | 6.928 | 7.986 | 7.870 |
| 19* | (3R,4S)-N-((R)-2,3-dihydro-1H-inden-1-yl)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 9.000 | 10.644 | 9.518 |
| 24* | 4-(((3R,4S)-3-benzyl-4-phenylpyrrolidin-1-yl)sulfonyl)-1-methyl-1H-imidazole. | 7.032 | 6.653 | 7.845 |

* Indicates the test set molecules

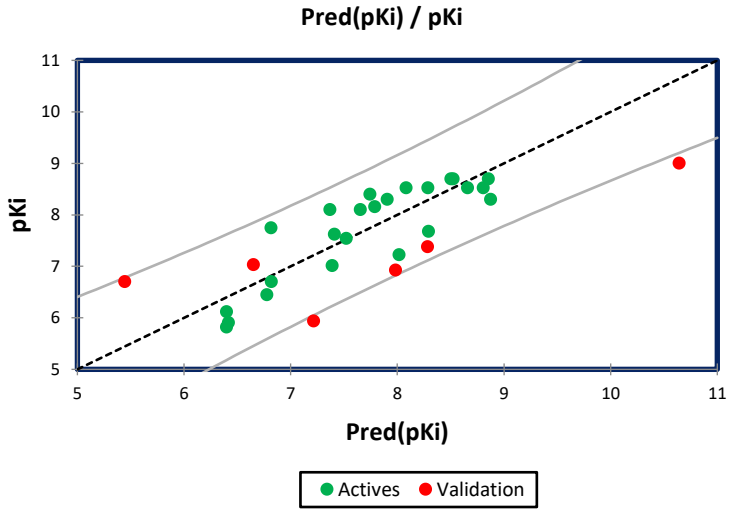



**Figure 44. Quantitative correlation between observed and predicted GluT1 activity using MLR technique**

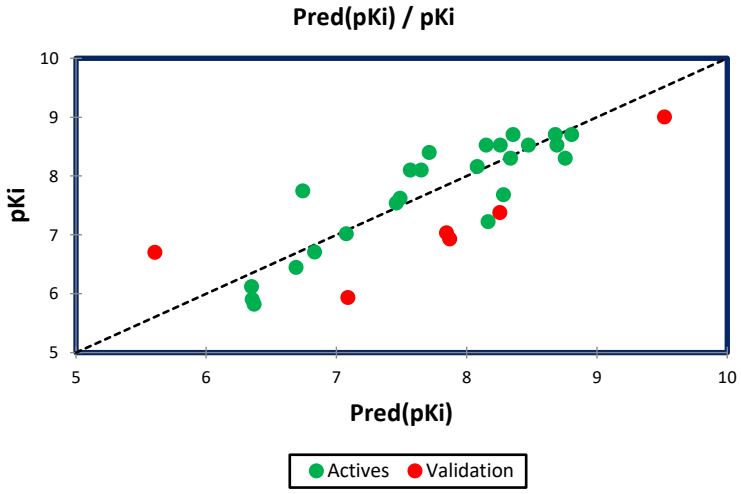



**Figure 45. Quantitative correlation between observed and predicted GluT1 activity using MNLR technique**

### 3.3.2. Internal Validation

A cross-validation (CV) technique was performed using the leave-one-out (LOO) procedure, for internal validation of the established QSAR models, testing each chemical compound once, modeling a new mathematical model on twenty-three molecules each time (N-

1 = 23) and predicting human GluT1 activity for each tested molecule [53]. The process was repeated for all molecules in the training set, as resulted in Table 31. Afterwards, the quadratic coefficient ($Q^2$ *cv*) of cross-validation, was calculated based on the following equation:

$$Q^2cv = 1 - \frac{\sum_i^n (Yob - Ypr)^2}{\sum_i^n (Yob - Ymean)^2} \qquad \text{Equation 3} \qquad [54]$$

As: Yob: the observed activity, Ypr: the predicted activity, and Ymean: the mean of observed activities.

According to Golbreikh and Tropsha's study, although the result of quadratic coefficient ($Q^2cv$=0.546) exceeds 0.5, so the developed QSAR model is internally validated with a good reliability and strong predictivity.

**Table 31. Observed and predicted activities using MLR, MNLR and CVLOO techniques**

| Molecules number | Molecules nomenclature | Observed pKi | Predicted pKi (MLR) | Predicted pKi (MNLR) | Predicted pKi (CVLOO) |
|---|---|---|---|---|---|
| 1 | 2-chloro-N-((3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-phenylpyrrolidin-3-yl)-3-(trifluoromethyl)benzamide. | 6.703 | 6.820 | 6.834 | 7.131 |
| 3 | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-phenyl-N-(o-tolyl)pyrrolidin-3-amine. | 5.821 | 6.399 | 6.367 | 6.648 |
| 4 | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-phenyl-N-(m-tolyl)pyrrolidin-3-amine. | 7.745 | 6.817 | 6.742 | 6.417 |
| 5 | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-phenyl-N-(p-tolyl)pyrrolidin-3-amine. | 5.903 | 6.419 | 6.353 | 6.634 |
| 6 | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-phenyl-N-(3-(trifluoromethyl)phenyl)pyrrolidin-3-amine. | 8.097 | 7.370 | 7.651 | 7.292 |
| 7 | (3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N-(3-(trifluoromethyl)phenyl)pyrrolidin-3-amine. | 8.523 | 8.085 | 8.150 | 8.018 |
| 8 | (3R,4S)-N-(3-chlorophenyl)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 8.523 | 8.288 | 8.259 | 8.247 |
| 9 | (3R,4S)-N-(4-fluoro-3-(trifluoromethyl)phenyl)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 8.301 | 8.876 | 8.758 | 9.085 |
| 12 | (3R,4S)-N-(3-chlorophenyl)-4-(2-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 7.678 | 8.296 | 8.282 | 8.394 |
| 14 | (3R,4S)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-(pyridin-3-yl)-N-(3-(trifluoromethyl)phenyl)pyrrolidin-3-amine. | 6.116 | 6.401 | 6.350 | 6.494 |

| | | | | | |
|---|---|---|---|---|---|
| 15 | (3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidin-3-amine. | 8.398 | 7.745 | 7.714 | 7.639 |
| 16 | N-((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)-4-(trifluoromethyl)pyridin-2-amine. | 8.097 | 7.656 | 7.568 | 7.576 |
| 17 | N-((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)-6-(trifluoromethyl)pyrimidin-4-amine. | 8.301 | 7.906 | 8.337 | 7.629 |
| 18 | (3R,4S)-N-(4-fluoro-3-(trifluoromethoxy)phenyl)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 8.699 | 8.504 | 8.357 | 8.477 |
| 20 | (3R,4S)-N-(4-fluoro-3-(trifluoromethyl)phenyl)-4-(4-fluorophenyl)-1-((1-methyl-1H-1,2,3-triazol-4-yl)sulfonyl)pyrrolidin-3-amine. | 8.523 | 8.806 | 8.694 | 8.896 |
| 21 | 4-(((3S,4R)-3-(4-fluorophenyl)-4-((3-(trifluoromethoxy)phenoxy)methyl)pyrrolidin-1-yl)sulfonyl)-1-methyl-1H-imidazole. | 8.155 | 7.790 | 8.082 | 7.649 |
| 22 | ((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)(3-(trifluoromethoxy)phenyl)methanone. | 7.538 | 7.522 | 7.459 | 7.515 |
| 23 | N-(((3S,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)methyl)-3-(trifluoromethoxy)aniline. | 7.222 | 8.017 | 8.164 | 8.108 |
| 25 | 4-(((3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)pyrrolidin-3-yl)oxy)-6-(trifluoromethyl)pyrimidine. | 7.013 | 7.390 | 7.077 | 7.583 |
| 26 | (3R,4S)-4-(4-fluorophenyl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidine-3-carboxamide. | 6.444 | 6.779 | 6.690 | 7.137 |
| 27 | (3R,4R)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-((R)-tetrahydro-2H-pyran-2-yl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidine-3-carboxamide. | 8.523 | 8.660 | 8.474 | 8.714 |
| 28 | (3R,4R)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-4-((S)-tetrahydrofuran-2-yl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidine-3-carboxamide. | 7.620 | 7.411 | 7.489 | 6.921 |
| 29 | (3R,4R)-4-(1,3-dioxepan-2-yl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidine-3-carboxamide. | 8.699 | 8.854 | 8.806 | 8.968 |
| 30 | (3R,4R)-4-(1,3-dioxan-2-yl)-1-((1-methyl-1H-imidazol-4-yl)sulfonyl)-N-(3-(trifluoromethoxy)phenyl)pyrrolidine-3-carboxamide. | 8.699 | 8.527 | 8.682 | 8.445 |

### 3.3.3. Validation Using Y-Randomisation Test

According to Golbreikh and Tropsha's study, the highest value of quadratic coefficient generated by cross validation technique with the leave one out procedure (CVLOO) is probable to be overestimated and may refers to a lucky correlation, so the internal validation is necessary but still not enough sufficient. For this reason, the Y-Randomization test is so required [55,56]. Using java Platform SE binary, we evaluated the predictive accuracy of developed QSAR model through one hundred randomizations as resulted in Table S9. The value of $cR^2p$ criteria was 0.621, which exceeds 0.5 threshold. Moreover, the resulted correlation, determination, and cross-validation with the leave-one-out procedure coefficients of the original model are so higher than the marked values for all 100 randomizations. So, the predicted GlyT1 activity by the originally developed model are not due to random chance.

### 3.3.4. Statistical Criterias of Golbreikh and Tropsha's study

Golbreikh and Tropsha's statistical study emphasizes that a certain number of postulated conditions must be verified before the quantitative structure-activity relationship model can be applied to other new molecules. In this part, we noted that the generated model by MLR technique satisfy with success all postulated threshold conditions of Tropsha and Golbraikh's theory, as presented in Table 32. Consequently, the developed model is qualified by an accurate and precise predictivity, which affirms again that human GlyT1 activity is really affected by the six selected descriptors.

**Table 32. Tropsha and Golbraikh conditions to examine the accuracy of generated MLR QSAR model**

| Statistical parameters | Developed Equations | Model scores | Threshold | Comment |
|---|---|---|---|---|
| $R^2$ | $R^2 = 1 - \frac{\sum ( Yobs - Y\,cal )^2}{\sum (Yobs - \overline{Yobs})^2}$ | 0.746 | > 0.6 | Accepted |
| $R^2adj$ | $R^2adj = \frac{(N-1)R^2 - p}{N - p - 1}$ | 0.657 | > 0.6 | Accepted |
| $R^2test$ | $R^2test = 1 - \frac{\sum ( Ycal(test) - Y\,obs(test) )^2}{\sum (Yobs(test) - \overline{Yobs}(train))^2}$ | 0.648 | > 0.6 | Accepted |
| $Q^2cv$ | $Q^2cv = 1 - \frac{\sum ( Ycal - Y\,obs )^2}{\sum (Yobs - \overline{Yobs})^2}$ | 0.529 | > 0.5 | Accepted |
| $R^2rand$ | The average of 100 R2rand (i) | 0.243 | $< R^2$ | Accepted |
| $Q^2cv$ 'LOO' rand | The average for 100 $Q^2cv$ 'LOO' rand (i) | -0.660 | $< Q^2cv$ | Accepted |
| $cR^2p$ | $cR^2p$ =R*$\sqrt{(R^2 - ( Average\ R_{rand} )^2)}$ | 0.621 | > 0.5 | Accepted |

### 3.4.In-silico Pharmacokinetics ADMET Prediction

We initiated the in-silico screening of thirty human GlyT1 inhibitors, using the BOILED-Egg model as a widely applied predictive technique in the field of medicinal chemistry, to discover drug candidates that could penetrate the central nervous system (CNS), as resulted in Figure 46. We noticed that most of the chemical compounds are part of white Egan Eggs, so they are predicted to be absorbed by the gastrointestinal tract, except the molecules marked as 2, 3, 4, 5, and 19, which are part of the yellow part of Egan-egg, thus they are predicted to be effluated from the central nervous system (CNS) by the P-glycoprotein, and predicted as potent inhibitors with an important ability to cross the blood-brain barrier (BBB). However, molecule 24 colored in red was predicted to not be effluated from the central nervous system, but it's capable to pass through the blood-brain barrier (BBB) [57]. Therefore, we conclude that just five glycine transporter type 1 inhibitors (C2, C3, C4, C5, and C19) which are predicted as central nervous system agents with the highest probability to cross the blood-brain barrier.

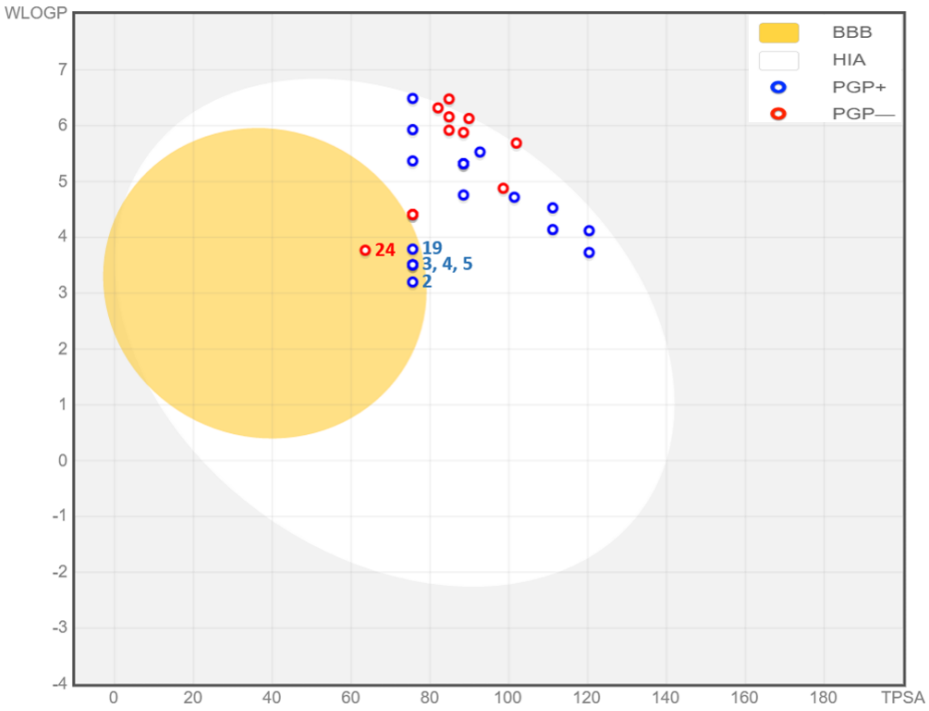



**Figure 46. Predictive model of BOILED-Egg for 3,4-disubstituted pyrrolidine sulfonamides**

Subsequently, we studied the physicochemical and pharmacokinetic properties of all central nervous system agents, as shown in Table 33 and Table 34, respectively. Based on the properties of nortriptyline as a positive control molecule co-crystallized with the DAT protein, we noticed that all predicted inhibitors are in good agreement with nortriptyline, where they respect all five rules of Lipinski (MW ≤ 500 g/mol, 40 ≤ MR ≤ 130, HBD < 5, Log P (Octanol/Water) < 5, HBA ≤ 10) [58], and satisfy the violations number of Veber, Ghose, Egan,

and Muegge [59]. Moreover, the pharmacokinetics properties reveal a good ADME-Toxicity profile, where all predicted molecules are defined by a good human intestinal absorption (HIA superior than 94 %), and the values of CNC and BBB permeabilities are included in (-2 to -3) Log PS, and higher than -1 Log BB, respectively. Thus, they have a good permeabilities to the central nervous system (CNS) and the blood-brain barrier (BBB). In addition, they are all predicted as substrates of 3A4 cytochrome. C4 and C24 compounds and C19 as the most active ligand, are predicted as potent inhibitors for 2C9 and 3A4 cytochromes. Also, they are defined by an excretion of a total clearance inferior than 1 Log ml/min/kg. According to the AMES toxicity test, all chemical compounds are predicted as non-toxic inhibitors of the human body, without any skin sensitization, but they have a known positive hepatotoxicity.

**Table 33. Prediction of physico-chemical parameters of central nervous system agents**

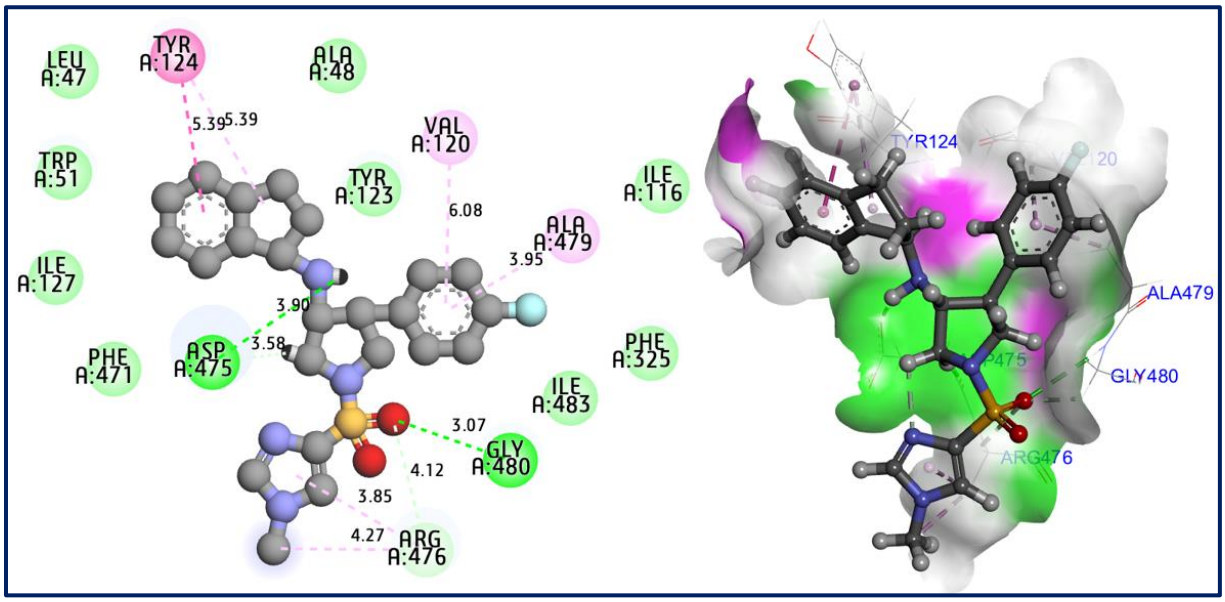
| Compounds number | Physical and chemical properties | | | | | | Lipinski violations | Veber violations | Egan violations | Ghose violations | Muegge violations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Molecular weight (g/mol) | Molar refractive index | Rotatable bonds Number | Log P (Octanol/ Water) | H-bond Acceptors Number | H-bond donors Number | Categorical (Yes/No) | | | | |
| Threshold | ≤500 | 40≤ MR ≤130 | <10 | <5 | ≤10 | <5 | | | | | |
| C2 | 382.48 | 109.01 | 5 | 3.03 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C3 | 396.51 | 113.98 | 5 | 2.65 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C4 | 396.51 | 113.98 | 5 | 3.24 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C5 | 396.51 | 113.98 | 5 | 3.00 | 4 | 1 | Yes | Yes | Yes | Yes | Yes |
| C19 | 440.53 | 119.90 | 5 | 3.37 | 6 | 1 | Yes | Yes | Yes | Yes | Yes |
| C24 | 381.49 | 109.48 | 5 | 2.55 | 4 | 0 | Yes | Yes | Yes | Yes | Yes |
| Nortriptylene | 263.38 | 86.06 | 3 | 3.22 | 1 | 1 | Yes | Yes | Yes | Yes | No |

**Table 34. Prediction of ADMET-Toxicity pharmacokinetic properties of CNS agents**

| Compounds number | Absorption | Distribution | | Metabolism | | | | | | | Excretion | Toxicity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intestinal Absorption (human) | BBB permeability | CNS permeability | Substrate | Inhibitor | | | | | | Total Clearance | AMES Toxicity | Hepatotoxicity | Skin Sensitization |
| | | | | Cytochromes | | | | | | | | | | |
| | | | | 2D6 | 3A4 | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 | | | | |
| | Numeric (% Absorbed) | Numeric (Log BB) | Numeric (Log PS) | Categorical (Yes/No) | | | | | | | Numeric (Log ml/min/kg) | Categorical (Yes/No) | | |
| C2 | 94.986 | 0.175 | -2.334 | No | Yes | No | Yes | Yes | No | No | 0.27 | Not toxic | Yes | No |
| C3 | 96.126 | 0.219 | -2.264 | No | Yes | No | Yes | No | No | Yes | 0.269 | Not toxic | Yes | No |
| C4 | 95.608 | 0.142 | -2.254 | No | Yes | No | Yes | Yes | No | Yes | 0.264 | Not toxic | Yes | No |
| C5 | 95.522 | 0.125 | -2.26 | No | Yes | No | Yes | No | No | Yes | 0.272 | Not toxic | Yes | No |
| C19 | 94.049 | 0.151 | -2.313 | No | Yes | No | No | Yes | No | Yes | 0.41 | Not toxic | Yes | No |
| C24 | 99.251 | 0.066 | -2.216 | No | Yes | Yes | Yes | Yes | No | Yes | 0.114 | Not toxic | Yes | No |
| Nortriptylene | 97.482 | 0.808 | -1.196 | No | Yes | Yes | No | No | Yes | No | 0.929 | Not toxic | Yes | Yes |

### 3.5.Molecular Docking

The results of molecular docking simulation exposed in the Figure 47, demonstrate the intermolecular interactions produced between the dopamine transporter (DAT) protein as a key therapeutic strategy to treat clinical depression and neuropathic pain [60,61], and the compound C19 as the most active ligand with a good ADME-Toxicity profile. 2D and 3D dimensional visualizations confirm that C19 ligand forms two hydrogen bonds towards ASP 475 and GLU 480 amino acids residues of DAT protein in A chain, with nuclear distances of 3.90 Å and 3.07 Å respectively, taking into account that the chemical bonds of the hydrogen bonding type make the (ligand-protein) complex so stable. Additionally, C19 reacts to VAL 120 and ALA 479 amino acids forming two chemical bonds of Pi-Alkyl type at 6.08 Å and 3.95 Å, respectively. And Pi-Pi T- shaped bond with TYR 124 amino acid at 5.39 Å. Consequently, ASP 475, GLU 487, VAL 120, ALA 479, and TYR 124 are the main sites of glycine transporter type 1 inhibition for the candidate drug. The resulted chemical bonds include intermolecular interactions similar to those detected towards Tyr 124, Ala 479, and Val 120 amino acid residues, which have been produced by two selective GluT1 inhibitors, after being bound to the same membrane protein of drosophila melanogaster dopamine transporter (DAT), which were also designed to improve memory performance while treating depression and the cognitive symptoms associated with schizophrenia [1,2,54].



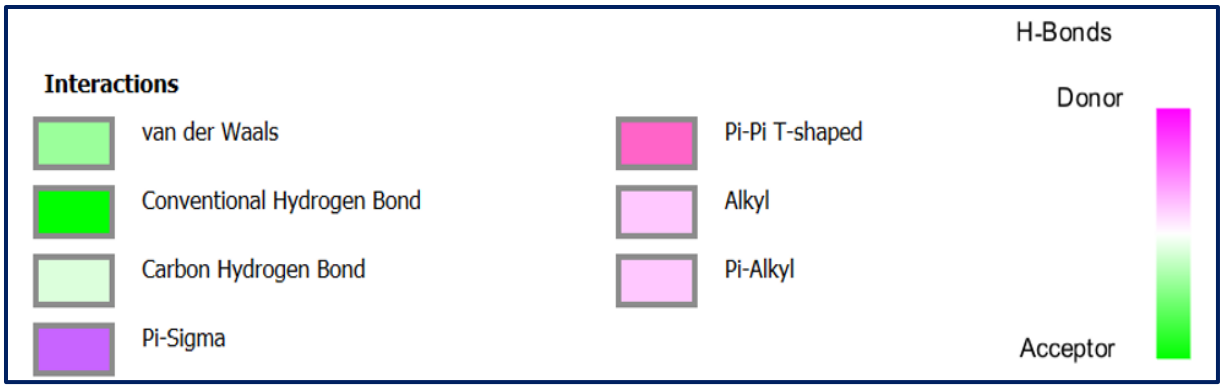



**Figure 47. 2D (left) and 3D (right) intermolecular interactions, produced for (C19 ligand - DAT protein) complex with binding energy of -8.96 Kcal/mol.**

### 3.6.Docking Validation Protocol

The process of molecular docking simulations was examined using docking validation protocol or re-docking methodology for the (C19 ligand – DAT protein) complex. In the first time, we have compared the intermolecular interactions produced for the studied complex with the active sites of nortriptyline as co-crystallized ligand bound to the same chain of targeted protein. Where, the active sites of nortriptyline are the residues of TYR 124, PHE 325, and PHE 43 amino acids, which are experimentally extracted using the ProteinsPlus online server [62], as resulted in Figure 48A, and TYR 124, PHE 325, PHE 43, VAL 120, and ALA 479 amino acids, which are theoretically produced using Autodock tools, as presented in Figure 48B. Therefore, we noticed that TYR 124, VAL 120, and ALA 479 amino acids are the same amino acids, which were obtained for (C19 - DAT protein) complex. In second time, we have used the superposition of co-crystallized and docked nortriptyline, and we have obtained a good overlay pose with a minimal root mean square deviation (RMSD =0.128) which is less than 2 Å as shown in Figure 46C. Therefore, we conclude that the candidate drug was docked in the active sites of targeted protein, and the process of molecular docking is successfully validated [63].

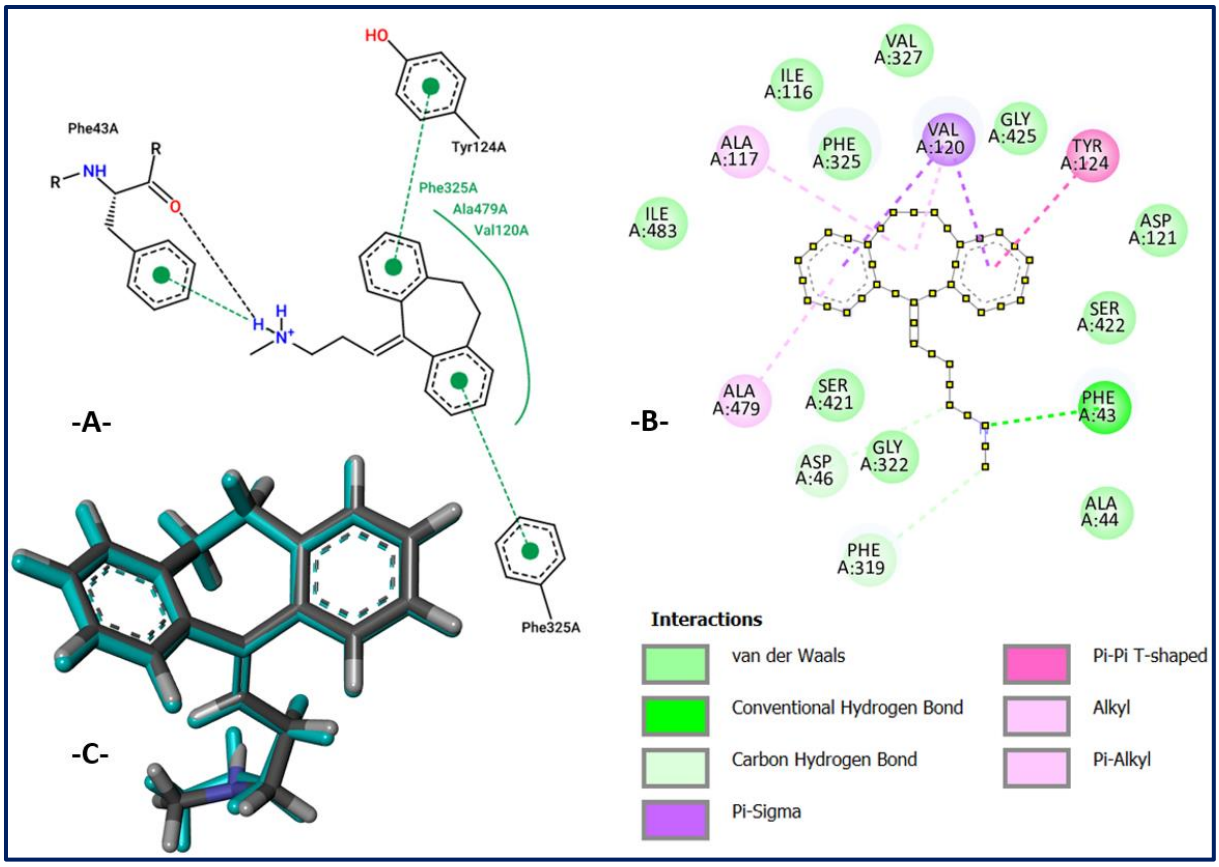



**Figure 48. A: Active sites of co-crystallized ligand of DAT protein in A-chain. B: 2D intermolecular interactions produced for the (docked nortriptyline - DAT protein) complex with a binding energy of −9.02 kcal/mol. C: Re-docking pose with a RMSD value of 0.128 Å (co-crystallized nortriptyline colored in cyan superposed on docked nortriptyline)**

### 3.7.Molecular Dynamic simulation

High-performance molecular dynamics simulation was performed to study the stability of intermolecular interactions produced for the (C19 ligand - DAT protein) complex during 100 nanoseconds [24,64]. The results of the conformational changes presented in Figure 49 show that the root mean square deviation (RMSD) of the candidate ligand oscillates in parallel with the RMSD values of the protein target during the first thirty nanoseconds, then destabilizes a little, and stabilizes again during the last thirty nanoseconds due to the effect of heavy atoms as oxygen and nitrogen which make the ligand-protein complex so stable. In addition, there are no deviations greater than 3 Å, which demonstrates that the simulation remains in equilibrium during the molecular dynamics simulation time without any noticeable deviation. Secondly, the root mean square fluctuation (RMSF) values show only two remarkable fluctuations for the residuals 470 and 530, which exceeded 3 Å, as shown in Figure 49. So, we conclude that the ligand does not diffuse far from its primary binding site.

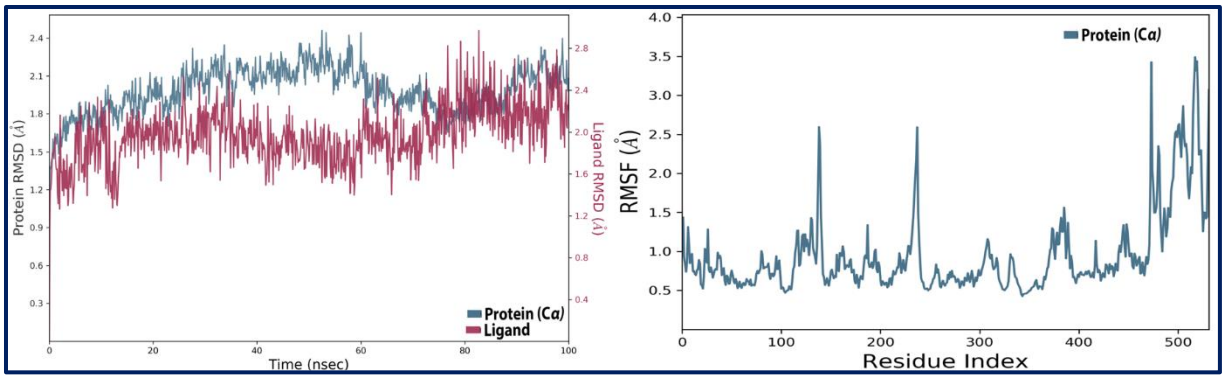



**Figure 49. Changes of Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) during 100 ns, for the dopamine transporter protein complexed with C19 ligand.**

In addition, we recorded negligible fluctuations for ligand properties, such as radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA), and polar surface area (PSA), which all oscillate in a very small range over 75 nanoseconds, making just one remarkable fluctuation at the same time, due to the influence of nitrogen, oxygen, sulfur, and fluorine, as typical heteroatoms characterizing this organic molecule [65,66]. However, they stabilize again until the rest of the molecular dynamics simulation time as displayed in Figure 50, which demonstrates minimal changes in the compactness of the candidate ligand. Consequently, the targeted protein is very flexible after being bound to the candidate ligand.

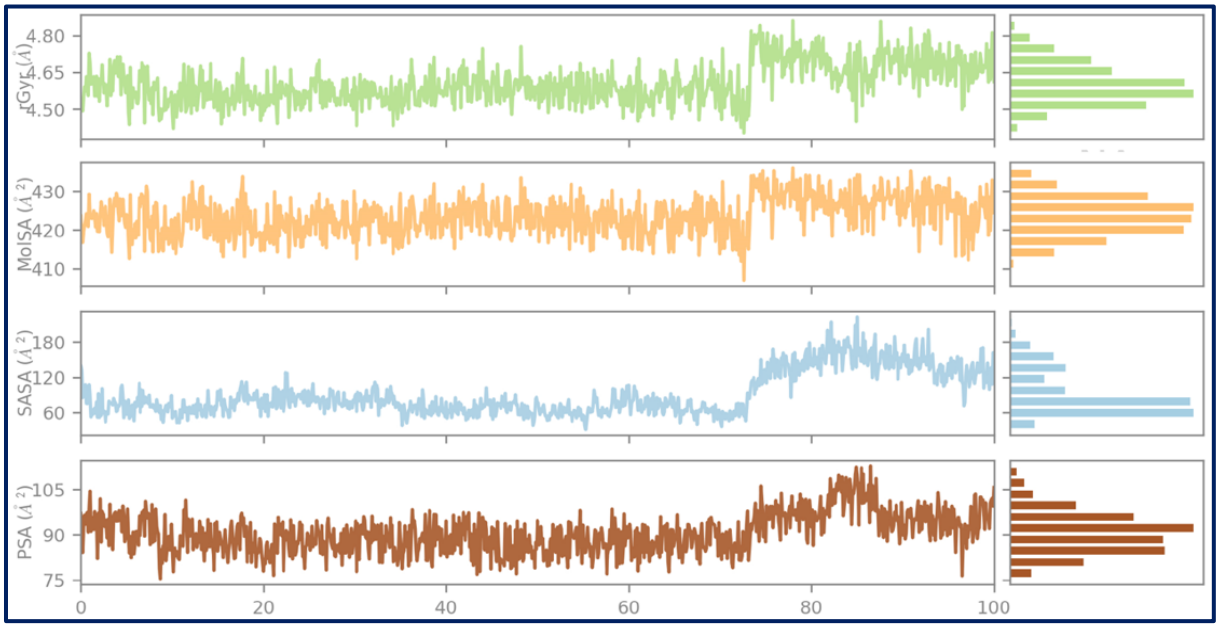



**Figure 50. Changes in radius of gyration (r Gyr), molecular surface area (MolSA), solvent accessible surface area (SASA), and polar surface area (PSA), over 100 ns of molecular dynamics simulation time**

Furthermore, the upper panel shown in Figure 51, reflects the specified number of contacts produced by the candidate ligand towards the targeted protein throughout molecular dynamics simulation time, while the lower panel displays the residues of amino acids, which reacted with the C19 compound along the trajectory. The scale on the right-hand side of the graph shows that the strongest connection is one that corresponds to a dark orange residue that has entered into more than one chemical reaction with the candidate ligand [67]. For this, ASP 475 amino acid residue has the main function in the stability of (protein-ligand) complex throughout 100 nanoseconds of MD simulation time, which presents an important interaction fraction of H-bonds as displayed on the left side of Figure 51.

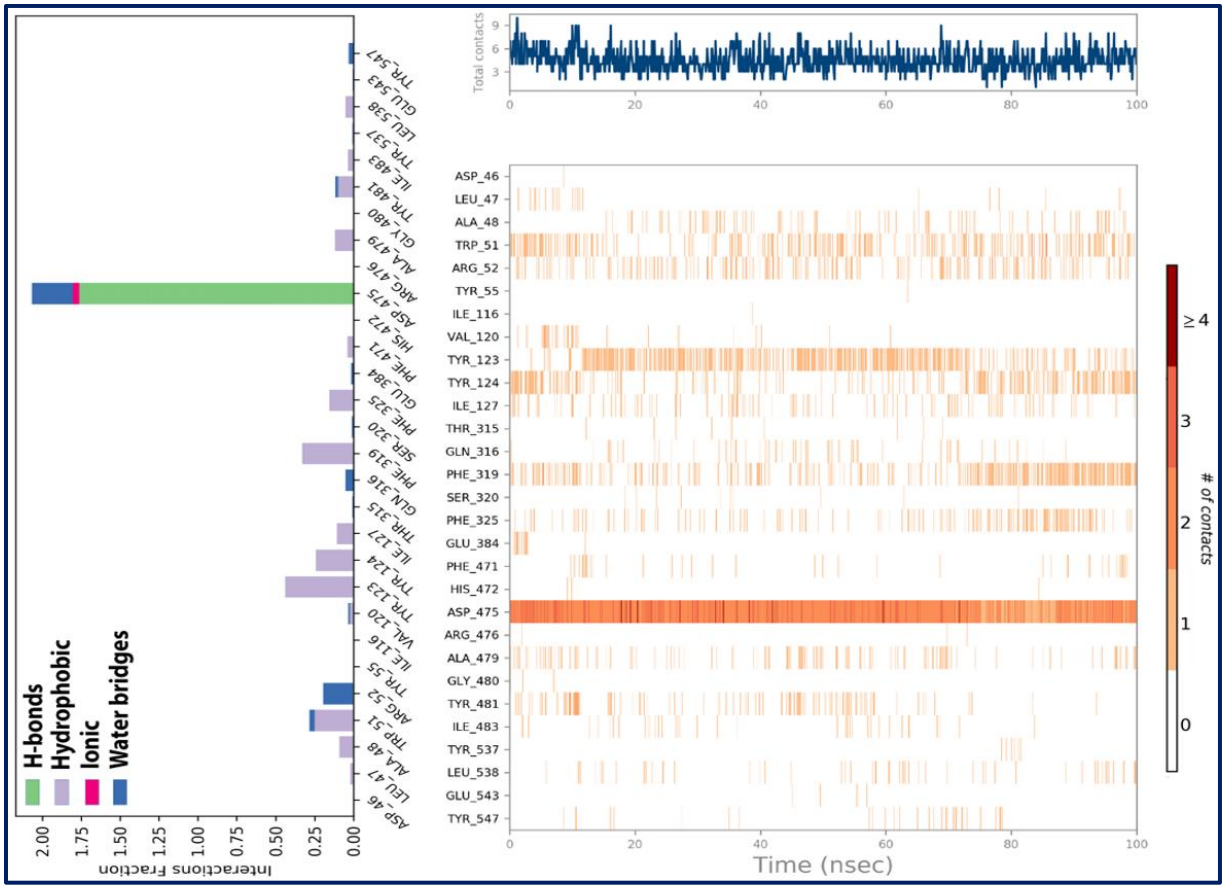



**Figure 51. Histogram of interaction fractions between DAT protein and C19 ligand (left), and the timeline depiction of intermolecular interactions and chemical contacts of candidate ligand in the active sites of targeted protein (right)**

The Molecular Mechanics with Generalized Born and Surface Area (MMGBSA) solvation was equally conducted to calculate the binding free energies of the C19 ligand in complex with DAT protein, as resulted in Table 35 [27,28]. The MMGBSA approach shows that the binding energies are negatively very large, indicating that the candidate drug have reacted with the protein target with low energies (minimal ΔG Bind scores), which makes the complex so stable with optimal energies during the molecular dynamic's simulation time.

**Table 35. Binding energies in Kcal/mol of the studied complex using MMGBSA approach.**

| Studied complex | MMGBSA ΔG Bind | MMGBSA ΔG Bind Coulomb | MMGBSA ΔG Bind Covalent | MMGBSA ΔG Bind Hbond | MMGBSA ΔG Bind Lipo | MMGBSA ΔG Bind Solv GB | MMGBSA ΔG Bind vdw |
|---|---|---|---|---|---|---|---|
| (C19 – protein) complex | **-51.417899** | **-56.518147** | **1.557139** | **-0.820922** | **-23.877425** | **73.360626** | **-45.119169** |

## 4. Conclusion

In summary, quantitative structure-activity relationships reveal that constitutional and geometric descriptors have a key function in the human GluT1 activity of the 3,4-disubstituted pyrrolidine sulfonamides family. In-silico screening indicates that C2, C3, C4, C5, and C19 were predicted as non-toxic inhibitors of the human body, and functioning as central nervous system agents with the highest probability to cross the blood-brain barrier, respecting Lipinski, Veber, Egan, Muegge, and Rhose conditions. The compound C19 having the highest activity and predicted with a good ADME-Tox profile, was chosen for molecular docking and molecular dynamics simulations , which indicate that this candidate ligand produced stable intermolecular interactions with ASP 475, GLU 480, VAL 120, ALA 479, and TYR 124 amino acids residues of dopamine transporter protein, because the studied properties of (C19 ligand – DAT protein) complex remained perfectly stable until 100 nanoseconds of molecular dynamic simulation time with lowest binding energies of MMGBSA solvation. Therefore, C19 ligand is strongly recommended as candidate drug to treat schizophrenia and similar disabilities in relation with glutaminergic N-methyl-D-aspartate receptor hypofunction.

**Références**

[1] A.R. Hudson, V.J. Santora, R.E. Petroski, T.A. Almos, G. Anderson, R. Barido, J. Basinger, C.L. Bellows, B.C. Bookser, N.J. Broadbent, C. Cabebe, C.-K. Chai, M. Chen, S. Chow, D.M. Chung, L. Heger, A.M. Danks, G.C. Freestone, D. Gitnick, V. Gupta, C. Hoffmaster, A.P. Kaplan, M.R. Kennedy, D. Lee, J. Limberis, K. Ly, C.C. Mak, B. Masatsugu, A.C. Morse, J. Na, D. Neul, J. Nikpur, J. Renick, K. Sebring, S. Sevidal, A. Tabatabaei, J. Wen, S. Xia, Y. Yan, Z.W. Yoder, D. Zook, M. Peters, J.G. Breitenbucher, Azetidine-based selective glycine transporter-1 (GlyT1) inhibitors with memory enhancing properties, Bioorganic & Medicinal Chemistry Letters. 30 (2020) 127214. https://doi.org/10.1016/j.bmcl.2020.127214.

[2] V.J. Santora, T.A. Almos, R. Barido, J. Basinger, C.L. Bellows, B.C. Bookser, J.G. Breitenbucher, N.J. Broadbent, C. Cabebe, C.-K. Chai, M. Chen, S. Chow, D.M. Chung, L. Crickard, A.M. Danks, G.C. Freestone, D. Gitnick, V. Gupta, C. Hoffmaster, A.R. Hudson, A.P. Kaplan, M.R. Kennedy, D. Lee, J. Limberis, K. Ly, C.C. Mak, B. Masatsugu, A.C. Morse, J. Na, D. Neul, J. Nikpur, M. Peters, R.E. Petroski, J. Renick, K. Sebring, S. Sevidal, A. Tabatabaei, J. Wen, Y. Yan, Z.W. Yoder, D. Zook, Design and Synthesis of Novel and Selective Glycine Transporter-1 (GlyT1) Inhibitors with Memory Enhancing Properties, J. Med. Chem. 61 (2018) 6018–6033. https://doi.org/10.1021/acs.jmedchem.8b00372.

[3] J.A. Lowe, X. Hou, C. Schmidt, F. David Tingley, S. McHardy, M. Kalman, S. DeNinno, M. Sanner, K. Ward, L. Lebel, D. Tunucci, J. Valentine, The discovery of a structurally novel class of inhibitors of the type 1 glycine transporter, Bioorganic & Medicinal Chemistry Letters. 19 (2009) 2974–2976. https://doi.org/10.1016/j.bmcl.2009.04.035.

[4] V. Eulenburg, W. Armsen, H. Betz, J. Gomeza, Glycine transporters: essential regulators of neurotransmission, Trends in Biochemical Sciences. 30 (2005) 325–333. https://doi.org/10.1016/j.tibs.2005.04.004.

[5] S. Wolkenberg, C. Sur, Recent Progress in the Discovery of Non-Sarcosine Based GlyT1 Inhibitors, CTMC. 10 (2010) 170–186. https://doi.org/10.2174/156802610790410974.

[6] S.E. Wolkenberg, Z. Zhao, D.D. Wisnoski, W.H. Leister, J. O'Brien, W. Lemaire, D.L. Williams, M.A. Jacobson, C. Sur, G.G. Kinney, D.J. Pettibone, P.R. Tiller, S. Smith, C. Gibson, B.K. Ma, S.L. Polsky-Fisher, C.W. Lindsley, G.D. Hartman, Discovery of GlyT1 inhibitors with improved pharmacokinetic properties, Bioorganic & Medicinal Chemistry Letters. 19 (2009) 1492–1495. https://doi.org/10.1016/j.bmcl.2009.01.015.

[7] B.N. Atkinson, S.C. Bell, M. De Vivo, L.R. Kowalski, S.M. Lechner, V.I. Ognyanov, C.-S. Tham, C. Tsai, J. Jia, D. Ashton, M.A. Klitenick, ALX 5407: A Potent, Selective Inhibitor of the hGlyT1 Glycine Transporter, Mol Pharmacol. 60 (2001) 1414–1420. https://doi.org/10.1124/mol.60.6.1414.

[8] D. Ouellet, S. Sutherland, T. Wang, P. Griffini, V. Murthy, First-Time-in-Human Study With GSK1018921, a Selective GlyT1 Inhibitor: Relationship Between Exposure and Dizziness, Clin Pharmacol Ther. 90 (2011) 597–604. https://doi.org/10.1038/clpt.2011.154.

[9] E. Pinard, A. Alanine, D. Alberati, M. Bender, E. Borroni, P. Bourdeaux, V. Brom, S. Burner, H. Fischer, D. Hainzl, R. Halm, N. Hauser, S. Jolidon, J. Lengyel, H.-P. Marty, T. Meyer, J.-L. Moreau, R. Mory, R. Narquizian, M. Nettekoven, R.D. Norcross, B. Puellmann, P. Schmid, S. Schmitt, H. Stalder, R. Wermuth, J.G. Wettstein, D. Zimmerli, Selective GlyT1 Inhibitors: Discovery of [4-(3-Fluoro-5-trifluoromethylpyridin-2-yl)piperazin-1-yl][5-methanesulfonyl-2-(( *S* )-2,2,2-trifluoro-1-methylethoxy)phenyl]methanone (RG1678), a Promising Novel

Medicine To Treat Schizophrenia, J. Med. Chem. 53 (2010) 4603–4614. https://doi.org/10.1021/jm100210p.
[10] W.P. Blackaby, R.T. Lewis, J.L. Thomson, A.S.R. Jennings, S.C. Goodacre, L.J. Street, A.M. MacLeod, A. Pike, S. Wood, S. Thomas, T.A. Brown, A. Smith, G. Pillai, S. Almond, M.R. Guscott, H.D. Burns, W. Eng, C. Ryan, J. Cook, T.G. Hamill, Identification of an Orally Bioavailable, Potent, and Selective Inhibitor of GlyT1, ACS Med. Chem. Lett. 1 (2010) 350–354. https://doi.org/10.1021/ml1001085.
[11] D. Alberati, J.-L. Moreau, J. Lengyel, N. Hauser, R. Mory, E. Borroni, E. Pinard, F. Knoflach, G. Schlotterbeck, D. Hainzl, J.G. Wettstein, Glycine reuptake inhibitor RG1678: A pharmacologic characterization of an investigational agent for the treatment of schizophrenia, Neuropharmacology. 62 (2012) 1152–1161. https://doi.org/10.1016/j.neuropharm.2011.11.008.
[12] W. Amberg, U.E.W. Lange, M. Ochse, F. Pohlki, B. Behl, A.L. Relo, W. Hornberger, C. Hoft, M. Mezler, J. Sydor, Y. Wang, H. Zhao, J.T. Brewer, J. Dietrich, H. Li, I. Akritopoulou-Zanze, Y. Lao, S.M. Hannick, Y.-Y. Ku, A. Vasudevan, Discovery of Novel Aminotetralines and Aminochromanes as Selective and Competitive Glycine Transporter 1 (GlyT1) Inhibitors, J. Med. Chem. 61 (2018) 7503–7524. https://doi.org/10.1021/acs.jmedchem.8b00300.
[13] K. Łątka, M. Bajda, Analysis of Binding Determinants for Different Classes of Competitive and Noncompetitive Inhibitors of Glycine Transporters, IJMS. 23 (2022) 8050. https://doi.org/10.3390/ijms23148050.
[14] E. Pinard, A. Alanine, D. Alberati, M. Bender, E. Borroni, P. Bourdeaux, V. Brom, S. Burner, H. Fischer, D. Hainzl, R. Halm, N. Hauser, S. Jolidon, J. Lengyel, H.-P. Marty, T. Meyer, J.-L. Moreau, R. Mory, R. Narquizian, M. Nettekoven, R.D. Norcross, B. Puellmann, P. Schmid, S. Schmitt, H. Stalder, R. Wermuth, J.G. Wettstein, D. Zimmerli, Selective GlyT1 Inhibitors: Discovery of [4-(3-Fluoro-5-trifluoromethylpyridin-2-yl)piperazin-1-yl][5-methanesulfonyl-2-(( *S* )-2,2,2-trifluoro-1-methylethoxy)phenyl]methanone (RG1678), a Promising Novel Medicine To Treat Schizophrenia, J. Med. Chem. 53 (2010) 4603–4614. https://doi.org/10.1021/jm100210p.
[15] H. Rosenbrock, M. Desch, G. Wunderlich, Development of the novel GlyT1 inhibitor, iclepertin (BI 425809), for the treatment of cognitive impairment associated with schizophrenia, Eur Arch Psychiatry Clin Neurosci. (2023). https://doi.org/10.1007/s00406-023-01576-z.
[16] Y. Wang, H. Zhao, J.T. Brewer, H. Li, Y. Lao, W. Amberg, B. Behl, I. Akritopoulou-Zanze, J. Dietrich, U.E.W. Lange, F. Pohlki, C. Hoft, W. Hornberger, S.W. Djuric, J. Sydor, M. Mezler, A.L. Relo, A. Vasudevan, De Novo Design, Synthesis, and Biological Evaluation of 3,4-Disubstituted Pyrrolidine Sulfonamides as Potent and Selective Glycine Transporter 1 Competitive Inhibitors, J. Med. Chem. 61 (2018) 7486–7502. https://doi.org/10.1021/acs.jmedchem.8b00295.
[17] S. Vilar, G. Cozza, S. Moro, Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery, CTMC. 8 (2008) 1555–1572. https://doi.org/10.2174/156802608786786624.
[18] N. Mostoufi, A. Constantinides, Linear and nonlinear regression analysis, in: Applied Numerical Methods for Chemical Engineers, Elsevier, 2023: pp. 403–476. https://doi.org/10.1016/B978-0-12-822961-3.00008-X.

[19] S. Tian, J. Wang, Y. Li, D. Li, L. Xu, T. Hou, The application of in silico drug-likeness predictions in pharmaceutical research, Advanced Drug Delivery Reviews. 86 (2015) 2–10. https://doi.org/10.1016/j.addr.2015.01.009.
[20] A. Serrano, B. Imbernón, H. Pérez-Sánchez, J.M. Cecilia, A. Bueno-Crespo, J.L. Abellán, QN-Docking: An innovative molecular docking methodology based on Q-Networks, Applied Soft Computing. 96 (2020) 106678. https://doi.org/10.1016/j.asoc.2020.106678.
[21] R.P.D. Bank, RCSB PDB - 4M48: X-ray structure of dopamine transporter elucidates antidepressant mechanism, (n.d.). https://www.rcsb.org/structure/4M48 (accessed January 6, 2023).
[22] M. El fadili, M. Er-rajy, H. Imtara, M. Kara, S. Zarougui, N. Altwaijry, O. Al kamaly, A. Al Sfouk, M. Elhallaoui, 3D-QSAR, ADME-Tox In Silico Prediction and Molecular Docking Studies for Modeling the Analgesic Activity against Neuropathic Pain of Novel NR2B-Selective NMDA Receptor Antagonists, Processes. 10 (2022) 1462. https://doi.org/10.3390/pr10081462.
[23] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct Biol. 9 (2002) 646–652. https://doi.org/10.1038/nsb0902-646.
[24] M. El fadili, M. Er-Rajy, M. Kara, A. Assouguem, A. Belhassan, A. Alotaibi, N.N. Mrabti, H. Fidan, R. Ullah, S. Ercisli, S. Zarougui, M. Elhallaoui, QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia, Pharmaceuticals. 15 (2022) 670. https://doi.org/10.3390/ph15060670.
[25] S.A. Hollingsworth, R.O. Dror, Molecular Dynamics Simulation for All, Neuron. 99 (2018) 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.
[26] X. Hu, Z. Zeng, J. Zhang, D. Wu, H. Li, F. Geng, Molecular dynamics simulation of the interaction of food proteins with small molecules, Food Chemistry. 405 (2023) 134824. https://doi.org/10.1016/j.foodchem.2022.134824.
[27] S. Faisal, S.L. Badshah, B. Kubra, M. Sharaf, A.-H. Emwas, M. Jaremko, M. Abdalla, Identification and Inhibition of the Druggable Allosteric Site of SARS-CoV-2 NSP10/NSP16 Methyltransferase through Computational Approaches, Molecules. 27 (2022) 5241. https://doi.org/10.3390/molecules27165241.
[28] C.J. Ononamadu, M. Abdalla, G.O. Ihegboro, J. Li, T.A. Owolarafe, T.D. John, Q. Tian, In silico identification and study of potential anti-mosquito juvenile hormone binding protein (MJHBP) compounds as candidates for dengue virus - Vector insecticides, Biochemistry and Biophysics Reports. 28 (2021) 101178. https://doi.org/10.1016/j.bbrep.2021.101178.
[29] T. Österberg, U. Norinder, Prediction of drug transport processes using simple parameters and PLS statistics The use of ACD/logP and ACD/ChemSketch descriptors, European Journal of Pharmaceutical Sciences. 12 (2001) 327–337. https://doi.org/10.1016/S0928-0987(00)00189-5.
[30] G.W.A. Milne, Software Review of ChemBioDraw 12.0, J. Chem. Inf. Model. 50 (2010) 2053–2053. https://doi.org/10.1021/ci100385n.
[31] C. Guerra Tort, V. Aguiar Pulido, V. Suárez Ulloa, F. Docampo Boedo, J.M. López Gestal, J. Pereira Loureiro, Electronic Health Records Exploitation Using Artificial Intelligence Techniques, in: 3rd XoveTIC Conference, MDPI, 2020: p. 60. https://doi.org/10.3390/proceedings2020054060.

[32] D. Groth, S. Hartmann, S. Klie, J. Selbig, Principal Components Analysis, Computational Toxicology. (2013) 527–547. https://doi.org/10.1007/978-1-62703-059-5_22.
[33] A. Maćkiewicz, W. Ratajczak, Principal components analysis (PCA), Computers & Geosciences. 19 (1993) 303–342. https://doi.org/10.1016/0098-3004(93)90090-R.
[34] M. Er-Rajy, M. El Fadili, N.N. Mrabti, S. Zarougui, M. Elhallaoui, QSAR, molecular docking, ADMET properties in silico studies for a series of 7-propanamide benzoxaboroles as potent anti-cancer agents, Chinese Journal of Analytical Chemistry. 50 (2022) 100163. https://doi.org/10.1016/j.cjac.2022.100163.
[35] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews. 23 (1997) 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.
[36] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.
[37] W.J. Egan, K.M. Merz,, J.J. Baldwin, Prediction of Drug Absorption Using Multivariate Statistics, J. Med. Chem. 43 (2000) 3867–3877. https://doi.org/10.1021/jm000292e.
[38] W.J. Egan, G. Lauri, Prediction of intestinal permeability, Advanced Drug Delivery Reviews. 54 (2002) 273–289. https://doi.org/10.1016/S0169-409X(02)00004-2.
[39] I. Muegge, S.L. Heald, D. Brittelli, Simple Selection Criteria for Drug-like Chemical Matter, J. Med. Chem. 44 (2001) 1841–1846. https://doi.org/10.1021/jm015507e.
[40] D.E.V. Pires, T.L. Blundell, D.B. Ascher, pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures, J. Med. Chem. 58 (2015) 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104.
[41] SwissADME, (n.d.). http://www.swissadme.ch/ (accessed January 7, 2023).
[42] D. Bassani, M. Pavan, G. Bolcato, M. Sturlese, S. Moro, Re-Exploring the Ability of Common Docking Programs to Correctly Reproduce the Binding Modes of Non-Covalent Inhibitors of SARS-CoV-2 Protease Mpro, Pharmaceuticals. 15 (2022) 180. https://doi.org/10.3390/ph15020180.
[43] R.P.D. Bank, RCSB PDB - 4M48: X-ray structure of dopamine transporter elucidates antidepressant mechanism, (n.d.). https://www.rcsb.org/structure/4M48 (accessed January 7, 2023).
[44] A. Kouranov, The RCSB PDB information portal for structural genomics, Nucleic Acids Research. 34 (2006) D302–D305. https://doi.org/10.1093/nar/gkj120.
[45] D. Systèmes, Free Download: BIOVIA Discovery Studio Visualizer, Dassault Systèmes. (2020). https://discover.3ds.com/discovery-studio-visualizer-download (accessed January 7, 2023).
[46] A.P. Norgan, P.K. Coffman, J.-P.A. Kocher, D.J. Katzmann, C.P. Sosa, Multilevel Parallelization of AutoDock 4.2, J Cheminform. 3 (2011) 12. https://doi.org/10.1186/1758-2946-3-12.
[47] J. Pyrhönen, K.K. Bansal, R. Bhadane, C.-E. Wilén, O.M.H. Salo-Ahen, J.M. Rosenholm, Molecular Dynamics Prediction Verified by Experimental Evaluation of the Solubility of Different Drugs in Poly(decalactone) for the Fabrication of Polymeric Nanoemulsions, Advanced NanoBiomed Research. 2 (2022) 2100072. https://doi.org/10.1002/anbr.202100072.

[48] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B. 105 (2001) 6474–6487. https://doi.org/10.1021/jp003919d.
[49] A. Bastianoni, E. Guastaldi, A. Barbagli, S. Bernardinetti, A. Zirulia, M. Brancale, T. Colonna, Multivariate Analysis Applied to Aquifer Hydrogeochemical Evaluation: A Case Study in the Coastal Significant Subterranean Water Body between "Cecina River and San Vincenzo", Tuscany (Italy), Applied Sciences. 11 (2021) 7595. https://doi.org/10.3390/app11167595.
[50] G.K. Uyanık, N. Güler, A Study on Multiple Linear Regression Analysis, Procedia - Social and Behavioral Sciences. 106 (2013) 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027.
[51] N. Kravić, J. Savosina, M. Agafonova-Moroz, V. Babain, A. Legin, D. Kirsanov, Nonlinear Multivariate Regression Algorithms for Improving Precision of Multisensor Potentiometry in Analysis of Spent Nuclear Fuel Reprocessing Solutions, Chemosensors. 10 (2022) 90. https://doi.org/10.3390/chemosensors10030090.
[52] K.B. de C. Rezende, A.J.L.A. da Cunha, J. Amim Junior, R.G. Bornia, External validation of the Fetal Medicine Foundation algorithm for the prediction of preeclampsia in a Brazilian population, Pregnancy Hypertension. 17 (2019) 64–68. https://doi.org/10.1016/j.preghy.2019.05.006.
[53] M. Rafało, Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis, ICT Express. 8 (2022) 183–188. https://doi.org/10.1016/j.icte.2021.05.001.
[54] M. El fadili, M. Er-rajy, H. Imtara, O.M. Noman, R.A. Mothana, S. Abdullah, S. Zerougui, M. Elhallaoui, QSAR, ADME-Tox, molecular docking and molecular dynamics simulations of novel selective glycine transporter type 1 inhibitors with memory enhancing properties, Heliyon. (2023) e13706. https://doi.org/10.1016/j.heliyon.2023.e13706.
[55] A. Golbraikh, A. Tropsha, Beware of q2!, Journal of Molecular Graphics and Modelling. 20 (2002) 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1.
[56] A.K. Halder, T. Jha, Validated predictive QSAR modeling of N-aryl-oxazolidinone-5-carboxamides for anti-HIV protease activity, Bioorganic & Medicinal Chemistry Letters. 20 (2010) 6082–6087. https://doi.org/10.1016/j.bmcl.2010.08.050.
[57] A. Daina, V. Zoete, A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules, ChemMedChem. 11 (2016) 1117–1121. https://doi.org/10.1002/cmdc.201600182.
[58] M. Er-rajy, M. El Fadili, H. Hadni, N.N. Mrabti, S. Zarougui, M. Elhallaoui, 2D-QSAR modeling, drug-likeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives, Struct Chem. 33 (2022) 973–986. https://doi.org/10.1007/s11224-022-01913-3.
[59] C.-Y. Jia, J.-Y. Li, G.-F. Hao, G.-F. Yang, A drug-likeness toolbox facilitates ADMET study in drug discovery, Drug Discovery Today. 25 (2020) 248–258. https://doi.org/10.1016/j.drudis.2019.10.014.
[60] A. Adamski, D. Kruszka, Z. Dutkiewicz, M. Kubicki, A. Gorczyński, V. Patroniak, Novel family of fused tricyclic [1,4]diazepines: Design, synthesis, crystal structures and molecular docking studies, Tetrahedron. 73 (2017) 3377–3386. https://doi.org/10.1016/j.tet.2017.05.015.

[61] A. Belhassan, H. Zaki, M. Benlyas, T. Lakhlifi, M. Bouachrine, Study of novel triazolo-benzodiazepine analogues as antidepressants targeting by molecular docking and ADMET properties prediction, Heliyon. 5 (2019) e02446. https://doi.org/10.1016/j.heliyon.2019.e02446.
[62] Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server, (n.d.). https://proteins.plus/ (accessed January 15, 2023).
[63] M. Abdullahi, G.A. Shallangwa, A. Uzairu, In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of H5N1 influenza A virus subtype, Beni-Suef Univ J Basic Appl Sci. 9 (2020) 2. https://doi.org/10.1186/s43088-019-0023-y.
[64] M. Er-rajy, M. El fadili, H. Imtara, A. Saeed, A. Ur Rehman, S. Zarougui, S.A. Abdullah, A. Alahdab, M.K. Parvez, M. Elhallaoui, 3D-QSAR Studies, Molecular Docking, Molecular Dynamic Simulation, and ADMET Proprieties of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer, Life. 13 (2023) 127. https://doi.org/10.3390/life13010127.
[65] A. AlAmiery, Gravimetric and Density Functional Theory Investigations on 4-Amioantipyrin Schiff base as an inhibitor for mild steel in HCl solution, Prog. Color Colorants Coat. (2023). https://doi.org/10.30509/pccc.2023.167077.1196.
[66] S. TANWER, S. Shukla, Cefuroxime: A Potential Corrosion Inhibitor for Mild Steel in Sulphuric Acid Medium, Prog. Color Colorants Coat. 16 (2023). https://doi.org/10.30509/pccc.2022.166974.1165.
[67] O. Ebenezer, N. Damoyi, M.A. Jordaan, M. Shapi, Unveiling of Pyrimidindinones as Potential Anti-Norovirus Agents—A Pharmacoinformatic-Based Approach, Molecules. 27 (2022) 380. https://doi.org/10.3390/molecules27020380.

# Conclusion générale

Répondant aux principaux objectifs de la présente thèse, visant à découvrir des composés chimiques biologiquement actifs, prédits en tant qu'agents puissants du système nerveux central (SNC) avec des profils ADMET favorables, capables de fournir des propriétés thérapeutiques efficaces similaires à celles des médicaments candidats destinés au traitement des perturbations et d'hypofonctionnement des récepteurs NMDA, sans engendrer d'effets secondaires d'hépatotoxicité et d'allergie cutanée sur le corps humain, tout en produisant des complexes présentant un excellent niveau de stabilité moléculaire envers les sites actifs de la cible protéique, avec des affinités de liaison largement négatives en ordre énergétique, nous avons appliqué nos investigations In-silico assistées par ordinateur, sur des bases de données purement expérimentales comprenant des séries de molécules bioactives, synthétisées organiquement, et évaluées biologiquement par des tests expérimentaux In-vivo et In-vitro.

❖ Les premières investigations ont été effectuées au moyen des études tridimensionnels des relations quantitatives structure-activité (3D-QSAR) sur une série de trente-deux inhibiteurs du sous unité GluN2B du récepteur NMDA, présentant des activités analgésiques contre la douleur neuropathique, à partir lesquelles nous sommes arrivé que les champs stériques, électrostatique, et accepteurs de liaison hydrogène ont un effet significatif sur l'activité analgésique contre la douleur neuropathique, justifiée par des effets favorables du groupe [Oxazol-2(3H)-one] et des groupements atomiques de type halogène en positions méta et para du cycle aromatique des dérivés de l'isoxazole, comme il était expliqué respectivement par des modèles prédictifs tridimensionnels, à savoir l'analyse comparative des champs moléculaires (CoMFA), et l'analyse comparative de l'indice de similarité moléculaire (CoMSIA). La deuxième évaluation a été réalisée sur la base des cinq règles de Lipinski, ainsi que les règles de violation de Muegge, Egan, Ghose, et Veber, qui ont été toutes satisfaites pour les ligands L1, L2, et L3, prédits avec des propriétés physicochimiques et pharmacocinétiques ADMET appropriées, en tant qu'agents non toxiques du système nerveux central avec la plus grande capacité à traverser la barrière hémato-encéphalique (BBB). Les résultats obtenus ont été renforcés par des simulations du Docking moléculaire, qui ont révélé que les isoxazoles L1, L2 et L3 réagissent spécifiquement avec les acides aminés Phe176, Glu235, Glu236, Gln110, Asp136 et Glu178 de la protéine de transport codée par 3QEL.pdb, ces interactions intermoléculaires montrent que les ligands candidats ont été dockés dans les trois sites actifs (Phe176, Glu236, et Gln110) des récepteurs NMDA en sous unités GluN1 et GluN2B, en chaine D dans laquelle se complexe le ligand co-crystallisé (Ifenprodil). Par conséquent, ils peuvent être appliqués comme thérapeutiques pour prévenir la douleur neuropathique.

❖ Les deuxièmes investigations ont été appliquées sur trente antagonistes des sous-unités GluN2B du récepteur NMDA pénétrant dans le cerveau comme agents thérapeutiques contre les accidents vasculaires cérébraux (AVC), à travers lesquelles nous sommes arrivés que les composés chimiques C22 et C13 ont été testés avec succès en tant qu'inhibiteurs non toxiques des cytochromes CYP2D6 et CYP3A4, avec un bon profil ADMET, et conçus pour être des agents efficaces pour le système nerveux central en raison de la plus grande probabilité à traverser la barrière hémato-encéphalique, ils répondent aux nombres de violations souhaitées d'Egan, Veber, Rhose et Muegge, ainsi qu'ils respectent les cinq règles de Lipinski, en formant des interactions intermoléculaires stables avec les résidus d'acides aminés de la protéine de transport codée par 3QEL.pdb. Les résultats du Docking moléculaire ont été renforcés par les simulations de la dynamique moléculaire couplée aux approches MM-GBSA, qui ont révélé des niveaux acceptables de la stabilité moléculaire le long de 200 nanosecondes du temps de simulation. Par conséquent, nous recommandons vivement les composés chimiques appelés C22 et C13 en tant que puissants médicaments contre les accidents vasculaires cérébraux.

❖ Les troisièmes investigations ont été appliquées sur quarante-quatre dérivés de bicyclo((aryl)méthyl) benzamide en tant qu'inhibiteurs du transporteur de glycine de type 1 (GlyT1) afin de découvrir des candidats antipsychotiques efficaces pour le traitement de la schizophrénie. Dans un premier temps, deux modèles QSAR ont été développés à l'aide des techniques MLR et MNLR et ont été examinés par validation externe et interne, domaine d'applicabilité (AD), test de randomisation en Y et critères de Golbreikh et Tropsha, indiquant un effet significatif du donneur de liaison hydrogène, de la polarisabilité, de la tension superficielle, des énergies d'étirement et de torsion et du diamètre topologique sur l'activité locomotrice (LMA). Par la suite, la prédiction pharmacocinétique ADMET in silico a révélé un profil favorable des ligands les plus actifs, où L6, L9, L30, L31 et L37 ont été prédits comme des inhibiteurs non toxiques pour le cytochrome 2D6, qui respectent les règles de Lipinski, Veber, Egan et Ghose, avec une excellente absorption dépassant 91% et la plus grande probabilité de pénétrer dans le système nerveux central (SNC). En revanche, le ligand L32, en tant que valeur aberrante dans l'étude QSAR, présente un profil ADMET défavorable avec la probabilité la plus élevée d'être absorbé par le tractus gastro-intestinal. Enfin, les résultats obtenus ont été renforcés et nuancés par des études d'amarrage et de dynamique moléculaire, qui confirment que les ligands L6, L9, L30, L31 et L37 réagissent spécifiquement avec les acides aminés Tyr124, Phe43, Phe325, Asp46, Phe319 et Val120 de la protéine membranaire du transporteur de dopamine (DAT) d'une manière qui bloque le transporteur de glycine de type 1 (GlyT1) en formant des complexes dynamiquement stables pendant 100 ns de temps de

simulation de la dynamique moléculaire. Ils pourraient donc être utilisés en médecine pour traiter la schizophrénie. Toutefois, ils doivent être soumis à des études in vitro et in vivo pour évaluer leur efficacité et leur sécurité en tant que médicaments contre la schizophrénie.

❖ Les investigations suivantes ont été réalisés sur une famille structurelle de quarante dérivés de pyrrolo [3,4-c] pyrazole et d'azétidine, agissant en tant qu'inhibiteurs sélectifs du transporteur de glycine type 1 (GlyT1), afin d'identifier les inhibiteurs puissants de GluT1 pour le traitement des perturbations de la mémoire liés aux récepteurs NMDA.

Tout d'abord, les études bidimensionnelles des relations quantitatives structure-activité (2D-QSAR), ont révélé que l'activité inhibitrice humaine du GluT1 est fortement et significativement corrélée avec des descripteurs constitutionnels, géométriques, physicochimiques et topologiques, d'après des modèles prédictifs linéaires et non linéaires strictement validés, conformément aux critères statistiques inclus dans l'étude de Golbreikh et Tropsha. De plus, les résultats de prédiction des propriétés physicochimiques et pharmacocinétiques ADMET, ont montré que les ligands L28 et L30 font des ressemblances importantes aux candidats médicaments, car ils respectent les seuils critiques postulés par Lipinski, ainsi qu'ils répondent aux conditions de violations proposées par Egan, Veber, Muegge, et Ghose. En outre, ils offrent des profils ADMET intéressants, dépourvus de toute hépatotoxicité ou allergie cutanée, ce qui les distingue des autres inhibiteurs actifs par leur capacité à pénétrer la barrière hémato-encéphalique, et sont donc de puissants agents du SNC.

Par ailleurs, les simulations du Docking moléculaire ont également confirmé que les deux petites molécules produisent des interactions intermoléculaires envers la cible protéique du transporteur de dopamine (DAT), faisant intervenir, en plus des résidus d'acides aminés : Phe319, Tyr123, Arg52, Asp475, Ala117, Ala479, Ile116 et Ile483, les sites actifs de protéine membranaire (DAT) suivants : Phe325, Phe43 et Tyr124, qui ont été détectés de manière similaire vers le ligand co-crystallisé (Nortriptyline). Ce type de liaisons chimiques a été examiné en termes de stabilité à l'échelle atomique par des simulations de la dynamique moléculaire couplée aux approches MM-GBSA, qui a montré des niveaux très élevés de stabilité des complexes moléculaires produits (L28, L30 -DAT) durant 50 nanosecondes du temps de simulation. Par conséquent, les composés chimiques étiquetés L28 et L30, ont été fortement recommandés pour le traitement des perturbations de la mémoire liés à l'hypofonctionnement des récepteurs NMDA.

❖ Les dernières investigations ont été appliquées sur une classe structurelle de trente dérivés de pyrrolidine sulfonamides 3,4-disubstitués, agissant en tant qu'inhibiteurs sélectifs et compétitifs du GlyT1, en vue de découvrir d'inhibiteurs efficaces du transporteur de glycine

type 1, qui pourraient introduire des traitements de schizophrénie et d'autres déficiences associées à l'hypofonctionnement des récepteurs NMDA.

Au début, les études des relations quantitatives établies entre la structure d'inhibiteurs examinés et leur activités biologiques (QSAR), ont montré que les descripteurs moléculaires de type géométriques et constitutionnels exercent des effets significatifs sur l'activité inhibitrice humaine de GlyT1, expliquées par des modèles mathématiques linéaires et non linéaires, rigoureusement certifiés sur la base d'étude statistique de Golbreikh et Tropsha.

Le composé marqué C19, a été évalué suffisamment en raison de leur activité biologique la plus grande par rapport aux autres inhibiteurs. Heureusement, le composé examiné fait des ressemblances significatives aux médicaments candidats, pour les raisons suivantes :

- ✓ Tout d'abord, leurs propriétés physicochimiques répondent aux cinq règles de Lipinski, et respectent les nombres de violation désirés de Veber, Egan, Muegge, et Ghose.
- ✓ Ensuite, les prédictions des propriétés pharmacocinétiques d'ADME, et de toxicité ont été bien désirées par rapport au ligand natif (Nortriptyline).
- ✓ Par ailleurs, le modèle prédictif d'œuf d'Egan, a montré que le composé étiqueté C19, a été prédit en tant qu'agent puissant du système nerveux central, avec une probabilité significative de traverser la barrière hémato-encéphalique.
- ✓ En outre, le mécanisme d'inhibition du ligand candidat a montré la présence des interactions intermoléculaires envers la cible protéique qui font intervenir des effets stabilisants comme des liaisons hydrogènes détectées avec les résidus d'acides aminés Asp475, et Gly480, en plus d'autres familles des liaisons chimiques incluant des sites actifs de la cible protéique comme résultant d'après la simulation du Docking moléculaire avec une affinité de liaison largement négative (-8.96 Kcal/mol).
- ✓ Enfin, le composé chimique C19 présente un haut niveau de stabilité moléculaire durant 100 ns de temps de simulation de dynamique moléculaire, ce qui est renforcé par les calculs des énergies de liaison au moyen de l'approche MM-GBSA, qui ont à leur tour révélé des valeurs largement négatives pour les affinités de liaison en kcal/mol.

❖ A travers ces cinq applications, on est arrivé aux recommandations suivantes :

I. Les ligands **L1**, **L2**, et **L3** peuvent être appliqués comme thérapeutiques pour prévenir la douleur neuropathique.

II. Les composés chimiques étiquetés **C22 et C13** sont vivement recommandés en tant que puissants candidats-médicaments contre les accidents vasculaires cérébraux.

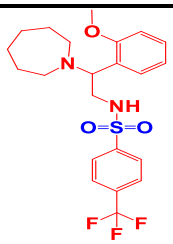
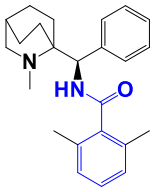
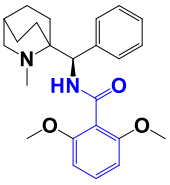
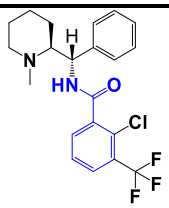
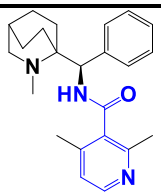
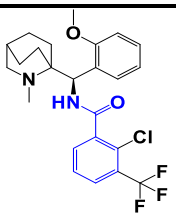
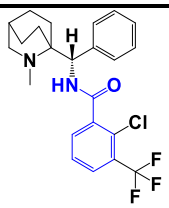
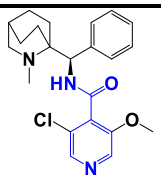
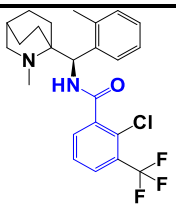
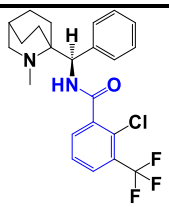
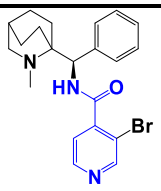
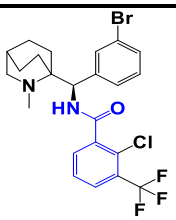
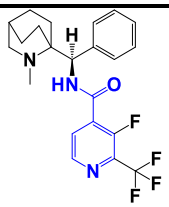
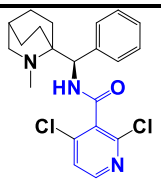
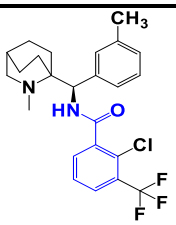
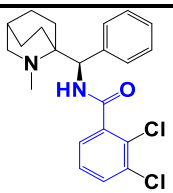
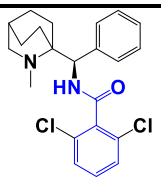
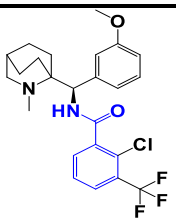
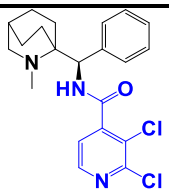
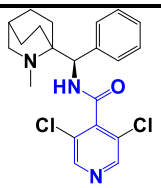
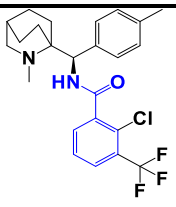
III. Les ligands marqués par **L6, L9, L30, L31 et L37** pourraient être utilisés en médecine pour traiter la schizophrénie.

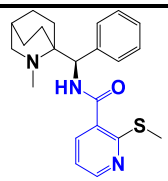
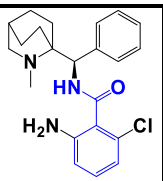
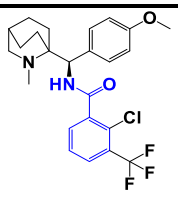
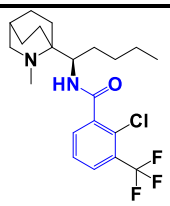
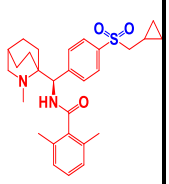
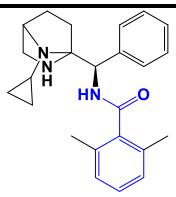
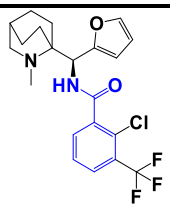
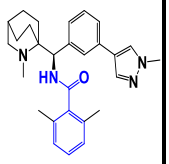
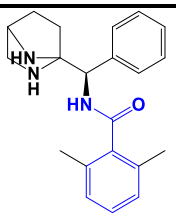
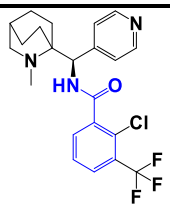
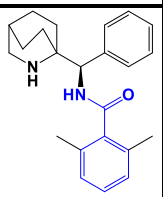
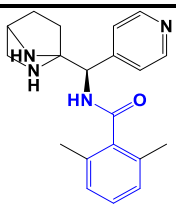
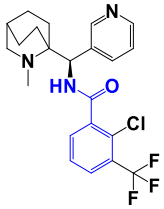
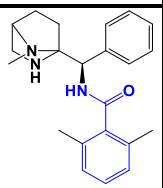
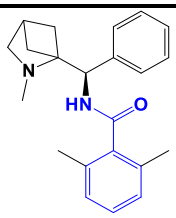
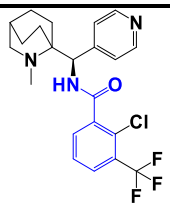
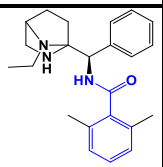
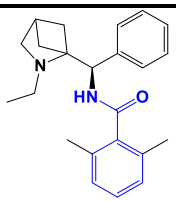
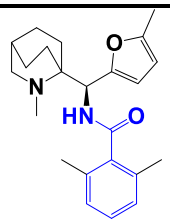
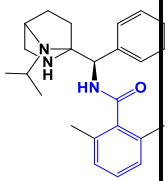
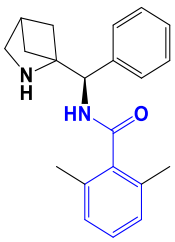
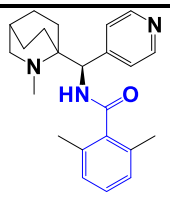
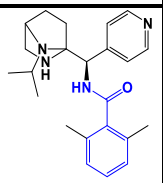
IV. Les ligands **L28 et L30**, sont fortement recommandés pour le traitement des perturbations de la mémoire liés à l'hypofonctionnement des récepteurs NMDA.

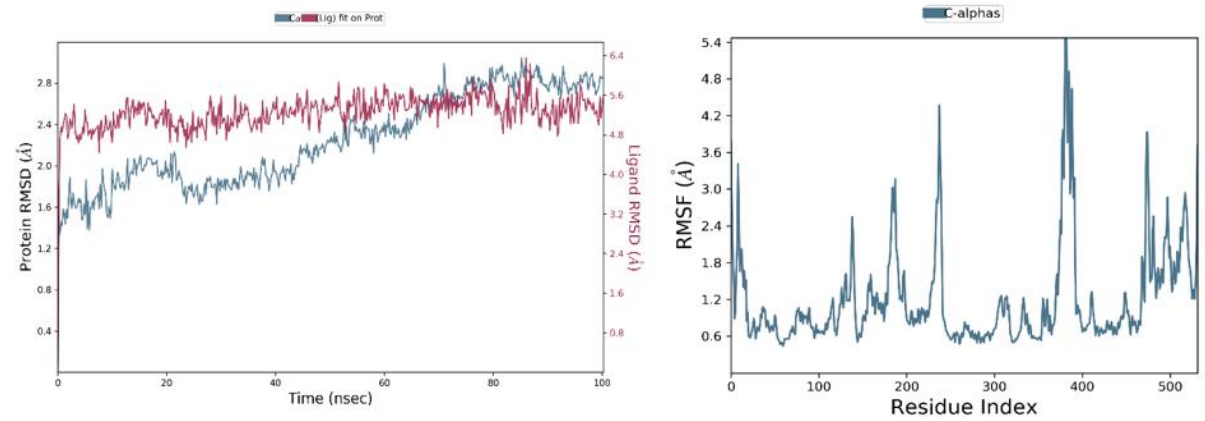
V. Le composé chimique **C19** est fortement conseillé pour le traitement de schizophrénie et d'autres déficiences associées à l'hypofonctionnement des récepteurs NMDA.

# Annexes

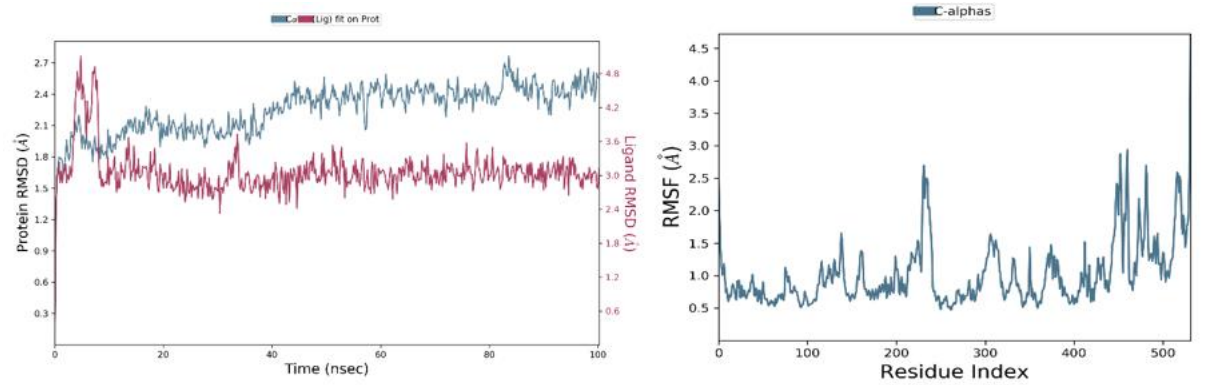
**Table S1. Forty-four molecules and their biological activities.**

| N° | Structure | $IC_{50}$ | $Log_{10}IC_{50}$ | N° | Structure | $IC_{50}$ | $Log_{10}IC_{50}$ | N° | Structure | $IC_{50}$ | $Log_{10}IC_{50}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 |  | 37 | 1.568202 | 9 |  | 1 | 0 | 17 |  | 6 | 0.778151 |
| 2 |  | 3 | 0.477121 | 10 |  | 494 | 2.69373 | 18 |  | 340 | 2.531479 |
| 3 |  | 41 | 1.612784 | 11 |  | 474 | 2.67578 | 19 |  | 514 | 2.710963 |
| 4 |  | 6 | 0.778151 | 12 |  | 75 | 1.87506 | 20 |  | 9 | 0.954242 |
| 5 |  | 92 | 1.963788 | 13 |  | 809 | 2.90795 | 21 |  | 29 | 1.462398 |
| 6 |  | 1 | 0 | 14 |  | 23 | 1.36173 | 22 |  | 177 | 2.247973 |
| 7 |  | 6 | 0.778151 | 15 |  | 45 | 1.65321 | 23 |  | 1703 | 3.231215 |

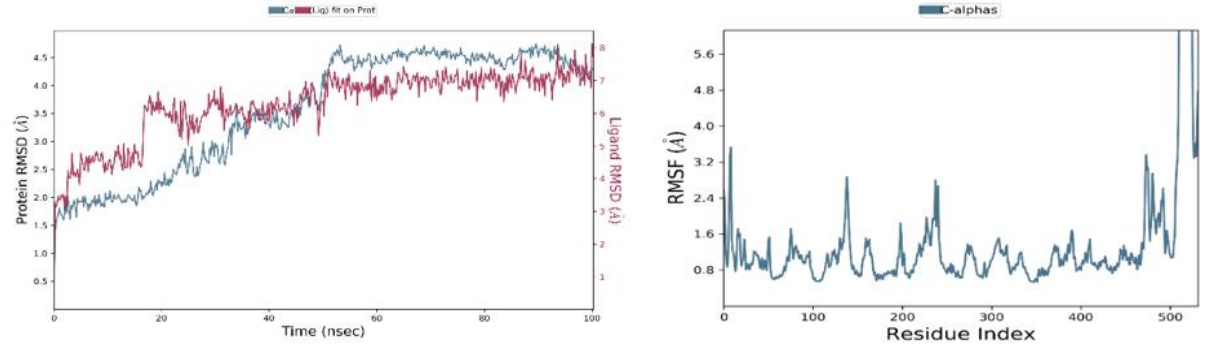
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 |  | 3 | 0.477121 | 16 |  | 8 | 0.90309 | 24 |  | 1759 | 3.245266 |
| 25 |  | 3362 | 3.526598 | 32 |  | 1 | 0 | 39 |  | 2 | 0.301030 |
| 26 |  | 78 | 1.892095 | 33 |  | 946 | 2.97589 | 40 |  | 58 | 1.763428 |
| 27 |  | 18 | 1.255272 | 34 |  | 3 | 0.47712 | 41 |  | 45 | 1.653212 |
| 28 |  | 211 | 2.324282 | 35 |  | 3 | 0.47712 | 42 |  | 7 | 0.845098 |
| 29 |  | 904 | 2.956168 | 36 |  | 4 | 0.60206 | 43 |  | 3 | 0.477121 |
| 30 |  | 1 | 0 | 37 |  | 1 | 0 | 44 |  | 168 | 2.225309 |
| 31 |  | 1 | 0 | 38 |  | 10 | 1 | | | | |



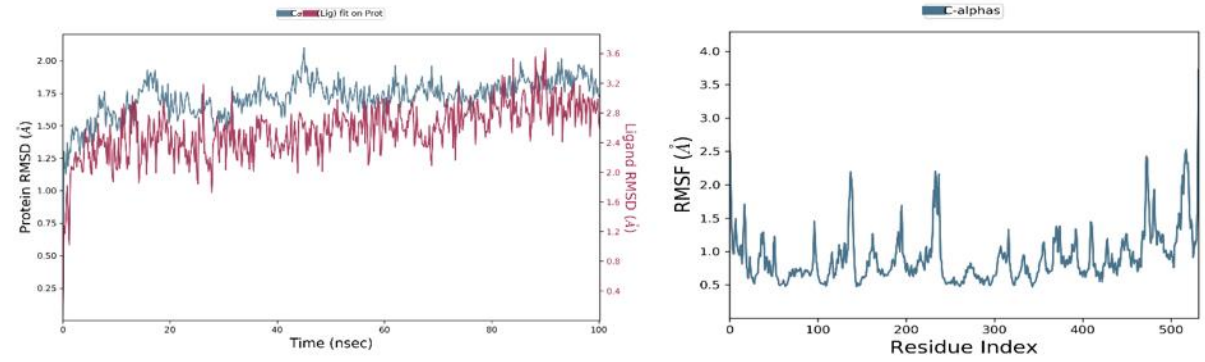

**C1**




**C3**

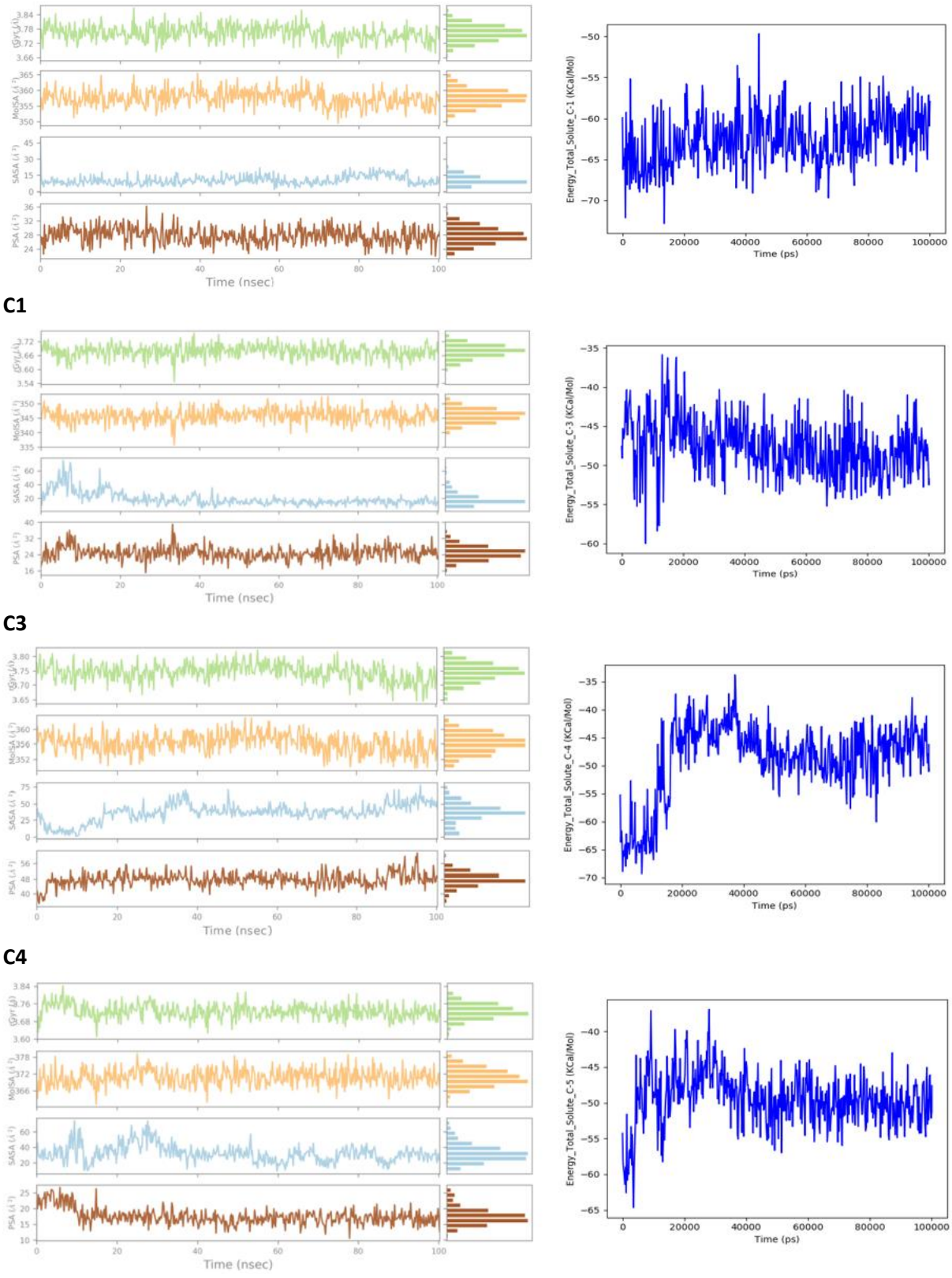



**C4**




**C5**

**Figure S1. RMSD and RMSF graphs for L6, L30, L31, L37 ligands complexed with the dopamine transporter membrane protein during 100 ns.**




**C1**

**C3**

**C4**

**C5**

**Figure S2. Rg, MolSA, SASA and PSA during 100 ns of MD simulation, and the variation of total free energy for L6, L30, L31 and L37 ligands complexed with DAT protein.**

**Table S2. The calculations of various molecular descriptors**

| N° | %H | %C | %N | %O | %S | %F | %Cl | %Br | æ | P |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.96 | 57.88 | 6.14 | 10.51 | 7.02 | 12.48 | 0 | 0 | 45.25 | 921.3 |
| 2 | 5.4 | 61.39 | 6.82 | 3.89 | 0 | 13.87 | 8.63 | 0 | 41.05 | 823.3 |
| 3 | 5.54 | 63.23 | 6.41 | 3.66 | 0 | 13.05 | 8.11 | 0 | 43.9 | 870.2 |
| 4 | 5.54 | 63.23 | 6.41 | 3.66 | 0 | 13.05 | 8.11 | 0 | 43.9 | 870.2 |
| 5 | 5.5 | 62.7 | 9.97 | 3.8 | 0 | 18.03 | 0 | 0 | 41.96 | 836 |
| 6 | 6 | 65.51 | 6.95 | 3.97 | 0 | 0 | 17.58 | 0 | 43.87 | 845.4 |
| 7 | 5.73 | 62.38 | 10.39 | 3.96 | 0 | 0 | 17.54 | 0 | 43.11 | 840.9 |
| 8 | 7.13 | 69.26 | 11.01 | 4.19 | 8.4 | 0 | 0 | 0 | 44.29 | 857.3 |
| 9 | 8.34 | 79.52 | 7.73 | 4.41 | 0 | 0 | 0 | 0 | 43.81 | 847.6 |
| 10 | 8.04 | 76 | 11.56 | 4.4 | 0 | 0 | 0 | 0 | 43.05 | 843.2 |
| 11 | 6.55 | 66.07 | 10.51 | 8 | 0 | 0 | 8.87 | 0 | 43.82 | 862.4 |
| 12 | 5.84 | 60.87 | 10.14 | 3.86 | 0 | 0 | 0 | 19.28 | 42.28 | 817.7 |
| 13 | 5.73 | 62.38 | 10.39 | 3.96 | 0 | 0 | 17.54 | 0 | 43.11 | 840.9 |
| 14 | 6 | 65.51 | 6.95 | 3.97 | 0 | 0 | 17.58 | 0 | 43.87 | 845.4 |
| 15 | 5.73 | 62.38 | 10.39 | 3.96 | 0 | 0 | 17.54 | 0 | 43.11 | 840.9 |
| 16 | 6.83 | 68.83 | 10.95 | 4.17 | 0 | 0 | 9.23 | 0 | 43.6 | 836.2 |
| 17 | 7.66 | 73.07 | 7.1 | 12.17 | 0 | 0 | 0 | 0 | 45.28 | 888.3 |
| 18 | 5.61 | 61.74 | 6 | 6.85 | 0 | 12.21 | 7.59 | 0 | 46.55 | 928.8 |
| 19 | 5.81 | 63.93 | 6.21 | 3.55 | 0 | 12.64 | 7.86 | 0 | 45.81 | 908.5 |
| 20 | 4.49 | 53.56 | 5.43 | 3.1 | 0 | 11.05 | 6.87 | 15.49 | 46.95 | 921.3 |
| 21 | 5.81 | 63.93 | 6.21 | 3.55 | 0 | 12.64 | 7.86 | 0 | 45.81 | 908.5 |
| 22 | 5.61 | 61.74 | 6 | 6.85 | 0 | 12.21 | 7.59 | 0 | 46.55 | 928.8 |
| 23 | 5.81 | 63.93 | 6.21 | 3.55 | 0 | 12.64 | 7.86 | 0 | 45.81 | 908.5 |
| 24 | 5.61 | 61.74 | 6 | 6.85 | 0 | 12.21 | 7.59 | 0 | 46.55 | 928.8 |
| 25 | 6.77 | 60.5 | 6.72 | 3.84 | 0 | 13.67 | 8.5 | 0 | 41.5 | 856.7 |
| 26 | 5.19 | 59.09 | 6.56 | 7.5 | 0 | 13.35 | 8.31 | 0 | 40.85 | 825.5 |
| 27 | 5.29 | 60.34 | 9.6 | 3.65 | 0 | 13.02 | 8.1 | 0 | 43.14 | 865.8 |
| 28 | 5.29 | 60.34 | 9.6 | 3.65 | 0 | 13.02 | 8.1 | 0 | 43.14 | 865.8 |
| 29 | 5.29 | 60.34 | 9.6 | 3.65 | 0 | 13.02 | 8.1 | 0 | 43.14 | 865.8 |
| 30 | 8.25 | 75.37 | 7.64 | 8.73 | 0 | 0 | 0 | 0 | 42.67 | 841.1 |
| 31 | 8.04 | 76 | 11.56 | 4.4 | 0 | 0 | 0 | 0 | 43.05 | 843.2 |
| 32 | 7.55 | 69.97 | 5.83 | 9.99 | 6.67 | 0 | 0 | 0 | 53.84 | 1054.3 |
| 33 | 7.74 | 75.98 | 12.66 | 3.61 | 0 | 0 | 0 | 0 | 52.92 | 959.7 |
| 34 | 8.1 | 79.27 | 8.04 | 4.59 | 0 | 0 | 0 | 0 | 41.86 | 818.4 |
| 35 | 7.79 | 75.61 | 12.02 | 4.58 | 0 | 0 | 0 | 0 | 41.48 | 795.8 |
| 36 | 8.04 | 76 | 11.56 | 4.4 | 0 | 0 | 0 | 0 | 43.31 | 835.8 |
| 37 | 8.28 | 76.35 | 11.13 | 4.24 | 0 | 0 | 0 | 0 | 45.11 | 873.9 |
| 38 | 7.99 | 72.98 | 14.8 | 4.23 | 0 | 0 | 0 | 0 | 44.35 | 869.4 |
| 39 | 7.78 | 76.76 | 11.19 | 4.26 | 0 | 0 | 0 | 0 | 44.32 | 844.4 |
| 40 | 7.51 | 75.19 | 12.53 | 4.77 | 0 | 0 | 0 | 0 | 39.53 | 771.1 |
| 41 | 7.19 | 71.4 | 16.65 | 4.76 | 0 | 0 | 0 | 0 | 38.77 | 765.3 |
| 42 | 7.84 | 79 | 8.38 | 4.78 | 0 | 0 | 0 | 0 | 40.15 | 767.5 |
| 43 | 8.1 | 79.27 | 8.04 | 4.59 | 0 | 0 | 0 | 0 | 41.98 | 807.6 |
| 44 | 7.55 | 78.71 | 8.74 | 4.99 | 0 | 0 | 0 | 0 | 38.2 | 738.3 |

Table S2. **Continued**

| N° | MV | d | NHA | NHD | NRB | LogP | MW | MR | E-S | E-B |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 363.3 | 1.256 | 7 | 1 | 8 | 5.924 | 456.521 | 114.15 | 2.6809 | 143.1218 |
| 2 | 326.7 | 1.257 | 5 | 1 | 6 | 5.106 | 410.86 | 104.427 | 2.609 | 8.1117 |
| 3 | 338 | 1.292 | 5 | 1 | 6 | 5.42 | 436.897 | 111.658 | 3.43 | 7.0783 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 338 | 1.292 | 5 | 1 | 6 | 5.42 | 436.897 | 111.658 | 3.7251 | 8.7788 |
| 5 | 330.2 | 1.302 | 7 | 1 | 6 | 4.426 | 421.431 | 104.54 | 3.3398 | 8.596 |
| 6 | 316.4 | 1.274 | 2 | 1 | 5 | 5.178 | 403.344 | 110.489 | 3.5148 | 8.4091 |
| 7 | 309.6 | 1.305 | 3 | 1 | 5 | 4.26 | 404.332 | 109.393 | 3.5012 | 8.9937 |
| 8 | 310.9 | 1.22 | 3 | 1 | 6 | 3.992 | 381.534 | 111.942 | 3.31 | 10.2287 |
| 9 | 325.1 | 1.115 | 2 | 1 | 5 | 4.748 | 362.507 | 110.961 | 3.6309 | 8.5105 |
| 10 | 318.3 | 1.141 | 3 | 1 | 5 | 3.995 | 363.495 | 108.663 | 3.581 | 9.2962 |
| 11 | 321.7 | 1.243 | 4 | 1 | 6 | 3.812 | 399.913 | 109.99 | 3.7574 | 10.2987 |
| 12 | 301.9 | 1.372 | 3 | 1 | 5 | 3.7 | 414.338 | 106.345 | 3.2349 | 9.3668 |
| 13 | 309.6 | 1.305 | 3 | 1 | 5 | 3.55 | 404.332 | 109.393 | 3.5181 | 9.5478 |
| 14 | 316.4 | 1.274 | 2 | 1 | 5 | 4.468 | 403.344 | 110.489 | 3.6081 | 8.9743 |
| 15 | 309.6 | 1.305 | 3 | 1 | 5 | 3.55 | 404.332 | 108.332 | 3.5393 | 9.3194 |
| 16 | 306.7 | 1.251 | 3 | 2 | 5 | 4.165 | 383.914 | 110.384 | 3.3159 | 8.3603 |
| 17 | 340.5 | 1.158 | 4 | 1 | 7 | 4.388 | 394.506 | 113.805 | 4.1245 | 11.7777 |
| 18 | 362 | 1.289 | 6 | 1 | 7 | 5.208 | 466.923 | 118.121 | 4.5556 | 11.7664 |
| 19 | 354.2 | 1.272 | 5 | 1 | 6 | 5.869 | 450.924 | 116.699 | 4.5198 | 10.6251 |
| 20 | 354.1 | 1.456 | 5 | 1 | 6 | 6.283 | 515.793 | 119.28 | 3.7527 | 8.7897 |
| 21 | 354.2 | 1.272 | 5 | 1 | 6 | 5.919 | 450.924 | 116.699 | 3.7965 | 8.943 |
| 22 | 362 | 1.289 | 6 | 1 | 7 | 5.208 | 466.923 | 118.121 | 4.0183 | 11.3092 |
| 23 | 354.2 | 1.272 | 5 | 1 | 6 | 5.919 | 450.924 | 116.699 | 3.7752 | 8.9455 |
| 24 | 362 | 1.289 | 6 | 1 | 7 | 5.208 | 466.923 | 118.121 | 3.9781 | 11.2314 |
| 25 | 344 | 1.211 | 5 | 1 | 8 | 5.529 | 416.908 | 105.519 | 3.5883 | 8.9904 |
| 26 | 320.8 | 1.33 | 6 | 1 | 6 | 4.596 | 426.859 | 104.08 | 3.3294 | 19.2107 |
| 27 | 331.2 | 1.321 | 6 | 1 | 6 | 3.923 | 437.885 | 109.501 | 3.6639 | 8.8619 |
| 28 | 331.2 | 1.321 | 6 | 1 | 6 | 3.923 | 437.885 | 109.501 | 3.682 | 8.9908 |
| 29 | 331.2 | 1.321 | 6 | 1 | 6 | 3.923 | 437.885 | 109.129 | 3.5318 | 9.0273 |
| 30 | 324.1 | 1.13 | 3 | 1 | 5 | 3.924 | 366.496 | 103.384 | 3.2701 | 19.1693 |
| 31 | 318.3 | 1.141 | 3 | 1 | 5 | 3.251 | 363.495 | 108.805 | 3.5677 | 8.5967 |
| 32 | 396.1 | 1.213 | 4 | 1 | 8 | 4.08 | 480.662 | 136.508 | 5.1759 | 140.1923 |
| 33 | 370.5 | 1.19 | 4 | 1 | 6 | 4.751 | 442.595 | 134.713 | 3.9397 | 19.0834 |
| 34 | 310 | 1.123 | 2 | 2 | 5 | 4.203 | 348.481 | 105.667 | 2.3774 | 4.8025 |
| 35 | 303.3 | 1.151 | 2 | 1 | 5 | 4.189 | 349.469 | 106.23 | 2.9116 | 15.7163 |
| 36 | 320.9 | 1.132 | 2 | 1 | 6 | 4.718 | 363.495 | 110.978 | 3.0768 | 17.1673 |
| 37 | 335.5 | 1.125 | 2 | 1 | 6 | 5.027 | 377.522 | 115.396 | 3.3677 | 22.5282 |
| 38 | 328.7 | 1.151 | 3 | 1 | 6 | 3.53 | 378.51 | 113.239 | 3.3293 | 22.644 |
| 39 | 310.2 | 1.21 | 2 | 1 | 6 | 4.552 | 375.506 | 113.44 | 2.9 | 22.6448 |
| 40 | 288.3 | 1.163 | 2 | 2 | 5 | 3.644 | 335.442 | 100.935 | 2.2337 | 13.5202 |
| 41 | 281.5 | 1.194 | 3 | 2 | 5 | 2.147 | 336.43 | 98.778 | 2.2124 | 13.6637 |
| 42 | 289.6 | 1.154 | 2 | 1 | 5 | 3.63 | 334.454 | 101.759 | 3.3316 | 39.2443 |
| 43 | 307.2 | 1.134 | 2 | 1 | 6 | 4.159 | 348.481 | 106.507 | 3.622 | 41.3819 |
| 44 | 274.6 | 1.166 | 2 | 2 | 5 | 3.085 | 320.428 | 96.465 | 2.8062 | 37.6277 |

Table S2. **Continued**

| N° | E-S,B | E-T | E-d,d | j | TD | MTI | SVD | W | E VDW | ɣ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.4467 | 39.2347 | 14.8556 | 1394171 | 13 | 19033 | 110.666 | 2752 | 25.851 | 41.3 |
| 2 | 0.551 | 12.9136 | -4.8438 | 819752 | 11 | 13449 | 99.777 | 1968 | 20.6431 | 40.3 |
| 3 | 0.5694 | 12.833 | -5.9795 | 890496 | 11 | 15615 | 105.777 | 2267 | 26.9175 | 43.9 |
| 4 | 0.5872 | 12.2837 | -5.8349 | 890496 | 11 | 15615 | 105.777 | 2267 | 26.9119 | 43.9 |
| 5 | 0.5733 | 10.5041 | -2.4643 | 890496 | 11 | 15442 | 114 | 2267 | 26.4976 | 44.6 |
| 6 | 0.5614 | 5.8562 | -4.7296 | 540374 | 10 | 12309 | 81.555 | 1682 | 27.8164 | 50.9 |
| 7 | 0.5429 | 10.9678 | -2.8645 | 540374 | 10 | 12145 | 83.555 | 1682 | 27.7334 | 54.3 |
| 8 | 0.5861 | 10.6827 | -7.9488 | 541514 | 10 | 12847 | 82.666 | 1686 | 25.6165 | 57.7 |
| 9 | 0.5228 | 10.3084 | -8.5481 | 534706 | 10 | 13114 | 82 | 1664 | 26.6179 | 46.2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0.5923 | 13.6092 | -8.4182 | 534706 | 10 | 12968 | 84 | 1664 | 27.1751 | 49.2 |
| 11 | 0.5659 | 16.8366 | -5.1807 | 633262 | 10 | 13566 | 89.777 | 1839 | 28.5364 | 51.6 |
| 12 | 0.5484 | 12.0135 | -6.3473 | 453229 | 10 | 11344 | 82 | 1516 | 26.7415 | 53.7 |
| 13 | 0.581 | 15.9864 | -2.8238 | 534706 | 10 | 12074 | 83.555 | 1664 | 26.3843 | 54.3 |
| 14 | 0.5911 | 12.0765 | -4.8274 | 534706 | 10 | 12220 | 81.555 | 1664 | 26.5006 | 50.9 |
| 15 | 0.5751 | 15.211 | -2.1403 | 534706 | 10 | 12055 | 83.555 | 1664 | 27.0067 | 54.3 |
| 16 | 0.5463 | 8.802 | -7.0325 | 534706 | 10 | 12518 | 83.777 | 1664 | 25.5126 | 55.2 |
| 17 | 0.5638 | 11.4242 | -7.3137 | 743707 | 10 | 15308 | 94 | 2020 | 30.5975 | 46.3 |
| 18 | 0.6054 | 11.1295 | -7.11 | 1176535 | 11 | 18037 | 113.777 | 2648 | 29.6276 | 43.3 |
| 19 | 0.6452 | 11.5098 | -5.9352 | 1021281 | 11 | 16888 | 107.777 | 2442 | 28.0307 | 43.2 |
| 20 | 0.6063 | 12.3354 | -5.9224 | 1031132 | 11 | 16464 | 107.777 | 2466 | 27.6462 | 45.7 |
| 21 | 0.5847 | 11.5767 | -5.8225 | 1031132 | 11 | 17064 | 107.777 | 2466 | 27.5379 | 43.2 |
| 22 | 0.6374 | 12.3649 | -6.0606 | 1197359 | 12 | 18341 | 113.777 | 2696 | 29.6039 | 43.3 |
| 23 | 0.588 | 11.5647 | -5.8409 | 1040706 | 12 | 17240 | 107.777 | 2490 | 27.5461 | 43.2 |
| 24 | 0.6452 | 12.3272 | -5.779 | 1217688 | 13 | 18645 | 113.777 | 2744 | 29.6453 | 43.3 |
| 25 | 0.6224 | 18.5378 | -5.695 | 813852 | 11 | 13341 | 93.777 | 1954 | 23.8063 | 38.4 |
| 26 | 0.4333 | 14.1068 | -5.0278 | 762828 | 11 | 13902 | 105.777 | 2071 | 21.5267 | 43.8 |
| 27 | 0.5882 | 18.0072 | -5.7737 | 890496 | 11 | 15420 | 107.777 | 2267 | 27.4757 | 46.6 |
| 28 | 0.593 | 18.0105 | -5.7925 | 890496 | 11 | 15444 | 107.777 | 2267 | 27.3742 | 46.6 |
| 29 | 0.6018 | 17.5352 | -6.3401 | 890496 | 11 | 15468 | 107.777 | 2267 | 27.7184 | 46.6 |
| 30 | 0.3753 | 11.6496 | -7.8286 | 449455 | 10 | 11603 | 82 | 1502 | 21.1925 | 45.3 |
| 31 | 0.5233 | 15.9695 | -8.5377 | 534706 | 10 | 12953 | 84 | 1664 | 27.184 | 49.2 |
| 32 | -1.2831 | 58.0893 | 0.7487 | 1410626 | 14 | 25581 | 104.666 | 3310 | 28.6971 | 50.1 |
| 33 | 0.3696 | 13.4879 | -9.9616 | 1227884 | 14 | 23316 | 103 | 3049 | 28.2098 | 45 |
| 34 | 0.1802 | 7.3285 | -8.8864 | 457656 | 10 | 12052 | 80 | 1530 | 23.814 | 48.5 |
| 35 | -0.1442 | 9.0526 | -8.2713 | 452023 | 10 | 11905 | 80 | 1511 | 19.9767 | 47.3 |
| 36 | -0.0978 | 9.4978 | -7.8896 | 535222 | 10 | 13132 | 82 | 1666 | 20.9472 | 45.9 |
| 37 | 0.0041 | 10.9824 | -7.8784 | 627536 | 10 | 14375 | 84 | 1823 | 21.7302 | 46 |
| 38 | 0.0029 | 16.6401 | -7.6927 | 627536 | 10 | 14206 | 86 | 1823 | 22.2699 | 48.9 |
| 39 | -0.0484 | 9.9783 | -8.2684 | 539544 | 10 | 14375 | 86 | 1822 | 20.5422 | 54.8 |
| 40 | -0.2291 | 7.6774 | -8.1638 | 383835 | 10 | 10883 | 78 | 1382 | 18.6928 | 51.1 |
| 41 | -0.2242 | 13.3491 | -8.0991 | 383835 | 10 | 10737 | 80 | 1382 | 19.2403 | 54.5 |
| 42 | -1.6471 | 8.3209 | -8.418 | 384086 | 10 | 10891 | 78 | 1383 | 18.4405 | 49.3 |
| 43 | -1.535 | 10.8208 | -8.286 | 458707 | 10 | 12086 | 80 | 1534 | 19.2144 | 47.7 |
| 44 | -1.7047 | 5.2114 | -8.8676 | 323172 | 10 | 9893 | 76 | 1257 | 17.1584 | 52.2 |

Table S2. **Continued**

| N° | n | E-H | E-L | ET | NRE | μ | χ | w | E-Gap | η |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.540 | -0.206 | -0.072 | -1 887.498 | 3 419.578 | 4.754 | 0.139 | 0.001 | -0.134 | 0.067 |
| 2 | 1.546 | -0.234 | -0.061 | -1 758.023 | 2 942.932 | 5.443 | 0.147 | 0.001 | -0.173 | 0.087 |
| 3 | 1.569 | -0.222 | -0.061 | -1 835.442 | 3 255.926 | 4.073 | 0.141 | 0.001 | -0.161 | 0.080 |
| 4 | 1.569 | -0.222 | -0.059 | -1 835.438 | 3 258.002 | 3.407 | 0.140 | 0.001 | -0.163 | 0.082 |
| 5 | 1.553 | -0.225 | -0.087 | -1 491.123 | 3 085.389 | 4.441 | 0.156 | 0.001 | -0.138 | 0.069 |
| 6 | 1.616 | -0.221 | -0.052 | -1 957.976 | 2 867.301 | 3.010 | 0.136 | 0.001 | -0.169 | 0.084 |
| 7 | 1.619 | -0.226 | -0.069 | -1 974.013 | 2 871.259 | 5.055 | 0.148 | 0.001 | -0.157 | 0.078 |
| 8 | 1.637 | -0.215 | -0.050 | -1 492.340 | 2 742.050 | 3.220 | 0.132 | 0.001 | -0.165 | 0.082 |
| 9 | 1.595 | -0.220 | -0.021 | -1 117.433 | 2 578.349 | 4.249 | 0.121 | 0.001 | -0.199 | 0.100 |
| 10 | 1.598 | -0.223 | -0.034 | -1 133.474 | 2 583.330 | 4.273 | 0.129 | 0.001 | -0.189 | 0.095 |
| 11 | 1.603 | -0.223 | -0.048 | -1 628.947 | 2 919.662 | 5.001 | 0.135 | 0.001 | -0.174 | 0.087 |
| 12 | 1.624 | -0.221 | -0.070 | -3 625.959 | 2 944.740 | 4.109 | 0.146 | 0.001 | -0.150 | 0.075 |
| 13 | 1.619 | -0.225 | -0.060 | -1 974.014 | 2 909.851 | 4.645 | 0.142 | 0.001 | -0.165 | 0.082 |
| 14 | 1.616 | -0.222 | -0.040 | -1 957.977 | 2 904.361 | 4.583 | 0.131 | 0.001 | -0.182 | 0.091 |
| 15 | 1.619 | -0.227 | -0.061 | -1 974.010 | 2 905.995 | 5.165 | 0.144 | 0.001 | -0.166 | 0.083 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 1.636 | -0.216 | -0.037 | -1 553.755 | 2 751.322 | 2.305 | 0.126 | 0.001 | -0.178 | 0.089 |
| 17 | 1.585 | -0.214 | -0.014 | -1 267.849 | 2 931.677 | 4.209 | 0.114 | 0.001 | -0.199 | 0.100 |
| 18 | 1.562 | -0.213 | -0.057 | -1 949.961 | 3 685.274 | 4.121 | 0.135 | 0.001 | -0.156 | 0.078 |
| 19 | 1.566 | -0.224 | -0.059 | -1 874.753 | 3 454.951 | 4.267 | 0.142 | 0.001 | -0.165 | 0.082 |
| 20 | 1.583 | -0.226 | -0.059 | -4 406.567 | 3 950.958 | 3.674 | 0.143 | 0.001 | -0.167 | 0.084 |
| 21 | 1.566 | -0.221 | -0.059 | -1 874.759 | 3 434.156 | 3.456 | 0.140 | 0.001 | -0.163 | 0.081 |
| 22 | 1.562 | -0.221 | -0.055 | -1 949.967 | 3 603.988 | 5.091 | 0.138 | 0.001 | -0.167 | 0.083 |
| 23 | 1.566 | -0.222 | -0.056 | -1 874.759 | 3 413.326 | 4.148 | 0.139 | 0.001 | -0.165 | 0.083 |
| 24 | 1.562 | -0.219 | -0.058 | -1 949.967 | 3 581.816 | 3.820 | 0.138 | 0.001 | -0.161 | 0.080 |
| 25 | 1.520 | -0.219 | -0.057 | -1 761.644 | 3 045.444 | 3.314 | 0.138 | 0.001 | -0.162 | 0.081 |
| 26 | 1.555 | -0.221 | -0.061 | -1 833.214 | 3 138.280 | 3.270 | 0.141 | 0.001 | -0.161 | 0.080 |
| 27 | 1.571 | -0.228 | -0.063 | -1 851.474 | 3 262.741 | 2.722 | 0.146 | 0.001 | -0.165 | 0.083 |
| 28 | 1.571 | -0.227 | -0.063 | -1 851.473 | 3 263.791 | 3.056 | 0.145 | 0.001 | -0.164 | 0.082 |
| 29 | 1.571 | -0.219 | -0.052 | -1 851.479 | 3 259.108 | 5.466 | 0.136 | 0.001 | -0.167 | 0.083 |
| 30 | 1.578 | -0.220 | -0.021 | -1 115.210 | 2 467.417 | 4.289 | 0.120 | 0.001 | -0.199 | 0.100 |
| 31 | 1.598 | -0.227 | -0.039 | -1 133.470 | 2 582.730 | 5.531 | 0.133 | 0.001 | -0.188 | 0.094 |
| 32 | 1.601 | -0.229 | -0.051 | -1 822.040 | 3 885.121 | 5.398 | 0.140 | 0.001 | -0.178 | 0.089 |
| 33 | 1.640 | -0.216 | -0.029 | -1 381.771 | 3 364.602 | 5.613 | 0.123 | 0.001 | -0.187 | 0.093 |
| 34 | 1.596 | -0.233 | -0.022 | -1 078.138 | 2 383.528 | 4.281 | 0.127 | 0.001 | -0.211 | 0.105 |
| 35 | 1.606 | -0.224 | -0.024 | -1 078.117 | 2 387.210 | 2.933 | 0.124 | 0.001 | -0.200 | 0.100 |
| 36 | 1.596 | -0.224 | -0.027 | -1 117.433 | 2 548.693 | 3.920 | 0.125 | 0.001 | -0.197 | 0.098 |
| 37 | 1.593 | -0.213 | -0.028 | -1 156.744 | 2 731.377 | 3.693 | 0.120 | 0.001 | -0.185 | 0.092 |
| 38 | 1.596 | -0.219 | -0.046 | -1 172.780 | 2 736.921 | 4.101 | 0.132 | 0.001 | -0.174 | 0.087 |
| 39 | 1.640 | -0.213 | -0.025 | -1 155.500 | 2 690.189 | 2.959 | 0.119 | 0.001 | -0.188 | 0.094 |
| 40 | 1.608 | -0.243 | -0.022 | -1 038.815 | 2 208.184 | 4.305 | 0.133 | 0.001 | -0.221 | 0.111 |
| 41 | 1.611 | -0.249 | -0.042 | -1 054.850 | 2 212.469 | 5.051 | 0.145 | 0.001 | -0.207 | 0.103 |
| 42 | 1.616 | -0.218 | -0.027 | -1 038.765 | 2 222.481 | 2.789 | 0.123 | 0.001 | -0.191 | 0.095 |
| 43 | 1.605 | -0.217 | -0.027 | -1 078.079 | 2 400.764 | 2.788 | 0.122 | 0.001 | -0.190 | 0.095 |
| 44 | 1.619 | -0.239 | -0.021 | -999.464 | 2 047.101 | 3.984 | 0.130 | 0.001 | -0.218 | 0.109 |

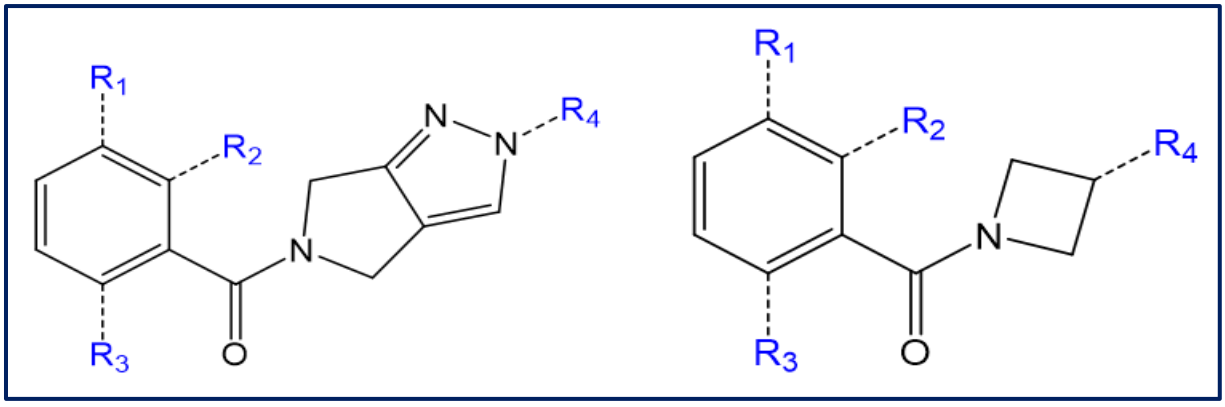



**Figure S3. Pyrrolo[3,4-c]pyrazole-based analogs 1-34 (left) and azetidine-based analogs 35-40 (right)**

**Table S3. Studied molecules with their inhibitory activities on the order of $IC_{50}$ and $pIC_{50}$**

| Compounds number | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $IC_{50}$ (µM) | observed $pIC_{50}$ |
|---|---|---|---|---|---|---|
| 1 | $SO_2Me$ | H |  |  | 15 | 7.824 |
| 2 | $SO_2Me$ | H |  |  | 8 | 8.097 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | $SO_2Me$ | H |  |  | 23 | 7.638 |
| 4* | $SO_2Me$ | H |  |  | 23 | 7.638 |
| 5 | $SO_2Me$ | H |  |  | 12 | 7.921 |
| 6 | $SO_2Me$ | H |  |  | 17 | 7.770 |
| 7 | $SO_2Me$ | H |  |  | 23 | 7.638 |
| 8 | $SO_2Me$ | H |  |  | 10 | 8.000 |
| 9 | $SO_2Me$ | H |  |  | 38 | 7.420 |
| 10* | $SO_2Me$ | H |  |  | 16 | 7.796 |
| 11 | $SO_2Me$ | H |  |  | 10 | 8.000 |
| 12 | $SO_2Me$ | H |  |  | 7 | 8.155 |
| 13 | $SO_2Me$ | H |  |  | 10 | 8.000 |
| 14* | $SO_2Me$ | H |  |  | 6 | 8.222 |
| 15 | $SO_2Me$ | H |  |  | 4 | 8.398 |
| 16 | $SO_2Et$ | H |  |  | 29 | 7.538 |
| 17* | $SO_2Et$ | H |  |  | 27 | 7.569 |
| 18* | $SO_2Et$ | H |  |  | 12 | 7.921 |
| 19* | $SO_2$ i-Pr | H |  |  | 406 | 6.391 |
| 20 | CN | H |  |  | 78 | 7.108 |
| 21 | CN | H |  |  | 69 | 7.161 |
| 22 | CN | H |  |  | 18 | 7.745 |
| 23 | CN | H |  |  | 210 | 6.678 |
| 24 | CN | H |  |  | 105 | 6.979 |

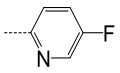
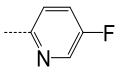
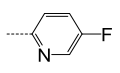
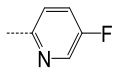
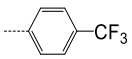
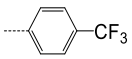
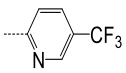
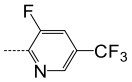
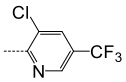
| | | | | | | |
|---|---|---|---|---|---|---|
| 25 | CN | H |  |  | 113 | 6.947 |
| 26 | CN | H |  |  | 21 | 7.678 |
| 27 | CN | H |  |  | 441 | 6.356 |
| 28 | CN | H |  |  | 20 | 7.699 |
| 29* | CN | H |  |  | 383 | 6.417 |
| 30 | CN | H |  |  | 55 | 7.260 |
| 31* | CN | H |  |  | 142 | 6.848 |
| 32 | CN | Cl |  |  | 33 | 7.481 |
| 33 | CN | OMe |  |  | 21 | 7.678 |
| 34 | $SO_2Me$ | OMe |  |  | 19 | 7.721 |
| 35 | $SO_2Me$ | OMe |  |  | 27 | 7.569 |
| 36 | $SO_2Me$ | H |  |  | 630 | 6.201 |
| 37 | $SO_2Me$ | OMe |  |  | 220 | 6.658 |
| 38 | $SO_2Me$ | OMe |  |  | 43 | 7.367 |
| 39 | $SO_2Me$ | OMe |  |  | 26 | 7.585 |
| 40 | $SO_2Me$ | OMe |  |  | 15 | 7.824 |

**Table S4. Calculations of various molecular descriptors**

| N° | %H | %C | %N | %O | %S | %F | %Cl | MW | æ | P |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.200 | 55.110 | 8.760 | 13.350 | 6.690 | 11.890 | 0.000 | 479.470 | 46.170 | 878.900 |
| 2 | 3.850 | 53.120 | 8.450 | 12.860 | 6.450 | 15.280 | 0.000 | 497.460 | 46.120 | 879.100 |
| 3 | 4.490 | 55.980 | 8.510 | 12.970 | 6.500 | 11.550 | 0.000 | 493.500 | 47.920 | 910.000 |
| 4 | 3.640 | 50.600 | 11.240 | 12.840 | 6.430 | 15.250 | 0.000 | 498.450 | 45.500 | 860.200 |
| 5 | 4.280 | 53.440 | 11.330 | 12.940 | 6.480 | 11.530 | 0.000 | 494.490 | 47.310 | 891.100 |
| 6 | 5.130 | 53.490 | 8.910 | 13.570 | 6.800 | 12.090 | 0.000 | 471.490 | 44.340 | 840.300 |
| 7 | 5.000 | 59.580 | 9.470 | 14.430 | 7.230 | 4.280 | 0.000 | 443.490 | 45.990 | 856.100 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 5.560 | 61.130 | 8.910 | 13.570 | 6.800 | 4.030 | 0.000 | 471.540 | 49.780 | 940.700 |
| 9 | 4.870 | 60.650 | 9.230 | 14.050 | 7.040 | 4.170 | 0.000 | 455.500 | 46.950 | 855.000 |
| 10 | 4.590 | 57.130 | 11.100 | 9.510 | 6.360 | 11.300 | 0.000 | 504.520 | 49.900 | 923.800 |
| 11 | 4.390 | 54.650 | 13.850 | 9.490 | 6.340 | 11.270 | 0.000 | 505.510 | 49.290 | 904.900 |
| 12 | 5.020 | 57.470 | 13.960 | 9.570 | 6.390 | 7.580 | 0.000 | 501.550 | 51.090 | 935.900 |
| 13 | 3.990 | 62.620 | 8.760 | 10.010 | 6.690 | 7.920 | 0.000 | 479.500 | 50.000 | 913.200 |
| 14 | 3.780 | 59.990 | 11.660 | 9.990 | 6.660 | 7.910 | 0.000 | 480.490 | 49.390 | 894.400 |
| 15 | 4.440 | 63.010 | 11.760 | 10.070 | 6.730 | 3.990 | 0.000 | 476.520 | 51.190 | 925.300 |
| 16 | 4.140 | 54.010 | 8.220 | 12.510 | 6.270 | 14.860 | 0.000 | 511.490 | 47.950 | 917.700 |
| 17 | 3.930 | 51.560 | 10.930 | 12.490 | 6.260 | 14.830 | 0.000 | 512.480 | 47.330 | 898.800 |
| 18 | 4.560 | 54.320 | 11.020 | 12.590 | 6.310 | 11.210 | 0.000 | 508.510 | 49.130 | 929.800 |
| 19 | 4.770 | 56.800 | 8.280 | 12.610 | 6.320 | 11.230 | 0.000 | 507.530 | 49.750 | 948.600 |
| 20 | 3.630 | 59.460 | 12.610 | 7.200 | 0.000 | 17.100 | 0.000 | 444.380 | 43.080 | 812.600 |
| 21 | 3.390 | 56.630 | 15.720 | 7.180 | 0.000 | 17.060 | 0.000 | 445.370 | 42.470 | 793.800 |
| 22 | 4.110 | 59.860 | 15.870 | 7.250 | 0.000 | 12.910 | 0.000 | 441.410 | 44.270 | 824.700 |
| 23 | 5.060 | 60.280 | 13.390 | 7.650 | 0.000 | 13.620 | 0.000 | 418.410 | 41.310 | 773.800 |
| 24 | 5.410 | 70.950 | 15.040 | 8.590 | 0.000 | 0.000 | 0.000 | 372.420 | 43.000 | 789.500 |
| 25 | 4.640 | 64.440 | 17.890 | 8.180 | 0.000 | 4.850 | 0.000 | 391.400 | 42.340 | 770.800 |
| 26 | 5.460 | 68.200 | 18.080 | 8.260 | 0.000 | 0.000 | 0.000 | 387.430 | 44.140 | 801.800 |
| 27 | 5.300 | 69.160 | 17.530 | 8.010 | 0.000 | 0.000 | 0.000 | 399.450 | 45.110 | 800.600 |
| 28 | 4.940 | 64.280 | 18.740 | 3.570 | 0.000 | 8.470 | 0.000 | 448.470 | 48.060 | 869.400 |
| 29 | 3.540 | 67.440 | 16.390 | 3.740 | 0.000 | 8.890 | 0.000 | 427.410 | 46.350 | 827.900 |
| 30 | 4.280 | 70.910 | 16.540 | 3.780 | 0.000 | 4.490 | 0.000 | 423.440 | 48.160 | 858.800 |
| 31 | 3.390 | 56.630 | 15.720 | 7.180 | 0.000 | 17.060 | 0.000 | 445.370 | 42.470 | 793.800 |
| 32 | 2.940 | 52.570 | 14.600 | 6.670 | 0.000 | 15.840 | 7.390 | 479.810 | 44.290 | 822.600 |
| 33 | 3.600 | 55.580 | 14.730 | 10.100 | 0.000 | 15.990 | 0.000 | 475.400 | 44.770 | 844.000 |
| 34 | 3.810 | 50.000 | 10.600 | 15.140 | 6.070 | 14.380 | 0.000 | 528.480 | 47.810 | 910.500 |
| 35 | 4.030 | 50.290 | 2.670 | 15.220 | 6.100 | 21.690 | 0.000 | 525.460 | 44.670 | 936.600 |
| 36 | 3.870 | 50.910 | 2.830 | 12.920 | 6.470 | 23.010 | 0.000 | 495.440 | 42.140 | 878.000 |
| 37 | 4.850 | 55.140 | 3.060 | 17.490 | 7.010 | 12.460 | 0.000 | 457.460 | 42.690 | 874.600 |
| 38 | 3.830 | 47.910 | 5.320 | 15.200 | 6.090 | 21.650 | 0.000 | 526.450 | 43.790 | 932.200 |
| 39 | 3.520 | 46.330 | 5.150 | 14.690 | 5.890 | 24.430 | 0.000 | 544.440 | 43.840 | 939.500 |
| 40 | 3.410 | 44.970 | 4.990 | 14.260 | 5.720 | 20.320 | 6.320 | 560.890 | 45.710 | 969.300 |

Table S4. **Continued**

| N° | MV | n | ɣ | d | E-S | E-B | E-S,B | E-T | E VDW | E-d,d |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 335.000 | 1.612 | 47.300 | 1.430 | 5.131 | 164.321 | -0.815 | 40.370 | 22.226 | 6.804 |
| 2 | 337.800 | 1.604 | 45.800 | 1.470 | 5.466 | 165.288 | -0.814 | 41.546 | 22.122 | 7.532 |
| 3 | 350.200 | 1.606 | 45.600 | 1.400 | 5.205 | 164.481 | -0.819 | 39.663 | 22.846 | 6.761 |
| 4 | 326.500 | 1.620 | 48.100 | 1.520 | 5.048 | 164.461 | -0.818 | 46.099 | 22.429 | 8.673 |
| 5 | 338.800 | 1.622 | 47.800 | 1.450 | 5.032 | 164.539 | -0.798 | 45.335 | 23.495 | 6.783 |
| 6 | 318.900 | 1.618 | 48.100 | 1.470 | 4.150 | 167.796 | -0.802 | 54.150 | 18.207 | 6.626 |
| 7 | 323.000 | 1.637 | 49.300 | 1.370 | 5.123 | 164.741 | -0.932 | 44.223 | 21.565 | 3.361 |
| 8 | 359.600 | 1.615 | 46.800 | 1.310 | 5.433 | 164.539 | -0.930 | 43.791 | 22.995 | 3.381 |
| 9 | 312.300 | 1.683 | 56.100 | 1.450 | 5.188 | 166.067 | -1.194 | 47.601 | 21.364 | 3.709 |
| 10 | 343.700 | 1.653 | 52.100 | 1.460 | 5.868 | 172.326 | -0.531 | 53.509 | 24.921 | 1.830 |
| 11 | 332.300 | 1.671 | 54.900 | 1.520 | 5.438 | 171.461 | -0.540 | 58.165 | 25.228 | 3.097 |
| 12 | 344.700 | 1.670 | 54.300 | 1.450 | 5.420 | 171.525 | -0.518 | 57.417 | 26.289 | 1.182 |
| 13 | 340.800 | 1.662 | 51.500 | 1.400 | 5.708 | 165.448 | -0.890 | 52.372 | 25.213 | 1.796 |
| 14 | 329.400 | 1.680 | 54.300 | 1.450 | 5.306 | 164.649 | -0.889 | 56.820 | 25.524 | 2.875 |
| 15 | 341.700 | 1.679 | 53.700 | 1.390 | 5.279 | 164.668 | -0.868 | 56.134 | 26.577 | 1.003 |
| 16 | 353.900 | 1.599 | 45.200 | 1.440 | 5.777 | 165.543 | -0.931 | 41.157 | 23.029 | 7.559 |
| 17 | 342.500 | 1.614 | 47.400 | 1.490 | 5.356 | 164.694 | -0.928 | 45.714 | 23.335 | 8.713 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 354.800 | 1.615 | 47.100 | 1.430 | 5.340 | 164.757 | -0.905 | 44.971 | 24.394 | 6.822 |
| 19 | 366.200 | 1.601 | 45.000 | 1.380 | 5.906 | 164.343 | -1.078 | 39.893 | 23.728 | 6.928 |
| 20 | 315.300 | 1.605 | 44.100 | 1.400 | 4.566 | 35.777 | 0.463 | -1.808 | 22.302 | -1.367 |
| 21 | 303.900 | 1.622 | 46.500 | 1.460 | 4.140 | 34.875 | 0.447 | 2.887 | 22.608 | -0.111 |
| 22 | 316.200 | 1.624 | 46.200 | 1.390 | 4.119 | 34.908 | 0.469 | 2.190 | 23.664 | -2.011 |
| 23 | 296.400 | 1.620 | 46.400 | 1.410 | 3.234 | 38.146 | 0.468 | 11.002 | 18.378 | -2.185 |
| 24 | 297.600 | 1.649 | 49.500 | 1.250 | 3.867 | 34.161 | 0.346 | -0.169 | 21.840 | -6.190 |
| 25 | 289.100 | 1.660 | 50.500 | 1.350 | 3.793 | 34.344 | 0.340 | 5.502 | 22.048 | -4.308 |
| 26 | 301.400 | 1.660 | 50.000 | 1.280 | 3.766 | 34.329 | 0.362 | 4.902 | 23.106 | -6.201 |
| 27 | 290.700 | 1.711 | 57.500 | 1.370 | 3.820 | 35.571 | 0.101 | 8.401 | 22.894 | -5.853 |
| 28 | 322.100 | 1.676 | 53.000 | 1.390 | 4.503 | 41.201 | 0.737 | 14.878 | 26.332 | -7.307 |
| 29 | 306.900 | 1.687 | 52.900 | 1.390 | 4.264 | 34.469 | 0.362 | 14.851 | 25.644 | -5.852 |
| 30 | 319.200 | 1.686 | 52.400 | 1.320 | 4.263 | 34.575 | 0.389 | 13.967 | 26.705 | -7.759 |
| 31 | 303.900 | 1.622 | 46.500 | 1.460 | 4.162 | 35.501 | 0.404 | 3.948 | 22.791 | -0.046 |
| 32 | 313.200 | 1.632 | 47.500 | 1.530 | 4.007 | 34.375 | 0.343 | 12.491 | 24.121 | 2.679 |
| 33 | 325.500 | 1.610 | 45.100 | 1.460 | 4.244 | 35.084 | 0.321 | 15.312 | 25.671 | -0.993 |
| 34 | 348.100 | 1.609 | 46.800 | 1.510 | 5.332 | 166.825 | -0.787 | 57.542 | 25.816 | 7.008 |
| 35 | 376.300 | 1.510 | 38.300 | 1.396 | 4.769 | 164.926 | -1.767 | 59.021 | 25.208 | 4.927 |
| 36 | 352.300 | 1.515 | 38.500 | 1.406 | 4.296 | 164.574 | -1.897 | 50.479 | 22.780 | 4.198 |
| 37 | 342.800 | 1.540 | 42.300 | 1.334 | 4.682 | 164.399 | -1.791 | 55.144 | 24.486 | 4.725 |
| 38 | 369.500 | 1.509 | 40.500 | 1.424 | 4.725 | 165.230 | -1.705 | 64.830 | 25.373 | 5.179 |
| 39 | 373.700 | 1.503 | 39.900 | 1.456 | 4.808 | 165.244 | -1.705 | 64.866 | 25.385 | 6.683 |
| 40 | 381.400 | 1.516 | 41.600 | 1.470 | 4.982 | 165.436 | -1.710 | 65.383 | 26.173 | 6.527 |

Table S4. **Continued**

| N° | NHA | NHD | NRB | Log P | MR | j | TD | MTI | PSA | SVD |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9 | 0 | 7 | 2.751 | 115.481 | 1554000 | 15 | 22775 | 79.280 | 128.667 |
| 2 | 10 | 0 | 7 | 2.980 | 115.698 | 1784327 | 15 | 23957 | 79.280 | 136.667 |
| 3 | 9 | 0 | 7 | 3.250 | 120.522 | 1797229 | 15 | 24989 | 79.280 | 130.667 |
| 4 | 11 | 0 | 7 | 1.836 | 113.541 | 1797229 | 15 | 23798 | 91.640 | 138.667 |
| 5 | 10 | 0 | 7 | 2.172 | 118.224 | 1797229 | 15 | 24687 | 91.640 | 132.667 |
| 6 | 9 | 0 | 7 | 2.284 | 112.116 | 1331825 | 14 | 20585 | 79.280 | 120.667 |
| 7 | 7 | 0 | 6 | 2.716 | 114.996 | 1171085 | 14 | 20405 | 79.280 | 112.667 |
| 8 | 7 | 0 | 7 | 3.734 | 123.972 | 1574139 | 15 | 24525 | 79.280 | 116.667 |
| 9 | 7 | 0 | 7 | 2.851 | 117.771 | 1173577 | 15 | 22563 | 79.280 | 116.667 |
| 10 | 9 | 0 | 5 | 2.248 | 125.397 | 1732128 | 16 | 26768 | 73.290 | 132.667 |
| 11 | 10 | 0 | 5 | 1.104 | 123.240 | 1744206 | 16 | 26604 | 85.650 | 134.667 |
| 12 | 9 | 0 | 5 | 1.440 | 127.924 | 1744206 | 16 | 27537 | 85.650 | 128.667 |
| 13 | 7 | 0 | 5 | 3.823 | 124.719 | 1504229 | 15 | 25485 | 70.050 | 128.667 |
| 14 | 8 | 0 | 5 | 2.678 | 122.562 | 1515255 | 15 | 25336 | 82.410 | 130.667 |
| 15 | 7 | 0 | 5 | 3.014 | 127.245 | 1515255 | 15 | 26219 | 82.410 | 124.667 |
| 16 | 10 | 0 | 8 | 3.509 | 120.521 | 2032993 | 15 | 26001 | 79.280 | 138.667 |
| 17 | 11 | 0 | 8 | 2.365 | 118.364 | 2047123 | 15 | 25833 | 91.640 | 140.667 |
| 18 | 10 | 0 | 8 | 2.701 | 123.047 | 2047123 | 15 | 26766 | 91.640 | 134.667 |
| 19 | 9 | 0 | 8 | 3.589 | 124.798 | 2019349 | 15 | 26739 | 79.280 | 132.667 |
| 20 | 9 | 0 | 7 | 3.941 | 107.906 | 1372345 | 15 | 20683 | 68.930 | 132.000 |
| 21 | 10 | 0 | 7 | 2.796 | 105.749 | 1383017 | 15 | 20541 | 81.290 | 134.000 |
| 22 | 9 | 0 | 7 | 3.132 | 110.432 | 1383017 | 15 | 21348 | 81.290 | 128.000 |
| 23 | 8 | 0 | 7 | 3.245 | 104.324 | 1007309 | 14 | 17634 | 68.930 | 116.000 |
| 24 | 5 | 0 | 6 | 3.448 | 106.988 | 742962 | 14 | 16385 | 68.930 | 100.000 |
| 25 | 7 | 0 | 6 | 2.532 | 105.047 | 879349 | 14 | 17262 | 81.290 | 110.000 |
| 26 | 6 | 0 | 6 | 2.868 | 109.730 | 879349 | 14 | 17937 | 81.290 | 104.000 |
| 27 | 6 | 0 | 7 | 3.003 | 112.506 | 893777 | 15 | 19984 | 81.290 | 108.000 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 8 | 0 | 5 | 2.410 | 120.132 | 1355895 | 16 | 24011 | 75.300 | 124.000 |
| 29 | 7 | 0 | 5 | 3.504 | 114.770 | 1166565 | 15 | 21977 | 72.060 | 126.000 |
| 30 | 6 | 0 | 5 | 3.840 | 119.453 | 1166565 | 15 | 22778 | 72.060 | 120.000 |
| 31 | 10 | 0 | 7 | 2.796 | 106.063 | 1393479 | 16 | 20674 | 81.290 | 134.000 |
| 32 | 10 | 0 | 7 | 3.394 | 110.867 | 1569554 | 16 | 21503 | 81.290 | 135.778 |
| 33 | 11 | 0 | 8 | 2.814 | 112.526 | 1775357 | 16 | 23355 | 90.520 | 142.000 |
| 34 | 12 | 0 | 8 | 1.380 | 120.317 | 2264336 | 16 | 26738 | 100.870 | 146.667 |
| 35 | 11 | 0 | 9 | 2.644 | 114.138 | 2487016 | 15 | 24178 | 72.910 | 144.667 |
| 36 | 10 | 0 | 8 | 3.099 | 107.674 | 1974405 | 15 | 21565 | 63.680 | 136.667 |
| 37 | 8 | 0 | 8 | 1.761 | 108.164 | 1338258 | 13 | 18077 | 72.910 | 118.667 |
| 38 | 12 | 0 | 9 | 1.294 | 111.402 | 2487016 | 15 | 23958 | 85.270 | 146.667 |
| 39 | 13 | 0 | 9 | 1.467 | 111.618 | 2800322 | 15 | 24993 | 85.270 | 154.667 |
| 40 | 12 | 0 | 9 | 2.037 | 116.207 | 2800322 | 15 | 24993 | 85.270 | 148.444 |

Table S4. **Continued**

| N° | WI | E-H | E-L | ET | NRE | μ | χ | w | E-Gap | η |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3307 | -0.240 | -0.062 | -2013.26 | 3497.68 | 4.654 | -0.151 | 0.127 | -0.178 | 0.089 |
| 2 | 3584 | -0.244 | -0.062 | -2112.50 | 3676.91 | 3.592 | -0.153 | 0.129 | -0.182 | 0.091 |
| 3 | 3611 | -0.237 | -0.061 | -2052.59 | 3644.57 | 5.084 | -0.149 | 0.126 | -0.176 | 0.088 |
| 4 | 3611 | -0.255 | -0.072 | -2128.53 | 3653.02 | 5.593 | -0.163 | 0.146 | -0.182 | 0.091 |
| 5 | 3611 | -0.254 | -0.067 | -2068.62 | 3650.09 | 5.316 | -0.160 | 0.137 | -0.187 | 0.094 |
| 6 | 3006 | -0.247 | -0.060 | -1977.57 | 3459.55 | 5.318 | -0.153 | 0.125 | -0.188 | 0.094 |
| 7 | 2807 | -0.240 | -0.052 | -1814.77 | 3125.23 | 7.449 | -0.146 | 0.114 | -0.187 | 0.094 |
| 8 | 3349 | -0.241 | -0.053 | -1893.40 | 3482.56 | 7.287 | -0.147 | 0.114 | -0.188 | 0.094 |
| 9 | 3090 | -0.240 | -0.052 | -1852.84 | 3230.47 | 7.700 | -0.146 | 0.114 | -0.187 | 0.094 |
| 10 | 3849 | -0.233 | -0.053 | -2070.81 | 3919.78 | 4.327 | -0.143 | 0.114 | -0.180 | 0.090 |
| 11 | 3877 | -0.237 | -0.068 | -2086.84 | 3895.13 | 6.699 | -0.153 | 0.138 | -0.169 | 0.084 |
| 12 | 3877 | -0.236 | -0.062 | -2026.93 | 3890.71 | 6.415 | -0.149 | 0.128 | -0.174 | 0.087 |
| 13 | 3530 | -0.243 | -0.073 | -1951.91 | 3497.17 | 5.237 | -0.158 | 0.148 | -0.169 | 0.085 |
| 14 | 3557 | -0.253 | -0.077 | -1967.94 | 3473.76 | 4.060 | -0.165 | 0.154 | -0.177 | 0.088 |
| 15 | 3557 | -0.252 | -0.076 | -1908.03 | 3469.82 | 4.064 | -0.164 | 0.153 | -0.176 | 0.088 |
| 16 | 3863 | -0.243 | -0.061 | -2151.82 | 3849.90 | 3.280 | -0.152 | 0.127 | -0.182 | 0.091 |
| 17 | 3891 | -0.254 | -0.072 | -2167.85 | 3825.35 | 5.543 | -0.163 | 0.146 | -0.182 | 0.091 |
| 18 | 3891 | -0.253 | -0.066 | -2107.94 | 3821.90 | 5.259 | -0.160 | 0.136 | -0.187 | 0.094 |
| 19 | 3839 | -0.239 | -0.059 | -2091.90 | 3882.01 | 4.698 | -0.149 | 0.124 | -0.180 | 0.090 |
| 20 | 3096 | -0.247 | -0.072 | -1616.85 | 3038.63 | 5.376 | -0.159 | 0.145 | -0.175 | 0.087 |
| 21 | 3121 | -0.257 | -0.075 | -1632.88 | 3015.46 | 4.319 | -0.166 | 0.152 | -0.182 | 0.091 |
| 22 | 3121 | -0.256 | -0.074 | -1572.97 | 3012.50 | 4.249 | -0.165 | 0.150 | -0.182 | 0.091 |
| 23 | 2572 | -0.250 | -0.069 | -1481.92 | 2835.29 | 6.603 | -0.160 | 0.140 | -0.181 | 0.091 |
| 24 | 2164 | -0.238 | -0.055 | -1219.88 | 2381.12 | 8.735 | -0.146 | 0.117 | -0.183 | 0.092 |
| 25 | 2395 | -0.238 | -0.061 | -1335.71 | 2538.94 | 9.936 | -0.150 | 0.127 | -0.177 | 0.088 |
| 26 | 2395 | -0.250 | -0.065 | -1275.24 | 2518.37 | 7.016 | -0.158 | 0.135 | -0.185 | 0.092 |
| 27 | 2659 | -0.250 | -0.065 | -1313.31 | 2619.85 | 7.062 | -0.158 | 0.135 | -0.185 | 0.093 |
| 28 | 3369 | -0.235 | -0.065 | -1531.28 | 3262.64 | 5.760 | -0.150 | 0.133 | -0.170 | 0.085 |
| 29 | 3071 | -0.256 | -0.085 | -1472.28 | 2870.97 | 5.118 | -0.170 | 0.169 | -0.171 | 0.085 |
| 30 | 3071 | -0.255 | -0.084 | -1412.38 | 2866.21 | 5.348 | -0.169 | 0.168 | -0.171 | 0.085 |
| 31 | 3146 | -0.247 | -0.068 | -1632.89 | 3055.06 | 5.264 | -0.157 | 0.138 | -0.179 | 0.090 |
| 32 | 3341 | -0.248 | -0.075 | -2092.48 | 3403.91 | 4.407 | -0.162 | 0.151 | -0.173 | 0.087 |
| 33 | 3569 | -0.245 | -0.066 | -1747.42 | 3419.32 | 3.059 | -0.156 | 0.136 | -0.179 | 0.090 |
| 34 | 4079 | -0.243 | -0.063 | -2243.06 | 4111.44 | 2.536 | -0.153 | 0.130 | -0.180 | 0.090 |
| 35 | 3884 | -0.263 | -0.056 | -2279.12 | 4048.20 | 3.461 | -0.159 | 0.122 | -0.207 | 0.104 |
| 36 | 3456 | -0.268 | -0.063 | -2164.61 | 3592.79 | 4.508 | -0.165 | 0.134 | -0.204 | 0.102 |
| 37 | 2646 | -0.256 | -0.052 | -1942.06 | 3400.17 | 1.666 | -0.154 | 0.116 | -0.204 | 0.102 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 38 | 3884 | -0.264 | -0.062 | -2295.72 | 4077.01 | 3.944 | -0.163 | 0.131 | -0.202 | 0.101 |
| 39 | 4137 | -0.259 | -0.079 | -2754.75 | 4285.35 | 1.913 | -0.169 | 0.158 | -0.180 | 0.090 |
| 40 | 4137 | -0.259 | -0.083 | -2394.40 | 4279.02 | 1.548 | -0.171 | 0.165 | -0.177 | 0.088 |

**Table S5. Selected descriptors and the inhibitory activities of $pIC_{50}$ order for forty molecules**

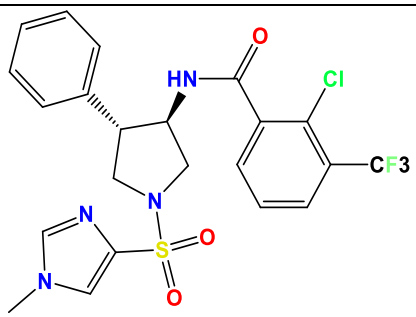
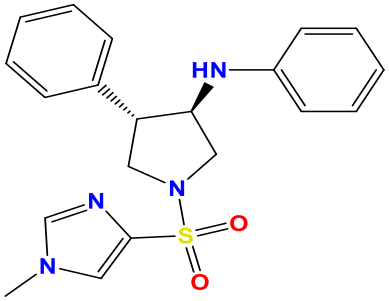
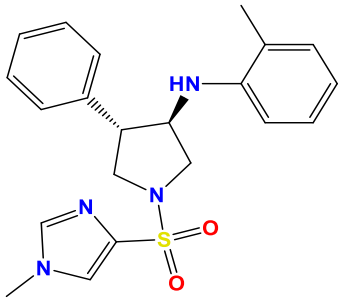
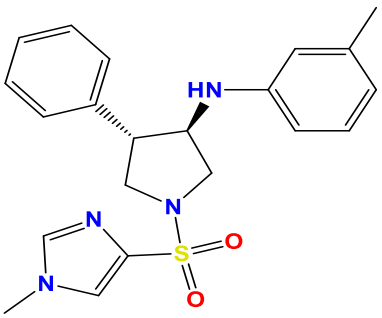
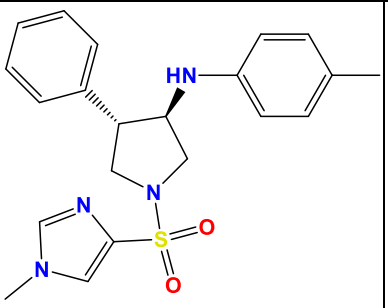
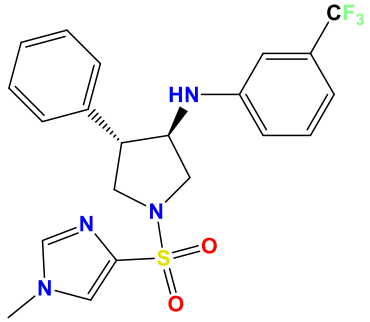
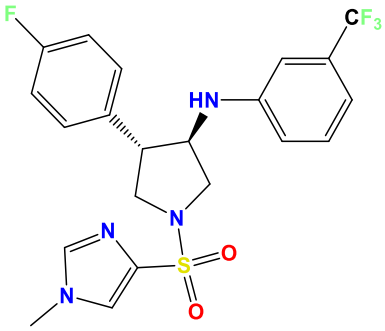
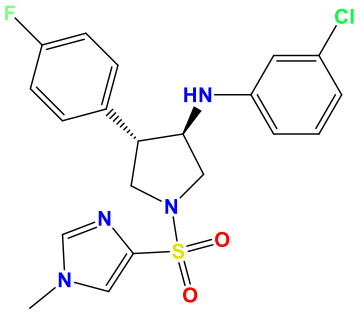
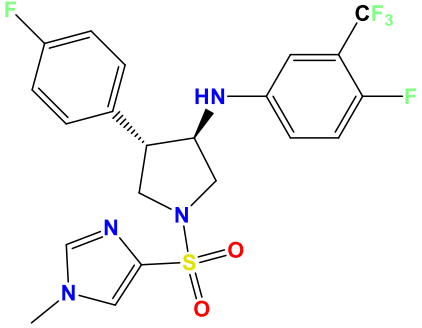
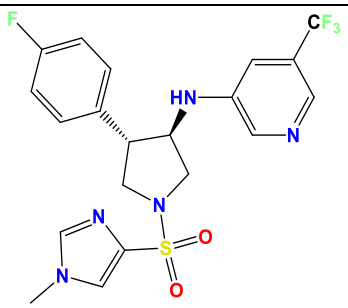
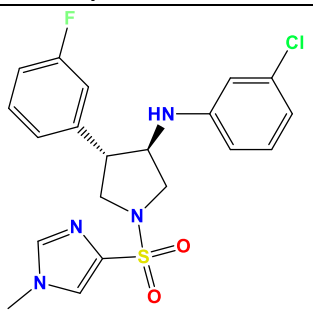
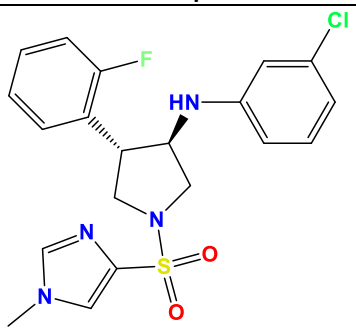
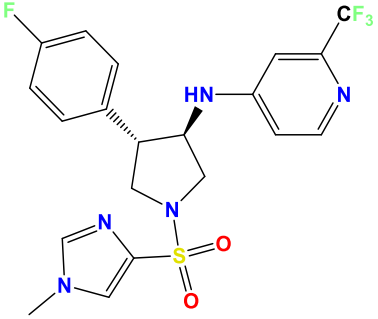
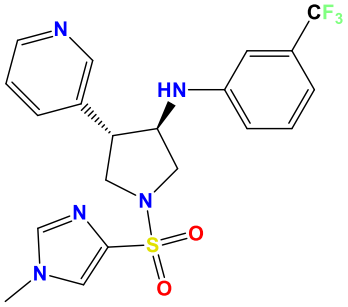
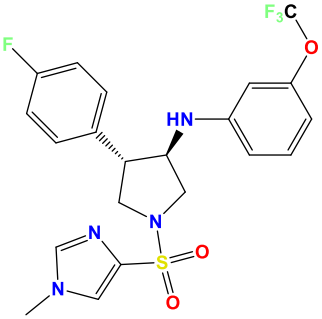
| N° | MW | NRB | Log P | TD | PSA | $pIC_{50}$ |
|---|---|---|---|---|---|---|
| 1 | 479.4700 | 7 | 2.7510 | 15 | 79.2800 | 7.8240 |
| 2 | 497.4600 | 7 | 2.9800 | 15 | 79.2800 | 8.0970 |
| 3 | 493.5000 | 7 | 3.2500 | 15 | 79.2800 | 7.6380 |
| 4* | 498.4500 | 7 | 1.8360 | 15 | 91.6400 | 7.6380 |
| 5 | 494.4900 | 7 | 2.1720 | 15 | 91.6400 | 7.9210 |
| 6 | 471.4900 | 7 | 2.2840 | 14 | 79.2800 | 7.7700 |
| 7 | 443.4900 | 6 | 2.7160 | 14 | 79.2800 | 7.6380 |
| 8 | 471.5400 | 7 | 3.7340 | 15 | 79.2800 | 8.0000 |
| 9 | 455.5000 | 7 | 2.8510 | 15 | 79.2800 | 7.4200 |
| 10* | 504.5200 | 5 | 2.2480 | 16 | 73.2900 | 7.7960 |
| 11 | 505.5100 | 5 | 1.1040 | 16 | 85.6500 | 8.0000 |
| 12 | 501.5500 | 5 | 1.4400 | 16 | 85.6500 | 8.1550 |
| 13 | 479.5000 | 5 | 3.8230 | 15 | 70.0500 | 8.0000 |
| 14* | 480.4900 | 5 | 2.6780 | 15 | 82.4100 | 8.2220 |
| 15 | 476.5200 | 5 | 3.0140 | 15 | 82.4100 | 8.3980 |
| 16 | 511.4900 | 8 | 3.5090 | 15 | 79.2800 | 7.5380 |
| 17* | 512.4800 | 8 | 2.3650 | 15 | 91.6400 | 7.5690 |
| 18* | 508.5100 | 8 | 2.7010 | 15 | 91.6400 | 7.9210 |
| 19* | 507.5300 | 8 | 3.5890 | 15 | 79.2800 | 6.3910 |
| 20 | 444.3800 | 7 | 3.9410 | 15 | 68.9300 | 7.1080 |
| 21 | 445.3700 | 7 | 2.7960 | 15 | 81.2900 | 7.1610 |
| 22 | 441.4100 | 7 | 3.1320 | 15 | 81.2900 | 7.7450 |
| 23 | 418.4100 | 7 | 3.2450 | 14 | 68.9300 | 6.6780 |
| 24 | 372.4200 | 6 | 3.4480 | 14 | 68.9300 | 6.9790 |
| 25 | 391.4000 | 6 | 2.5320 | 14 | 81.2900 | 6.9470 |
| 26 | 387.4300 | 6 | 2.8680 | 14 | 81.2900 | 7.6780 |
| 27 | 399.4500 | 7 | 3.0030 | 15 | 81.2900 | 6.3560 |
| 28 | 448.4700 | 5 | 2.4100 | 16 | 75.3000 | 7.6990 |
| 29* | 427.4100 | 5 | 3.5040 | 15 | 72.0600 | 6.4170 |
| 30 | 423.4400 | 5 | 3.8400 | 15 | 72.0600 | 7.2600 |
| 31* | 445.3700 | 7 | 2.7960 | 16 | 81.2900 | 6.8480 |
| 32 | 479.8100 | 7 | 3.3940 | 16 | 81.2900 | 7.4810 |
| 33 | 475.4000 | 8 | 2.8140 | 16 | 90.5200 | 7.6780 |
| 34 | 528.4800 | 8 | 1.3800 | 16 | 100.8700 | 7.7210 |
| 35 | 525.4600 | 9 | 2.6440 | 15 | 72.9100 | 7.5690 |
| 36 | 495.4400 | 8 | 3.0990 | 15 | 63.6800 | 6.2010 |
| 37 | 457.4600 | 8 | 1.7610 | 13 | 72.9100 | 6.6580 |
| 38 | 526.4500 | 9 | 1.2940 | 15 | 85.2700 | 7.3670 |
| 39 | 544.4400 | 9 | 1.4670 | 15 | 85.2700 | 7.5850 |
| 40 | 560.8900 | 9 | 2.0370 | 15 | 85.2700 | 7.8240 |

* Indicates the test set molecules

**Table S6. Results of one hundred randomizations using Y-Randomization test**

| Model | *R* | *R^2* | *Q^2* | *Model* | *R* | *R^2* | *Q^2* |
|---|---|---|---|---|---|---|---|
| Original | 0.852511 | 0.726776 | 0.54939 | Random N°51 | 0.518712 | 0.269062 | -0.05893 |
| Random N°1 | 0.289288 | 0.083688 | -0.4153 | Random N°52 | 0.397863 | 0.158295 | -0.30626 |
| Random N°2 | 0.218451 | 0.047721 | -0.51687 | Random N°53 | 0.23627 | 0.055824 | -0.52824 |
| Random N°3 | 0.336045 | 0.112926 | -0.25546 | Random N°54 | 0.549873 | 0.30236 | -0.11778 |
| Random N°4 | 0.32185 | 0.103587 | -0.35123 | Random N°55 | 0.200885 | 0.040355 | -0.42519 |
| Random N°5 | 0.384968 | 0.1482 | -0.4219 | Random N°56 | 0.194728 | 0.037919 | -0.71804 |
| Random N°6 | 0.380524 | 0.144798 | -0.25259 | Random N°57 | 0.348293 | 0.121308 | -0.48519 |
| Random N°7 | 0.398988 | 0.159191 | -0.35729 | Random N°58 | 0.385678 | 0.148747 | -0.39898 |
| Random N°8 | 0.518043 | 0.268369 | -0.2249 | Random N°59 | 0.167215 | 0.027961 | -0.30552 |
| Random N°9 | 0.387422 | 0.150096 | -0.24648 | Random N°60 | 0.394593 | 0.155704 | -0.29152 |
| Random N°10 | 0.616245 | 0.379757 | 0.076575 | Random N°61 | 0.32596 | 0.10625 | -0.3653 |
| Random N°11 | 0.310023 | 0.096114 | -0.3883 | Random N°62 | 0.256014 | 0.065543 | -0.42633 |
| Random N°12 | 0.195321 | 0.03815 | -0.52226 | Random N°63 | 0.572911 | 0.328227 | 0.063455 |
| Random N°13 | 0.50862 | 0.258695 | -0.24645 | Random N°64 | 0.277236 | 0.07686 | -0.45402 |
| Random N°14 | 0.447397 | 0.200164 | -0.32124 | Random N°65 | 0.393112 | 0.154537 | -0.42287 |
| Random N°15 | 0.26287 | 0.069101 | -0.35577 | Random N°66 | 0.419765 | 0.176203 | -0.39675 |
| Random N°16 | 0.499976 | 0.249976 | -0.26118 | Random N°67 | 0.484096 | 0.234349 | -0.11161 |
| Random N°17 | 0.431383 | 0.186091 | -0.1933 | Random N°68 | 0.513383 | 0.263562 | -0.01946 |
| Random N°18 | 0.214761 | 0.046122 | -0.44086 | Random N°69 | 0.282557 | 0.079838 | -0.48155 |
| Random N°19 | 0.367734 | 0.135228 | -0.35575 | Random N°70 | 0.492596 | 0.242651 | -0.17683 |
| Random N°20 | 0.357737 | 0.127975 | -0.29203 | Random N°71 | 0.320948 | 0.103008 | -0.27816 |
| Random N°21 | 0.456529 | 0.208419 | -0.12905 | Random N°72 | 0.43973 | 0.193362 | -0.22201 |
| Random N°22 | 0.336251 | 0.113065 | -0.29033 | Random N°73 | 0.449733 | 0.20226 | -0.22051 |
| Random N°23 | 0.178759 | 0.031955 | -0.4679 | Random N°74 | 0.162709 | 0.026474 | -0.4942 |
| Random N°24 | 0.238333 | 0.056803 | -0.35035 | Random N°75 | 0.388229 | 0.150721 | -0.29389 |
| Random N°25 | 0.31313 | 0.09805 | -0.38015 | Random N°76 | 0.513209 | 0.263383 | -0.21509 |
| Random N°26 | 0.372394 | 0.138677 | -0.33561 | Random N°77 | 0.207093 | 0.042888 | -0.53707 |
| Random N°27 | 0.243272 | 0.059181 | -0.32137 | Random N°78 | 0.48733 | 0.237491 | -0.23485 |
| Random N°28 | 0.476057 | 0.22663 | -0.15845 | Random N°79 | 0.485017 | 0.235242 | -0.30532 |
| Random N°29 | 0.490166 | 0.240263 | -0.10811 | Random N°80 | 0.224843 | 0.050555 | -0.37172 |
| Random N°30 | 0.308319 | 0.095061 | -0.25374 | Random N°81 | 0.401996 | 0.161601 | -0.24787 |
| Random N°31 | 0.384067 | 0.147508 | -0.27224 | Random N°82 | 0.23471 | 0.055089 | -0.43202 |
| Random N°32 | 0.596179 | 0.35543 | -0.00805 | Random N°83 | 0.410004 | 0.168103 | -0.20028 |
| Random N°33 | 0.552442 | 0.305192 | -0.06271 | Random N°84 | 0.620674 | 0.385236 | -0.10598 |
| Random N°34 | 0.60402 | 0.36484 | 0.077637 | Random N°85 | 0.28267 | 0.079902 | -0.42226 |
| Random N°35 | 0.15598 | 0.02433 | -0.55712 | Random N°86 | 0.373135 | 0.139229 | -0.2539 |
| Random N°36 | 0.297145 | 0.088295 | -0.47628 | Random N°87 | 0.432306 | 0.186888 | -0.29103 |
| Random N°37 | 0.603349 | 0.36403 | -0.15549 | Random N°88 | 0.40657 | 0.165299 | -0.1807 |
| Random N°38 | 0.292283 | 0.08543 | -0.47085 | Random N°89 | 0.273884 | 0.075012 | -0.30891 |
| Random N°39 | 0.459094 | 0.210767 | -0.22007 | Random N°90 | 0.297711 | 0.088632 | -0.30545 |
| Random N°40 | 0.350976 | 0.123184 | -0.28731 | Random N°91 | 0.521434 | 0.271894 | -0.08235 |
| Random N°41 | 0.417663 | 0.174442 | -0.27241 | Random N°92 | 0.271847 | 0.073901 | -0.45944 |
| Random N°42 | 0.490464 | 0.240555 | -0.24261 | Random N°93 | 0.631859 | 0.399245 | 0.063478 |
| Random N°43 | 0.181515 | 0.032948 | -0.38556 | Random N°94 | 0.467017 | 0.218105 | -0.36379 |
| Random N°44 | 0.433603 | 0.188011 | -0.21744 | Random N°95 | 0.256243 | 0.065661 | -0.47306 |
| Random N°45 | 0.325239 | 0.105781 | -0.4834 | Random N°96 | 0.320073 | 0.102447 | -0.27401 |
| Random N°46 | 0.455845 | 0.207795 | -0.37583 | Random N°97 | 0.183087 | 0.033521 | -0.32641 |
| Random N°47 | 0.549193 | 0.301613 | -0.25478 | Random N°98 | 0.276311 | 0.076348 | -0.35485 |
| Random N°48 | 0.32562 | 0.106028 | -0.47706 | Random N°99 | 0.501475 | 0.251477 | -0.16823 |
| Random N°49 | 0.222447 | 0.049483 | -0.35725 | RandomN°100 | 0.498164 | 0.248167 | -0.13354 |
| Random N°50 | 0.413546 | 0.17102 | -0.29646 | | | | |

**Table S7. 2D structures of chemical compounds and their biological activities of pKi order.**

| | | |
|---|---|---|
|  1 pKi=6.703 |  2 pKi=6.701 |  3 pKi=5.821 |
|  4 pKi=7.745 |  5 pKi=5.903 |  6 pKi=8.097 |
|  7 pKi=8.523 |  8 pKi=8.523 |  9 pKi=8.301 |
|  10 pKi=5.936 |  11 pKi=7.377 |  12 pKi=7.678 |
|  13 pKi=6.928 |  14 pKi=6.116 |  15 pKi=8.398 |

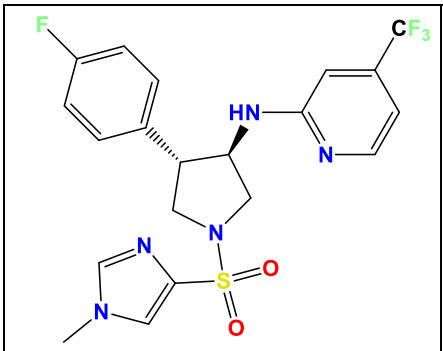
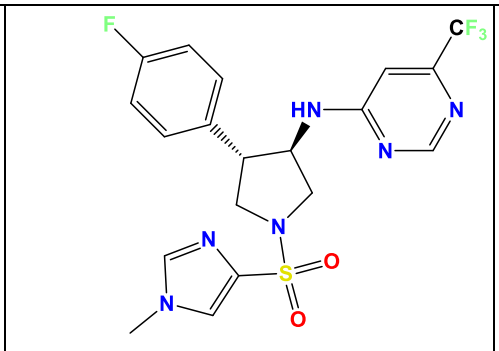
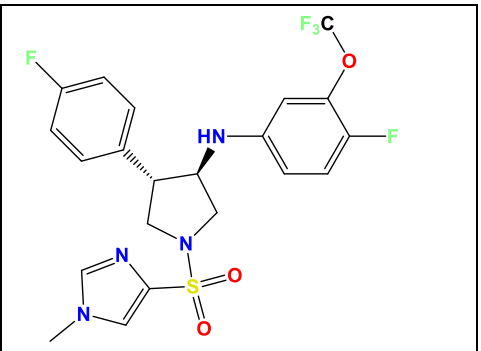
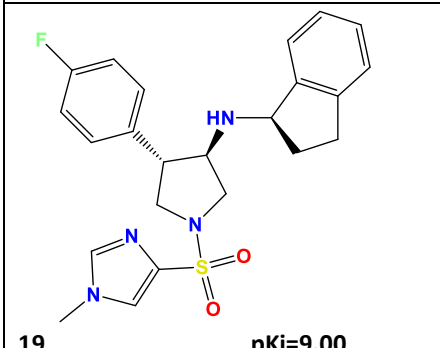
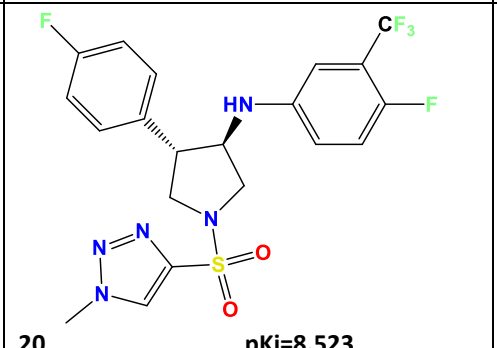
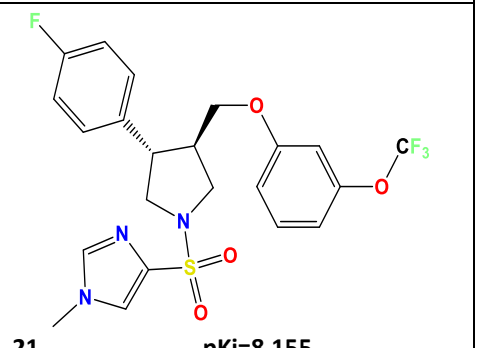
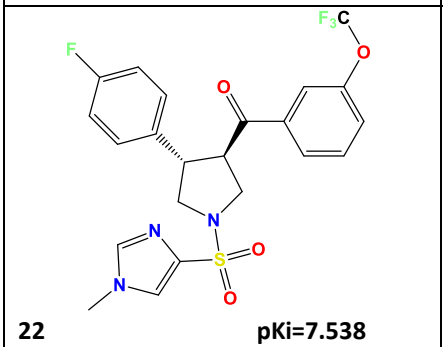
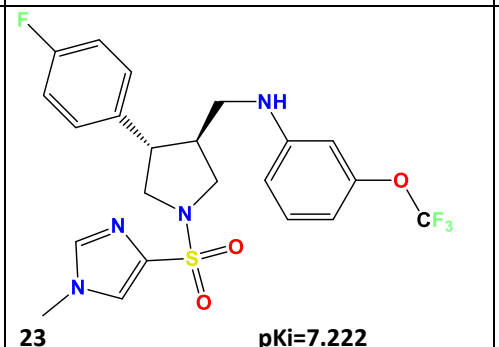
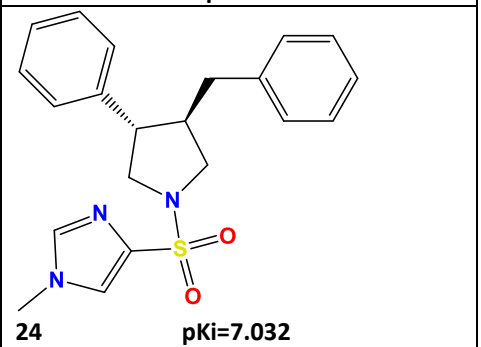
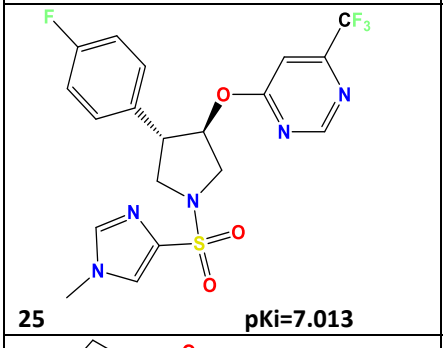
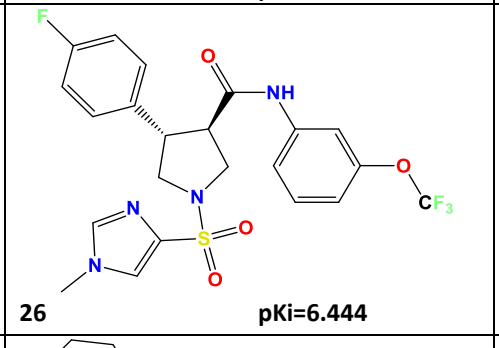
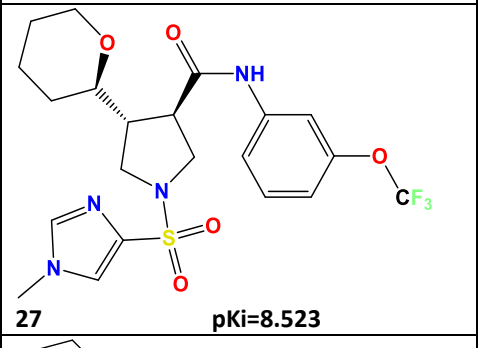
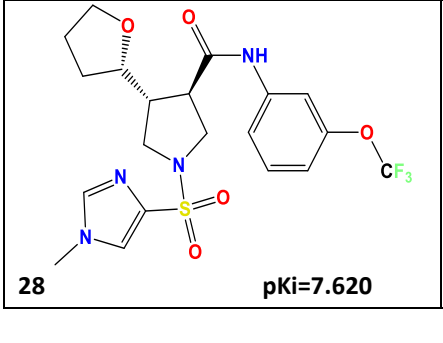
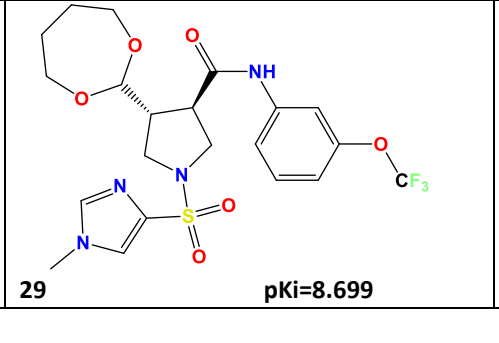
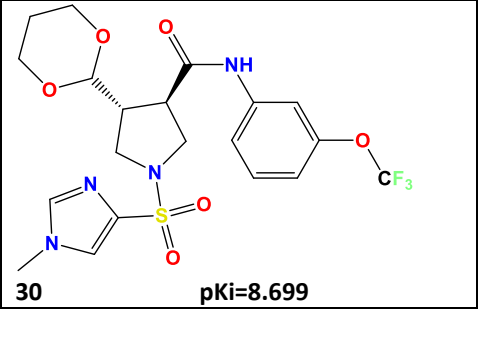
| | | |
|---|---|---|
|  16 pKi=8.097 |  17 pKi=8.301 |  18 pKi=8.699 |
|  19 pKi=9.00 |  20 pKi=8.523 |  21 pKi=8.155 |
|  22 pKi=7.538 |  23 pKi=7.222 |  24 pKi=7.032 |
|  25 pKi=7.013 |  26 pKi=6.444 |  27 pKi=8.523 |
|  28 pKi=7.620 |  29 pKi=8.699 |  30 pKi=8.699 |

**Table S8. Selected descriptors of studied compounds with inhibitory activities of pKi order**

| N° | % H | % C | % N | D | T E | VDW E | pKi |
|---|---|---|---|---|---|---|---|
| 1 | 3.930 | 10.920 | 9.360 | 1.490 | 58.678 | 19.795 | 6.703 |
| 2* | 5.800 | 14.650 | 8.370 | 1.310 | 35.378 | 18.322 | 6.701 |
| 3 | 6.100 | 14.130 | 8.070 | 1.290 | 36.789 | 19.478 | 5.821 |
| 4 | 6.100 | 14.130 | 8.070 | 1.290 | 34.719 | 18.945 | 7.745 |
| 5 | 6.100 | 14.130 | 8.070 | 1.290 | 34.680 | 18.964 | 5.903 |
| 6 | 4.700 | 12.440 | 7.100 | 1.400 | 39.322 | 18.992 | 8.097 |
| 7 | 4.300 | 11.960 | 6.830 | 1.440 | 39.315 | 18.886 | 8.523 |
| 8 | 4.640 | 12.880 | 7.360 | 1.430 | 35.361 | 18.733 | 8.523 |
| 9 | 3.940 | 11.520 | 6.580 | 1.480 | 40.323 | 18.962 | 8.301 |
| 10* | 4.080 | 14.920 | 6.820 | 1.500 | 44.897 | 19.276 | 5.936 |
| 11* | 4.640 | 12.880 | 7.360 | 1.430 | 35.363 | 18.728 | 7.377 |
| 12 | 4.640 | 12.880 | 7.360 | 1.430 | 35.991 | 18.930 | 7.678 |
| 13* | 4.080 | 14.920 | 6.820 | 1.500 | 41.241 | 19.404 | 6.928 |
| 14 | 4.470 | 15.510 | 7.090 | 1.450 | 44.863 | 19.457 | 6.116 |
| 15 | 4.160 | 11.560 | 9.910 | 1.460 | 35.610 | 20.071 | 8.398 |
| 16 | 4.080 | 14.920 | 6.820 | 1.500 | 45.988 | 19.668 | 8.097 |
| 17 | 3.860 | 17.860 | 6.800 | 1.560 | 44.147 | 20.191 | 8.301 |
| 18 | 3.810 | 11.150 | 9.550 | 1.500 | 39.296 | 20.541 | 8.699 |
| 19* | 5.720 | 12.720 | 7.260 | 1.400 | 40.735 | 19.063 | 9.000 |
| 20 | 3.720 | 14.370 | 6.560 | 1.540 | 40.651 | 18.898 | 8.523 |
| 21 | 4.240 | 8.410 | 12.810 | 1.430 | 38.144 | 22.613 | 8.155 |
| 22 | 3.850 | 8.450 | 12.860 | 1.460 | 39.408 | 22.598 | 7.538 |
| 23 | 4.450 | 11.240 | 9.630 | 1.430 | 35.626 | 20.613 | 7.222 |
| 24* | 6.080 | 11.010 | 8.390 | 1.260 | 37.612 | 19.037 | 7.032 |
| 25 | 3.630 | 14.860 | 10.180 | 1.550 | 45.449 | 22.519 | 7.013 |
| 26 | 3.930 | 10.930 | 12.490 | 1.490 | 37.856 | 19.281 | 6.444 |
| 27 | 5.010 | 11.150 | 15.920 | 1.510 | 50.346 | 20.051 | 8.523 |
| 28 | 4.750 | 11.470 | 16.380 | 1.550 | 54.607 | 16.254 | 7.620 |
| 29 | 4.600 | 11.110 | 19.030 | 1.560 | 50.711 | 23.759 | 8.699 |
| 30 | 4.600 | 11.110 | 19.030 | 1.560 | 50.846 | 22.774 | 8.699 |

* Indicates the test set molecules

**Table S9. Statistical results for 100 randomizations**

| Model | *R* | *R^2* | *Q^2* | Model | *R* | *R^2* | *Q^2* |
|---|---|---|---|---|---|---|---|
| ***Original*** | ***0.864*** | ***0.746*** | ***0.529*** | Random N°51 | 0.678 | 0.460 | -0.035 |
| Random N°1 | 0.480 | 0.231 | -0.538 | Random N°52 | 0.443 | 0.196 | -0.454 |
| Random N°2 | 0.509 | 0.259 | -0.525 | Random N°53 | 0.373 | 0.139 | -1.193 |
| Random N°3 | 0.312 | 0.097 | -1.937 | Random N°54 | 0.628 | 0.395 | -0.172 |
| Random N°4 | 0.444 | 0.197 | -1.806 | Random N°55 | 0.380 | 0.144 | -0.836 |
| Random N°5 | 0.551 | 0.304 | -0.706 | Random N°56 | 0.384 | 0.148 | -0.877 |
| Random N°6 | 0.396 | 0.157 | -1.030 | Random N°57 | 0.489 | 0.239 | -0.676 |
| Random N°7 | 0.635 | 0.403 | -0.147 | Random N°58 | 0.621 | 0.385 | -0.195 |
| Random N°8 | 0.487 | 0.237 | -1.430 | Random N°59 | 0.603 | 0.364 | -0.365 |
| Random N°9 | 0.239 | 0.057 | -0.983 | Random N°60 | 0.433 | 0.188 | -1.198 |
| Random N°10 | 0.509 | 0.259 | -0.329 | Random N°61 | 0.633 | 0.401 | -0.534 |
| Random N°11 | 0.625 | 0.391 | -0.393 | Random N°62 | 0.423 | 0.179 | -0.626 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Random N°12 | 0.630 | 0.397 | -0.325 | Random N°63 | 0.575 | 0.330 | -0.297 |
| Random N°13 | 0.399 | 0.160 | -0.610 | Random N°64 | 0.341 | 0.117 | -1.395 |
| Random N°14 | 0.531 | 0.282 | -0.433 | Random N°65 | 0.533 | 0.284 | -1.631 |
| Random N°15 | 0.525 | 0.275 | -0.718 | Random N°66 | 0.537 | 0.288 | -0.405 |
| Random N°16 | 0.457 | 0.209 | -1.052 | Random N°67 | 0.300 | 0.090 | -1.331 |
| Random N°17 | 0.461 | 0.212 | -0.572 | Random N°68 | 0.301 | 0.091 | -1.246 |
| Random N°18 | 0.402 | 0.162 | -0.636 | Random N°69 | 0.559 | 0.312 | -0.433 |
| Random N°19 | 0.498 | 0.248 | -0.575 | Random N°70 | 0.481 | 0.231 | -0.631 |
| Random N°20 | 0.359 | 0.129 | -0.689 | Random N°71 | 0.519 | 0.269 | -0.555 |
| Random N°21 | 0.360 | 0.129 | -1.309 | Random N°72 | 0.431 | 0.185 | -0.476 |
| Random N°22 | 0.596 | 0.356 | -0.254 | Random N°73 | 0.587 | 0.345 | -0.247 |
| Random N°23 | 0.388 | 0.151 | -0.805 | Random N°74 | 0.661 | 0.436 | -0.024 |
| Random N°24 | 0.553 | 0.306 | -0.280 | Random N°75 | 0.326 | 0.106 | -0.920 |
| Random N°25 | 0.467 | 0.218 | -0.551 | Random N°76 | 0.490 | 0.240 | -0.382 |
| Random N°26 | 0.407 | 0.166 | -0.496 | Random N°77 | 0.473 | 0.224 | -0.471 |
| Random N°27 | 0.598 | 0.357 | -0.196 | Random N°78 | 0.655 | 0.430 | -0.000 |
| Random N°28 | 0.411 | 0.169 | -0.526 | Random N°79 | 0.388 | 0.151 | -0.682 |
| Random N°29 | 0.500 | 0.250 | -0.549 | Random N°80 | 0.359 | 0.129 | -1.036 |
| Random N°30 | 0.602 | 0.362 | -0.318 | Random N°81 | 0.633 | 0.401 | -0.616 |
| Random N°31 | 0.540 | 0.291 | -0.907 | Random N°82 | 0.545 | 0.297 | -0.527 |
| Random N°32 | 0.476 | 0.227 | -1.029 | Random N°83 | 0.381 | 0.145 | -0.621 |
| Random N°33 | 0.535 | 0.286 | -0.267 | Random N°84 | 0.614 | 0.377 | -0.388 |
| Random N°34 | 0.294 | 0.086 | -0.877 | Random N°85 | 0.734 | 0.539 | 0.001 |
| Random N°35 | 0.381 | 0.145 | -0.634 | Random N°86 | 0.329 | 0.108 | -0.672 |
| Random N°36 | 0.452 | 0.205 | -0.551 | Random N°87 | 0.655 | 0.429 | -0.300 |
| Random N°37 | 0.471 | 0.222 | -0.889 | Random N°88 | 0.374 | 0.140 | -0.670 |
| Random N°38 | 0.620 | 0.385 | -0.337 | Random N°89 | 0.279 | 0.078 | -0.518 |
| Random N°39 | 0.523 | 0.273 | -0.278 | Random N°90 | 0.726 | 0.528 | 0.055 |
| Random N°40 | 0.629 | 0.395 | 0.012 | Random N°91 | 0.341 | 0.116 | -1.155 |
| Random N°41 | 0.239 | 0.057 | -0.960 | Random N°92 | 0.459 | 0.211 | -0.353 |
| Random N°42 | 0.482 | 0.233 | -0.791 | Random N°93 | 0.435 | 0.189 | -0.666 |
| Random N°43 | 0.449 | 0.202 | -0.517 | Random N°94 | 0.339 | 0.115 | -2.102 |
| Random N°44 | 0.400 | 0.160 | -1.292 | Random N°95 | 0.655 | 0.429 | -0.417 |
| Random N°45 | 0.428 | 0.184 | -0.829 | Random N°96 | 0.288 | 0.083 | -1.104 |
| Random N°46 | 0.398 | 0.159 | -0.538 | Random N°97 | 0.264 | 0.070 | -0.982 |
| Random N°47 | 0.692 | 0.479 | -0.063 | Random N°98 | 0.396 | 0.156 | -1.175 |
| Random N°48 | 0.385 | 0.148 | -0.990 | Random N°99 | 0.464 | 0.216 | -0.511 |
| Random N°49 | 0.408 | 0.166 | -0.569 | Random N°100 | 0.494 | 0.244 | -0.755 |
| Random N°50 | 0.699 | 0.489 | -0.141 | | | | |