ANTÔNIO PEDRO LEMOS DE MESQUITA

THERMODYNAMIC POTENTIAL CALCULATIONS VIA CLASSICAL MOLECULAR DYNAMICS: A VIRTUAL SCREENING STRATEGY FOR IDENTIFYING mGluR7/8 AND 5-HT_{1A} AGONISTS FOR AUTISM SPECTRUM DISORDER

LAVRAS – MG
2025

ANTÔNIO PEDRO LEMOS DE MESQUITA

THERMODYNAMIC POTENTIAL CALCULATIONS VIA CLASSICAL MOLECULAR DYNAMICS: A VIRTUAL SCREENING STRATEGY FOR IDENTIFYING mGluR7/8 AND 5-HT_{1A} AGONISTS FOR AUTISM SPECTRUM DISORDER

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-Graduação Multicêntrico em Química de Minas Gerais, área de concentração em Química, para a obtenção do título de Mestre.

Orientador
Prof. DSc. Teodorico de Castro Ramalho

Coorientadora
Prof.ª DSc. Elaine Fontes Ferreira da Cunha

LAVRAS – MG
2025

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Lemos de Mesquita, Antônio Pedro
Thermodynamic Potential Calculations via Classical Molecular Dynamics: A Virtual Screening Strategy for Identifying mGluR7/8 and 5-HT_{1A} Agonists for Autism Spectrum Disorder / Antônio Pedro Lemos de Mesquita. – Lavras : UFLA, 2025.
158 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2025.
Orientador: Teodorico de Castro Ramalho.
Coorientadora: Elaine Fontes Ferreira da Cunha.
Bibliografia.

ANTÔNIO PEDRO LEMOS DE MESQUITA

CÁLCULOS DE POTENCIAIS TERMODINÂMICOS POR DINÂMICA MOLECULAR CLÁSSICA: UMA ESTRATÉGIA DE TRIAGEM VIRTUAL NA IDENTIFICAÇÃO DE AGONISTAS DE mGluR7/8 E 5-HT_{1A} PARA O TRANSTORNO DO ESPECTRO AUTISTA

THERMODYNAMIC POTENTIAL CALCULATIONS VIA CLASSICAL MOLECULAR DYNAMICS: A VIRTUAL SCREENING STRATEGY FOR IDENTIFYING mGluR7/8 AND 5-HT_{1A} AGONISTS FOR AUTISM SPECTRUM DISORDER

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-Graduação Multicêntrico em Química de Minas Gerais, área de concentração em Química, para a obtenção do título de Mestre.

APROVADA em 18 de julho de 2025.
DSc. Fernando Martins dos Santos Júnior UFF
DSc. Fernando Henrique Ferrari Alves UFLA

Orientador
Prof. DSc. Teodorico de Castro Ramalho

Coorientadora
Prof.ª DSc. Elaine Fontes Ferreira da Cunha

LAVRAS – MG
2025

A todos os estudantes do mundo que, por guerra, perseguição, morte e outras violações de direitos humanos não puderam concluir seus estudos e pesquisas. Uma homenagem minúscula de alguém que gostaria de fazer muito mais.

Dedico

AGRADECIMENTOS

À minha família base, minha mãe e meu irmão, pelo apoio incondicional, suporte e amor;

Ao Pedro, meu namorado, pelo suporte e companheirismo em todos os aspectos da minha jornada de pós graduação;

À Adriana, minha colega de laboratório que se tornou minha amiga, que me abriu portas de colaboração e me ensinou muito sobre estatística e QSAR. Obrigado por todo apoio, conversas e trocas de experiências;

À Ingrid, que me convidou para colaborar em um dos seus projetos de doutorado, que culminou na minha primeira publicação científica. Guardo esse trabalho com muito carinho;

À Letícia Miranda, pelos bons momentos compartilhados no laboratório, amizade, parceria, trabalho e disciplinas que pudemos compartilhar;

Aos meus orientadores, professor Teodorico e professora Elaine, que me acompanharam em todas as etapas, elucidando dúvidas, questionamentos, dando apoio e orientação;

À banca, professores Fernando Martins e Fernando Ferrari pela participação e contribuição. Aos membros suplentes, Thais e Matheus, pela disponibilidade e atenção;

Aos meus amigos Kika, Marcos, Gleicy, Frida e Stefany pelas conversas, desabafos e apoio emocional;

Às minhas tias maternas, Clara, Rita e Celena e às minhas tias paternas Margarida e Lígia por todo suporte e carinho;

Ao professor Matheus Puggina com quem pude colaborar em meu segundo artigo da dissertação e em mais outros dois com suas alunas de doutorado;

Ao programa de Pós-graduação Multicêntrico em Química de Minas Gerais pela oportunidade de realizar meu mestrado;

Ao professor Márcio Pozzobon, coordenador do programa;

À CAPES (código 001), pela concessão da bolsa de estudos que possibilitou minha jornada acadêmica até então;

À Universidade Federal de Lavras e ao Departamento de Química (Instituto de Ciências Naturais) pela infraestrutura e qualidade de ensino e pesquisa.

Muito obrigado!

“Tudo no mundo começou com um sim. Uma molécula disse sim a outra molécula e nasceu a vida. Mas antes da pré-história havia a pré-história da pré-história e havia o nunca e havia o sim. Não sei o quê, mas sei que o universo jamais começou.”

(Clarice Lispector)

RESUMO

A utilização de recursos computacionais para cálculos termodinâmicos foi uma revolução nas ciências químicas e apresenta aplicações indispensáveis em sistemas biológicos macromoleculares, como receptores proteicos/enzimáticos e ligantes pequenos. O emprego de cálculos rápidos, baseados em solvente implícito e aproximações de ensembles estatísticos por dinâmica molecular clássica vem ganhando destaque por seus resultados numericamente viáveis e custo computacional drasticamente reduzido em comparação a métodos padrões, como integração termodinâmica e cálculos perturbativos. A aplicação central proposta neste trabalho é dada em protocolos de triagem virtual, em que são necessários cálculos computacionais rápidos e eficientes, para que sejam testadas propostas de diversos ligantes farmacologicamente viáveis em seus respectivos alvos biológicos. A metodologia empregada envolve busca por ligantes estrutural e/ou farmacoforicamente similares a compostos de referência em bancos de moléculas de acesso público, cálculos de ancoramento molecular estatisticamente validados, com caráter discriminatório de atividade biológica, construção de modelos de QSAR baseado em machine learning, predições de propriedades biológicas e físico-químicas de absorção, distribuição, metabolismo, excreção e toxicologia (ADMET), simulações de dinâmica molecular clássica e cálculos de energia livre por QM/MM-PB/GBSA e paisagens de energia livre. Neste trabalho foram propostos dois manuscritos científicos utilizando essas metodologias, com aplicação à descoberta de agonistas de mGluR7/8 e 5-HT_{1A}, proteínas altamente correlacionadas com a patogênese do transtorno do espectro autista em humanos, uma condição neuropsiquiátrica que afeta 1 a cada 100 crianças no mundo, com sintomas heterogêneos que se manifestam geralmente em comportamentos sociais restritos, gestos repetitivos, dificuldades de fala e cognição, com comorbidades graves, como depressão, ansiedade e ideação suicida. A aplicação de métodos computacionais rápidos e eficazes são de extrema importância, dado o caráter urgente da problemática e ao fato da inexistência de tratamentos quimioterápicos seguros. Como resultados desses estudos, identificamos 3 agonistas duais de mGluR7/8 com atividade computacional estabelecida e três agonistas seletivos de 5-HT_{1A}, sendo dois deles já em estágios avançados de testes clínicos.

Palavras-chave: dinâmica molecular; termodinâmica; simulações computacionais; autismo; desenvolvimento de fármacos; triagem virtual.

ABSTRACT

The use of computational resources for thermodynamic calculations has been a revolution in chemical sciences and plays an indispensable role in the study of macromolecular biological systems, such as protein/enzyme receptors and small ligands. The implementation of rapid calculations, based on implicit solvent models and approximations of statistical ensembles through classical molecular dynamics, has gained prominence due to its numerically feasible results and drastically reduced computational cost compared to standard methods such as thermodynamic integration and perturbative calculations. The central application proposed in this work lies in virtual screening protocols, which require fast and efficient computational calculations to test various pharmacologically viable ligands against their respective biological targets. The methodology employed involves the search for ligands that are structurally and/or pharmacophorically similar to reference compounds in publicly accessible molecular libraries, statistically validated molecular docking calculations with discriminatory power for biological activity, construction of QSAR models based on machine learning, predictions of biological and physicochemical ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, classical molecular dynamics simulations, and free energy calculations using QM/MM-PB/GBSA and free energy landscapes. In this work, two scientific manuscripts were proposed using these methodologies, with application to the discovery of agonists for mGluR7/8 and 5-HT_{1A}, proteins highly correlated with the pathogenesis of autism spectrum disorder (ASD) in humans – a neuropsychiatric condition affecting 1 in 100 children worldwide, characterized by heterogeneous symptoms such as restricted social behavior, repetitive gestures, speech and cognitive difficulties, and severe comorbidities like depression, anxiety, and suicidal ideation. The application of fast and effective computational methods is of utmost importance, given the urgent nature of this issue and the lack of safe chemotherapeutic treatments. As a result of these studies, we identified three dual agonists of mGluR7/8 with established computational activity and three selective 5-HT_{1A} agonists, two of which are already in advanced stages of clinical testing.

Keywords: molecular dynamics; thermodynamics; computational simulations; autism; drug development; virtual screening.

INDICADORES DE IMPACTO

Este trabalho apresenta impactos relevantes em potencial nas áreas tecnológica, social e de saúde, com ênfase na aplicação de métodos computacionais à descoberta de novos fármacos voltados ao tratamento de condições neuropsiquiátricas, como o transtorno do espectro autista (TEA). A utilização de protocolos eficientes de triagem virtual e cálculos termodinâmicos via simulações de dinâmica molecular com métodos clássicos e híbridos QM/MM-PB/GBSA, permite acelerar significativamente o processo de identificação de candidatos a fármacos, reduzindo os custos financeiros e o tempo normalmente demandado por métodos experimentais convencionais. Do ponto de vista tecnológico, os protocolos de triagem resultaram na identificação de seis candidatos a agonistas com alto potencial terapêutico (três duais para mGluR7/8 e três seletivos para 5-HT_{1A}), sendo dois deles já presentes em fases avançadas de testes clínicos, com potencial de reposicionamento. Tal contribuição pode ser incorporada por centros de pesquisa, universidades e indústrias farmacêuticas que atuam na triagem de compostos neuroativos. Em relação aos impactos sociais e de saúde, destaca-se a importância da pesquisa voltada ao TEA, condição que afeta cerca de 1 em cada 100 crianças no mundo, segundo estimativas da OMS. Os resultados obtidos contribuem para a construção de soluções terapêuticas mais seguras e direcionadas, suprindo lacunas críticas na farmacoterapia atual do autismo, que carece de tratamentos eficazes e com baixos efeitos colaterais. Potencialmente, esses impactos beneficiam populações em situação de vulnerabilidade, incluindo famílias de crianças autistas, profissionais da saúde e comunidades escolares. O trabalho se alinha prioritariamente à área temática 6 – Saúde, e secundariamente à área 7 – Tecnologia e Produção, conforme definido pela Política Nacional de Extensão Universitária. Quanto ao alinhamento com os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas, este trabalho contribui diretamente com os seguintes: ODS 3 – Saúde e Bem-Estar: ao promover o desenvolvimento de alternativas terapêuticas seguras e eficazes para o TEA; ODS 9 – Indústria, Inovação e Infraestrutura: ao propor novos modelos computacionais aplicáveis na indústria farmacêutica; ODS 10 – Redução das Desigualdades: por buscar soluções terapêuticas acessíveis para populações com TEA.

IMPACT INDICATORS

This work presents significant potential impacts in the technological, social, and health domains, with emphasis on the application of computational methods to the discovery of new drugs targeting neuropsychiatric disorders, such as autism spectrum disorder (ASD). The use of efficient virtual screening protocols and thermodynamic calculations via molecular dynamics simulations and classical and hybrid QM/MM-PB/GBSA methods enables a substantial acceleration in the identification of drug candidates, reducing both financial costs and the time typically required by conventional experimental approaches. From a technological perspective, the screening protocols led to the identification of six agonist candidates with high therapeutic potential (three dual agonists for mGluR7/8 and three selective for 5-HT_{1A}), two of which are already in advanced stages of clinical trials, with repurposing potential. This contribution may be integrated into the workflows of research centers, universities, and pharmaceutical industries involved in the screening of neuroactive compounds. In terms of social and health impacts, the focus on ASD is particularly relevant, as it is a condition affecting approximately 1 in every 100 children worldwide, according to WHO estimates. The results contribute to the development of safer and more targeted therapeutic strategies, addressing critical gaps in current autism pharmacotherapy, which still lacks effective treatments with minimal side effects. Potentially, these impacts benefit vulnerable populations, including families of autistic children, healthcare professionals, and school communities. This work aligns primarily with Thematic Area 6 – Health, and secondarily with Thematic Area 7 – Technology and Production, as defined by the National Extension Policy for Brazilian Universities. Regarding the alignment with the United Nations Sustainable Development Goals (SDGs), this work contributes directly to the following: SDG 3 – Good Health and Well-being: by promoting the development of safe and effective therapeutic alternatives for ASD; SDG 9 – Industry, Innovation, and Infrastructure: by proposing new computational models applicable to the pharmaceutical industry; SDG 10 – Reduced Inequalities: by seeking accessible therapeutic solutions for populations affected by ASD.

SUMMARY

FIRST PART

1 INTRODUCTION

The rational design of drugs demands extensive theoretical investigation into the efficiency of bioactive compounds. This broad process integrates the contributions of chemists, biologists, and healthcare professionals. The study of the structural attributes of small molecules is fundamental in virtual screening protocols, where, based on a reference structure, one seeks other molecules with similar characteristics capable of performing relevant biological activities (KILLAM *et al.*, 2024). In addition to docking calculations, ADMET (Absorption, Distribuition, Metabolism, Excretion, Toxicology) property estimations, and QSAR-based (Quantitative Structure Relationship) predictions, molecular dynamics simulations are key steps in the computational characterization of biological activity, as they operate in higher degrees of freedom, with time as a critical variable that describes systems evolving toward equilibrium under ensemble conditions that more realistically represent drug–receptor energetic interactions (VIDAL-LIMON; AGUILAR-TOALÁ; LICEAGA, 2022).

The integration of robust molecular dynamics methodologies with free energy calculations is essential for the theoretical validation of activity, showing strong correlation with experimental data (CRABB *et al.*, 2020). Fast free energy calculation methods based on implicit solvent models are employed to optimize computational and financial efficiency, as they can be applied in early stages of experimental workflows, saving resources and avoiding waste generation (ORMEÑO; GENERAL, 2024). In this work, rapid post-dynamics thermodynamic potential calculations were coupled to propose, within virtual screening protocols, agonists with selective affinity for mGluR7/8 (metabotropic glutamate receptors) and 5-HT_{1A} (5-hydroxytryptamine) receptors, both of which are implicated in the pathogenesis of autism (HODGES; FEALKO; SOARES, 2020).

mGluR7/8 receptors modulate glutamatergic synapses, which are essential for neural homeostasis. When dysregulated, they lead to synaptic hyperexcitation, which is reflected in restricted social behavior, cognitive deficits, and hypersensitivity to external stimuli – core symptoms of autism (OYA *et al.*, 2023). 5-HT_{1A} receptors, key components of serotonergic synapses and located in raphe neurons, act by inhibiting synaptic transmission in response to excess free serotonin (SAITOW; TAKUMI; SUZUKI, 2020). When impaired, they result in serotonergic hypersignaling, manifesting as mood disorders and socio-affective deficits, also strongly associated with autism.

In this dissertation, Part 1 includes Section 2, which corresponds to the theoretical framework, divided into two main parts. The first (Sections 2.1 to 2.3) addresses the theoretical foundations of the computational simulations used, starting with the principles of analytical mechanics – describing molecular systems through Lagrangian and Hamiltonian functions – followed by molecular dynamics, which involves the formulation and integration of the equations of motion. It also presents the basis of classical electrostatics, which supports the implicit solvent electrostatic potentials employed in PB/GBSA (Poisson-Boltzmann/Generalized Born Surface Area) free energy calculations. Additional elements of statistical thermodynamics are introduced, especially ensemble theory, which underpins the computation of thermodynamic potentials.

In a second part (Sections 2.4 and 2.5), the biomolecular aspects surrounding the microenvironment in which the simulations were carried out are explored, aligned with current literature that identifies mGluR7/8 and 5-HT_{1A} receptors as key pharmacological targets for mitigating the adverse effects of autism – a condition affecting millions of people worldwide, often impairing individual well-being and associated with severe comorbidities such as depression, anxiety, and suicide (MAYES *et al.*, 2013; NYRENIUS *et al.*, 2023; UPTHEGROVE *et al.*, 2018), in addition to the social stigma faced daily by this population. Furthermore, the latter part addresses the fundamentals of computational medicinal chemistry, highlighting core elements such as virtual screening, QSAR model generation, and protein receptor homology modeling.

Finally, in Part 2, two manuscripts currently under submission to specialized journals are presented, demonstrating the efficiency of these methodologies and identifying six agonists for mGluR7/8 and 5-HT_{1A} as potential chemotherapeutic agents for the core symptoms of autism. These results highlight the efficiency of classical and hybrid (Quantum Mechanics/Molecular Mechanics) computational methods as viable, cost-effective, and experimentally consistent alternatives for drug development.

2 THEORETICAL FRAMEWORK

In this section, a first part will present the theoretical foundations that underlie molecular simulations, addressing the implications of analytical mechanics, the Hamiltonian description of systems, their applications in molecular dynamics, and the modeling of classical force fields. Some principles of electrostatics necessary for the characterization of implicit solvents, the Poisson-Boltzmann equation, and the generalized Born model will also be addressed. These foundations are essential for understanding the methods for fast free energy calculations by QM/MM-PB-GBSA applied in virtual screening protocols aiming advances in drug discovery. Topics in statistical thermodynamics, focused on the *NVT* and *NpT* ensembles, are also present, with applications in standard free energy calculations, molecular equilibration processes, and construction of surfaces by free energy landscape.

In a second stage, the systems in which the simulations will occur will be presented. The autistism spectrum disorder, its genetic and molecular bases, involving the mGluR and 5-HT_{1A} pathways, as well as the biochemical panorama that permeates the metabolism of these receptors will be addressed. The fundamentals of computational medicinal chemistry will also be highlighted, with a focus on virtual screening protocols and drug modeling, addressing aspects of QSAR and homology modeling.

2.1 Theoretical principles of molecular simulation

This subsection presents the main fundamentals of molecular simulation, focusing on macromolecular systems from the perspective of classical analytical mechanics. The Lagrangian of a multiparticle system will be described, which results in a molecular Hamiltonian function, used to elucidate the total energy of the systems of interest, considering the conservative case and without explicit time dependence in the Hamiltonian. In addition, the main velocity integration schemes used in molecular dynamics will also be discussed. The potential energy functions will be described in terms of classical force fields, highly applicable to the systems used in this work.

2.1.1 Classical mechanics applied to molecular systems

In the study of molecular systems consisting of numerous interacting particles, particularly those under the influence of conservative forces, Lagrangian and Hamiltonian

formalisms present fundamental methods of comprehension and prediction of the behavior of the system (CICCOTTI; DECHERCHI; MELONI, 2025) (derivations are found in Appendix C.1). A conservative system is one whose total mechanical energy does not vary with time. This means that the particle forces can be derived from a potential energy function, and there are no dissipative motions like friction or heat exchange with the surroundings (GUHA; GHOSE-CHOUDHURY, 2025).

The Lagrangian provides a generalized framework to define motion in terms of coordinates that are specific to the system's geometry and constraints. It facilitates the derivation of equations of motion in many degrees of freedom systems, such as molecules composed of dozens or even hundreds of atoms (ARADI; NIKLASSON; FRAUENHEIM, 2015). The Hamiltonian offers the same but occasionally more convenient framework for computational modeling. It is expressed in terms of conjugate momenta and generalized coordinates, and is found widely used in simulations of molecular dynamics due to the direct relation to energy conservation and phase space structure (CICCOTTI; DECHERCHI; MELONI, 2025).

A characteristic feature of conservative molecular systems is that the Hamiltonian typically does not explicitly vary with time (FANG; LIU; SARKAR, 2025). This occurs when the kinetic and potential energy functions themselves are time-independent — that is, forces and system configuration do not change with time due to external forces. In such cases, the Hamiltonian function remains constant during evolution of the system. This time independence of the Hamiltonian reflects the physical truth that the total energy of the system remains constant (BOSSE *et al.*, 2025). Because the motion equations are derived only from position and momentum variables and contain no time-dependent term, the system is evolving in a way that maintains the total energy unchanged.

This such absence of explicit time dependence not only confirms energy conservation but also makes analytical and numerical research easier. In molecular simulation, it allows the use of symplectic integrators and other energy-conserving schemes that are valid on large timescales, providing more stable and accurate models of molecular dynamics.

2.1.2 Molecular dynamics: integration and equations of motion

Molecular mechanics can be defined as the application of classical mechanics, either analytical (Hamiltonian) or Newtonian, to systems at the atomic-molecular level (RÁCZ *et al.*, 2022). Each component of the system is described by its position (**q**) and momentum (**p**) in a phase space, so that the temporal evolution (t) can be determined and, therefore, the trajectories

of each particle can be extracted. This approach makes it possible to simplify the obtaining of velocities and forces acting on any portion of the trajectory and, as shown now, the Hamiltonian corresponds directly to Newton's second law (TUCKERMAN; MARTYNA, 2000). In a conservative system of N particles, the total energy can be expressed in terms of the molecular Hamiltonian function (2.1), (GUHA; GHOSE-CHOUDHURY, 2025).

$$\mathcal{H}(\mathbf{p}_i, \mathbf{q}_i) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} + \sum_i^N \mathcal{V}(\mathbf{q}_i) \tag{2.1}$$

which describes, under the influence of momentum, in relation to kinetic energy, and position, expressed by potential energy, as central components of the total energy. Developing this relationship, we can arrive at simpler expressions that revisit Newton's second law, with fundamental applications in the dynamic description of molecular systems. Applying Hamilton's first equation, we obtain

$$\dot{\mathbf{q}}_i = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} = \sum_i^N \frac{\partial}{\partial \mathbf{p}_i} \frac{\mathbf{p}_i^2}{2m_i} = \sum_i^N \frac{\mathbf{p}_i}{m_i} \tag{2.2}$$

Therefore,

$$\sum_i^N \mathbf{p}_i = \sum_i^N \dot{\mathbf{q}}_i m_i \tag{2.3}$$

a known result, where $\dot{\mathbf{q}}$ is the velocity ($d\mathbf{q}/dt$), such as the classical momenta of a system composed of a set of particles. From Hamilton's second equation, we have

$$-\dot{\mathbf{p}}_i = \frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} \tag{2.4}$$

$$\dot{\mathbf{p}}_i = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} = -\sum_i^N \frac{\partial \mathcal{V}}{\partial \mathbf{q}_i} = -\nabla \mathcal{V}(\mathbf{q}_i) = \mathbf{F_i} \tag{2.5}$$

But, by the derivative of the momentum, we obtain

$$\dot{\mathbf{p}}_i = \frac{d\mathbf{p}_i}{dt} = m_i \frac{d^2\mathbf{q}_i}{dt^2} = \mathbf{F_i} \tag{2.6}$$

and by the definition of acceleration $\mathbf{a} = d^2\mathbf{q}_i/dt^2$, we finally have the classical expression of Newton's second law from the molecular Hamiltonian:

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad \therefore \quad \sum_i^N \mathbf{F}_i = \sum_i^N m_i \mathbf{a}_i \tag{2.7}$$

In this way, the equivalence of the Hamiltonian with the Newtonian description is shown. Given the simplification of obtaining velocities and forces through the Hamiltonian, this information is fundamental for describing how the system evolves energetically, but it is

not yet possible to determine analytically the exact evolution of the trajectories, for this we use algorithms for integrating the equations of motion (LEACH, 2001).

The Verlet algorithm, proposed in 1967, is one of the most widely used numerical integration schemes in classical molecular dynamics simulations (SPREITER; WALTER, 1999). It is particularly valued for its simplicity, numerical stability, and ability to conserve energy over long simulations of conservative systems. The method is based on a second-order Taylor expansion of the particle positions (Appendix C.2), taking into account only the acceleration (second derivative of position with respect to time), which is related to the force acting on each particle (GRUBMÜLLER *et al.*, 1991). The Verlet integration does not explicitly use velocities, which can be calculated separately if needed. The general equation is

$$\mathbf{q}_i(t+\delta t) = 2\mathbf{q}_i(t) - \mathbf{q}_i(t-\delta t) + \delta t^2 \frac{d^2\mathbf{q}_i}{dt^2} \tag{2.8}$$

This formulation makes the Verlet algorithm particularly suitable for systems where forces are known as functions of positions, such as in classical mechanics. Since it does not require velocity information to update positions, it avoids numerical instabilities associated with direct velocity estimation and is time-reversible, a desirable property in simulating conservative systems (BUFOLO; SOBRAL, 2024). Furthermore, the quality of the algorithm application depends on the time interval between one position and another δt (GROMACS DEVELOPMENT TEAM, 2024); the smaller this temporal evolution, the more accurate the integration will be. In general, the literature indicates an ideal interval of 2 fs in molecular dynamics simulations to ensure good data resolution and a good description of velocities.

While the initial Verlet algorithm (1967) is a simple and stable equation-of-motion integrator, it does not provide an immediate access to particle velocities. This limitation led to the development of two other schemes: the Leap-Frog and Velocity-Verlet. Both schemes retain the attractive energy conservation capabilities of the original Verlet approach but provide mechanisms to track velocities in better way and with less computational cost (MARTYS; MOUNTAIN, 1999). The Leap-Frog algorithm (1970), calculates positions and velocities at staggered time steps. The velocities are evaluated at smaller integer steps (the half compared to Verlet), while positions are updated at full time steps. The first equation represents the updated position as a function of the current position, the velocity at time t, and a correction term to the force (acceleration) upon the particle (TANG, D. *et al.*, 2021):

$$\dot{\mathbf{q}}_i(t) = \frac{1}{2}\left(\frac{d}{dt}\mathbf{q}(t) + \frac{1}{2}\delta t \frac{d^2\mathbf{q}_i}{dt^2} + \frac{d}{dt}\mathbf{q}(t) - \frac{1}{2}\delta t \frac{d^2\mathbf{q}_i}{dt^2}\right) \tag{2.9}$$

In contrast, the Velocity-Verlet algorithm, developed in 1982, updates both positions and velocities at synchronized time steps, making it more convenient for thermodynamic calculations (MARTYS; MOUNTAIN, 1999). The position is updated using both the velocity and the acceleration at time t, then, the velocity at the next time step is calculated as the initial velocity plus the average of the acceleration at time t and at time $t + \delta t/2$, which requires the force to be evaluated again at the new position:

$$\dot{\mathbf{q}}_i(t+\delta t) = \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t\left(\frac{d^2\mathbf{q}_i(t)}{dt^2} + \frac{d^2\mathbf{q}_i(t+\delta t)}{dt^2}\right) \tag{2.10}$$

An auxiliary expression shows the intermediate form of the velocity update, useful in practice for force evaluation:

$$\dot{\mathbf{q}}_i(t+\frac{1}{2}\delta t) = \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t\frac{d^2\mathbf{q}_i}{dt^2} \tag{2.11}$$

The Velocity-Verlet algorithm is particularly suited for simulations where both kinetic and potential contributions to the energy must be evaluated at the same time step, and its time-symmetric structure preserves the conservative nature of the dynamics. Finally, to determine the initial velocities, one way to reliably estimate, with a realistic distribution of the kinetic energy of the system, is by using the Maxwell-Boltzmann statistics (GROMACS DEVELOPMENT TEAM, 2024), by the equation (2.12), in which k_B is the Boltzmann constant, T is the temperature and m_i is the mass of the atom.

$$P(\dot{\mathbf{q}}_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2}\exp\left[-\frac{m_i\dot{\mathbf{q}}_{ix}^2}{2k_B T}\right] \tag{2.12}$$

2.1.3 Classical force fields

Now that we have determined how energy evolves with the Hamiltonian and how to integrate velocities to obtain trajectories, we need to explore the potential energy of the system and, consequently, the way forces act on each particle, following the relation

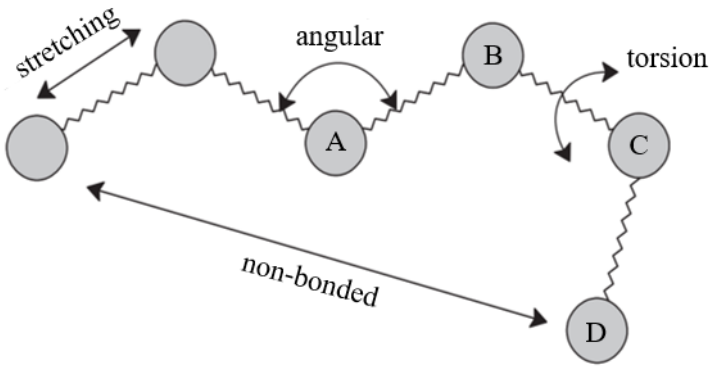
$$\mathbf{F}(\mathbf{r}) = -\vec{\nabla}\mathcal{V}(\mathbf{r}) = -\left(\frac{\partial\mathcal{V}}{\partial\mathbf{r}_1}, \frac{\partial\mathcal{V}}{\partial\mathbf{r}_2}, \frac{\partial\mathcal{V}}{\partial\mathbf{r}_3}, \ldots\right) \tag{2.13}$$

To do this, we need to use force fields. Force fields are potential energy functions ($\mathcal{V}$) that describe a system by summing intramolecular and intermolecular parameters, such that the total energy can be expressed by equation (2.14) (CHIACCHIO *et al.*, 2023). Force fields can be purely classical, or semi-empirical, incorporating both classical parameters and those derived from simplified Hamiltonians to aid in the description of electronic behavior (KISS *et al.*, 2022).

$$\mathcal{V}(\mathbf{q}) = \sum_{i}^{N} \mathcal{V}(\mathbf{q}_i)$$
$$\mathcal{V}(\mathbf{q}) = \sum_{intra} \mathcal{V}(\mathbf{q}_i) + \sum_{inter} \mathcal{V}(\mathbf{q}_i) \tag{2.14}$$

From the classical perspective, intramolecular potential energies are derived from the degrees of freedom of molecular movement (XU, Y.; VIRGIN, 2019) (Figure 1), which may include stretching, angular deformation, and torsional distortions. Interactions between non-bonded atoms, which may be within the same molecule or in sufficiently close neighboring molecules, are described by the potential energies associated with non-bonded atoms, related in the term $\mathcal{V}_{intra}$ in equation (2.14).

Figure 1 – Illustration of inter and intramolecular parameters



Source: (JENSEN, 2007)

The energy terms can be expanded into potential energies associated with stretching, angular deformations, torsion, and interactions between non-bonded atoms, as shown in equation (2.15).

$$\mathcal{V}(\mathbf{q}) = \sum_{stretching} \mathcal{V}(\mathbf{q}_i) + \sum_{angular} \mathcal{V}(\mathbf{q}_i) + \sum_{torsion} \mathcal{V}(\mathbf{q}_i) + \sum_{non-bonded} \mathcal{V}(\mathbf{q}_i) \tag{2.15}$$

The bond stretching and angular deformations can be expressed through harmonic potentials, following Hook's law

$$\mathcal{V}(\mathbf{q}) = \frac{1}{2} k(\mathbf{q} - \mathbf{q}_i) \tag{2.16}$$

relating an equilibrium position (parametrized) and an non-equilibrium position, due to molecular movement. The generalized coordinate can refer to the angle θ of angular deformation and the final expressions for those potentials is given by

$$\mathcal{V}_{angular} = \mathcal{V}(\theta) = k_\theta(\theta - \theta_i) \tag{2.17}$$

and

$$\mathcal{V}_{bond} = \mathcal{V}(\mathbf{q}) = k_b(\mathbf{q} - \mathbf{q}_i) \tag{2.18}$$

where k_θ and k_b are Hook's constants parametrized to bond and angular stretching.

Regarding intramolecular terms, four atoms bonded in sequence can generate torsions in the bonds that connect them. A well-defined example is the presence of dihedral angles, as shown in Figure 1, between the atoms ABCD. The potential energy can be expressed as a Fourier series (LIM, 2002) $f(t)$ (equation 2.19), where a_0 and a_j are Fourier coefficients dependent on the physical description of the system. w is the summation index, $w_0 = 2\pi/T_0$, where T_0 is the period of the function, t is the time, and θ is the phase, representing the frequency spectrum of the series.

$$f(t) = \frac{a_0}{2} + \sum_{j=1}^{\infty} \frac{a_j}{2} \cos(w_j t - \theta_j) \tag{2.19}$$

Including the values of the Fourier integrals a_0 and a_j in equation (2.19), considering $a_0 = 0$, we obtain the dihedral contribution to the potential energy:

$$\mathcal{V}(\mathbf{q}_j, t) = \sum_{diedrals} \sum_{j=1}^{4} \frac{1}{2L} \int_{c}^{c+2L} \mathbf{q}_j(t) \cos\left(\frac{j\pi t}{L}\right) dt \left[1 - (-1)^j \cos(\mathbf{q} - \mathbf{q}_j)\right] \tag{2.20}$$

replacing the generalized coordinate by the dihedral angle ϕ between the four atoms, we obtain

$$\mathcal{V}_{diedrals} = \mathcal{V}(\phi) = \sum_{diedros} \sum_{j=1}^{4} \frac{a_j}{2} \left[1 - (-1)^j \cos(\phi - \phi_j)\right] \tag{2.21}$$

This being the term of the dihedral angles used by the OPLS force field. The Fourier series are used as a fitting method to describe the oscillatory potential energy function related to the tortional bond parameters and dihedral angles (HOPKINS; ROITBERG, 2014).

Corresponding to the last term in equation (2.15), which is related to the energy of non-bonded atoms, two fundamental potentials are considered: Lennard-Jones potential, and electrostatic interactions. Given that the atoms are treated as charged rigid spheres, the interactions between them are Coulombic. Equation (2.22) (HAPALA *et al.*, 2016) refers to the interaction energy between atoms i and j, separated by a distance r_{ij}, with charges represented by variables q_i and q_j. ε is the permittivity constant.

$$\mathcal{V}_{coulomb} = \mathcal{V}(\mathbf{q}) = \frac{1}{2\pi\varepsilon} \frac{q_i q_j}{\mathbf{q}_{ij}} \tag{2.22}$$

The combination of van der Waals attraction and repulsion potentials origin the Lennard-Jones potential, as shown in (2.23) (SCHAUPERL *et al.*, 2020) , where r_{ij} is the distance between atoms i and j, σ_{ij} is the Lennard-Jones constant, and ε_{ij} is the minimum energy parameter.

$$\mathcal{V}_{LJ} = \mathcal{V}(\mathbf{q}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\mathbf{q}_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{\mathbf{q}_{ij}}\right)^{6}\right] \tag{2.23}$$

The constant σ_{ij} graphically indicates the intersection on the **q** axis, representing the minimum distance at which the two atoms can approach each other while still maintaining viable values of potential energy.

Finally, the general equation of a force field is the sum of the potentials discussed so far, combining equations above for all the atoms in the system, as illustrated in the functional form of an OPLS force field (WANG, J. *et al.*, 2004) in the following equation (2.24):

$$\begin{aligned}\mathcal{V}(\mathbf{q},\theta,\phi) = {} & \sum_{bonds} k_b(\mathbf{q}-\mathbf{q}_i) + \sum_{angles} k_\theta(\theta-\theta_i) + \sum_{diedrals}\sum_{j=1}^{4}\frac{a_j}{2}\left[1-(-1)^j\cos(\phi-\phi_j)\right] + \\ & \sum_{LJ} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\mathbf{q}_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{\mathbf{q}_{ij}}\right)^{6}\right] + \sum_{Coulomb}\frac{1}{2\pi\varepsilon}\frac{q_i q_j}{\mathbf{q}_{ij}}\end{aligned} \tag{2.24}$$

A variety of force fields have been reported in the literature for studying proteins and biomolecules. Examples include OPLS (DOHERTY *et al.*, 2017), CHARMM (HUANG, J. *et al.*, 2017), GAFF (AMBER) (SPRENGER; JAEGER; PFAENDTNER, 2015), GROMOS (SCHMID *et al.*, 2011), MM3 (SHANNON *et al.*, 2019) and MMFF (TOSCO; STIEFL; LANDRUM, 2014). Each force field should be applied according to the specific needs of the system and parameterized based on the required description of study.

2.1.4 Molecular docking

Typically, molecular docking involves computationally fitting a specific ligand into a macromolecule and predicting the optimal conformation the ligand can achieve relative to the macromolecule based on potential energy calculations and scoring functions (SHARMA; YENNAMALLI, 2023). Additionally, it allows the determination and quantification of ligand-receptor interactions. Molecular docking has become a widely adopted technique in medicinal chemistry and drug development because docking analysis is crucial for understanding the binding modes and interactions between the drug and its biological receptor (BLANES-MIRA *et al.*, 2023).

Molecular docking facilitates the rapid, cost-effective, and computationally efficient analysis of various pharmacologically relevant molecules, providing accurate and statistically reliable results. Docking programs return the best poses based on interaction energies between the ligand and the macromolecule, derived from scoring or fitness functions (PARIMALA, 2023). While these functions may not have a direct physical interpretation, they are effective

for ranking ligand poses in the active site (DU, L. *et al.*, 2023). The primary method for assessing the quality of the analysis is redocking, where a ligand with an existing experimental co-crystallized structure is re-docked, and the resulting poses are compared to the experimental compound using statistical criteria. The most commonly used metric is RMSD (Root Mean Square Deviation), which evaluates the accuracy of the predictions (equation 2.25). Generally, RMSD values below 2.0 Å (FLACHSENBERG *et al.*, 2024) are considered acceptable. However, it is also essential to carefully examine the spatial overlap between the experimental and virtual structures, paying attention to torsion angles and specific molecular groups of each molecule.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i}^{N}(\mathbf{q}_i - \bar{\mathbf{q}})^2} \tag{2.25}$$

Molegro Virtual Docker (MVD) (THOMSEN; CHRISTENSEN, 2006) is a software that performs molecular docking, and its scoring algorithm is derived from force fields. The parameters used by the program closely resemble those presented in last section, with the total scoring energy (E_{score}) obtained by summing intermolecular and intramolecular energies (equation 2.26) (MOLEXUS, 2019):

$$E_{score} = E_{inter} + E_{intra} \tag{2.26}$$

the intermolecular energy is calculated using equation (2.27), where the term E_{PLP} is defined as the piecewise linear potential, which encompasses two distinct parameters: one related to steric terms between atoms (van der Waals) and another, stronger, related to hydrogen bonds. The absolute value of 332.0 is used to convert the electrostatic potential units to kilocalories per mole (kcal/mol). The parameters q_i and q_j refer to the partial charges of atoms i and j, as in equation (2.22) for the Coulombic potential (MOLEXUS, 2019). The term r_{ij} represents the distance between the two atoms, and the algorithm accounts for its quadratic term.

$$E_{inter} = \sum_{i \in ligand} \sum_{j \in protein} \left[E_{PLP}(\mathbf{q}_{ij}) + 332.0 \frac{q_i q_j}{4\mathbf{q}_{ij}^2} \right] \tag{2.27}$$

The term for intramolecular energy is shown in (2.28), which is composed by the previously mentioned E_{PLP} term and the molecular torsion parameters, resembling the dihedral angle energy component of equation (2.28) (MOLEXUS, 2019).

$$E_{intra} = \sum_{i \in ligand} \sum_{j \in protein} E_{PLP}(\mathbf{q}_{ij}) + \sum_{flex.bonds} A[1 - \cos(m\theta - \theta_0)] + E_{clash} \tag{2.28}$$

The equation for intramolecular energy also includes an additional term (E_{clash}), defined as the parameter that imposes a penalty of 1000 units if a heavy atom, defined as any atom other than

carbon and hydrogen, is located within 2 Å. The constant A is the force constant derived from the Fourier series, θ is the torsion angle, and θ_0 is the equilibrium torsion angle. The parameter m indicates multiplicity. The torsional parameters A, m, and θ_0 have values that vary depending on hybridization. For sp^2-sp^3 hybridizations, the values of θ_0, m, and A are 0, 6, and 1.5, respectively. For sp^3-sp^3 hybridizations, the values are π, 3, and 3. The sp^2-sp^2 hybridization is not activated by default in the program, but if activated, it returns values of 0, 2, and 3 (MOLEXUS, 2019).

2.2 Electrostatic fundamentals and implicit solvents

Classically, a system of particles is treated as rigid and electrified spheres, without considering quantum electronic effects. Despite the deficiency of the quantum description, classical models are widely applied to systems of multiple particles, in which the quantum treatment becomes unfeasible due to the high computational cost (SOMMA *et al.*, 2003). In this section, we will recapitulate key elements of electrostatics, whose applicability lies in the derivation of the Poisson-Boltzmann equation and in the generalized Born model, in addition to the definition of implicit solvent potentials.

Given the geometric description of atoms, according to Moyses (NUSSENZVEIG, 2003), as conducting spheres charged on their surface with positive or negative charges (q), the resulting electric field will point outwards from the sphere if $q > 0$ and will point inwards from the sphere if $q < 0$. By positioning a spherical Gaussian surface of radius $\mathbf{r}$ that encompasses the entire sphere from its surface, so that $\mathbf{r} > R$, where R is the radius of the sphere itself, the electric field will be nonzero, since only inside conducting objects is the electric field zero. Detailed derivation can be found in Appendix C.3. The electric field of a charged sphere involved by a gaussian surface of radius $\mathbf{r}$ is given by (WARSHEL *et al.*, 2006):

$$E(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}_i - \mathbf{r_0}|^2} \tag{2.29}$$

Where ε_0 is the vaccum permittivity constant, q_i is the charge of the sphere and $|\mathbf{r}_i - \mathbf{r_0}|$ is the module of the radius vector. Now it is important do make some definitions. The first is that the electric potential $\varphi(\mathbf{r})$ is the electric potential energy per unit charge, so that (NUSSENZVEIG, 2003)

$$\varphi(\mathbf{r}) = \frac{\mathcal{V}(\mathbf{r})}{q} \tag{2.30}$$

the second is that the electric field of an electrically charged body is the quotient of force by the charge, so that (NUSSENZVEIG, 2003)

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{q} \tag{2.31}$$

Given definition 1 (2.30), the difference between two electric potentials $\varphi_2 - \varphi_1$ will be

$$\varphi_2 - \varphi_1 = \frac{1}{q}(\mathcal{V}_2 - \mathcal{V}_1) = \frac{\Delta\mathcal{V}}{q} \tag{2.31}$$

The electric force can be elucidated by integration, given the definition electric work. Integrating equation (2.29) and considering that $\varphi_2(\infty) = 0$ and $\varphi_1 = \varphi(\mathbf{r})$, we have the electric potential of a charged sphere, where the formalism is shown in Appendix C.3:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}_i - \mathbf{r_0}|} \tag{2.32}$$

There is an important connection with thermodynamics, since, from the electrostatic potential of the solution, it is possible to calculate the Gibbs energy of this process, in accordance with Castellan (CASTELLAN, 1983). Starting from the primordial definition of Gibbs energy,

$$-\Delta G \geqslant w_a \tag{2.33}$$

where w_a is the work other than the expansion of the system. If we consider that the system is subjected only to an electric potential, at a constant volume, that is, without expansion, the Gibbs energy will be equal to the work, in this way

$$\Delta G = -w \tag{2.34}$$

as previously defined, electrical work is equal to the negative of the variation in electrostatic potential, with this,

$$\Delta G = -(-\Delta\mathcal{V}) = \mathcal{V}(\mathbf{r}) - \mathcal{V}(\mathbf{r}_0) \tag{2.35}$$

Using definition 1 (2.30) of electric potential, the potential energy can be written as follows, for a set of N particles subjected to the same potential. However, the sum for all atoms considers the interaction of charges q_iq_j and q_jq_i twice. A necessary correction factor is to consider only half of the interactions:

$$\mathcal{V}(\mathbf{r}_i) = \frac{1}{2}\sum_i^N q_i\varphi(\mathbf{r}_i) \tag{2.36}$$

substituting (2.36) into (2.35), we obtain the following equation, considering the solvent and vacuum potentials. Simplifying, we have

$$\Delta G_{el} = \frac{1}{2}\sum_i^N q_i[\varphi^{solv}(\mathbf{r}_i) - \varphi^{vac}(\mathbf{r}_i)] \tag{2.37}$$

The final expression of the Gibbs energy of polar solvation, considering the discrete case, in which a set of particles have constant and independent charges. In the case of systems

in which the charges vary when they come into contact with others, the charges can be replaced by a charge density and the summation term becomes a definite integral in the entire space where the electric potential acts (DU, Q.; BEGLOV; ROUX, 2000):

$$\Delta G_{el} = \frac{1}{2}\int_{-\infty}^{\infty} \rho(\mathbf{r})[\varphi^{solv}(\mathbf{r}_i) - \varphi^{vac}(\mathbf{r}_i)]\, d\mathbf{r} \tag{2.38}$$

Given the important dependence of the Gibbs energy on the electric potential of the solutions, equations that determine the potential are necessary for the thermodynamic quantity to be calculated. For this, there are two possibilities: one in which the electric potential is elucidated for each solvent molecule, an approach known as explicit solvent (ZHANG *et al.*, 2017), but computationally expensive, as it involves the solution of extensive differential equations for each component of the solvent (RINGE *et al.*, 2022). For example, in a molecular dynamics simulation of a protein in water, a medium-sized protein with about 300 amino acid residues can be placed in a solvation box containing 50 to 100 thousand water molecules, that is, the application of explicit solvent results in the solution of 50 to 100 thousand differential equations per simulation frame. Considering that a 50 ns simulation has an average of 2500 frames, at a resolution of 2 fs, this is equivalent to solving hundreds of millions of equations. To correct this effect, implicit solvent models will be shown below, in which the approach to calculating the solvation free energy is simplified from the solution of only one differential equation per frame.

2.2.1 Generalized Born (GB) model

The application of implicit solvent methods has revolutionized molecular modeling by providing reliable and rapid approaches to studying solutions and calculating thermodynamic potentials in protein-ligand systems (THE AMBER DEVELOPERS, 2024). Unlike explicit solvent methods, implicit methods do not consider the concrete existence of solvent molecules or particles, but model the medium in terms of a continuous and homogeneous electric potential, which makes it possible to investigate the effects of solvation without explaining each molecule individually.

The Generalized Born (GB) model is a continuous and implicit approximation for the description of the effects of electrostatic solvation in biomolecular systems, notably developed to circumvent the high computational costs of explicit methods and the direct solution of the Poisson-Boltzmann (PB) equation. Historically, the need to represent the influence of the solvent on molecular systems dates back to the first efforts in statistical mechanics and

computational chemistry, in which it was realized that the inclusion of the aqueous medium was fundamental for the accuracy of free energy calculations (ONUFRIEV; BASHFORD; CASE, 2000). Explicit solvent models, although more detailed, require a high number of degrees of freedom and long simulations to ensure statistical convergence, especially when combined with methods such as molecular dynamics or Monte Carlo (KRAUSCH *et al.*, 2019).

As an alternative, implicit solvent models have emerged, which treat the solvent as a continuous medium with constant dielectric permittivity. Among them, the Poisson-Boltzmann (PB) model gained notoriety for describing the electrostatic potential as a function of the charge distribution in the solution. However, even the PB, in its nonlinear or linearized form, requires intensive numerical solutions, being limited in dynamic simulation contexts (THE AMBER DEVELOPERS, 2024). In this context, the Generalized Born model was proposed as an approximate analytical form of the original Born equation (1920) (SILVA *et al.*, 2024), which described the solvation energy of a spherical ion immersed in a dielectric medium. The GB formulation aims to extend this idea to systems with multiple atoms and arbitrary geometries, while maintaining computational efficiency.

Mathematically as stated by Andrew Leach (LEACH, 2001), the GB model acts on a set of particles with a Born radius a_i and charges q_i, with the total energy of electrostatic interaction being defined by

$$G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\varepsilon \mathbf{r}_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2.39}$$

the first set of summations can be expressed according to the electrostatic Coulomb contribution in vacuum

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\varepsilon \mathbf{r}_{ij}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\mathbf{r}_{ij}} - \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\mathbf{r}_{ij}} \tag{2.40}$$

leading to the complete expression

$$G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\mathbf{r}_{ij}} - \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\mathbf{r}_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2.41}$$

Finally, the free energy difference (ΔG_{elec}) can be computed from the subtraction of (2.39) and (2.41), which corresponds to the contribution in dielectric solution medium subtracted from the terms referring to the contribution in vacuum, giving rise to the final expression

$$\Delta G_{elec} = \left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\mathbf{r}_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2.42}$$

The GB model has been widely used in both molecular mechanics and quantum-mechanical approaches, natively implemented in semi-empirical models. Advantages include computational efficiency, the possibility of integration with classical force fields, and applicability to large-scale systems. However, some limitations are important, such as dependence on ideal parameterization, geometric simplification, and the absence of solvent-specific electrical effects (CHRISTENSEN *et al.*, 2016).

2.2.2 Poisson-Boltzmann (PB) equation

In terms of implicit solvent modeling, the Poisson-Boltzmann (PB) equation plays a fundamental role, due to its intrinsic dependence on the geometry of the system. This allows to capture more rigorously the electrostatic effects of solvation (BLOSSEY, 2023). Initially, the Poisson equation can be obtained from the differential form of Gauss's law (NUSSENZVEIG, 2003), given the application of the divergence theorem in its integral form shown in Appendix C .4, leading to the equation

$$\nabla^2 \varphi(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\varepsilon} \tag{2.43}$$

Boltzmann's contribution is given from the distribution of the electrostatic potential in terms of ion concentrations. This relationship can be extracted from standard chemical potentials, if we consider a system subjected to an external electric field, whose potential energy is given by $\mathcal{V}_{el}$ (SAFARIPOUR; ANAND; SNOEYINK, 2023). The total chemical potential μ can be written as:

$$\mu = \mu^0 + k_B T \ln \frac{c_i}{c_i^0} + z_i \mathbf{e} \varphi(\mathbf{r}) \tag{2.44}$$

where c is the concentration of the ion, z_i is the valence and $\mathbf{e}$ is the elementary charge. The equilibrium condition of this system requires equality of chemical potentials, that is $\mu = \mu^0$, solving the equation for c_i, we obtain the Boltzmann distribution for electrolytes:

$$c_i = c_i^0 e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} \tag{2.45}$$

the final expression that corresponds to the Boltzmann distribution of the electric potential of an electrolyte solution. To combine the Poisson (2.43) and Boltzmann (2.45) equations, we start from the definition of the charge density $\rho(\mathbf{r})$, which depends on the concentration of electrolytes in solution, given by the Boltzmann distribution. The substitution into the original Poisson equation, returns the Poisson-Boltzmann equation (2.46) (BLOSSEY, 2023):

$$\nabla^2\varphi(\mathbf{r}) = \sum_i^N \frac{z_i \mathbf{e} c_i^0}{\varepsilon} e^{-z_i \mathbf{e}\varphi(\mathbf{r})/k_B T} \tag{2.46}$$

Another way of expressing the PB equation is in cases of systems composed of binary electrolytes and symmetrically opposite charges, such as NaCl, for example, the concentrations of the Boltzmann equation must be modeled in terms of the charges of these two components, with this the equation becomes, written in its hyperbolic form, as

$$\nabla^2\varphi(\mathbf{r}) = -\frac{2\mathbf{e}c_i^0}{\varepsilon} \sinh\left[\frac{\mathbf{e}\varphi(\mathbf{r})}{k_B T}\right] \tag{2.47}$$

The PB equation is therefore a nonlinear, second-order partial differential equation. Despite the possibility of geometric consideration of the system of interest, the nonlinear form allows few analytical solutions, even in simple geometries. In order to simplify the solutions, Peter Debye and Erich Hückel, in 1923 proposed the linearization of PB, followed by the analytical solution according to a spherical geometry (HERRERO; JOLY, 2024). This is one of the few exact solutions available, but one of the most important. Debye and Hückel approximated the entire exponential term of the PB equation in terms of a Taylor series for low electric potentials (BLOSSEY, 2023), giving the solution for a spherical system:

$$\varphi(\mathbf{r}) = \frac{q_i}{4\pi\varepsilon_0} \frac{e^{-\kappa \mathbf{r}}}{\mathbf{r}} \tag{2.48}$$

Where r is the radius and κ is a constant term defined by the Debye length:

$$\kappa^2 = \sum_i^N \frac{z_i^2 \mathbf{e}^2 c_i^0}{\varepsilon k_B T} \tag{2.49}$$

The electric potential obtained by the Debye-Hückel solution is practical for systems composed of electrically charged ions or colloids and can be extended with reservations to more complex molecular systems, under conditions of low electric potentials and low ionic strengths. The electric potential, in this case, tends to decrease with the increase in the distance between two charged particles and with the increase in the effective concentration of these species, therefore, it is not useful in describing slightly diluted systems. In addition to the spherical solution, the linearized equation can also be solved in cylindrical coordinates, in which the resulting equation is, in cylindrical coordinates

$$\varphi(\mathbf{r}) = \varphi(a) \frac{\sqrt{\frac{\pi}{2\kappa \mathbf{r}}} e^{-\kappa \mathbf{r}}}{\sqrt{\frac{\pi}{2\kappa a}} e^{-\kappa a}} \tag{2.50}$$

where $\varphi(a)$ is the electric potential of a cylinder of radius a. The solution can also be obtained more trivially by considering a charged plane surface. The differential equation's solution becomes (BLOSSEY, 2023):

$$\varphi(\mathbf{r}) = \varphi(0)e^{-\kappa \mathbf{r}} \tag{2.51}$$

With this, it is seen that the analytical solutions are easily attributed to the PB equation in its linear form. However, for systems with higher concentration and ionic strength, the approximation is not properly applied, and it is necessary to use the nonlinear equation. Analytical solutions are extremely restricted to simple geometries and there is still no exact solution for application in systems such as proteins, for example (COOPER; BARBA, 2016). For this, the computational numerical solution is used, which allows the modeling of systems with varied geometries, in different chemical environments, in terms of different concentrations and electric potentials.

A widely used numerical approach, according to Andrew Leach (LEACH, 2001), is the finite difference method, in which the solute is wrapped in a cubic grid, surrounded by the solvent. The greater the number of grids, the more accurate the solution will be, giving the possibility of covering any geometry. The electric potential of the grid φ_0 (2.52) is calculated by

$$\varphi_0 = \frac{\sum_i^N \varepsilon_i \varphi_i + 4\pi \frac{q_0}{h}}{\sum_i^N \varepsilon_i + k_o'^2 f(\varphi_0)} \tag{2.52}$$

where q_0 is the charge of the solute on the grid, h is the size of the cubic edge and $f(\varphi_0)$ determines whether the PB solution will be in the linearized form or not. For the linear case, this function takes the value of Debye-Hückel's κ^2, for the nonlinear case, it takes the form of a power series written in terms of the hyperbolic portion. Having calculated the electric potential, by whatever means (linear, nonlinear, numerical PB or GB), these results can then be substituted into the corresponding (ΔG) expressions by equations (2.37) or (2.38) so that the polar solvation free energies can be calculated.

2.3 Free energy calculations and other thermodynamic potentials

2.3.1 Classical statistical ensembles: *NVT* e *NpT*

Classical statistical mechanics uses statistical and probabilistic knowledge to study the properties of microscopic systems and how these properties are translated into macroscopic observables, having an important correlation with thermodynamics (SALINAS, 1997). Classically, systems are treated continuously in a phase space, which can be defined as a set of

all possible microstates that can be accessed, in which each point in the phase space consists of a complete configuration of the microstate (CRAMER, 2004), given the description from the momentum **p** and the position **q**. The microstates that reflect a thermodynamic observable are described by 3 momentum coordinates $\mathbf{p}_x$, $\mathbf{p}_y$, $\mathbf{p}_z$ and by three generalized position coordinates $\mathbf{q}_x$, $\mathbf{q}_y$, $\mathbf{q}_z$, or simply (x, y, z). By combining the three momentum coordinates and the three position coordinates, a phase space is six-dimensional, but it also depends on the number of particles that compose it (MCQUARRIE, 2000). For a system of N particles, for example, there will be $6N$ dimensions.

The most important property of the phase space to be explored in this work is the probabilistic description of accessing a microstate of energy $E + dE$ with a certain configuration reflected by a defined momentum and position. Given a Hamiltonian function that represents a continuous and conservative system $\mathcal{H}(\mathbf{p},\mathbf{q})$, the probability density function $\rho(\mathbf{p},\mathbf{q})$ that scans all possible configurations of accessible microstates is given by the integration with respect to all momentum and position coordinates, which equals 1, since it is a function that guarantees the normalization of the total probability.

The probability density function, in practical terms, defines the "weight" of each microstate in relation to all others that exist in the phase space. This function is perfectly described by a Boltzmann distribution (MCQUARRIE; SIMON, 1997). Given an ensemble, that is, a set of N copies of the system, which has the characteristic energy of the microstate that can be accessed by the system, considering that the number of particles, the volume and the temperature of the ensemble are constant. This description is called a canonical ensemble, or NVT (MORTIMER, 2008), which can be classically defined as

$$\mathcal{Z}_{NVT} = \frac{1}{N!h^{3N}} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{2.53}$$

where the correction factor $1/h^{3N}$ is used (SALINAS, 1997), with h being Planck's constant, so that the classical values conform to the quantum ones in (2.53). Furthermore, this normalization factor removes the units of measurement, which are necessary for the physical interpretation of the partition function, since it represents the sum of all accessible states, not the energy of these states (LAURENDEAU, 2005). In the end, we obtain the corrected value for the continuous classical system.

Given the classical description of microscopic states, the connection with microscopic observables occurs through the postulate of statistical mechanics that assumes that these observables are the result of the weighted average of all microstates. Mathematically, an observable $\langle A \rangle$ can be obtained by equation (2.54)

$$\langle A \rangle = \frac{1}{N!h^{3N}} \frac{1}{Z} \iint A e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{2.54}$$

In other words, the observable A is multiplied by the sum of the microstates with energy $\mathcal{H}$, weighting this distribution and divided by all the possibilities of microstates represented by the partition function (ZHAI; ALEXANDROVA, 2016). Two observables are important in the canonical ensemble for determining the Gibbs energy, according to Levine (LEVINE, 2009), namely pressure and internal energy, that leads the calculation of entropy, given by $dS = dU/T + p/TdV$. Once all these potentials are defined, the Helmholtz energy can be calculated as $A = U - TS$. Therefore,

$$A = k_B T \ln \mathcal{Z}_{NVT} \tag{2.55}$$

with Helmholtz energy (2.55) being the central thermodynamic potential of the canonical ensemble, since it is described by the same properties at constant composition, volume and temperature. Finally, to determine the Gibbs energy, we start from the definition $G = A + pV$, and the resulting equation is:

$$G = k_B T \ln \mathcal{Z}_{NVT} + k_B T \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial \ln V} \right)_{T,N} \tag{2.56}$$

Another ensemble of great importance for this work is the isothermal-isobaric ensemble, which describes microstates of constant composition, pressure and temperature (*NpT*). Since the Gibbs energy is also defined with these same properties, it is the direct thermodynamic consequence of the ensemble. As in the canonical case, the phase space is also described in relation to the generalized momenta and coordinates, with the difference that volume fluctuations can be computed between one microstate and another. All the postulates of statistical mechanics remain valid, including the probability density function and the definition of macroscopic observables.

The partition function *NpT*, to compute a microstate with energy $\Delta E + p\Delta V$ can be obtained from the canonical ensemble if we consider that the volumetric fluctuations are proportional to a Boltzmann distribution based on the pressure and volume, integrating over every possible differential change dV:

$$\mathcal{Z}_{NpT} = \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} e^{pV/k_B T} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV$$

$$\mathcal{Z}_{NpT} = \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} e^{-[\mathcal{H}(\mathbf{p},\mathbf{q}) + pV]/k_B T} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \tag{2.57}$$

an observable for this ensemble can be obtained from an expression analogous to the canonical ensemble. To obtain the expression for the Gibbs energy, a shorter sequence of steps is

sufficient, which consists of defining the volume observable, followed by the derivation of the partition function with respect to pressure. The observable will be

$$V = -k_B T \left(\frac{\partial \ln \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T} \tag{2.58}$$

The connection with thermodynamics can be established by the Gibbs energy itself, $G = f(N, p, T)$. If temperature and composition remain constant, we have the following identity:

$$\left(\frac{\partial G}{\partial p} \right)_{T,N} = V \tag{2.59}$$

Therefore, integrating (2.59) from an initial state of zero to a final state f and substituting the volumetric observable, we obtain

$$G = -k_B T \ln \mathcal{Z}_{NpT} \tag{2.60}$$

The formalism of the ensembles NVT and NPT can be found in Appendix C.5. *NVT* and *NpT* ensembles are essential in molecular dynamics simulations to ensure thermodynamic reliability. The *NVT* ensemble is used in equilibration steps, in which thermostats are implemented to control particle velocities in accordance with the Maxwell-Boltzmann distribution (ORMEÑO; GENERAL, 2024). This step ensures that the average kinetic energy of the particles is compatible with the simulation temperature, in addition to avoiding unrealistic thermal fluctuations. The *NpT* ensemble is applied in association with barostats and thermostats that provide realistic temperatures and pressures to the system. This equilibration step also avoids density and pressure artifacts, which can distort the conformation of the system, in addition to negatively influencing diffusion and hydrophobic interaction effects, such as van der Waals (CRABB *et al.*, 2020). The energies can also be determined from these ensembles, as in Monte Carlo methods with *NpT*, or they can be re-estimated with *NVT*, as in the case of free energy landscapes, as will be seen later, in which the partition function is approximated from a probability density function fed with principal components that reflect the conformational microstates accessible to the system.

2.3.2 A brief introduction to standard free energy calculations

Along with modeling solvation effects, calculating absolute free energy differences is one of the great challenges of computational chemistry in the 21st century, in accordance with Andrew Leach (LEACH, 2001). Thermodynamic calculations require high methodological accuracy and depend on systems that are strictly in chemical equilibrium. One of the greatest difficulties, regarding calculations between proteins and ligands, is obtaining an effective

conformational sample of the system (DE RUITER; OOSTENBRINK, 2011). Considering the extreme flexibility of proteins, exact calculations require extensive simulations, with the need to obtain subtle metastates, with biased methodologies that are even more computationally expensive. Solvent treatment also has a great influence, considering the need for explicit treatment for more relevant results. All these obstacles mean the application of rigorous and computationally expensive methodologies to correct these effects inherent to the molecular system of interest (BHATI *et al.*, 2017).

Two standard methodologies stand out in this arduous task: thermodynamic integration and perturbative methods. Thermodynamic integration consists of gradually transforming a system from an initial state, such as an unbound protein (APO), into a final state corresponding to the protein-ligand complex (HOLO) using a coupling function λ, which varies homogeneously between a value of zero and one (GROMACS DEVELOPMENT TEAM, 2024). The coupling function acts in the gradual interpolation between the initial and final states, or in general terms, it represents the insertion of the ligand into the protein in well-defined steps, resulting in a series of simulations for each value of λ (GROMACS DEVELOPMENT TEAM, 2024). The general equation for the model (2.61) consists of the integration of the rate of change of the molecular Hamiltonian in the desired ensemble with respect to the partially increasing values of the coupling function.

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(\mathbf{p}^N, \mathbf{q}^N, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{2.61}$$

As stated, the method requires several simulations in several partial states between the beginning and the end of the thermodynamic cycle. The greater the number of simulations, the more accurate the method. In general, about 10 simulations are used for precise values, each with λ varying by 0.1 in 10 steps. These simulations involve long times (200 ns +, on average), with extensive minimization and equilibration steps and with well-defined configurations, in addition to the extra computational cost of numerically integrating the Hamiltonian function in terms of λ. As the ligand is partially interpolated into the protein, simulations are configured that consider only electrical interactions, others with only weak van der Waals interactions, and others with both contributions, to ensure the partial interaction of all energy parameters available for the formation of the final complex. Studies of this kind often depend on alchemical thermodynamic cycles, that is, long cycles that are physically impossible to occur (DE RUITER; BORESCH; OOSTENBRINK, 2013), but that have a physically probable final and initial state. This approach is allowed because the path independence of the thermodynamic

state functions is considered (WADE *et al.*, 2022), with the only two points that matter being the final and the initial state.

The second method worth mentioning in this brief introduction is perturbative models. Based on perturbation theory (DE J. RÍOS-ROLDÁN *et al.*, 2024), it is implied that the free energy difference between two states *X* and *Y* can be determined by increasing small perturbations in state *X*, until *Y* is reached. This implies the construction of a partition function of the type (LEACH, 2001):

$$\Delta A = -k_B T \ln \left\{ \frac{\iint d\mathbf{p}^N d\mathbf{q}^N e^{-\mathcal{H}_Y(\mathbf{p}^N,\mathbf{q}^N)/k_B T} e^{+\mathcal{H}_X(\mathbf{p}^N,\mathbf{q}^N)/k_B T} e^{-\mathcal{H}_X(\mathbf{p}^N,\mathbf{q}^N)/k_B T}}{\iint d\mathbf{p}^N d\mathbf{q}^N e^{-\mathcal{H}_X(\mathbf{p}^N,\mathbf{q}^N)/k_B T}} \right\} \qquad (2.62)$$

where $\mathcal{H}_Y$ consists of the Hamiltonian of state *Y* and $\mathcal{H}_X$ the Hamiltonian of state *X*. This requires two simulations in both states to collect representative ensemble averages. For large and flexible systems, these simulations must be even longer to obtain more absolute values. After determining these simulations, a perturbative potential is applied that connects the final and initial states, resulting in the associated free energy value. As with thermodynamic integration, the insertion of intermediate steps may be necessary to adapt the perturbation potential, requiring even more simulations, which may also go through complex alchemical cycles, especially in systems comprising proteins and ligands, which significantly increases the calculation time and, consequently, the computational cost.

In addition to the obstacles already mentioned, protein systems are strongly dependent on the pK_a values of amino acid residues, which may vary depending on the simulation and the achievement of equilibrium (OLSSON *et al.*, 2011). All these precautions must be taken in standard calculations so that the results converge as closely as possible to the experimental values. Given the difficulty and time associated with these methodologies, their application becomes unfeasible in virtual screening protocols, in which several molecular dynamics simulations are necessary to ensure the effectiveness of a ligand complexed in a protein. Based on this need, fast approximation methods (DE RUITER; BORESCH; OOSTENBRINK, 2013; DE RUITER; OOSTENBRINK, 2011) are widely applied based on single dynamics trajectories of the complexes. These fast methods consist of important simplifications, considering implicit solvent models that dramatically reduce the computational cost and calculation time, as will be seen in the next section. It is important to emphasize that, despite being approximate, when fast calculations are well applied, they have the potential to quantitatively estimate the free energy of interaction, with high correlation to experimental data, as long as care such as simulation

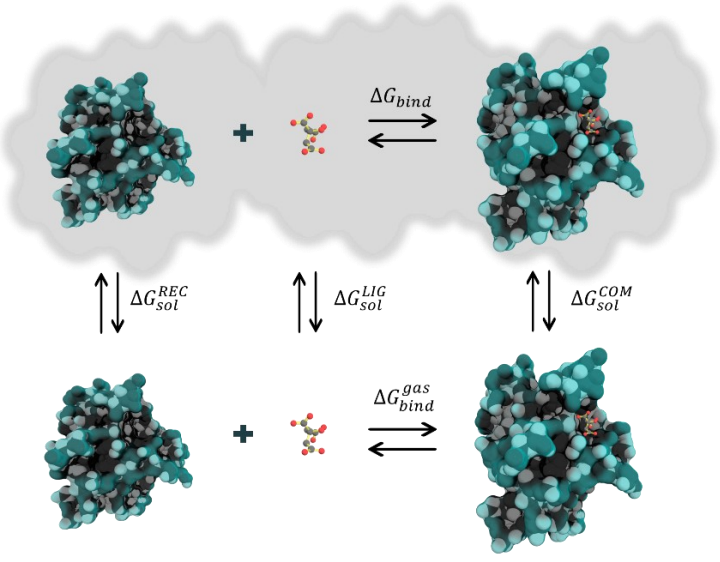
time, minimizations, equilibrations, application of the dielectric constant and quality of the frames are taken into account.

2.3.3 Fast calculations in implicit solvents

2.3.3.1 MM-PBSA and MM-GBSA methods

Molecular Mechanics Poisson-Boltzmann/Generalized-Born Surface Area (MM-PB/GBSA) methodologies are free energy calculation methods based on implicit solvent, applied to non-covalently bound molecular complexes, with emphasis on those composed of proteins and ligands, a system in which the method has found its greatest applicability. In addition, it uses molecular mechanics calculations, based on dynamic simulations, as an end-state method. This free energy calculation approach is simplified, as it does not consider the partition functions of the system components, but it is still quite realistic, especially when the aim is to list a series of complexes and compare the energy affinity values internally. In studies involving several complexes, it is not feasible to apply standard free energy methods, since the computational cost is high and implicit solvent approaches are capable of replacing these approaches with a significant level of confidence (YAU *et al.*, 2019). The standard thermodynamic, cycle used in the algorithms considers the complexation of a ligand to a protein in two stages: in a vacuum, without the effect of the solvent and in solution, according to the reactions in Figure 2.

Figure 2 – Thermodynamic cycle used in MM-PB/GBSA calculations



Source: (VALDÉS-TRESANCO *et al.*, 2021)

Thus, the binding free energy can be calculated from the averages of each frame used in the molecular dynamics trajectory as follows

$$\Delta G_{bind} = \langle G_{prot-lig} \rangle - \langle G_{prot} \rangle - \langle G_{lig} \rangle \tag{2.63}$$

in which each free energy term can be calculated considering some thermodynamic assumptions. The total free energy is given by the contribution of the energy of the solute (protein, ligand and complex) and the solvent:

$$G = G_{solu} + G_{solv} \tag{2.64}$$

By straightforward definition for solutes, $G_{solu} = H_{solu} - TS_{solu}$ and the enthalpy is $H_{solu} = U_{solu} + pV$. In molecular simulations, the volume of the solvation box is extremely small, making the enthalpy term $pV \approx 0$ and the enthalpy contribution reduced to the internal energy of the system, $H_{solu} \approx U_{solu}$. The internal energy of solutes can further be approximated as the potential energy calculated by the force fields E_{MM}, thus $U_{solu} \approx E_{MM}$ and $H_{solu} \approx E_{MM}$. With these considerations, $G_{solu} = E_{MM} - TS_{solu}$. Thus, the total free energy is calculated by

$$G = E_{MM} - TS_{solu} + G_{solv} \tag{2.65}$$

or, considering that the entropy of the solute is the total entropy of the system, since the application of implicit solvent disregards the physical form of the solvent, considering only the dielectric medium triggered by it, we have the equation (2.66) (VALDÉS-TRESANCO *et al.*, 2021)

$$G = E_{MM} + G_{solv} - TS \tag{2.66}$$

This equation is applied to all stages of the thermodynamic cycle, remembering that in the complexation stages in vacuum, there is no solvent, therefore $G_{solv} = 0$.

The contribution of the solvent is calculated from the polar and nonpolar solvation energy:

$$\Delta G_{sol} = \Delta G_{polar} + \Delta G_{non-polar} \tag{2.67}$$

The polar solvation part is calculated according to the Poisson-Boltzmann equation (MM-PBSA) or from the generalized Born model (MM-GBSA), as already extensively discussed in the previous topics, from the continuous electric potentials and their relation with the free energy. The nonpolar contribution term is given by

$$\Delta G_{non-polar} = \Delta G_{disp} + \Delta G_{cavity} = \Delta G_{disp} + (\gamma \cdot SASA + \beta) \tag{2.68}$$

where γ is the cavity tension term and β is the cavity displacement term. SASA is the surface area accessible to the solvent (BORJIAN BOROUJENI *et al.*, 2021). In general terms, it is the portion of the system (usually the protein) that can interact with water molecules, for example. The calculation of SASA corresponds to the determination of the surface area of the protein

capable of touching another surface area relative to water. The cavity tension is related to the energy required per unit area to create a cavity in the solvent. The cavity displacement consists of the energy required for the solvent to move to allocate the solute in the formed cavity. Finally, the dispersion term refers to the non-electrostatic interaction energy between the solute and solvent in the solvation process (SHENG; CHEN, 2021).

We only lack the entropic contribution, which is extremely important for the reliable estimation of the receptor-ligand binding energy. There are several methods already reported in the literature to quickly calculate the entropy of large and complex systems, without the direct application of quantum mechanics. One of the most accurate methods, but with the highest computational cost, is the calculation of Quasi-Harmonic entropy (QH) (PEREIRA; CECCHINI, 2021), given by the equation below:

$$S_{QH} = k_B \sum_{i=1}^{3N-6} \left[\ln\left(\frac{k_B T}{h\nu_i}\right) + 1 \right] \tag{2.69}$$

where ν_i is the vibrational frequency of the atoms that make up the system. This method is based on the atomic vibrations of atoms that are interconnected by small imaginary springs and estimates the vibrational entropy of the system. The greater the displacement of the atoms in the system and the greater the thermal vibration, the greater the entropy related by the QH method.

Another method for calculating entropy, faster and with lower computational cost, is the C2 method (EKBERG; RYDE, 2021), which estimates the conformational entropy of the complex:

$$S_{C2} = -\frac{1}{2} k_B \ln[(2\pi e^n) \det \mathbf{C}] \tag{2.70}$$

where **C** is the covariance matrix of the system. This approach is based on the assumption that there is no harmonic complex and is an approximation of the standard classical statistical entropy, based on a second-order cumulant expansion, in which the probability density function is replaced by the covariance matrix.

Finally, the application of MM-PB/GBSA calculations requires some well-founded caution. Starting with the choice of the internal dielectric constant of the system, which consists mainly of the contribution of the protein (VALDÉS-TRESANCO *et al.*, 2021). The solvent maintains a continuous dielectric constant of 80, but the solute must be very well characterized in this sense, with constants varying, ideally from 1 to 4 (LEACH, 2001). In reality, there is no specific dielectric constant for proteins, since this value is not static, but rather variable in different portions of the macromolecule. To do this, it is necessary to know the system of

interest well, apply extra calculations to determine the average dielectric constant, or use a standard complex, with known energy, to determine, by optimization, the best dielectric constant that results in the energy value that most closely resembles the experimental value.

In addition, it is also important to consider the charge density of the system, if the protein, for example, has regions highly charged by ionic amino acids, or if the total valence is too positive or negative (THE AMBER DEVELOPERS, 2024). This results in the excessive addition of ions to balance the system's charges, which influences strenuous fluctuations in the dielectric constant, making the calculations less accurate. The choice of the simulation interval that generates the calculation frames is extremely important. Priority should be given to frames that represent the stable complex, with low energy and conformational fluctuations, as close as possible to chemical equilibrium (ORMEÑO; GENERAL, 2024), since the theoretical construction of the model starts from equilibrium. One way to detect these frames is by analyzing energy or structural convergence, such as RMSD or radius of gyration, extracted from molecular dynamics simulations. Studies (CRABB *et al.*, 2020) have already shown that the quality of frames is better than the quantity, since good frames in smaller quantities represent more realistic free energy values than frames in larger quantities, but far from chemical equilibrium. Finally, given the critical characteristic of choosing the right frames, in the second article presented in this dissertation, a Free Energy Landscape analysis was applied before the free energy calculations, to obtain more stable conformations, so that they could be selected for later calculations (AL-KHAFAJI; TASKIN TOK, 2020). In addition, it is not possible to ignore the relevance of the implicit solvent method chosen. For calculations with molecular mechanics, the PB equation is preferred, as it better represents the geometry of the system and estimates the electrostatic potential more relevantly, despite the higher computational cost. The use of the nonlinear equation is preferred for even more feasible results (MILLER *et al.*, 2012). For more qualitative calculations, the GB method can be coupled with molecular mechanics for rapid results, with the aim of internal comparison between complexes. The GB method is better assimilated to hybrid molecular mechanics approaches, as will be seen below.

2.3.3.2 Hybrid semi-empirical approaches (QM/MM-GBSA)

Implicit solvent approaches are commonly applied in conjunction with hybrid methods involving molecular and quantum mechanics, with the application of semi-empirical Hamiltonians, which ensure an ideal balance between calculation accuracy and computational cost (THE AMBER DEVELOPERS, 2024). In protein-ligand systems, the active site residues

and the ligand are treated quantum mechanically by means of the semi-empirical molecular Hamiltonian operator, while the rest of the protein is treated by the classical Hamiltonian function (MM) and the solvent is implicitly considered by the Born model (QM/MM-GBSA). The total Gibbs energy is written in terms of the quantum and classical contributions (2.71):

$$G = (E_{MM} + E_{QM}) + G_{solv} - TS \tag{2.71}$$

this ensures that the critical portion of the system (ligand and active site) are treated more robustly, considering the energies related to the electronic effects of the components, thus generating even more accurate results.

Most of the semi-empirical models used are derived from NDDO (Neglect of Diatomic Differential Overlap), which analytically considers only some two-electron integrals, such as those related to the same atom in an approximate solution of the Schrödinger equation (CHRISTENSEN *et al.*, 2016). The more complex and computationally expensive integrals, such as resonance, exchange and Coulomb integrals, are assigned empirically adjusted parameters, or previously calculated by more consistent ab initio methods, such as Hartree-Fock and DFT. The first major model of this generation of semi-empirical methods was the MNDO (ONG; CAO; KWEK, 2023) (Modified Neglect of Diatomic Overlap), parameterized for a series of organic molecules, but with inconsistencies related to molecular geometries and dispersion terms of weak interactions (van der Waals). For this, the AM1 (DEWAR *et al.*, 1985) model, based on the MNDO, proposed a correction with the increase of Gaussian functions in the Hamiltonian, to model weak interactions. The PM3 and PM6 (ŘEZÁČ *et al.*, 2009) models emerged later to increase more robust parameterizations for practically the entire periodic table, beyond the s and p blocks.

However, the PM6 model did not quantitatively consider the dispersions related to hydrogen bonds, which consist of a combination of van der Waals and electrostatic interactions, and subsequent models were developed that consider these corrections, such as the PM6-DH+ (URQUIZA-CARVALHO; FRAGOSO; ROCHA, 2016). The Hamiltonian of the PM6-DH+ model (Appendix C.6) consists of the Hamiltonian of the PM6 method, based on the parameterized Fock operator, with a correction term for the dispersion and hydrogen bonding energy, in which the exchange integrals $(\mu\sigma|\lambda\mu)$, and the orbital interactions of different atomic centers, such as $(\mu\nu|\lambda\sigma)$ are parameterized, the overlap integrals between different atoms $\langle\phi_\mu|\phi_\nu\rangle$ are considered zero and the integrals involving four or more orbitals, of different atomic centers are totally neglected (STEWART, 2013). The Coulomb integrals for the same atom are

maintained, just like diagonal monatomic integrals. The integrals of the type (*ss*|*ss*), (*sp*|*sp*), (*pp*|*pp*), etc. are also maintained (LEACH, 2001) (Appendix C.6). Thus, the integration of semi-empirical methods in QM/MM-GBSA free energy calculations are relevant in the sense of providing a more solid approach, considering the electronic effects present in the active site, which corresponds to more robust calculations, with an accessible computational cost.

2.3.4 Free energy landscapes

Free energy landscapes (FELs) are considered free energy approximations from the *NpT* ensemble for the qualitative characterization of low-energy conformational states in molecular dynamics simulations. Like implicit solvent-based methods, this is an end-state methodology, in which the already finalized molecular dynamics trajectories are used. FEL graphs containing free energy surfaces are created from structural descriptors, such as RMSD and radius of gyration, which provide information about the movement, flexibility, and compactness of the system (ABOUZIED *et al.*, 2024). Despite being widely used descriptors in the scientific literature, they are often unable to detect subtle conformational fluctuations, allowing metastates with important influence on the system to go unnoticed (CAPELLI *et al.*, 2019). One way to generate more reliable FELs is to use principal components (PCs), which reduce the conformational dimensionality of the system, so that it is possible to detect smaller fluctuations and encompass a range of better described microstates (PAPALEO *et al.*, 2009). The PCs are obtained from the decomposition into eigenvalues and eigenvectors of the covariance matrix of the atomic positions throughout the simulation. This matrix represents how different regions of the molecule move together, and its analysis allows us to identify the modes of movement most relevant to the conformational dynamics.

From the calculations of the first two principal components, the Gibbs energy is then calculated by defining the *NpT* ensemble, with an adaptation of the partition function.

$$\mathcal{Z}_{NpT} \approx P(PC1, PC2) = \frac{1}{n h_{PC1} h_{PC2} (2\pi)} \times \sum_{i=1}^{n} \exp\left(-\frac{1}{2} \left[\left(\frac{x_{PC1} - x_{PC1,i}}{h_{PC1}} \right)^2 + \left(\frac{y_{PC2} - y_{PC2,i}}{h_{PC2}} \right)^2 \right] \right)$$

$$G \approx -k_B T \ln P(PC1, PC2) \tag{2.72}$$

where $P(PC1, PC2)$ represents the kernel density function (DELAIGLE; GIJBELS, 2004) estimated at point (x, y), n is the total number of samples obtained from the simulation, and (x_i, y_i) are the values of the variables PC1 and PC2 for each sample i. The parameter h corresponds

to the bandwidths (DELAIGLE; GIJBELS, 2004) used to smooth the distribution along each dimension. The kernel function adopted in this case is an isotropic bivariate Gaussian (NOH; AND LEE, 2013), which ensures symmetric smoothing around each sampled point.

P(*PC1*, *PC2*) represents the probability density function constructed from the principal components. Several methodologies can be used to normalize and smooth the projected data, such as kernel density estimates, Gaussian functions or approximations based on two-dimensional histograms. The choice of density estimation method directly impacts the shape of the free energy surface, influencing both the width and depth of the energy wells, in addition to affecting the overall smoothness of the energy profile (KRIVOV, 2011). In scenarios where statistical noise is present, it is common to apply Gaussian functions in order to smooth the distribution. However, this smoothing should be performed with caution, as there is a risk of masking relevant transient meta-states, confusing them with random noise and, thus, compromising the thermodynamic interpretation of the system (CHONG; HAM, 2019).

2.4 Introduction to biological and pharmacological context

Understanding the biological system is of utmost importance in the application of external computational methodologies for the development of new drugs. In this context, it is necessary to understand autism as a multifactorial disorder, with an important genetic and hereditary root that depends on several metabolic pathways mediated by several receptors. Therefore, this subsection will examine details of the pathology related to autism, the biochemical mechanisms that allow this condition and the pharmacological interaction with target receptors that can help to mitigate the effects of this disorder.

2.4.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD), also known as autism, is a genetic and hereditary neuropsychiatric condition (LORD *et al.*, 2020). Although it is an exceedingly heterogeneous disorder – manifesting differently in every patient – it does have some basic symptoms, such as social interaction and communication impairment, repetitive behavior and restricted interests (HODGES; FEALKO; SOARES, 2020). The diagnosis is clinical, based on behavioral assessment primarily carried out by neurologists and/or psychiatrists (JIANG *et al.*, 2023). Autism is typically diagnosed in children, though late diagnoses in adulthood are not uncommon (NYRENIUS *et al.*, 2023).

The occurrence of autism across the globe has been increasing. According to the World Health Organization (WHO), the condition is found in approximately one in every hundred children, with males having a higher incidence (SALGADO *et al.*, 2022). The rise can be explained by greater public awareness and more sophisticated diagnostic criteria. In its more severe forms, autism can profoundly impact the development of cognition, leading to disabilities in areas of communication and language (PICKLES *et al.*, 2022). Total absence of speech is observed in certain individuals. Autism is also very strongly associated with Fragile X Syndrome (FXS), caused by mutations in the FMR1 gene (BRAŠIĆ *et al.*, 2021), which is documented in the literature as a monogenic cause of autism in humans.

Individuals with autism are a high-risk population for the comorbidities of depression, anxiety, phobias, and suicide – risks far higher than for neurotypical individuals (MENEZES *et al.*, 2020; MUSKETT *et al.*, 2019). These psychological symptoms, which may also involve panic and generalized anxiety disorders, can be the result of social exhaustion, overstimulation, and stigma experienced by autistic individuals, leading to isolation and, in some cases, suicidal tendencies (MAYES *et al.*, 2013; NYRENIUS *et al.*, 2023; UPTHEGROVE *et al.*, 2018) – even among children. The levels of autism are defined by the need for support. Level 1 characterizes the need for occasional support with less intense social and cognitive impairment (SPAIN; ZIVRALI YARAR; HAPPÉ, 2020). Level 2 is more severe, but not constantly, needing support as a consequence of moderate deficits, which can still facilitate some independence in daily functioning (GRANIERI *et al.*, 2020). Level 3 characterizes the most challenging social-cognitive impairments and behavioral fixity, requiring extensive and life-long support (HAMMOND; HOFFMAN, 2014). Individuals at this level typically have issues with basic activities such as good hygiene and nutrition.

The etiology of ASD remains controversial in both the medical and scientific fields, though there is a consensus that it is caused by a combination of environmental and genetic factors (TROPICALE *et al.*, 2021). There are over 150 genes identified to be associated with autism, particularly those involving synaptic function, neuroplasticity, and brain development (SERAJEE *et al.*, 2003) – indicating its genetic component. A high occurrence of autistic children being born to parents who are also on the spectrum has been documented in research (ALYMOV; KAPITSA; VORONINA, 2021). Environmental risk factors, such as prenatal exposure to toxins and nutritional deficiencies, are particularly relevant during prenatal and perinatal phases (CHIEN *et al.*, 2019). Epigenetic alterations, involving DNA methylation and histone acetylation, are also believed to influence the neurodevelopmental processes underlying autism (OZTENEKECIOGLU *et al.*, 2021).

Two hypotheses relating to glutamatergic and serotonergic neurotransmission have been among the most notable of the genetic hypotheses. In the former, metabotropic glutamate receptors (mGluRs) play a role in ASD pathogenesis, considering their role in neural plasticity, learning, and memory (CAIA *et al.*, 2016). Glutamatergic signaling disruption is strongly linked to ASD-like phenotypes in animals. The serotonergic hypothesis, on the other hand, deals with the 5-hydroxytryptamine (5-HT) receptors, neural proteins implicated in clinical manifestations of the disorder (KURAHASHI *et al.*, 2025). Serotonin is also a key neuromodulator that influences mood, sensory experiences, and social communication. In the following sections, the involvement of these two neurotransmitter systems in the neurogenesis of ASD will be reviewed in greater detail.

2.4.2 mGluRs

The glutamatergic system is a significant player in the pathophysiology ASD (KOLODNY *et al.*, 2020), and mounting evidence indicates the dysregulation of glutamate signaling as being implicated in the disorder's neurodevelopmental dysanatomy (CAIA *et al.*, 2016). Glutamate is the primary excitatory neurotransmitter of the central nervous system (CNS) and is crucial for synaptic plasticity (ANASHKINA; ERLYKINA, 2021), learning, memory (LV *et al.*, 2024), and overall neuronal communication (NIKITIN *et al.*, 2023). Excitatory glutamatergic and inhibitory GABAergic (gamma-aminobutyric acid) signaling is essential for normal brain function, and disruption of the balance is thought to contribute to the cardinal signs of ASD (SEARS; HEWETT, 2021).

Disruption of glutamate receptor expression and function has been found in ASD (KOLODNY *et al.*, 2020). Specifically, N-methyl-D-aspartate (NMDA) (WANG, X. *et al.*, 2022) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) (NIESCIER; LIN, 2021) receptor dysfunctions, the central glutamate receptors involved in synaptic transmission and plasticity, have been documented (RUSAKOV; STEWART, 2021). Dysfunction of these receptors can cause hyperexcitability or hypoexcitability in certain areas of the brain, disrupting neural networks for social communication, cognitive processing, and behavior regulation (EBRAHIMI-GHIRI; KHAKPAI; ZARRINDAST, 2020).

Early studies involving the use of high-performance liquid chromatography revealed that there were elevated plasma Glu concentrations and reduced plasma glutamine levels in ASD patients (LEE *et al.*, 2023) as well as their first-degree relatives. These findings have been corroborated by subsequent studies, which also ascertained that there was a relationship

between increased plasma Glu concentrations and the severity of symptoms of ASD (OYA *et al.*, 2023).

Glutamate expression changes, reduced excitatory transmission, and NMDA receptor, (NMDAR)-mediated synaptic plasticity (HAIN; MOSER, 2024) and metabotropic glutamate receptor-mediated signaling are responsible for the cognitive and behavioral deficits in ASD (DUFOUR *et al.*, 2023). Metabotropic glutamate receptors (mGluRs) are G protein-coupled receptors that modulate synaptic transmission and neuronal excitability (MCCULLOCK; KAMMERMEIER, 2021). They are divided into three groups: Group I (GHOSH DASTIDAR *et al.*, 2020a) (mGluR1 and mGluR5), which mediate excitatory neurotransmission and are involved in synaptic plasticity (EGBENYA; AIDOO; KYEI, 2021); and Groups II (mGluR2 and mGluR3) (LI et al., 2020) and III (mGluR4, mGluR6, mGluR7, and mGluR8), which inhibit the release of neurotransmitters and maintain neuronal equilibrium (RABEH *et al.*, 2023). mGluRs are directly involved in the pathophysiology of ASD (TEIXEIRA; RAMALHO, 2021), particularly through their implication in the synaptical plasticity and modulation of excitatory neurotransmission (MOXON-EMRE *et al.*, 2021).

They take part in modulating the strength of the synapses and the excitability of the neurons (WU et al., 2024), processes of great importance for the acquisition and memory. Dysregulation of mGluR signaling has been implicated in the cognitive and behavioral deficits characteristic of ASD (EBRAHIMI-GHIRI; KHAKPAI; ZARRINDAST, 2020). Hyperactivated mGluR5 signaling was present in certain animal models of ASD and resulted in the disruption of the excitatory/inhibitory transmission (MCCULLOCK; KAMMERMEIER, 2021). The disruption results in altered synaptic plasticity, like increased long-term depression (GHOSH DASTIDAR *et al.*, 2020a), which has been observed in animal models as well as in human studies of ASD. Therapeutic strategies that focus on the mGluR signal transduction pathways, such as mGluR5 (BRAŠIĆ *et al.*, 2021) antagonists, are explored to attempt to restore normal synaptic function and alleviate ASD symptoms (JONG *et al.*, 2023), but to date, no pharmacologically relevant molecule has been revealed with suitable toxicological profile (DUPONT *et al.*, 2023; GALINEAU *et al.*, 2022), thus the quest is still imperative. Genetic studies identified correlations between autism and the GRM8 gene (LIMON *et al.*, 2011) that encodes the metabotropic glutamate receptor 8 (mGluR8) of group III (KESSI *et al.*, 2022).

It is localized at or near presynaptic sites and, upon activation, inhibits glutamate release from the presynaptic terminal (FISCHER *et al.*, 2022). Hence, malfunction of mGluR8 could result in increased neuronal activity (BUKINA *et al.*, 2021). Rather surprisingly, mGluR8 receptors are present at high density in specific populations of presynaptic GABAergic and

glutamatergic terminals that project to GABAergic interneurons in the rat hippocampus (BRAFA MUSICORO *et al.*, 2024). If this type of distribution is discovered in the human hippocampus, abnormal mGluR8 expression may interfere with GABAergic transmission, impairing the coordination of synchronized activity generated by inhibitory interneurons (LIMON *et al.*, 2011; PARENT; NISWENDER, 2024). With the high homology between metabotropic glutamate receptor 7 (mGluR7) and mGluR8 (GOEL; PENG; LU, 2020), it is not unrealistic to speculate mGluR7 as a candidate in biochemical pathways linked to ASD. mGluR7, like mGluR8, is localized mainly at presynaptic terminals (DOMIN, 2022), where it controls release of neurotransmitter, namely Glu. The structural and functional analogy between both receptors leads to the hypothesis that mGluR7 may be involved in the regulation of synaptic transmission and neuronal excitability (FREITAS; NISWENDER, 2023) in manners that resemble mGluR8. It is highly meaningful for modelling new molecules with pharmacological potential to know the structures and the biochemical pathways of both receptors.

2.4.2.1 Structure, biochemistry and general mechanisms

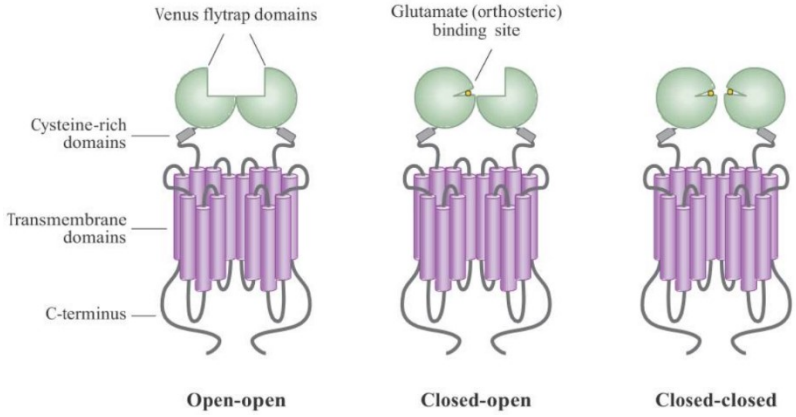
Structurally, the mGluRs have three fundamental subunits: the glutamate-binding component, the Venus Flytrap (VFT) domain, which constitutes the orthosteric binding site, to be centered; cysteine-rich domains (CRDs); and the transmembrane domain, which divides the proteins into intra and extracellular halves as shown in Figure 3. As indicated, VFD is large extracellular N-terminal domain, which has the glutamate-binding site (NISWENDER; CONN, 2010). Structural analysis indicates that two lobes of the VFD bind glutamate in a cleft between them (SCHKERYANTZ et al., 2018).

Evidence indicates that two back-to-back dimerize VFDs and agonist binding induces large conformational changes. There are three main VFD dimer states: open-open (inactive), open-closed, and closed-closed. The open-open conformation is stabilized by antagonists (HENTER; PARK; ZARATE, 2021), and the other conformations are agonist-binding related (LUESSEN; CONN, 2022). Mutations that affect VFD function have the potential to switch receptor pharmacology to agonists from antagonists. In addition, the VFD also binds divalent cations, including magnesium and calcium, which could affect receptor activity (KOEHL *et al.*, 2019).

Ligand binding mGluRs causes conformational shifts cascading from CRDs to the heptahelical domain (HD) and C-terminal tail. Such signal transduction is accomplished by the

CRD, which is made of pivotal cysteine residues linked by disulfide bridges. Studies (MUTO *et al.*, 2007) have shown cysteine interaction in the CRD and VFD are critical to proper signaling, as illustrated by mutations that affect signal propagation but not ligand binding (RONDARD *et al.*, 2006).

Figure 3 – General illustration of the three main domains of mGluRs



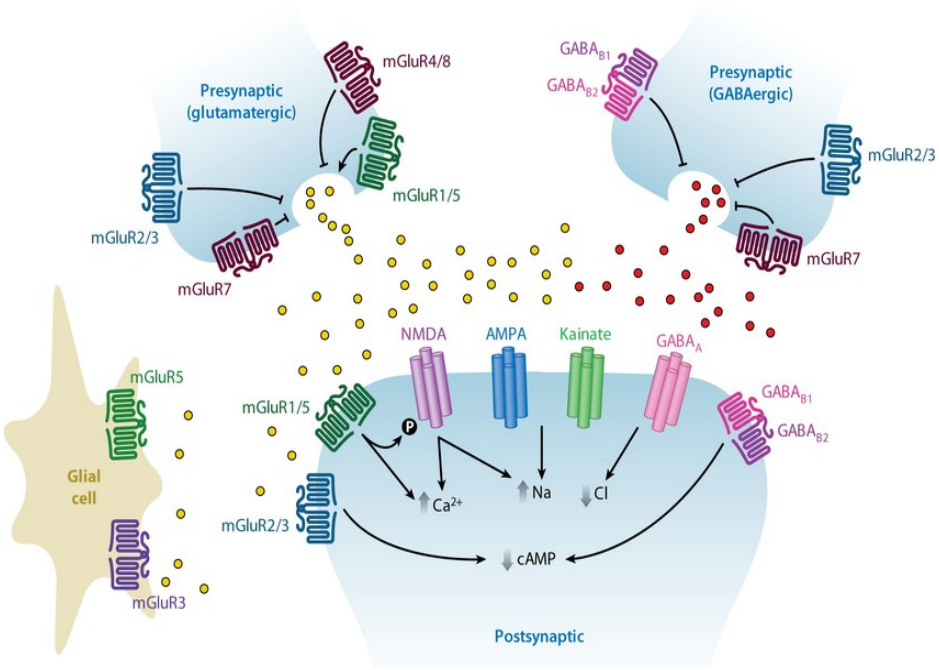
Source: (NISWENDER; CONN, 2010)

Metabotropic glutamate receptors are members of the G-protein-coupled receptor (GPCR) superfamily, the largest receptor gene family within the human genome (CALEBIRO *et al.*, 2021). GPCRs are membrane proteins that are ligand-activated by a large diversity of extracellular ligands, such as light, peptides, and neurotransmitters (VOSS; MÜLLER, 2022). The ligand binding results in a conformational change in the receptor, which engages intracellular G proteins. This interaction triggers a cascade of signaling pathways by switching guanosine-diphosphate (GDP) (WANG, X. *et al.*, 2022) with guanosine-triphosphate (GTP) on the G protein (HUANG, M.; WANG, 2021), thereby influencing various cellular processes, including enzyme activity and ion channel function (PLESTED; POULSEN, 2021). In general, group I mGluRs tend to be postsynaptic (GHOSH DASTIDAR *et al.*, 2020b), while group II and III mGluRs tend to be presynaptic (LI *et al.*, 2020), with some exceptions. Group II (mGluRs 2 and 3) and group III (mGluRs 4, 7, and 8) receptors have the effect of inhibiting the release of the neurotransmitters glutamate and GABA (BRAFA MUSICORO *et al.*, 2024), whereas the group I receptors (mGluRs 1 and 5) promote the release of neurotransmitters (BOLNEO *et al.*, 2022) (Figure 4). In postsynaptic terminal, glutamate-gated ion channels such as NMDA, AMPA, and kainate receptors increase intracellular sodium or calcium, thereby cell excitability is enhanced (WU *et al.*, 2024).

mGluR7 is diffusely distributed throughout the brain (MATSUNAGA; ARUGA, 2021) and has very low affinity for glutamate, strongly localized to active zones of the synapse

(BOCCELLA *et al.*, 2019). It is thought to act as a low-pass filter during neurotransmission, activated only when glutamate levels are high, thereby preventing overstimulation (DOMIN, 2022). Additionally, mGluR7 has been associated with central nervous system conditions, including anxiety (MAZZITELLI *et al.*, 2022) and depression (KADRIU *et al.*, 2021). mGluR8, while less highly expressed at low levels of Glu than mGluR4 and mGluR7, and found throughout the brain (RABEH et al., 2023), primarily presynaptic but also found in some postsynaptic sites and peripheral tissue (EBRAHIMI-GHIRI; KHAKPAI; ZARRINDAST, 2020a).

Figure 4 – mGluRs neuronal localization and function



Source: (NISWENDER; CONN, 2010)

Higher levels of glutamate within neurons exert action on group I of mGluRs, i.e., mGluR5 (STOPPEL *et al.*, 2021), that starts the series of signals mediating glutamatergic synapses to facilitate the transfer of glutamate to postsynaptic neurons. Glutamate release in presynaptic neurons is regulated by group III of mGluRs (YANG, Z.-Q., 2005), namely mGluR7 and mGluR8, that bind glutamate, thereby inhibiting its release and synaptic transmission. Group III dysfunction leads to decreased inhibition, thereby causing an increase in synaptic activity (DOMIN, 2022). This can lead to pathological processes of social development, hypersensitization, and cognitive impairments (KOLODNY et al., 2020).

Hence, Group III receptor agonist administration, such as mGluR7 and mGluR8, would enhance the reuptake of glutamate, inhibit overactive synaptic activity, and restore neurotransmitter homeostasis (ELTOKHI *et al.*, 2021), thereby mitigating glutamate

dysregulation complications. Dual agonist research of mGluR7/8 can be done through a virtual screening process, a useful drug discovery tool in medicinal chemistry research.

2.4.3 5-HT$_{1A}$

The 5-hydroxytryptamine, or serotonin, 1A (5-HT$_{1A}$) receptor, a member of the G-protein coupled receptors (GPCRs), has the typical seven transmembrane helices (TM1 to TM7) joined by extracellular (ECL) and intracellular (ICL) loops, and a cytoplasmic carboxy-terminal tail – typical features of the structural topology of GPCRs (XU, P. *et al.*, 2021). The three-dimensional structure of the receptor is very dynamic, taking active or inactive conformational states in response to the presence of ligand, lipids, and coupling with the G protein. Functionally, 5-HT$_{1A}$ is also classified into two general populations: autoreceptors within the raphe nucleus-harboring serotonergic neurons and heteroreceptors within non-serotonergic neurons of regions such as the hippocampus, prefrontal cortex, and amygdala. Structurally, these versions are not distinguishable in primary sequence but are variant in location and lipid microenvironment, influencing their functional configuration (BORROTO-ESCUELA *et al.*, 2018).

The active receptor conformation is stabilized by occupation of the orthosteric binding site between TM3, TM5 and TM6 helices by serotonin or by synthetic agonist. Furthermore, electron cryomicroscopy structures revealed that even in APO form, 5-HT$_{1A}$ could be in a pre-activated conformation that is stabilized by structured water molecules resembling serotonin, which is responsible for its high basal activity compared to other 5-HT subtypes (XU, P. *et al.*, 2021). Inside the intracellular space, 5-HT$_{1A}$ interacts predominantly with G$_{i/o}$-type G proteins. Inside raphe serotonergic neurons, it is most commonly paired with the G$_{i3}$ subunit, while in hippocampal neurons it has a preference for G$_{i2}$, to confer differences between desensitization and activation profiles of downstream signaling (ALBERT; VAHID-ANSARI, 2019).

The mechanisms of these actions include inhibition of adenylyl cyclase, GIRK-type potassium channel opening and VDCC-type calcium channel closing, and activation of signal cascades such as ERK1/2 and PI3K-Akt, especially when the receptor is bound in complexes with tyrosine kinase receptors such as FGFR1 and PDGFRβ. The structural regulation of 5-HT$_{1A}$ is also dependent on membrane lipids (XU, P. *et al.*, 2021). A current study (ZWECKSTETTER; DITYATEV; PONIMASKIN, 2021) have proven that binding to phosphatidylinositol-4-phosphate (PtdIns4P) at the G-protein coupling locus and cholesterol

molecule availability affect the receptor and ligand conformational stability and affinity, for instance, to antipsychotic aripiprazole.

In the case of ligands, the receptor has high binding affinity to serotonin, which interacts with conserved residues such as aspartate and Tryptophan. Besides endogenous agonists, drugs like 8-OH-DPAT, buspirone and aripiprazole modulate its activity selectively. But most of these ligands are unable to distinguish between autoreceptors and heteroreceptors, which can produce pharmacologically opposite effects (BANTICK *et al.*, 2004). The progress in crystallography and molecular modeling in recent times has made it possible to identify key residues for selectivity, e.g., Glutamate and Lysine, opening the doors to the design of more selective drugs. Therefore, 5-HT_{1A} is a receptor whose form is solidly anchored in its function, with static (helices and conserved residues) and dynamic aspects (conformational state, lipid presence and coupling to different G proteins), the latter playing a key role in its function in the regulation of neuropsychiatric functions (BAUER, 2015).

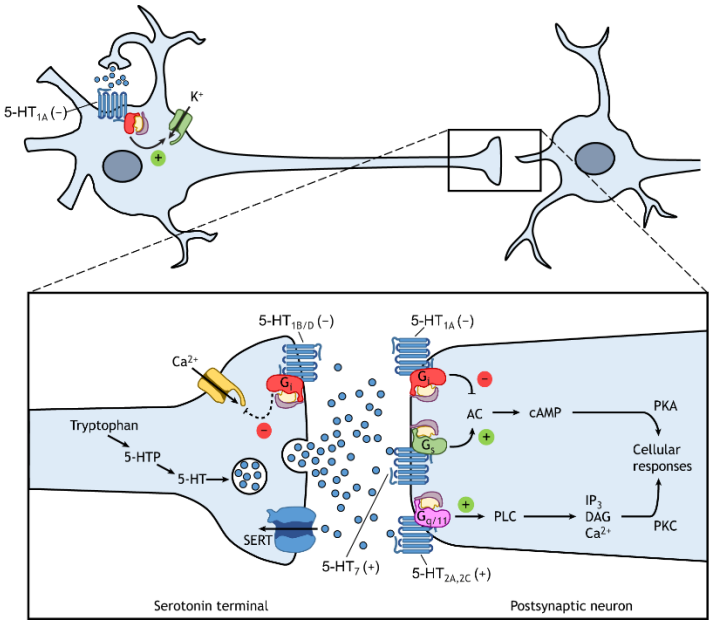
In terms of neuron localization Figure 5, the serotonergic neuron (serotonin terminal), the 5-HT_{1A} receptor is presynaptic in the membrane, where it also acts as an autoreceptor. It negatively regulates the release of serotonin by inhibiting adenylyl cyclase (AC) activity via the G_i protein, lowering cAMP levels and leading to the opening of potassium (K^+) channels, hyperpolarization of the membrane, and a reduction in neuron excitability (ALBERT; VAHID-ANSARI, 2019). In the postsynaptic neuron, 5-HT_{1A} receptor also acts as a heteroreceptor, partnered with the G_i protein, with the same inhibitory effect on adenylyl cyclase. This results in reduced intracellular signaling via PKA, modulating some cellular responses, such as regulation of excitability, plasticity and synaptic activity. Therefore, 5-HT_{1A} can play inhibitory functions both in the modulation of the release of serotonin (autoreceptor) and in the response of the target neuron to released serotonin (heteroreceptor), demonstrating its importance in the homeostasis of serotonergic neurotransmission (BUBAK *et al.*, 2020).

The 5-HT_{1A} serotonin receptor is also crucial in the modulation of neural circuits that lie within social, emotional and repetitive behavior, which are the axes that get disrupted in the case of autism spectrum disorder (ASD). These regions of the brain include areas such as the hippocampus and cortex. 5-HT_{1A} modulates neuron excitability, release of neurotransmitters and plasticity of the synapse, directly affecting the dynamics of emotion and emotion from the beginning of development (WANG, C.-C. *et al.*, 2013).

Functional alterations of the serotonergic system, and especially 5-HT_{1A}, are associated with abnormality in regulation of anxiety, repetitive behavior pattern and cognitive disturbance in autism. Hypofunction of this receptor may be accountable for increased rigidity of behavior

and social stimulus avoidance, two key characteristics of autism spectrum (KURAHASHI *et al.*, 2025). On the other hand, modulation of 5-HT$_{1A}$ signaling tends to inhibit repetitive behavior and promote more balanced exploratory responses, so that there could be a direct relationship between the functionality of this receptor and behavioral flexibility. In addition to being involved in its autocrine function, 5-HT$_{1A}$ also has a functional interaction with the 5-HT$_7$ receptor, with which it heterodimerizes to regulate its signaling, internalization, and expression. This interaction, in regulating the sensitivity of the serotonergic system, can significantly influence neurochemical homeostasis, especially in pathological conditions with dysregulation of inhibitory and excitatory mechanisms, for instance, in autism (MARTIN *et al.*, 2025).

Figure 5 – 5-HT$_{1A}$ neuronal localization



Source: (BUBAK *et al.*, 2020)

The other interaction axis of interest is between 5-HT$_{1A}$ and the neurotrophic system, specifically with the TrkB receptor, through which the actions of BDNF (brain-derived neurotrophic factor) are mediated. Balanced signaling through both 5-HT$_{1A}$ and TrkB is required for neurogenesis, maturation of synaptic networks, and emotional modulation (KONDAUROVA *et al.*, 2023). The inhibition of this axis can influence cortical and hippocampal development, leading to the behavior patterns seen in ASD. The reduction in TrkB expression, with preservation of BDNF levels, suggests a qualitative impairment in the process of neurotrophic transduction, conceivably linked to central serotonergic dysfunctions (ALBERT; VAHID-ANSARI, 2019).

Thus, the 5-HT_{1A} receptor emerges as a structurally and functionally significant element in autism pathogenesis, influencing core facets of behavior through interactions with additional serotonergic receptors and neural plasticity mechanisms. Its roles in anxiety modulation, repetition of behavior and synaptic architecture make it an important target for therapeutic interventions designed at normalizing neurochemical and behavioral asynchrony in ASD (KURAHASHI *et al.*, 2025).

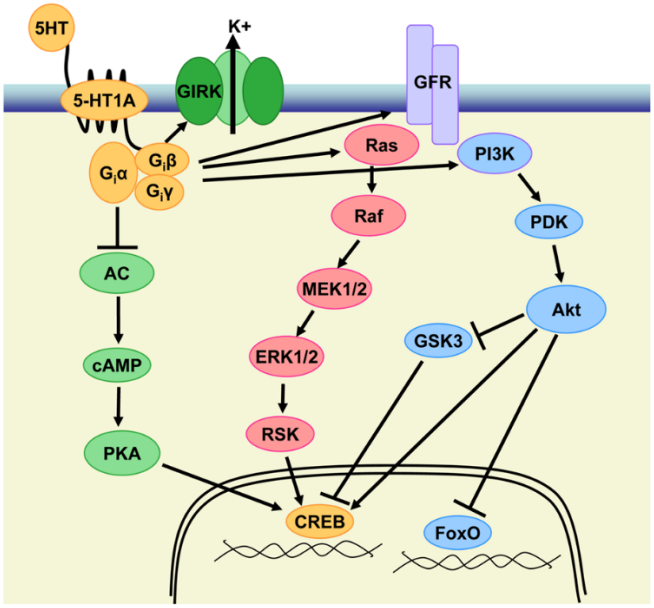
2.4.3.1 Metabolic pathways

5-HT_{1A} receptor metabolism is a complex set of biochemical processes that control its expression, activation, desensitization and intracellular signaling, and has a central role in brain physiology and in neuropsychiatric pathophysiology, in particular for autism spectrum disorder (ASD). As a G protein-coupled receptor (GPCR), 5-HT_{1A} is maximally activated by G_i/G_o proteins and its activation governs a variety of metabolic cascades that directly influence the neuronal excitability, synaptic plasticity and cell survival (POLTER; LI, 2010). Activation of 5-HT_{1A} happens following binding of the orthosteric receptor site with serotonin or drug agonists such that GDP to GTP exchange on the α subunit of the G_i protein is catalyzed. The $G_{\alpha i}$ subunit is inhibitory to adenylyl cyclase (AC) and reduces intracellular cyclic AMP (cAMP) and thus activity of protein kinase A (PKA) (Figure 6). The conventional pathway effectively reduces neuronal excitability and release of neurotransmitter and opening of potassium channels (GIRK) and closing of N-type calcium channels, hyperpolarizing the cell (COURTNEY; FORD, 2016).

Analysing the signal pathway of Figure 6, when activated, PKA can phosphorylate nuclear proteins such as the transcription factor CREB, modulating gene expression. In parallel, the βγ subunits of the G_i protein activate G protein-regulated potassium channels (GIRK), promoting the efflux of K^+ ions and hyperpolarization of the cell membrane. These subunits also participate in the activation of other signaling pathways, such as the MAPK/ERK pathway and the PI3K/Akt pathway (POLTER; LI, 2010). In the MAPK/ERK pathway, activation of the Ras protein leads to a cascade of phosphorylation of Raf, MEK1/2, and ERK1/2, culminating in the activation of the RSK kinase, which also acts on CREB in the cell nucleus. Simultaneously, activation of PI3K leads to phosphorylation of PDK and subsequent activation of the Akt protein. This, in turn, inhibits proteins such as GSK3 and the FoxO transcription factor, promoting effects related to cell survival and regulation of gene expression. Thus,

signaling via 5-HT$_{1A}$ coordinates multiple intracellular pathways involved in neuroplasticity, neuroprotection, stress response, and mood modulation (ALBERT; VAHID-ANSARI, 2019).

Figure 6 – 5HT$_{1A}$ metabolic pathway



Source: (POLTER; LI, 2010)

In addition to this canonical pathway, 5-HT$_{1A}$ also triggers other pathways via the βγ subunit of the G protein to activate phospholipase C β2/3 (PLCβ), which generates inositol triphosphate (IP3) and diacylglycerol (DAG). These messengers increase intracellular calcium and activate protein kinase C (PKC), which is a positive regulator of signaling pathways for synaptic plasticity and neuronal development, such as calcium-calmodulin-dependent kinases (CAMKII) and the PI3K/Akt/mTOR pathway (POLTER; LI, 2010). In autism, these biochemical processes have a particular significance. There is some evidence in autistic mouse models, i.e., BTBR mice, that functional 5-HT$_{1A}$ signaling is disrupted with decreased cellular response and unchanged expression levels of the receptor. The observation suggests engagement of functional desensitization or uncoupling from intracellular transduction processes, especially in the hippocampus – a structure of special interest in contextual memory modulation as well as modulation of repetitive behavior (ALBERT; VAHID-ANSARI, 2019). The 5-HT$_{1A}$ interaction with BDNF activation of the TrkB receptor adds another level of metabolic control that is significant in ASD pathophysiology. Activation of 5-HT$_{1A}$ reduces the expression level of TrkB receptors, an action which controls trophic response to BDNF and thus has immediate implications for synaptic plasticity and the formation of functional neural networks. Overexpression of 5-HT$_{1A}$ in the autistic model brain was discovered to reduce TrkB but not BDNF levels, demonstrating a derangement of the balance of function of the serotonin–

BDNF system to cause development and remodeling of synapses (KONDAUROVA *et al.*, 2023).

Further, 5-HT_{1A} is even known to heterodimerize with the 5-HT_7 receptor, a process that suppresses 5-HT_{1A} signaling itself. Such heterodimerization corrupts the proper activation of the aforementioned metabolic events and a compensatory upregulation of 5-HT_7 expression was also detected upon 5-HT_{1A} overexpression in BTBR mice, recapitulating an adaptive process liable to propagating synaptic dysfunction in autism. Finally, modifications of the 5-HT_{1A} receptor also impact nervous system mechanisms of neuroprotection and neurogenesis. Neural progenitor proliferation and differentiation both increase under chronic activation of the receptor, and its signaling through PI3K/Akt and MAPK is used in neuronal survival and the proper organization of circuits (KONDAUROVA *et al.*, 2023). Dysruptions of these pathways of brain development (either through genetic, epigenetic, or functional alteration of the receptor) may result in aberrant formation of synaptic profiles and in normal behavior of ASD. Thus, metabolism of 5-HT_{1A} receptor is a part of intrinsic biochemical pathway of regulation of synaptic homeostasis, and its dysregulation leads to pathogenesis of neuropsychiatric disease such as autism directly. The molecular processes of such signaling must be understood in order to create targeted treatments that may restore the functional balance of broken brain circuits (SAITOW; TAKUMI; SUZUKI, 2020).

2.5 Fundamentals of computational medicinal chemistry

Medicinal chemistry is a multidisciplinary science at the border of chemistry, biology, and pharmacology with the intention of designing, synthesis, and developing bioactive entities for medical therapy (AUBERSON *et al.*, 2023). It is the art of using chemical principles to the research and management of biological systems with the objective of discovering new drugs and improving existing ones (BORN *et al.*, 2021). This section includes the identification of new drug targets, optimization of drug lead candidates for activity as well as safety, and identification of mechanisms of action.

Advances in computational chemistry, molecular biology, and experimental protein identification methods have made it easier to identify a number of disease-related molecular targets and determine their three-dimensional structures (TANG, Y.; MORETTI; MEILER, 2024). This set of conditions, combined with the increasing need for drug-receptor interaction characterization at the molecular level and the demand to develop drugs in increasingly short times, has led to the paradigm known as Structure-Based Drug Design (SBDD).

Computational techniques are significant in medicinal chemistry by enhancing the effectiveness and precision of the drug discovery process (MELLER *et al.*, 2024). One significant application is in lead compound identification and optimization. Techniques such as molecular docking and virtual screening allow chemists to rapidly screen large compound libraries to approximate their binding capacity to biological targets. SBDD applies high-resolution protein target structures, obtained by techniques like x-ray crystallography (BACELLAR; CHERGUI, 2022), magnetic nuclear resonance (SHAHRAJABIAN; SUN, 2024) or cryo-electron microscopy (KIKKAWA; YANAGISAWA, 2022), to create molecules that bind precisely into the active site.

Molecular dynamics simulations next fine-tune these designs by predicting drug molecule and target dynamics over time, revealing stability and binding interactions (CHEN *et al.*, 2024). Cheminformatics tools and artificial intelligence (NIAZI; MARIAM, 2024) algorithms also support the identification of novel chemical entities in addition to predicting their pharmacokinetic and toxicological properties (YANG, Y. *et al.*, 2024), thus streamlining the drug development process and enhancing the likelihood of success in clinical trials. Therefore, once the disease is identified, the first action is to identify the molecular target responsible for its initiation and that can be treated chemotherapeutically.

Having obtained that structure, the second action is computational screening or *in silico* screening (SAVA, 2023). During this stage, molecular docking must be achieved quickly and selectively (BLANES-MIRA *et al.*, 2023), employing compound libraries that hold the structures of different small molecules. It attempts to identify the highest possible number of compounds with the optimal interaction with the binding site of the macromolecule effectively by virtual screening.

2.5.1 Virtual Screening

Virtual screening is a powerful computational technique used to identify promising new molecules from massive libraries of compounds for potential therapeutic applications (KILLAM *et al.*, 2024). The technique efficiently truncates vast collections of chemical compounds by predicting the quality of interaction between them and specific biological targets (LIMA *et al.*, 2024), thereby accelerating the process of drug discovery. Target preparation involves preparing a chemically diverse compound library for screening. These libraries might contain a wide range of chemical entities, from small molecules to complex natural products (OLMEDO *et al.*, 2023). Major sources of compound libraries include the ZINC Database

(TINGLE *et al.*, 2023), which includes commercially accessible compounds; PubChem (KIM et al., 2016), a large collection of chemical molecules and biological activity; Enamine (GRYGORENKO, 2021), providing lead-like and drug-like molecules; and other available libraries for free access.

The spirit of virtual screening is docking simulations (DU, L. *et al.*, 2023), in which all compounds within the library are theoretically docked into the target protein binding site. Docking algorithms predict the optimal binding orientations and conformations of the compounds and produce docked poses that describe how each molecule would bind to the target (VIDAL-LIMON; AGUILAR-TOALÁ; LICEAGA, 2022). The highest-scoring ones are selected for theoretical toxicologic studies to ensure that the identified compounds are safe and effective in development.

Theoretical ADMET analysis is a computer-based technique used to predict the pharmacokinetic and safety profiles of lead drug candidates in early stages of the drug development pipeline (MITEVA; VILLOUTREIX, 2017). Using various computational models and software, it is possible to make estimations of how the drug will work within the body and therefore select lead compounds and reduce late-stage failure. To predict how well a drug will be absorbed when administered orally, *in silico* models are used to assess factors such as solubility, permeability, and lipophilicity (PARIMALA, 2023). Techniques like Lipinski's Rule of Five (LIPINSKI, 2004) evaluate the molecular size, lipophilicity, and hydrogen bonding capabilities, which influence oral bioavailability. Quantitative Structure-Activity Relationship (QSAR) models and specialized software tools like Swiss ADME (DAINA; MICHIELIN; ZOETE, 2017) provide predictions of gastrointestinal absorption based on the structure of the compound.

Theoretical drug distribution analysis is about predicting how a drug will distribute in the body and crossing biological barriers. It entails evaluation of the potential for crossing the blood-brain barrier (BBB) through the use of prediction models that include molecular size and lipophilicity (TRAN; TAYARA; CHONG, 2023). In order to realize how a drug is being metabolized, computational models are crucial in forecasting interactions with the liver enzymes, which are mainly cytochrome P450 (CYP) enzymes (TRAN; TAYARA; CHONG, 2023). Such models are useful in approximating the metabolic stability of the drug and suggesting possible metabolic routes. Through forecasting which enzymes can metabolize the drug, one can foresee how efficacy and safety of the drug may be affected.

Excretion forecasts involve projecting the way a drug is eliminated from the body. These involve estimating renal excretion, where models calculate the clearance of the drug through

the kidneys based on its nature and interaction with transporters (KAZMI *et al.*, 2019). Hepatic metabolism is also considered, providing insights into the metabolism of the drug in the liver before excretion (IZAT; SAHIN, 2021). Also, it is critical to forecast potential toxicity for drug safety and to make the compounds eligible to proceed to further phases, including in vitro and in vivo studies.

2.5.2 QSAR and Similarity Analyses

Quantitative Structure-Activity Relationship (QSAR) modeling is one of the cornerstones of computational medicinal chemistry. QSAR aims at modeling the biological activity or physicochemical properties of chemical compounds from their molecular structure (GRAMATICA, 2020). The approach relies on the similarity principle, which supposes that molecules with similar structures will possess similar properties or biological activities. QSAR translates molecular structure information into numerical descriptors, input parameters in statistical or mathematical models (AMBURE *et al.*, 2015).

The long-term goal is the development of predictive models for estimating the behavior of untested molecules, thereby expediting drug discovery, material design, and toxicity prediction (FERREIRA; ANDRICOPULO, 2019). The mathematical basis of the QSAR model entails the development of functions between molecular descriptors and target activities. Early developments were dominated by simple linear methods such as multiple linear regression (MLR) (ROY; KAR; AMBURE, 2015). However, as cheminformatics matured, more potent techniques such as support vector machines (SVM), artificial neural networks (ANN) (SHAYANFAR; SHAYANFAR, 2022), random forests (RF), and more recently deep learning architectures gained popularity in the field. These approaches utilize descriptors at different levels: 0D (e.g., molecular weight, number of atoms), 1D (e.g., logP, number of functional groups), 2D (e.g., topological descriptors, fingerprints), 3D (e.g., conformation-dependent geometric descriptors), and even 4D (e.g., molecular field descriptors across several conformations) (GRAMATICA, 2020).

Feature selection and dimensionality reduction – through methods like principal component analysis (PCA) – are necessary to avoid overfitting and to maintain model generalizability. QSAR is widely applied in virtual screening of bioactive molecules, toxicology prediction (eco-toxicology and human toxicology), and pharmacokinetic property prediction (ADMET) (WOOD *et al.*, 2013). However, for a model to be reliable, it must undergo rigorous validation. Internal validation is most commonly performed through cross-

validation procedures such as k-fold or leave-one-out (LOO) (CHIRICO; GRAMATICA, 2011), whereas external validation with independent test sets is still the gold standard. The most common performance metrics are the coefficient of determination (r^2), root-mean-square error (RMSE), and external prediction coefficients such as q^2 and $RMSE_{ext}$ (CONSONNI; BALLABIO; TODESCHINI, 2009). Additionally, activity cliff detection, outlier detection, and applicability domain assessment calculations are essential to provide model interpretability and stability.

The future of QSAR consists of developing into hybrid models, greater use of artificial intelligence, exploration of big data, and greater standardization of the data, as per the paper reviewed. All these advances cement QSAR's position as a tool of choice at the intersection of theoretical chemistry, bioinformatics, and big-data driven health and environmental science (GRAMATICA, 2020).

2.5.3 Homology modeling of protein targets

Homology modeling, also known as comparative modeling, is an informatics technique to predict the three-dimensional (3D) structure of a protein based on its sequence homology with one or more proteins of known experimentally determined structure (STUDER *et al.*, 2021). The hypothesis is that similar amino acid sequences are likely to assume similar 3D conformations. This approach is especially useful when experimental methods like x-ray crystallography or cryo-electron microscopy are not feasible because of time, expense, or technical restraints (BERTONI *et al.*, 2017).

The homology modeling process usually includes a number of important steps. To begin with, the target protein sequence is paired with one or more template proteins – proteins whose structures have been determined and contain enough sequence identity (WATERHOUSE *et al.*, 2018). Correct alignment of the sequences is crucial since it directly contributes to the quality of the structural prediction. Following alignment, the target protein backbone is constructed by duplicating the respective coordinates of the template structure.

Low-similarity segments, particularly loops and insertions, are constructed using fragment libraries or *ab initio* techniques. Side chains are then appended and refined to prevent steric clash and preserve genuine rotamer conformations. Lastly, the entire model undergoes energy minimization and, in some cases, molecular dynamics refinement to correct geometrical flaws and improve structural reasonability (STUDER *et al.*, 2021). Homology modeling is an

important tool in structural biology and drug design, particularly in target-based design where the receptor structure is needed for molecular docking and virtual screening.

The accuracy of the output model largely depends on sequence identity between target and template. Typically, sequence identities above 50% give good models suitable for most downstream uses (WATERHOUSE *et al.*, 2018). Nevertheless, problems such as alignment errors, bad loop conformations, and low-quality templates may compromise the reliability of the prediction. Algorithmic method enhancements and availability of structural databases (e.g., PDB, AlphaFold Protein Structure Database) are further enhancing homology modeling's reliability and usability as a central methodology of computational structural biology (JHA *et al.*, 2022).

SECOND PART

Artigo 1 - Redigido conforme norma do periódico científico - Versão preliminar	
Título do artigo:	*In silico* screening of dual mGluR7/8 agonists as a plausible potential treatment for ASD: virtual screening, molecular docking, ADMET predictions, molecular dynamics, and MM-PBSA calculations
Autores:	Antônio Pedro Lemos de Mesquita Elaine Fontes Ferreira da Cunha Teodorico de Castro Ramalho
Periódico:	Journal of Computational Biophysics and Chemistry
ISSN	2737-4165 (print); 2737-4173 (online)
DOI	https://doi.org/10.1142

PAPER 1 – *In silico* screening of dual mGluR7/8 agonists as a plausible potential treatment for ASD: virtual screening, molecular docking, ADMET predictions, molecular dynamics, and MM-PBSA calculations

Antônio Pedro L. Mesquita[1]

Department of chemistry, Federal University of Lavras
Lavras, Minas Gerais – 37200-000, Brazil
antoniolemos.mesquita@gmail.com

Elaine F. F. da Cunha

Department of chemistry, Federal University of Lavras
Lavras, Minas Gerais – 37200-000, Brazil
elaine_cunha@ufla.br

Teodorico Castro Ramalho

Department of chemistry, University of Hradec Kralove
Hradec Kralove, Czech Republic
teo@ufla.br

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)
Published (Day Month Year)

Autism spectrum disorder is a neurodevelopmental condition that, while highly heterogeneous in individual manifestations, still shares a number of common symptoms such as delayed or no speech, impaired social interaction, cognitive impairments, and hypersensitivity to external stimuli. The theory of glutamatergic neurotransmission positions the metabotropic glutamate receptors (mGluRs) as key neural proteins associated with the development of ASD. This work presents an extended *in silico* study aimed at modeling dual agonists for mGluR7/8 receptors with possible efficacy in mitigating symptoms in ASD individuals. In the current study, a structure-based virtual screening was executed, taking **L-AP4** as the reference molecule, a well-known group III agonist of the mGluRs. The tridimensional structure of mGluR7 was modelled by homology, and molecular docking studies were carried out based on screening results, aside absorption, distribution, metabolism, excretion and toxicology (ADMET) predictions in both receptors. Two molecules with favorable structural and energetic profiles and acceptable ADMET parameters were selected for subsequent molecular dynamics simulations and Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) calculations. Eventually, compound **412** - 3-[(3-cyanophenyl)methylsulfonylamino]-2-hydroxypropanoic acid - emerged as a promising candidate with computationally validated activity at both receptors.

Keywords: Autism spectrum disorder; glutamate receptors; mGluRs; dual agonists; molecular dynamics; implicit solvation; MM-PBSA; *in silico*.

[1]Present address; corresponding author.

1. Introduction

Autism Spectrum disorder (ASD) is a neurobehavioral condition characterized by impairments in social communication, social interaction, and restricted, repetitive patterns of behavior, interests, or activities [1, 2]. The disorder is highly heterogeneous; there is great variation among individuals regarding the severity and manifestation of the symptoms. Diagnosis commonly depends on clinical evaluations, basing judgments on behavior and developmental history-usually evident during early childhood [2, 3]. Prevalence of ASD is rising globally, 1 in every 100 children has been affected [4]. Indeed, studies have identified that individuals with ASD are at higher risk of comorbid mental health conditions such as depression, anxiety, and suicide ideation [5]. Comorbid mental health conditions are notably more prevalent among the ASD population than among the general population.

In terms of ASD's pathophysiology, the dysregulation in glutamatergic neurotransmission, which affects neural functioning and is associated with social behavior, alongside other features of the disorder [6]. Metabotropic glutamate receptors or mGluRs are G-protein-coupled receptors primarily located in the presynaptic and postsynaptic neurons of the cerebral cortex, and have low concentrations in glial cells. mGluRs are classified into three different groups, namely, group I (mGluR1/5), group II (mGluR 2/3), and group III (mGluR4/6/7/8) [7], based on their similarities in structure and metabolic characteristics. Structurally, there are three determining domains in mGluRs, the: extracellular domain- or Venus Flytrap, which contains the orthosteric binding site, closing after capturing glutamate molecules; transmembrane domain, contains seven α-helices and an intracellular C-terminal domain also having α-helices conformations that interacts with other proteins existing in the neuronal cytoplasm [8, 9]. The mGluRs also have a metabolic coupling to the ionotropic glutamate receptors (iGlus), in particularly with group I, because it modulates the activity of the NMDA and AMPA receptors, but also plays important roles in the sodium and potassium ion channels that control the release of neurotransmitter, and again acts in the transport and metabolism of glutamate and GABA [10]. mGluRs are critically implicated in ASD via their role in synaptic plasticity and modulation of excitatory neurotransmission [11].

Dysregulation of signaling through these receptors has been linked to the cognitive and behavioral deficits characteristic of ASD. An overactivation might lead to an imbalance in excitatory/inhibitory transmission, thus contributing to altered synaptic plasticity, such as exaggerated long-term depression observed in ASD models. Mutation in the GRM8 gene encoding mGluR8 has been found in autistic individuals and may lead to dysregulation in glutamate neurotransmission [12, 13]. In the largest group, group III (mGluR4/6/7/8) [14], mGluR7 and mGluR8 are highly homologous, these proteins act synergistically to mediate glutamate uptake in presynaptic neurons. Furthermore, the remarkable number of homologous members present in this protein group further complicates *in vitro* and *in vivo* tests due to the difficulty of studying each receptor separately [9]. Therefore, the search for dual agonists, that is, compounds capable of simultaneously activating two metabolically homologous receptors [15], is not only an outstanding approach for autism treatment, but it can also be a valuable tool that could facilitate studies involving these two receptors separately, since they are also implicated in other conditions, including psychiatric conditions, such as depression and panic disorder [16, 17].

Therefore, dysregulation of the group III mGluRs - such as mGluR7 and mGluR8 - may cause reduced inhibition and increased synaptic activity, thereby giving rise to social, cognitive, and behavioral abnormalities relevant to ASD. Targeting these receptors with specific agonists may enhance the reuptake of glutamate [18], reduce excessive synaptic activity, and restore neurotransmitter balance, thereby reducing complications associated with glutamate dysregulation. In this regard, the present study has focused on the design of an *in silico* approach to identify dual agonists for mGluR7 and mGluR8 receptors, aiming at reducing symptoms related to autism spectrum disorder by means of a structure-based virtual screening approach, combined with theoretical studies on the absorption, distribution, metabolism, excretion and toxicity (ADMET). Molecular dynamics studies were also carried out with Molecular Mechanics – Poisson-Boltzmann Surface Area (MM-PBSA) calculations.

2. Materials and Methods

2.1. *Preparation of protein systems and homology modelling*

The crystal structure of mGluR8, identified by PDB-ID 6BT5 and cocrystallized with the **L-AP4** molecule, was downloaded from Protein Data Bank (PDB) website. The structure chosen for mGluR8 is justified by the fact that the reference ligand used for pharmacophore extraction was already complexed to the protein, also preserving the conformation of the orthosteric site. The structural determination technique used was X-Ray diffraction, with a resolution of 2.92 Å. The incomplete residues were reconstructed using the CHARMM-GUI [19] server with the CHARMM force field. The receptor was protonated using the PDB2PQR [20] server. pKa calculations were performed using PropKA [21], implemented within PDB2PQR. Hydrogens were added using the AMBER force field, based on the estimated pKa results at the physiological pH of 7.4. The mGluR7 template PDB structure, identified by PDB-ID 3MQ4 and co-crystallized with LY341495 antagonist, was downloaded from the Protein Data Bank. The choice of the crystal with the complexed antagonist was strategic so that the exact region of the orthosteric site responsible for the interaction with the ligands could be accessed. Due to the incompleteness of some amino acid residues, a homology modeling was performed using the SwissModel server [22, 23] . The complete FASTA sequence of the human mGluR7 was obtained from Uniprot (Entry Q14831) server and submitted to SwissModel for template search and model construction. The template selection was performed using the same crystal (3MQ4) so that incomplete amino acid residues could be re-modeled, preserving the binding site of the original crystal, which was already well characterized. The parameters of Global Model Quality Estimate (GMQE), Qualitative Model Energy Analysis – Distance Constraints (QMEANDisco Global), amino acid sequence similarity and Ramachandran plot angles were considered for template validation.

2.2. *Virtual screening*

The **L-AP4** ligand was used as a template to search for multiple molecules with equivalent pharmacophores using the Pharmit [24] online server. The pharmacophores were identified by inspecting the intermolecular interactions with mGluR8 residues at the binding site. The crystal conformation was preserved when submitting the ligand to Pharmit, and the pharmacophores evaluated were the same as those shown in the interaction diagram of the original receptor. The carbonyl oxygens were classified as hydrogen bond acceptors due to the resonance, while the nitrogen atoms were classified as both acceptors and donors. The oxygens attached to phosphorus were classified as hydrogen bond acceptors and negative ions. The search settings for hit screening were set to obey Lipinski's rule [25]: a molecular weight range of 0 to 500 g mol^{-1}, with rotational bonds ranging from 0 to 10, log *P* from 0 to 5, polar surface area (PSA) from 0 to 140, hydrogen bond acceptors (HBA) from 0 to 10, and hydrogen bond donors (HBD) from 0 to 5. The *hit reduction* parameter was set to a maximum of 1 conformational hit per molecule.

2.3. *Preliminary toxicology tests*

SwissADME [26] server was utilized for a preliminary toxicology test to reduce the hit compounds identified in the Pharmit stage. After filtering using OpenBabel (v. 3.1.0) [27] to exclude possible duplicate structures, the SMILES of the molecules were submitted to the SwissADME server. The results were analyzed according to Lipinski [28], PAINS [29], and Brenk [30] rules.

2.4. *Molecular docking and ROC curve validation*

Five reference ligands with experimental values of half-maximal effective concentrations (EC_{50}) were used in this work. The ligands are **L-AP4** [31], **L-CCG-I** [31], **L-SOP4** [31], **(RS)-PPG** [31], and **DCG-IV** [32]. All of the three-dimensional structures were designed using BIOVIA's Discovery Studio software (version 3.1.1) in MOL2 formats. The charges were assigned using Open Babel. The decoys

were generated using the SMILES structure of each reference ligand on the DUD-E server [33]. A total of 50 decoys were generated for each ligand, resulting in a total of 250 decoys.

All 250 decoys and the five reference ligands were subjected to molecular docking simulations using Molegro Virtual Docker (MVD) [34] software (version 2011.4.3), using the guided differential evolution MOLDOCK algorithm. The SMILES structures of the decoys were converted to MOL2 format, and the three-dimensional structures were generated using OpenBabel software. Additionally, all molecules were protonated at pH 7.4, and partial charges were assigned using OpenBabel. For the mGluR8 docking simulations, the charges were calculated using the Electronegativity Equalization Method (EEM) based on B3LYP/6-31G*/MPA calculations. In the case of mGluR7 dockings, the charge calculation method employed was EEM-BA, based on B3LYP/6-311G/AIM calculations. This difference in the charge calculation method is justified by the fact that the MVD algorithm is sensitive to charges, therefore, it was later found through the ROC curve that the EEM-BA charges better described the interactions of the ligands with mGluR7. Two sets of molecular dockings were performed in MVD. For mGluR8, the docking setup involved isolating the active site by creating an 8 units sphere around the co-crystallized **L-AP4** ligand. A set of 47 amino acid residues within the active site were configured to be flexible. Docking runs consisted of 100 steps for each pose, and only the minimum energy pose was accepted for each run. For mGluR7 docking with the same decoys and active ligands, the only difference in the protocol was the charge calculation method. The mGluR7 compounds were assigned EEM-BA charges, with a restraint sphere of 12 units and 56 amino acid residues set as flexible. Both protein charges were calculated using the Gasteiger method. To plot the ROC curve and calculate the AUC we used OriginPro software (v. 2024).

2.5. *Quantitative toxicology tests*

The structures from docking were subjected to a quantitative toxicology assessment using the DeepPK server for ADMET [35] analysis, as well as SwissADME. DeepPK provided predictions for lethal dose (LD_{50}), Blood Brain Barrier (BBB) permeation, half-life, carcinogenesis, liver injury and inhibition of Cythochrome P450 3A4 predictions. In parallel we also conducted a chirality analysis to select the most suitable stereochemical structures for drug design. A final analysis was performed using SwissADME, evaluating some ideal physical chemistry proprieties, such as molecular mass ($200 < MM < 600$), rottable bonds (RotBonds < 15), H-bond acceptors (< 10) and donnors (< 5), solubility ($\log S > 0$), topological surface area (TPSA < 140 $Å^2$), $\log P$ (< 5.88), gastrointestinal absortionn (high as possible) and skin permeability coefficient ($\log Kp$ low as possible). The results are presented on supplementary material in **Table S1**.

2.6. *Molecular dynamics simulations*

All molecular dynamics simulations were performed using Gromacs [36] (version 2021.4). The atomic coordinates of the ligands were extracted from the lowest-energy poses obtained from docking, and their topologies were generated using ACPYPE [37] from the AmberTools23 [38] package, as well as MKTOP [39]. Ligand preparation was conducted using the OPLS/AA force field parameterizations. The proteins were submitted to pKa prediction and protonation calculations at physiological pH (7.4) using the PDB2PQR server. The receptor topologies were prepared directly in Gromacs, also employing the OPLS/AA force field. The simulations were set up in a cubic box with a distance of 1.5 nm from the protein to the box edges. A total of 48,534 water molecules were added to the systems using the TIP4P water model. Positive ions were added to the system to neutralize the charges. Steepest descent energy minimizations, both with and without position restraints, were performed until the system's maximum force reached a value below 10 kJ/mol. Temperature and pressure equilibrations were carried out using the NVT and NPT ensembles, respectively. The time interval *dt* applied in this step was 2 fs, with constraints on the existing bonds with hydrogen atoms. The NVT equilibration was conducted at 310 K for 100 ps using the leap-frog integrator and Verlet scheme. The NPT equilibration also lasted 100 ps, with a reference pressure of 1 bar, using the same integrator as in the NVT stage. An extra water dynamics step was applied for 500 ps for solvent relaxation. The molecular dynamics simulations were configured to run for 200 ns using Gromacs on a high-performance GPU, with the leap-frog integrator and Verlet scheme. Graphical analyses of the MD simulations were conducted across the entire

simulation trajectory, totaling 10,000 frames, using the XMGRACE (v. 5.1.22) program. Visual analyses and intermolecular interaction studies were performed using Visual Molecular Dynamics (VMD) (v. 1.9.4) and Discovery Studio by BIOVIA (version 3.1.1).

2.7. *MM-PBSA calculations*

The MM-PBSA calculations were performed using the gmx_mmpbsa software (v. 1.6.4) [40], which is based on the mmpbsa.py script from the AMBER simulator and is compatible with all versions of Gromacs. These calculations were conducted on the last 50 ns of the simulation, using 250 frames out of a total of 2500, i.e., with an interval of every 10 frames. The internal dielectric constant values were optimized through multiple calculations until a suitable value for the systems was achieved. A value of 6 was used for mGluR7 and 3 for mGluR8, based on the reference ligand. The external dielectric constant was set to 80, which is consistent with the water molecules in the system. The temperature was configured at 310 K, and the entropy calculation method employed was Interaction Entropy [41].

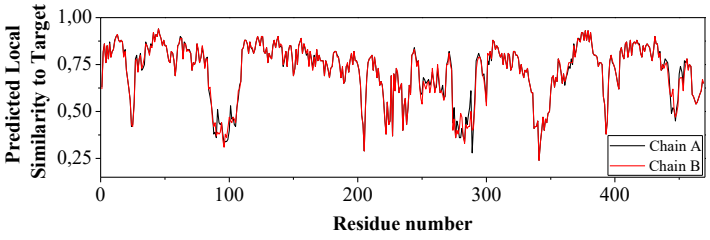
3. Results

3.1. *Virtual screening*

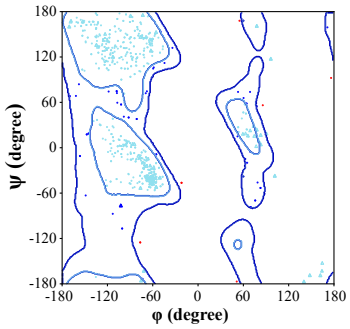
For the ligand-structure based virtual screening, using the cocrystallized **L-AP4** molecule as template, all of the molecular data bases available on Pharmit server were used for a wide search with over 500 million molecules, including CHEMBL32 (39 matches), ChemDiv (4 matches), ChemSpace (166 matches), Enamine (32 matches), MCULE (49 matches), MCULE-ULTIMATE (5 matches), MolPort (23 matches), NCI open Chemical Repository (0 matches), PubChem (650 matches), LabNetwork (7 matches) and ZINC (57 matches). The total amount of compounds obtained in this stage was 1032, this large number of molecules could be screened through a preliminary qualitative toxicology test to exclude non-drug-like structures, thereby enhancing the accuracy of the analysis in the search for new medicines. After the toxicology analysis in SwissADME through Lipinski's, Brenk's and PAINS rules, the number of compounds reduced to 447. All of the remaining molecules were submitted to the docking screening assay.

3.2. *Homology modeling of the mGluR7 structure*

The incompleteness of amino acid residues in mGluR7 tridimensional structure were enough to make a homology modeling convenient, so Swiss Model server was used for this task. From the FASTA sequence of human mGluR7 available on Uniprot, we modeled a new structure with 100% coverage of sequence identity and a 0.77 GMQE, using 3MQ4 structure as template. Most of the residues on the model were localized with almost 80% of similarity compared with the template. The binding site, the most important region for the docking analysis has 100% of sequence similarity and 80% of alignment. **Fig. 1 (a)** shows the graph corresponding to the local QMEANDisCo of the model, in terms of each amino acid residue. The Global QMEANDisCo of the model was evaluated at 0.70, which represents a good validation of the structure. **Fig. 1 (b)** shows the Ramachandran plot, in which Cyan, blue and red (dots/triangles) represent torsion angles of favored, allowed and disallowed regions respectively; dot represents residues other than glycine and triangles represents glycine.



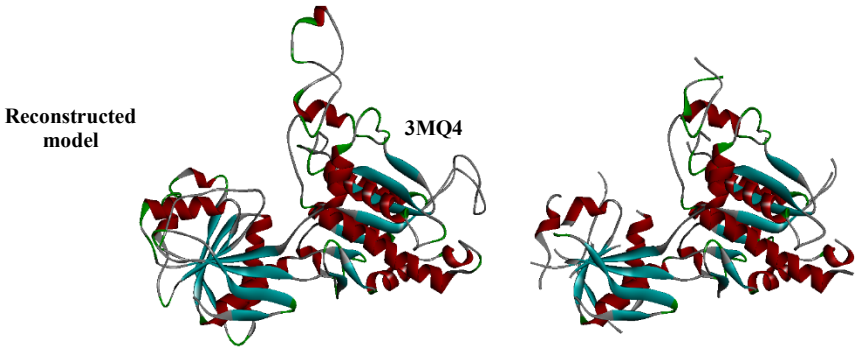
(a)



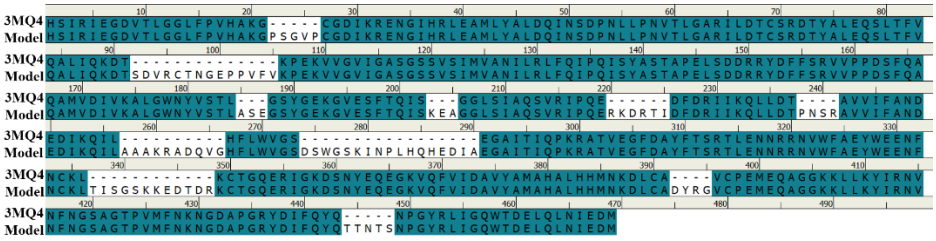
(b)

Fig. 1. Homology modeling validation. (a) Predicted local similarity to target plot; (b) Ramachandran plot [42].

By analyzing the Ramachandran plot, it can be seen that most of the amino acid residues had accpetble angle values equivalent to the conformation of β-sheets and α-helices, with only 6 residues with forbidden angles, none of which were present in the binding site. **Fig. 2 (a)** illustrates the three-dimensional structure of the generated model and the original crystal, where the incomplete regions that were reconstructed are identified. More explicitly, **Fig. 2 (b)** presents the sequences of the model and the template, with the reconstructed residues illustrated in white and the identical sequences present in both structures in blue.



(a)



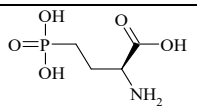
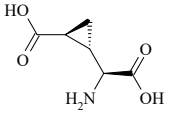
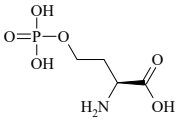
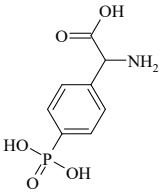
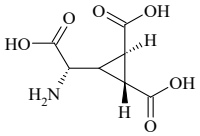
(b)

Fig. 2. Model and template from SwissModel. (a) 3D structures; (b) Sequence alignment and identity.

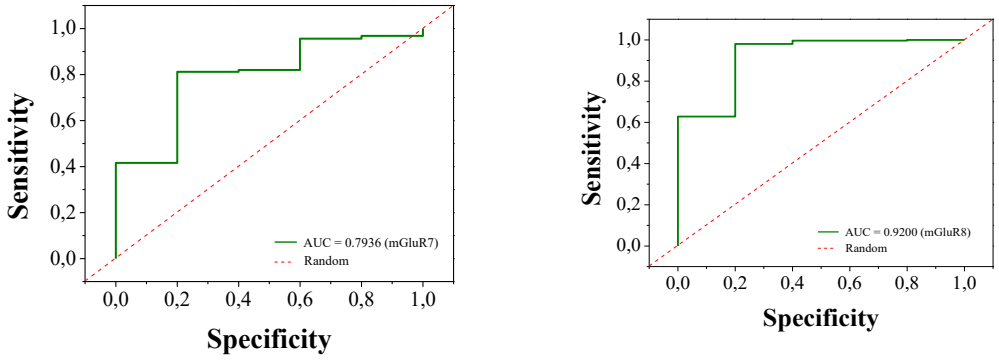
3.3. *Molecular docking screening*

The molecular docking parameters were stablished from the validation step through the plot of a ROC curve. Decoys from DUD-E server were generated using five active ligands (**Table 1**) with experimental agonist data (EC_{50}) evaluated.

Table 1. Active ligands agonism and structure data

Reference compound	EC_{50} (μM)		Structure
	mGluR7	mGluR8	
L-AP4	175	0.061	
L-CCG-I	230	3.4	
L-SOP4	31	0.27	
(RS)-PPG	185	0.21	
DCG-IV	40.1	32	

In this study, we conducted an evaluation of Ligand Efficiency (LE) parameters. These parameters are determined by calculating the ratio of MolDock Scores to the number of heavy atoms (non-hydrogen atoms) present in each molecule. This metric is particularly valuable as it provides an assessment of the binding affinity relative to the molecular size, facilitating comparisons across different ligands. **Fig. 3** illustrates the Receiver Operating Characteristic (ROC) curves that were generated for the validation docking analyses performed on the mGluR7 and mGluR8 receptors. These ROC curves offer a graphical representation of the true positive rate versus the false positive rate, thereby used as a reliable indicator of the predictive accuracy of the docking models for these specific receptor subtypes.



(a) (b)

Fig. 3. ROC curve for mGluRs. (a) ROC curve for the mGluR7 dockings; (b) ROC curve for the mGluR8 dockings.

Recent literature shows the efficiency of LE analysis in virtual screening protocols [43–45], so the docking scoring results on **Table 2** were based on this metric. The redocking study was also made in both receptors. The RMSD for the superimposition was calculated on Discovery Studio, resulting in 1.3443 and 0.9186 Å for mGluR7 and mGluR8, respectively. Both results show RMDS lower than 2.0000 Å, as shown in **Fig. S7** in supplementary material, so the docking could be validated by redocking [46].

Table 2. Docking results for the 27 dual ligands selected

Compound	Ligand Efficiency (kcal/mol)		CID PubChem
	mGluR8	mGluR7	
338	-10.0844	-6.21707	72750714
343	-9.6275	-5.37477	123764026
417	-9.3843	-5.45212	102984030
413	-9.20147	-6.45295	106108446
341	-9.08939	-5.11578	131543322
222	-8.95302	-5.1705	82659296
268	-8.9452	-5.54476	10561964
278	-8.93801	-5.54267	66167195
203	-8.88396	-5.5792	3999504
318	-8.88263	-5.80183	82402970
379	-8.78771	-5.74993	87289610
366	-8.73425	-5.42208	12548074
427	-8.68798	-6.06708	163958624
374	-8.64649	-6.2467	104936257
1	-8.62385	-6.18105	58994918*
362	-8.58996	-5.78559	82402370
412	-8.46657	-5.56029	66165652
292	-8.45879	-5.28944	CSCS00010449501**
421	-8.41993	-5.42437	106108873
400	-8.35865	-5.51728	10978323
398	-8.34452	-5.90418	66167173
416	-8.18336	-5.21226	83903472
252	-8.03216	-6.10576	82396057
259	-8.02712	-5.24605	70945612
326	-7.997	-5.79771	66167108
244	-7.98153	-5.48274	87929530
428	-7.97649	-5.31365	106108579

*ChemSpider ID; **ChemSpace ID.

Since the docking parameters were set and validated, 447 structures from the Pharmit and SwissADME analysis were docked into the mGluR8 and mGluR7 structures. Only the energies lower than the one obtained by the worst active ligand were considered (-7.9754 kcal/mol for mGluR8 and -5.10948 kcal/mol for mGluR7). The 27 structures of the dual compounds, whose energies are presented in **Table 2**, can be found in **Fig. S8** of the supplementary material.

The intermolecular energy terms of the ligands interacting on receptor's binding site is also a very important factor to predict the efficiency of binding. Generally, the energy associated with mGluR8 is lower than that of mGluR7. This difference may be attributed to a greater number of intermolecular interactions between the ligand and receptor 8 compared to receptor 7. The observed hydrogen bonds (HB) and salt bridges range from approximately -0.2 to -1.8 kcal/mol. The highest interaction energies occur when a single amino acid residue forms multiple hydrogen bonds with the ligand. The threonine residue in the active site of both receptors demonstrates a strong interaction with all ligands it contacts, as seen in ligands **222**, **341**, **400** and **412** for mGluR7, and ligands **222**, **400** and **412** for mGluR8. For mGluR7, Gly126 obtained an interaction energy of -0.74 kcal/mol for the ligand **412**, the second lowest value after Thr150. The Ser127 residues also demonstrated low interaction energies with several ligands, namely **L-AP4**, **222**, **362**, **400** and **412**. For mGluR8, two other key residues of relevant interactions are Ser156, with a value of -1.19 kcal/mol for the reference ligand, and the interactions are also repeated for

ligands **222**, **341**, **362**, **400** and **412**. Ala177 was also an interacting residue that appeared recurrently in all compounds, with energies ranging from -0.60 to -0.93 kcal/mol. Furthermore, the interaction distances are within the average hydrogen bond distances described in the literature, ranging from 2.5 to 3.5 angstroms [47]. The table containing information of interacting energy and HB distances can be found in the supplementary material (**Table S4**).

3.4. *Quantitative toxicology tests*

The 27 docking structures underwent toxicology prediction tests using DeepPK and SwissADME. Ten ligands (**Table 3**) passed, all with LD_{50} > 2.08 g/kg and a single chiral center. **Table 4** highlights key ADMET parameters: five ligands (**416**, **417**, **400**, **412**, **222**) cross the blood brain barrier (BBB), an important prediction, considering that mGluRs are proteins mostly present in neurons of the cerebral cortex, with **417** at medium confidence. Most have a half-life < 3h, except **412**. All are non-carcinogenic, though **222** and **400** have medium confidence. Ligands **416**, **417**, **400**, and **362** show potential liver toxicity. None inhibit cytochrome P450 3A4. Only **222** and **412** meet every safety criteria, making them ideal for molecular dynamics and MM-PBSA calculations.

Table 3. Toxicity and chirality results for the docking ligands

Compound	Docking score		Chiral Centers	LD_{50} (g/kg)
	mGluR8	mGluR7		
L-AP4	-11.7364	-6.13892	1	1.44
L-CCG-I	-11.9699	-5.85761	3	1.63
L-SOP4	-10.6485	-6.44166	1	1.63
PPG	-7.97540	-5.77349	0	1.66
DCG-IV	-10.6773	-5.10948	3	2.08
341	-9.08939	-5.11578	1	2.86
318	-8.88263	-5.80183	1	2.86
417	-9.3843	-5.45212	1	2.59
416	-8.18336	-5.21226	1	2.55
400	-8.35865	-5.51728	1	2.45
252	-8.03216	-6.10576	1	2.45
412	-8.46657	-5.56029	1	2.31
222	-8.95302	-5.1705	1	2.29
428	-7.97649	-5.31365	1	2.27
362	-8.58996	-5.78559	1	2.24

Table 4. DeePK results

Ligand	BBB	BBB (qualitative)	Half-life	Carcinogenesis	Liver injury	CYP450 3A4
341	-3.74	Non-penetrable (HC)	< 3h	Safe (HC)	Safe (MC)	Non-inhibitor (HC)
428	-3.79	Non-penetrable (LC)	< 3h	Safe (HC)	Safe (MC)	Non-inhibitor (HC)
416	-2.89	Penetrable (HC)	< 3h	Safe (HC)	Toxic MC	Non-inhibitor (HC)
318	-3.19	Non-penetrable (HC)	< 3h	Safe (HC)	Safe (LC)	Non-inhibitor (HC)
417	-3.32	Penetrable (MC)	< 3h	Safe (HC)	Toxic (LC)	Non-inhibitor (HC)
400	-3.6	Penetrable (HC)	< 3h	Safe (MC)	Toxic (LC)	Non-inhibitor (HC)
252	-4.64	Non-penetrable (HC)	< 3h	Safe (HC)	Safe (HC)	Non-inhibitor (HC)
412	-3.65	Penetrable (HC)	≥ 3h	Safe (HC)	Safe (MC)	Non-inhibitor (HC)
222	-3.66	Penetrable (HC)	< 3h	Safe (MC)	Safe (LC)	Non-inhibitor (HC)
362	-3.54	Non-penetrable (HC)	< 3h	Safe (HC)	Toxic (LC)	Non-inhibitor (HC)

HC = high confidence; MD = medium confidence; LC = low confidence.

3.5. *Intermolecular interaction analysis*

In this topic, we analyzed hydrogen bond interactions of ligands **222** and **412** with mGluR7 and mGluR8. In mGluR7, ligand **222**, **Fig. 4 (a)**, formed hydrogen bonds with Ser127 and Thr150, interacting with a primary amine and a pyrazine nucleus. Ligand **412**, **Fig. 4 (b)** showed the highest number of hydrogen bonds, interacting with multiple residues, including Arg76, Asn42, and Gly126, and forming five bonds via its sulfonamide group. For mGluR8, ligand **222**, **Fig. 4 (c)**, formed three hydrogen bonds with Ser156, Thr179, and Ala177, alongside electrostatic and pi-ion interactions. Ligand **412**, **Fig. 4 (d)** formed the most hydrogen bonds (nine), primarily involving carboxyl and sulfonamide groups. The sulfonamide group in ligand **412** demonstrated strong stabilization potential due to its extensive hydrogen bonding. Thus, it is clear that the screening ligands obtained highly interacting pharmacophores. The next step involved molecular dynamics simulations to study ligand-receptor interactions over time.

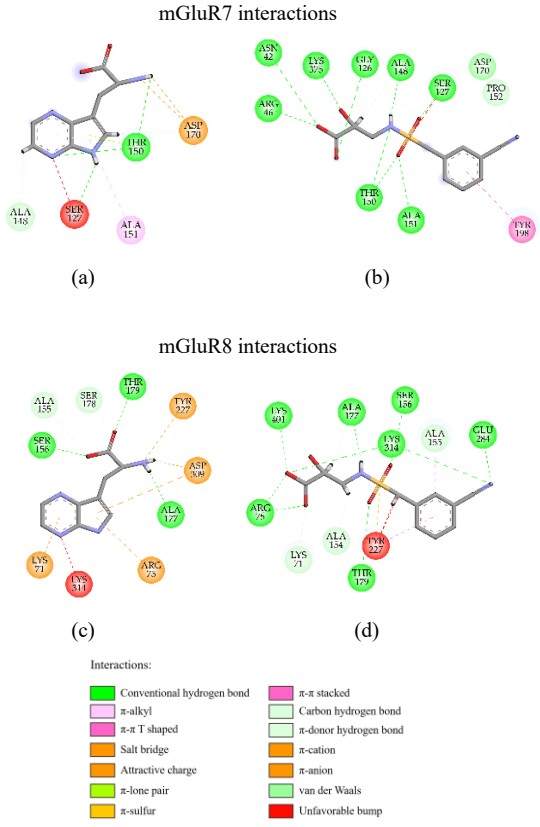


Fig. 4. 2D molecular interaction diagram for 222 and 412 in both receptors. (a) 222 interacting with mGluR7; (b) 412 interacting with mGluR7; (c) 222 interacting with mGluR8; (d) 412 interacting with mGluR8.

3.6. *Molecular dynamics simulations*

Ligands **222** and **412**, which passed through the stages of virtual screening and ADMET, were selected for molecular dynamics (MD) simulations in both receptors since they possess pharmacophoric groups equivalent to those of **L-AP4**, exhibit favorable interactions with mGluR7/8 in docking analyses, and demonstrate a promising toxicological, metabolic, and pharmacokinetic profile. In this study, the MD simulations contributed to understanding the interaction dynamics between mGluR7/8 and the two ligands approved in the virtual screening stage, in addition to **L-AP4**, the reference ligand.

Figs. 5 (a) and **5 (b)** show the RMSD versus time plots for the simulated ligands. For mGluR7, in **Fig. 5 (a)**, the RMSD ranged between 0.05 and 0.15 nm, which corresponds to 0.5 to 1.5 Å. All ligands for mGluR7 had average RMSD values lower than that of the reference ligand (**L-AP4**) over the entire MD simulation, with the lowest being ligand **222**. Although ligand **412** exhibited fluctuations, they were

all within the range of 0.025 nm (~ 0.1 – 0.125 nm), indicating stability in the active site of mGluR7. Additionally, ligand **412** demonstrated high stability between approximately 100 and 175 ns, with an average RMSD of around 0.125 nm. **Fig. 5 (b)** shows the RMSD versus time plot for the ligands complexed with mGluR8. **L-AP4** and ligand **222** showed extremely low fluctuations throughout the simulation period, with average RMSD values of 0.045 and 0.0375 nm, respectively. For ligand **412**, the RMSD fluctuated slightly, between 0.0525 and 0.06 nm.

Figs. 5 (c) and **5 (d)** present the RMSD plots for the proteins over time. The structure of mGluR7 exhibited maximum fluctuations of approximately 0.5 nm. Initial fluctuations are common as the ligands induce conformational changes in the protein until an equilibrium conformation is established. After 100 ns of simulation, the RMSD of mGluR7 tended to stabilize for all ligands, with the 150 to 200 ns interval showing maximum system stabilization, as seen in **Fig. 5 (c)**. **Fig. 5 (d)** shows the RMSD variation of mGluR8 complexed with each ligand. Complexes with **222**, **412**, and the reference ligand stabilized at approximately 0.3 – 0.4 nm, with ligand **412** having a slightly higher RMSD compared to the other three. A comparison of the RMSDs for mGluR7 and mGluR8 shows that all complexes tended to achieve conformational stability after 100 ns, with mGluR8 exhibiting lower average RMSD fluctuations compared to mGluR7.

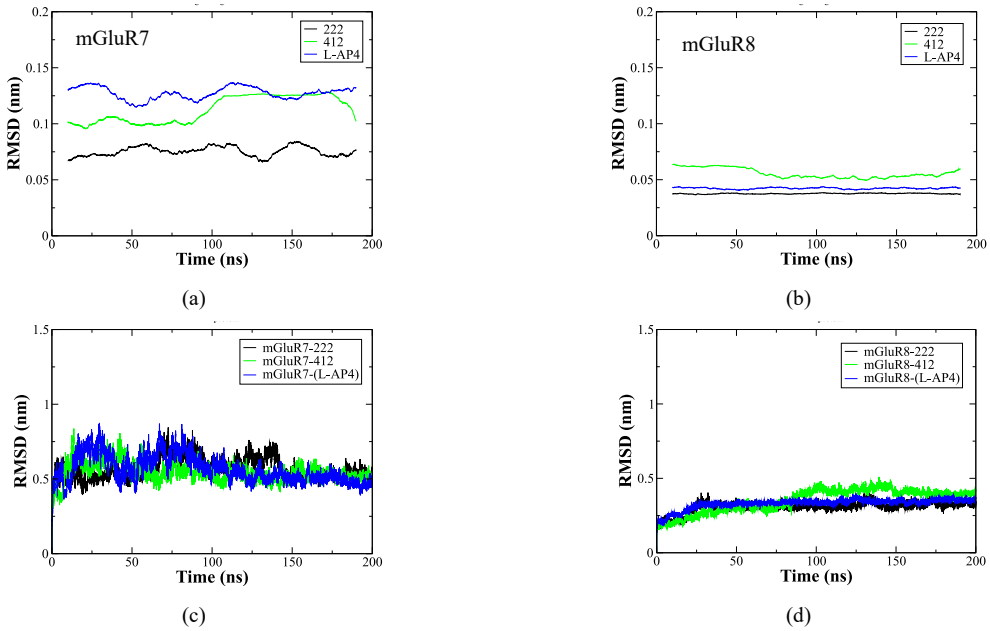


Fig. 5. RMSD of ligands complexed and RMSD of the proteins. (a) RMSD for the ligands complexed with mGluR7; (b) RMSD for the ligands complexed with mGluR8; (c) Protein-protein RMSD for mGluR7; (d) Protein-protein RMSD for mGluR8;

Figs. 6 (a)-(c) show the number of hydrogen bonds over time for mGluR7. **L-AP4** exhibited fluctuations in the number of hydrogen bonds from 0 to 10 during the first 50 ns, maintaining an average of approximately 6 hydrogen bonds after 50 ns. Ligand **222** had the lowest average number of hydrogen bonds for mGluR7, with intervals where no bonds were formed, as observed between 100-110 ns and 140-185 ns. Ligand **412**, around 150 ns, showed minimal fluctuations with about 5 hydrogen bonds, maintaining this relatively constant value until the end of the simulation. On average, ligands **412** and **L-AP4** exhibited similar numbers of hydrogen bonds from the midpoint of the simulation onwards.

Figs. 6 (d)-(f) show the hydrogen bonds formed with the ligands in mGluR8. **L-AP4** maintained a constant number of approximately 4 hydrogen bonds for 125 ns of simulation, which then decreased to an average of 2 bonds until the end of the simulation. Ligands **222** and **412** had slightly higher average counts than **L-AP4**, and these ligands were also associated with lower RMSF values (details on supplementary material, **Fig. S3**), indicating that both were able to more effectively stabilize the residues in the active site – especially ligand **412**, which achieved the highest average number of hydrogen bonds.

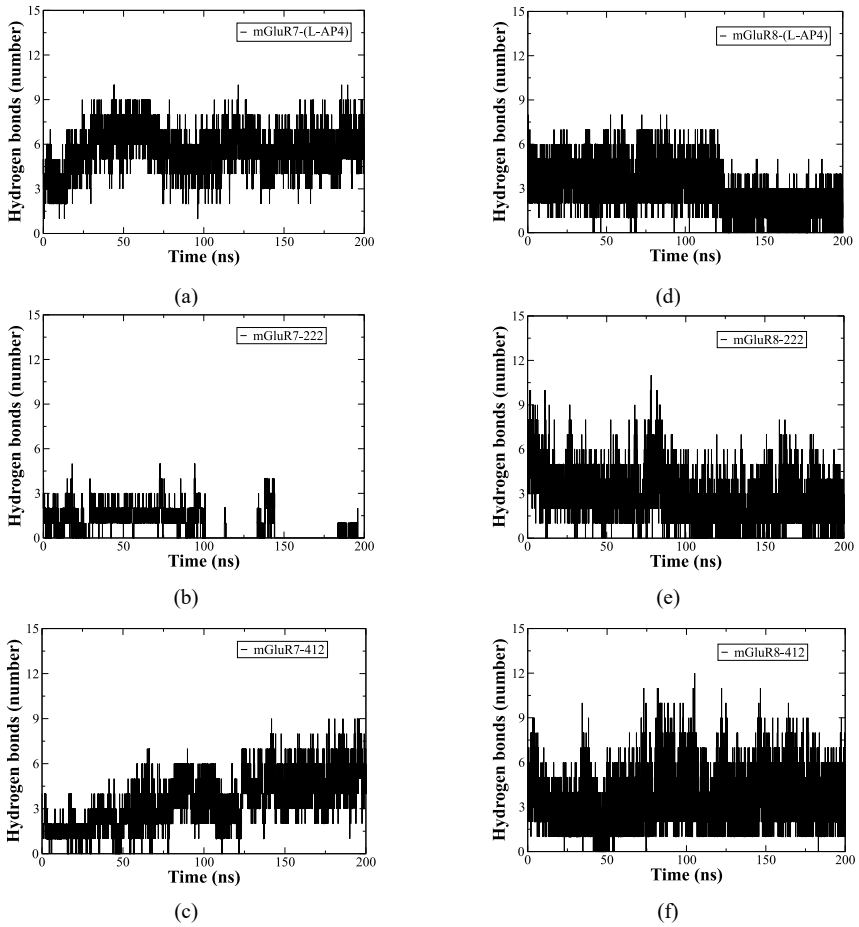


Fig. 6 Hydrogen bonds between the simulated ligands and the active site of mGluR7/8. (a) Hydrogen bonds for the complex (L-AP4)-mGluR7; (b) Hydrogen bonds for the complex 222-mGluR7; (c) Hydrogen bonds for the complex 412-mGluR7; (d) Hydrogen bonds for the complex (L-AP4)-mGluR8; (e) Hydrogen bonds for the complex 222-mGluR8; (f) Hydrogen bonds for the complex 412-mGluR8;

3.7. *MM-PBSA calculations*

The software gmx_mmpbsa [40] performs calculations based on a thermodynamic cycle divided into two stages, which involve the complexation of a ligand with a receptor in the gas phase (without solvent) and in solution. The Gibbs free energy of binding, ΔG_{bind}, is calculated using the average Gibbs energy values of the complex ($G_{complex}$), the receptor ($G_{receptor}$), and the ligand (G_{ligand}). **Table 5** presents the data obtained from the MM-PBSA calculations. The van der Waals and electrostatic energies (ΔE_{VDW} and ΔE_{EL}) were calculated from the same force field used in MD simulations. The Gibbs free energy values in the gas phase and in solvent, along with the binding free energy ($\Delta G_{GAS\text{-}SOL}$ and $\Delta G_{binding}$), were calculated using the Poisson-Boltzmann equation (PB), with the entropic term determined using the interaction entropy method. The experimental value was estimated from the EC_{50} of **L-AP4** presented in **Table 1**, where the Gibbs free energy was approximated as $\Delta G_{exp} \approx RT \ln EC_{50}$ [48]. The errors in the experimental measurement remain unknown.

Table 5. Thermodynamic results from MM-PBSA calculations.

Receptor	Compound	MMPBSA Energies (kcal/mol)				
		ΔE_{VDW}	ΔE_{EL}	$\Delta G_{GAS\text{-}SOL}$	$\Delta G_{binding}$	$\Delta G_{experimental}$
mGluR7	L-AP4	-2.88 ± 0.52	-11.89 ± 0.03	-10.21 ± 0.52	-8.10 ± 3.79	-5.33
	222	-2.09 ± 0.61	-0.25 ± 0.13	-2.11 ± 0.92	-1.57 ± 0.92	-
	412	-26.22 ± 3.38	0.15 ± 0.02	-28.10 ± 3.77	-17.98 ± 3.77	-
mGluR8	L-AP4	-2.40 ± 0.68	-111.59 ± 6.16	-24.53 ± 4.27	-13.33 ± 4.27	-10.23
	222	-18.51 ± 2.98	-35.70 ± 2.83	-22.59 ± 2.76	-16.64 ± 2.76	-
	412	-31.68 ± 4.67	-18.70 ± 4.12	-32.08 ± 3.39	-27.36 ± 3.39	-

In terms of van der Waals and electrostatic contributions, for mGluR7, the reference ligand was the only one with a higher electrostatic contribution than van der Waals. For the other compounds (**222** and **412**), the van der Waals energies were lower, with compound **412** achieving the lowest values. The free energy results for the complexes in gas and solution phases reveal the contribution of solvation and the effect of solvent water molecules. All results were negative for mGluR7, with the lowest value associated with compound **412**, around -28 kcal/mol, while the highest was for compound **222**. The second lowest binding free energy was related to the reference ligand, and for ligand **222**, the result was slightly negative at -1.57 kcal/mol. The best energy binding in mGluR7 is related to **412** (-17.98 ± 3.77 kcal/mol).

For mGluR8, the electrostatic energy contribution was more pronounced for compounds **L-AP4** and **222**, with a significant difference for the reference ligand, where its electrostatic energy was approximately 46 times lower compared to its van der Waals energy. The van der Waals contribution was greater than the electrostatic contribution for compound **412**. The gas-solution-phase free energy was negative for all compounds, with **L-AP4** and **222** ranging from -22.59 to -24.53 kcal/mol. Compound **412** stood out with an energy value of -32.08 kcal/mol, about 40% lower than the other ligands. Thus, the total binding free energy for compound **412** was the lowest, totaling -27.36 ± 3.39 kcal/mol.

In addition to the molecular dynamics analyses previously presented, and considering the best energetic performance of ligand **412**, a study focused on the interaction dynamics with key residues identified in molecular docking was conducted, using minimum distance measurements between the ligand and residues (**Fig. S4** in the supplementary material). For mGluR7, residues Gly126 and Ser127 showed stable interactions with **412** and low RMSF values, suggesting less flexibility and stronger interactions. In particular, Gly126 remained on an average minimum distance of 0.25 – 0.3 nm, compatible with stable hydrogen bonding throughout almost the entire simulation **(Fig. S4 (b) – (c)**, while Ser127 presented larger fluctuations, suggesting more distant interactions as electrostatic. In mGluR8, residues Ser156, Ala177, and Thr179, also with low interaction energies and reduced RMSF, remained close to the ligand (approximately 0.25 nm) during the entire 200 ns of simulation, indicating consistent hydrogen bonding – **Fig. S4 (a), (d)–(e)**.

4. Discussion

The results of the virtual screening were effective by searching for pharmacophores equivalent to the reference ligand (**L-AP4**) among millions of molecules in the Pharmit database, the number was reduced to 1,032 compounds. These were further filtered using Lipinski's, Brenk's, and PAINS rules, resulting in 447 compounds. The Lipinski rule evaluates physicochemical parameters that facilitate the prediction of oral absorption of compounds. Brenk filtering, in turn, identifies potentially undesirable substructures, such as toxic groups or those linked to low bioavailability. The PAINS filter acts to identify promiscuous molecules, i.e., compounds that tend to interact nonspecifically with multiple targets and generate false positives in biological assays [49]. The combined application of these filters is essential to increase selectivity and reduce the risk of failure in the later stages of drug development. Using the three-dimensional structures of mGluR7 (homology-modeled) and mGluR8, docking studies statistically validated by ROC curves were performed [50], selecting 27 potential dual ligands with compatible or superior energy parameters compared to reference compounds (**Table 1**). ADMET studies complemented the conformational and energetic analysis of molecular docking, selecting ligands that were not only compatible but also suitable for the neural environment. Ligands **222** and **412**, in particular, demonstrated high blood-brain barrier penetration accuracy. Moreover, these two ligands also displayed safe metabolism, excretion, and toxicity profiles, with none predicted to inhibit cytochrome P450 3A4, an estimated half-life of approximately 3 hours, and no carcinogenic or hepatotoxic risks.

Molecular dynamics (MD) simulations provided insights into the interaction mechanisms of the approved ligands over time, using the reference ligand for comparison. RMSD analyses confirmed the stabilization of ligands within the active sites, as well as receptor stabilization in aqueous environments. The RMSD values for mGluR7 structures exhibited greater fluctuations than mGluR8, which can be explained by the fact that mGluR8 was co-crystallized with the reference ligand, meaning its structure was already conformationally suited to the ligands, unlike mGluR7. The average fluctuation of RMSD values (between 0.25 to 0.3 nm) presented by mGluR8 is consistent with other molecular dynamics

results, as found in the work of Zhai and collaborators (2022) [51]. In terms of hydrogen bonds, for mGluR7 only ligand **222** exhibited significant periods without these bonds, while the others maintained adequate quantities compared to the reference. The hydrogen bond results from MD for ligand **222** were consistent with the docking results, as it showed the highest intermolecular energy among the ligands. This could be due to its greater effect on the protein's RMSF, increasing residue flexibility and altering the position of interacting amino acid groups, thus reducing or preventing hydrogen bond formation. For the other ligands, the average number of hydrogen bonds throughout the trajectory was consistent with the docking results.

Furthermore, the molecular dynamics results obtained for mGluR8 showed greater stability compared to mGluR7, which can be justified by multiple structural and biological factors. First, the mGluR7 structure used was obtained by homology modeling, which, although it is a robust approach for structural prediction, does not fully replace the accuracy of a crystallographic structure or one obtained by high-resolution cryomicroscopy [22, 52]. Even models with high sequence identity can present relevant local differences that impact the dynamics and stability of the system during prolonged simulations. Furthermore, metabolically, mGluR7 is known to have considerably lower sensitivity to glutamate and other agonists compared to other group III receptors [9], which was reflected in the molecular dynamics results, evidencing lower conformational stability and more fluctuating dynamics — consistent with its low-affinity physiology. Thus, the computational findings reinforce both the limitations inherent to the structural model and the known biological characteristics of the receptor, providing greater robustness to the interpretations made.

For the mGluR7 receptor, the reference ligand **L-AP4** presented a binding energy of –8.10 ± 3.79 kcal/mol, close to the estimated experimental value (–5.33 kcal/mol), with a predominance of the electrostatic contribution over the van der Waals interactions. Ligand **222** exhibited weaker affinity ($\Delta G_{binding}$ = –1.57 ± 0.92 kcal/mol), reflecting less expressive intermolecular interactions. Ligand **412** stood out for presenting the best binding energy for mGluR7 (–17.98 ± 3.77 kcal/mol), with a dominant contribution from the van der Waals interactions (ΔE_{vdW} = –26.22 ± 3.38 kcal/mol) and a slightly unfavorable electrostatic effect (ΔE_{el} = +0.15 ± 0.02 kcal/mol). In the case of mGluR8, ligand **L-AP4** again showed considerable affinity ($\Delta G_{binding}$ = –13.33 ± 4.27 kcal/mol), supported almost exclusively by the strong electrostatic interaction (ΔE_{el} = –111.59 ± 6.16 kcal/mol), typical of specific polar interactions. Ligand **222** presented intermediate binding energy (–16.64 ± 2.76 kcal/mol), with relevant contributions from both electrostatic and van der Waals forces, indicating a balanced binding profile.

Ligand **412**, once again, presented the best computational binding energy among all compounds analyzed for mGluR8 ($\Delta G_{binding}$ = –27.36 ± 3.39 kcal/mol). This affinity is supported by strong van der Waals interactions (ΔE_{vdW} = –31.68 ± 4.67 kcal/mol) and additional electrostatic contributions (ΔE_{el} = –18.70 ± 4.12 kcal/mol). The high hydrophobic contribution of compound **412** suggests a favorable fit in the receptor active site, potentially favored by nonpolar interactions and conformational complementarity. This feature may be especially relevant in lipophilic environments such as the central nervous system, in which hydrophobic interactions significantly contribute to the stability of the ligand-receptor complex. This profile reinforces the potential of ligand **412** as an effective candidate for dual modulation of mGluR7 and mGluR8 receptors.

The internal dielectric constant applied in molecular dynamics simulations reflects the ability of the protein environment to "dampen" electrostatic interactions between charges. Higher dielectric constant values suggest that the charges in the system are more exposed and undergo greater electrostatic shielding, while lower values indicate a more hydrophobic environment or a more internal and compact organization of the charges [53, 54]. In the present study, the dielectric constant applied for mGluR7 was 6, while for mGluR8 it was 3, which suggests that the mGluR7 system presents greater exposure of electric charges or a structural arrangement that favors more moderate electrostatic interactions. The Optimized Potentials for Liquid Simulations (OPLS) force field presents several advantages for protein molecular dynamics simulations followed by MM-PBSA calculations. Its careful parameterization for non-covalent interactions, such as hydrogen bonds and hydrophobic forces, ensures an accurate description of protein-ligand interactions [55, 56]. Furthermore, OPLS is optimized for both explicit and implicit solvent simulations, which favors consistency with continuum-based solvent free energy models, such as MM-PBSA [57]. The broad coverage of organic molecules and bioactive compounds in the most recent versions (such as OPLS3 and OPLS4) facilitates the modeling of protein systems complexed with diverse drugs [58]. Simulations performed with OPLS tend to preserve the experimental

geometries observed in proteins and complexes, favoring the generation of realistic conformations for subsequent energy analysis. Another important point is the incorporation of polarization corrections and optimized charges in OPLS, which increases the accuracy of the electrostatic potential calculations used in MM-PBSA [59].

Theoretical studies, such as molecular dynamics simulations, molecular docking and free energy calculations, have been consolidated as indispensable tools in the rational development of drugs and biomolecules, allowing the screening and optimization of compounds before expensive laboratory steps [60, 61]. This background becomes especially relevant regarding the group III metabotropic glutamate receptors (mGluRs), specifically the mGluR7 and mGluR8 subtypes, whose study is still in its preliminary stages and would require additional experimental study. Such earlier studies indicate that mGluR dysfunctions could be implicated in neurological states of disease, e.g., autism, where the aberrant signaling of such receptors is discovered to underlie synaptic as well as behavioral dysfunctions. Robust computational screenings can strategically guide these preliminary studies, prioritizing the most promising targets and molecules to optimize resource allocation in subsequent confirmatory experiments. These computational tactics can potentially hasten the discovery of targeted treatments for neuropsychiatric diseases, such as autism, by being able to screen and optimize molecules before the more expensive experimental phases.

5. Conclusions

This study aimed to conduct a comprehensive *in silico* investigation for the discovery of new dual agonists for mGluR7/8, given that these receptors have a significant correlation with the development of autism spectrum disorder symptoms in humans. The stages of this research included virtual screening using various publicly accessible online molecule databases, followed by studies of molecular docking, absorption, distribution, metabolism, excretion, and toxicity (ADMET). Two compounds were selected after these analyses and were further subjected to molecular dynamics simulations and MM-PBSA calculations to elucidate interaction energies and stabilization features within the active sites of both receptors. Ligand number **412** - 3-[(3-Cyanophenyl)methylsulfonylamino]-2-hydroxypropanoic acid - demonstrated the best performance based on energetic and conformational analyses, emerging as a promising molecule with agonistic properties computationally demonstrated in this work, which may be further reassessed in subsequent studies to obtain experimental data.

Statement of Usage of Artificial Intelligence

The authors declare that OpenAI's ChatGPT (version GPT-4, accessed via ChatGPT in 2025) was used solely to assist in **translating** the manuscript from Portuguese to English and for **grammar refinement** during the revision process. **The content, scientific analysis, and interpretations are entirely the work of the authors**.

Data Availability

Data is contained within the article and supplementary material.

Author Contributions

Conceptualization, A. P. L. Mesquita and T. C. Ramalho; formal analysis, E. F. F. Cunha and T. C. Ramalho; writing—original draft preparation, A. P. L. Mesquita; writing—review and editing, A. P. L. Mesquita, E. F. F. Cunha and T. C. Ramalho; supervision, E. F. F. Cunha; project administration, T. C. Ramalho.

Conflict of Interest

The authors declare no conflicts of interest.

Funding Information

This work was funded by CAPES (001), CNPq (36-/2014-9), FAPEMIG (PPM-00831-15), and the University of Hradec Kralove (Faculty of Science, VT2019-2021).

Ethics Statement

The authors declare that this study did not involve any experiments with human participants or animals. All research presented herein is entirely theoretical and computational in nature, conducted using in silico methods and simulations. Therefore, no ethical approval was required.

ORCID

Antônio Pedro Lemos de Mesquita - https://orcid.org/0009-0003-0410-8094

Elaine Fontes Ferreira da Cunha - https://orcid.org/0000-0002-3845-3801

Teodorico Castro Ramalho - https://orcid.org/0000-0002-7324-1353

References

1. Lord C, Brugha TS, Charman T, et al (2020) Autism spectrum disorder. Nature Reviews Disease Primers 2020 6:1 6:1–23. https://doi.org/10.1038/s41572-019-0138-4
2. Hodges H, Fealko C, Soares N (2020) Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. Transl Pediatr 9:S55. https://doi.org/10.21037/TP.2019.09.09
3. Upthegrove R, Abu-Akel A, Chisholm K, et al (2018) Autism and psychosis: Clinical implications for depression and suicide. Schizophr Res 195:80–85. https://doi.org/10.1016/J.SCHRES.2017.08.028
4. Jiang M, Lu T, Yang K, et al (2023) Autism spectrum disorder research: knowledge mapping of progress and focus between 2011 and 2022. Front Psychiatry 14:. https://doi.org/10.3389/FPSYT.2023.1096769/FULL
5. Nyrenius J, Waern M, Eberhard J, et al (2023) Autism in adult psychiatric out-patients: self-reported suicidal ideation, suicide attempts and non-suicidal self-injury. BJPsych Open 9:e167. https://doi.org/10.1192/BJO.2023.553
6. Montanari M, Martella G, Bonsi P, Meringolo M (2022) Autism Spectrum Disorder: Focus on Glutamatergic Neurotransmission. International Journal of Molecular Sciences 2022, Vol 23, Page 3861 23:3861. https://doi.org/10.3390/IJMS23073861
7. Mozafari R, Karimi-Haghighi S, Fattahi M, et al (2023) A review on the role of metabotropic glutamate receptors in neuroplasticity following psychostimulant use disorder. Prog Neuropsychopharmacol Biol Psychiatry 124:110735. https://doi.org/https://doi.org/10.1016/j.pnpbp.2023.110735
8. Parent HH, Niswender CM (2024) Therapeutic Potential for Metabotropic Glutamate Receptor 7 Modulators in Cognitive Disorders. Mol Pharmacol 105:348–358. https://doi.org/10.1124/MOLPHARM.124.000874
9. Niswender CM, Conn PJ (2010) Metabotropic Glutamate Receptors: Physiology, Pharmacology, and Disease. Annu Rev Pharmacol Toxicol 50:295. https://doi.org/10.1146/ANNUREV.PHARMTOX.011008.145533
10. Alagarsamy S, Saugstad J, Warren L, et al (2005) NMDA-induced potentiation of mGluR5 is mediated by activation of protein phosphatase 2B/calcineurin. Neuropharmacology 49:135–145. https://doi.org/https://doi.org/10.1016/j.neuropharm.2005.05.005
11. Edfawy M, Guedes JR, Pereira MI, et al (2019) Abnormal mGluR-mediated synaptic plasticity and autism-like behaviours in Gprasp2 mutant mice. Nat Commun 10:1431. https://doi.org/10.1038/s41467-019-09382-9
12. Limon A, Reyes-Ruiz JM, Miledi R, et al (2011) GABA and Glutamate Receptors of the Autistic Brain. Autism - A Neurodevelopmental Journey from Genes to Behaviour. https://doi.org/10.5772/20304
13. Gholipour P, Ebrahimi Z, Mohammadkhani R, et al (2024) Effects of (S)-3,4-DCPG, an mGlu8 receptor agonist, on hippocampal long-term potentiation at perforant pathway-dentate gyrus synapses in prenatal valproic acid-induced rat model of autism. Sci Rep 14:. https://doi.org/10.1038/S41598-024-63728-Y
14. Rabeh N, Hajjar B, Maraka JO, et al (2023) Targeting mGluR group III for the treatment of neurodegenerative diseases. Biomedicine & Pharmacotherapy 168:115733. https://doi.org/10.1016/J.BIOPHA.2023.115733

15. Li Y, Zhou Q, Dai A, et al (2023) Structural analysis of the dual agonism at GLP-1R and GCGR. Proceedings of the National Academy of Sciences 120:e2303696120. https://doi.org/10.1073/pnas.2303696120
16. Ebrahimi-Ghiri M, Khakpai F, Zarrindast MR (2020) Combined treatment of scopolamine and group III mGluR antagonist, CPPG, exerts antidepressant activity without affecting anxiety-related behaviors. Physiol Behav 224:113034. https://doi.org/10.1016/J.PHYSBEH.2020.113034
17. Ebrahimi-Ghiri M, Khakpai F, Zarrindast MR (2020) Combined treatment of scopolamine and group III mGluR antagonist, CPPG, exerts antidepressant activity without affecting anxiety-related behaviors. Physiol Behav 224:113034. https://doi.org/10.1016/J.PHYSBEH.2020.113034
18. Caia J, Ding L, Zhang JS, et al (2016) Elevated plasma levels of glutamate in children with autism spectrum disorders. Neuroreport 27:272–276. https://doi.org/10.1097/WNR.0000000000000532
19. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/JCC.20945
20. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665–W667. https://doi.org/10.1093/NAR/GKH381
21. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins 61:704–721. https://doi.org/10.1002/PROT.20660
22. Waterhouse A, Bertoni M, Bienert S, et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296. https://doi.org/10.1093/NAR/GKY427
23. Bienert S, Waterhouse A, De Beer TAP, et al (2017) The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res 45:D313–D319. https://doi.org/10.1093/NAR/GKW1132
24. Sunseri J, Koes DR (2016) Pharmit: interactive exploration of chemical space. Nucleic Acids Res 44:W442. https://doi.org/10.1093/NAR/GKW287
25. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25. https://doi.org/10.1016/S0169-409X(96)00423-1
26. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports 2017 7:1 7:1–13. https://doi.org/10.1038/srep42717
27. O'Boyle NM, Banck M, James CA, et al (2011) Open Babel: An Open chemical toolbox. J Cheminform 3:1–14. https://doi.org/10.1186/1758-2946-3-33/TABLES/2
28. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1:337–341. https://doi.org/10.1016/J.DDTEC.2004.11.007
29. Baell JB, Nissink JWM (2018) Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. ACS Chem Biol 13:36–44. https://doi.org/10.1021/acschembio.7b00903
30. Brenk R, Schipani A, James D, et al (2008) Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. ChemMedChem 3:435–444. https://doi.org/10.1002/CMDC.200700139
31. Yang Z-Q (2005) Agonists and antagonists for group III metabotropic glutamate receptors 6, 7 and 8. Curr Top Med Chem 5:913–918. https://doi.org/10.2174/1568026054750272
32. Brabet I, Parmentier ML, De Colle C, et al (1998) Comparative effect of l-CCG-I, DCG-IV and γ-carboxy-l-glutamate on all cloned metabotropic glutamate receptor subtypes. Neuropharmacology 37:1043–1051. https://doi.org/10.1016/S0028-3908(98)00091-4
33. Mysinger MM, Carchia M, Irwin JohnJ, Shoichet BK (2012) Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J Med Chem 55:6582–6594. https://doi.org/10.1021/jm300687e
34. Thomsen R, Christensen MH (2006) MolDock: A new technique for high-accuracy molecular docking. J Med Chem 49:3315–3321. https://doi.org/10.1021/JM051197E/SUPPL_FILE/JM051197ESI20060314_081922.PDF
35. Myung Y, De Sá AGC, Ascher DB (2024) Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. Nucleic Acids Res 52:W469–W475. https://doi.org/10.1093/NAR/GKAE254
36. Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001
37. Sousa da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367. https://doi.org/10.1186/1756-0500-5-367
38. Case DA, Aktulga HM, Belfon K, et al (2023) AmberTools. J Chem Inf Model 63:6183–6191. https://doi.org/10.1021/ACS.JCIM.3C01153/ASSET/IMAGES/LARGE/CI3C01153_0002.JPEG
39. Ribeiro AAST, Horta BAC, De Alencastro RB (2008) MKTOP: a program for automatic construction of molecular topologies. J Braz Chem Soc 19:1433–1435. https://doi.org/10.1590/S0103-50532008000700031

40. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E (2021) Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. J Chem Theory Comput 17:6281–6291. https://doi.org/10.1021/ACS.JCTC.1C00645/ASSET/IMAGES/LARGE/CT1C00645_0005.JPEG
41. Duan L, Liu X, Zhang JZH (2016) Interaction entropy: A new paradigm for highly efficient and reliable computation of protein-ligand binding free energy. J Am Chem Soc 138:5722–5728. https://doi.org/10.1021/JACS.6B02682/ASSET/IMAGES/JA-2016-026824_M008.GIF
42. Kumar M, Rathore RS (2025) RamPlot: a webserver to draw 2D, 3D and assorted Ramachandran (, ψ) maps. J Appl Crystallogr 58:630–636. https://doi.org/10.1107/S1600576725001669
43. Ke Y-Y, Selvaraj Coumar M, Shiao H-Y, et al (2014) Ligand efficiency based approach for efficient virtual screening of compound libraries. https://doi.org/10.1016/j.ejmech.2014.06.029
44. Bolteau R, Duroux R, Laversin A, et al (2022) High ligand efficiency quinazoline compounds as novel A2A adenosine receptor antagonists. Eur J Med Chem 241:. https://doi.org/10.1016/j.ejmech.2022.114620
45. Jain P, Satija J, Sudandiradoss C (2023) Discovery of andrographolide hit analog as a potent cyclooxygenase-2 inhibitor through consensus MD-simulation, electrostatic potential energy simulation and ligand efficiency metrics. Sci Rep 13:. https://doi.org/10.1038/s41598-023-35192-7
46. Aziz M, Ejaz SA, Zargar S, et al (2022) Deep Learning and Structure-Based Virtual Screening for Drug Discovery against NEK7: A Novel Target for the Treatment of Cancer. Molecules 2022, Vol 27, Page 4098 27:4098. https://doi.org/10.3390/MOLECULES27134098
47. MacLeod JM, Rosei F (2011) Directed Assembly of Nanostructures. Comprehensive Nanoscience and Technology 1–5:13–68. https://doi.org/10.1016/B978-0-12-374396-1.00098-2
48. Vilseck JZ, Ding X, Hayes RL, Brooks CL (2021) Generalizing the Discrete Gibbs Sampler-Based λ-Dynamics Approach for Multisite Sampling of Many Ligands. J Chem Theory Comput 17:3895–3907. https://doi.org/10.1021/ACS.JCTC.1C00176/ASSET/IMAGES/LARGE/CT1C00176_0005.JPEG
49. Bruns RF, Watson IA (2012) Rules for Identifying Potentially Reactive or Promiscuous Compounds. J Med Chem 55:9763–9772. https://doi.org/10.1021/jm301008n
50. Empereur-Mot C, Guillemain H, Latouche A, et al (2015) Predictiveness curves in virtual screening. J Cheminform 7:. https://doi.org/10.1186/S13321-015-0100-8
51. Zhai J, He X, Sun Y, et al (2022) In Silico Binding Affinity Prediction for Metabotropic Glutamate Receptors Using Both Endpoint Free Energy Methods and A Machine Learning-Based Scoring Function. Phys Chem Chem Phys 24:18291. https://doi.org/10.1039/D2CP01727J
52. Phillips BP, Miller EA (2021) Membrane protein folding and quality control. Curr Opin Struct Biol 69:50–54. https://doi.org/https://doi.org/10.1016/j.sbi.2021.03.003
53. Roux B, Simonson T (1999) Implicit solvent models. Biophys Chem 78:1–20. https://doi.org/https://doi.org/10.1016/S0301-4622(98)00226-9
54. Warshel A, Sharma PK, Kato M, et al (2006) Electrostatic Basis for Enzyme Catalysis. Chem Rev 106:3210–3235. https://doi.org/10.1021/cr0503106
55. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc 118:11225–11236. https://doi.org/10.1021/ja9621760
56. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864
57. Banks JL, Beard HS, Cao Y, et al (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem 26:1752–1780. https://doi.org/https://doi.org/10.1002/jcc.20292
58. Roos K, Wu C, Damm W, et al (2019) OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. J Chem Theory Comput 15:1863–1874. https://doi.org/10.1021/acs.jctc.8b01026
59. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105:6474–6487. https://doi.org/10.1021/jp003919d
60. Cournia Z, Allen BK, Beuming T, et al (2020) Rigorous Free Energy Simulations in Virtual Screening. J Chem Inf Model 60:4153–4169. https://doi.org/10.1021/acs.jcim.0c00116
61. Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational Methods in Drug Discovery. Pharmacol Rev 66:334–395. https://doi.org/10.1124/PR.112.007336

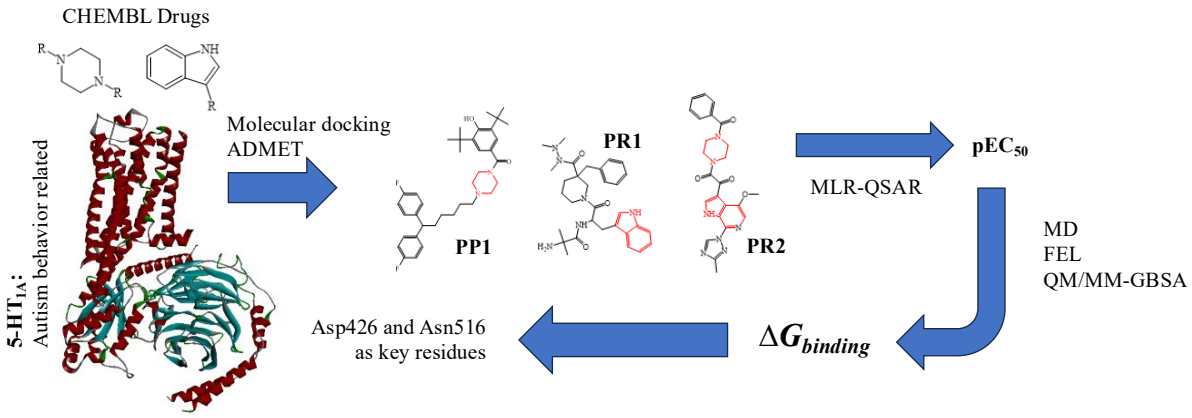
Artigo 2 - Redigido conforme norma do periódico científico - Versão preliminar	
Título do artigo:	Machine Learning, Molecular Dynamics, and QM/MM Free Energy Calculations Reveal the Potential of Repositioned Compounds as 5-HT_{1A} Agonists for ASD
Autores:	Antônio Pedro Lemos de Mesquita Adriana Cássia de Faria Matheus Puggina de Freitas Elaine Fontes Ferreira da Cunha Teodorico de Castro Ramalho
Periódico:	Journal of Biomolecular Structure and Dynamics
ISSN	0739-1102 (print); 1538-0254 (online)
DOI	https://doi.org/10.1080

PAPER 2 – Machine Learning, Molecular Dynamics, and QM/MM Free Energy Calculations Reveal the Potential of Repositioned Compounds as 5-HT$_{1A}$ Agonists for ASD

Antônio Pedro L. Mesquita[1], Adriana C. de Faria[1], Matheus P. Freitas[1], Elaine F. F. da Cunha[1,*] and Teodorico Castro Ramalho[1,2±]

[1] Department of Chemistry, Federal University of Lavras. Lavras, Brazil.; antoniolemos.mesquita@gmail.com
[2] Department of Chemistry, University of Hradec Kralove. Hradec Kralove, Czech Republic.
*Correspondence: elaine_cunha@ufla.br; Tel.: +55 (35) 3829-5129; ± teo@ufla.br

GRAPHICAL ABSTRACT



ABSTRACT

5-Hydroxytryptamine type 1A (5-HT$_{1A}$) receptors are broadly expressed in the central nervous system and play a key role in serotonin signaling. Due to their growing association with autism spectrum disorder (ASD) neurogenesis and the lack of safe, targeted pharmacological treatments, this study aimed to virtually screen CHEMBL compounds containing Piperazine-(1,4R) and 1H-Indole-(3R) scaffolds. The goal was to identify potential 5-HT$_{1A}$ agonists and investigate their interactions to rationalize future drug design. Using computational approaches – including molecular docking, molecular dynamics, PCA-based free energy landscapes, QM/MM-GBSA and MM-PBSA energy calculations, and MLR-QSAR modeling – we identified 39 candidates with favorable energetic, structural, physicochemical, and pharmacokinetic profiles (via ADMET). Three compounds with the highest predicted activity were analyzed in detail: one preclinical ligand (PP1: CHEMBL2031962; ΔG = –29.06 kcal/mol) and two in clinical stages (PR1: Anamorelin, ΔG = –13.97 kcal/mol; PR2: Temsavir, ΔG = –33.61 kcal/mol). All outperformed Buspirone, a reference 5-HT$_{1A}$ agonist. Furthermore, two amino acid residues (Asp246 and Asn516) were consistently involved in stable interactions with the selected ligands, highlighting their importance for receptor activation. These findings support the potential of repurposing known compounds as effective 5-HT$_{1A}$ agonists in the context of ASD therapy.

INTRODUCTION

The 5-hydroxytryptamine type 1A (5-HT_{1A}) receptors belong to the G protein-coupled receptor (GPCR) family and are widely distributed in the central nervous system, with significant presence in the prefrontal cortex, hippocampus, septum, and raphe nuclei [1]. Predominantly located on the plasma membrane of neurons, 5-HT_{1A} receptors participate in metabolic pathways involving the inhibition of adenylate cyclase and, consequently, the reduction of intracellular cyclic adenosine monophosphate (cAMP) levels [2]. Structurally, they are composed of seven transmembrane domains, a typical characteristic of GPCRs. Their activation by $G_{i/o}$ protein-coupled agonists leads to the inhibition of adenylate cyclase, resulting in decreased intracellular cAMP [3]. This process triggers the opening of potassium channels and the closing of calcium channels, promoting membrane hyperpolarization and a reduction in neurotransmitter release, including serotonin itself [1]. As autoreceptors, 5-HT_{1A} receptors negatively regulate serotonergic release; as postsynaptic receptors, they modulate functions related to mood, anxiety, and cognition. Agonist molecules binding to these receptors modulate physiological processes such as anxiety, mood regulation, thermoregulation, and stress response [4].

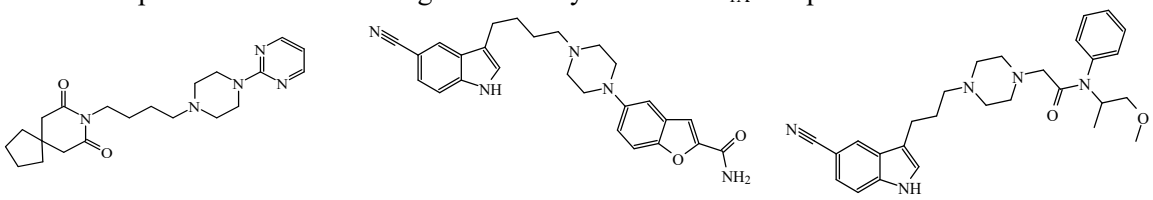
Recently, there has been growing interest in investigating the correlation between 5-HT_{1A} receptor function and neurodevelopmental disorders, particularly autism spectrum disorder (ASD) [5–7]. Autism is a complex condition characterized by persistent deficits in social communication and restricted, repetitive behavioral patterns. Neurobiological evidence indicates that dysfunction in serotonergic signaling, notably involving 5-HT_{1A} receptors, may be implicated in alterations in synaptic development, neuronal plasticity, and sociability circuits – often disrupted in individuals with ASD [6]. Autistic individuals may exhibit deficits ranging from subtle to severe in these areas, frequently accompanied by comorbidities such as anxiety, depression, suicide [8], attention-deficit/hyperactivity disorder (ADHD), and epilepsy [9]. These manifestations significantly impact patients' quality of life, autonomy, and social inclusion, while also posing an ongoing challenge for families and caregivers. Despite advances in educational strategies and behavioral therapies, available pharmacological treatments remain limited and do not address the core symptoms of the disorder. Given this, the identification of new molecular targets and the development of more effective, safe, and specific drugs capable of modulating central neurobiological pathways involved in ASD and providing broader therapeutic benefits has become urgent.

Among compounds with recognized agonist activity at the 5-HT_{1A} receptor, those featuring structural cores of piperazine and indole – such as the molecules Buspirone [10] and Vilazodone [11] (**Figure 1**), both clinically approved – stand out. These functional groups are widely considered critical pharmacophores, as they directly interact with key residues in the receptor's orthosteric site, enhancing affinity, selectivity, and stabilization of the 5-HT_{1A} active conformation. Based on this, the hypothesis was proposed that other ligands containing these same chemical substructures could similarly exhibit potential agonist activity, motivating a targeted virtual screening approach to identify new bioactive candidates. Previous rational drug development efforts, such as those by [12–15] reinforce this approach, highlighting the fundamental role of indolealkylpiperazine-like structures and their derivatives in functional activity. Notably, molecule 9_24, described by [14], exhibited the lowest EC_{50} value among the analyzed compounds, serving as a structural reference for new molecular designs.

In this study, an integrated approach combining machine learning, molecular docking, molecular dynamics, and free energy calculations based on quantum mechanics/molecular mechanics with the generalized Born surface area (QM/MM-GBSA) model was conducted to evaluate the binding profile and conformational stability of potential agonists identified through an extensive virtual screening of CHEMBL Drugs, isolating compounds containing substituted Piperazine-(1,4R) and 1H-Indole-(3R) groups, whether in preclinical development or more advanced phases (1 to 4). The adopted strategy aims to accelerate the rational discovery of drugs targeting the modulation of the 5-HT_{1A}

receptor, focusing on the repurposing [16] of existing molecules and the identification of key molecular substructures and amino acid residues essential for ligand recognition by the serotonin receptor, contributing to the rational development of future innovative therapies aimed at mitigating the clinical manifestations of autism.

Figure 1: Compounds with validated agonist activity at the 5-HT_{1A} receptor.



Buspirone
EC_{50} = 15.14 nM
K_i = 3.9 nM

Vilazodone
EC_{50} = 0.79 nM
K_i = 0.3 nM

9_24
EC_{50} = 0.059 nM
K_i = 5 nM

MATERIALS AND METHODS

Homology Modeling

Due to the absence of structural regions (mostly loops) in the available experimental structure of the protein, homology modeling was chosen as a strategy to complement the structure. This approach allowed the generation of a complete three-dimensional model, based on a homologous protein with adequate sequence identity, ensuring the continuity of the polypeptide chain and the structural quality necessary for subsequent docking and molecular dynamics studies. Using the Swiss Model online server [17,18], the FASTA sequence of the crystallized structure of 5-HT_{1A} (PDB-ID 8FYX), obtained from the Protein Data Bank (PDB), was registered, containing the amino acid residues of the side chain D, where the binding site is complexed with a molecule of Buspirone. The model was then built based on the 8FYT structure, complexed with a molecule of lysergic acid diethylamide (LSD), preserving the structure of the binding site. After the process, the new three-dimensional structure was downloaded in PDB, as well as the model validation report, containing the sequence similarity, the global Qualitative Model Energy Analysis and Distance Constraints (QMEANDisCo) and the Ramachadran plot. Extra analyses were made with RamPlot, to generate 3D Ramachandran plots with frequencies and ProSA web server to calculate de Z-score and energy per residue of the model. Results can be found in supplementary material, in **Figure S1.**

Virtual screening and ADMET predictions

The subsequent phase of the study involved the virtual screening of ligands containing key molecular substructures, with the objective of investigating their interactions with the 5-HT_{1A} receptor. Two Piperazine-1,4R and 1H-indole(3R) cores were drawn in CHEMBL Drugs server [19,20], where the search for other molecules with the same substructure was configured, resulting in a total of 29,446 compounds, excluding duplicates. Absorption, distribution, metabolism, excretion and toxicity (ADMET) prediction tests were performed by DeepPK [21] and SwissADME [22] servers. The main descriptor used for compound exclusion was the ability to penetrate the blood-brain barrier. Other descriptors used for ligand quality assessment were: human oral bioavailability 50%, inhibition of cytochrome P450 (CYP450) 3A4, half-life of drug, liver injury, Lipinki's, Brenk's rule [23] and PAINS [24]. Ligands in preclinical stage (0) and in phases 1, 2, 3, 4 were screened.

Molecular docking protocol and validation

To evaluate the activity of the virtual screening compounds on 5-HT_{1A}, a molecular docking protocol was applied, with validation performed using a Receiver Operating Characteristic (ROC) curve. Ten molecules with biological activities experimentally elucidated by Zhu et al. [12], with low EC_{50} values were drawn in BIOVIA Discovery Studio (2024) and used as reference ligands. For each molecule, 50 decoys were generated in DUDE server [25], totaling 500. Hydrogen atoms were added at physiological pH (7.4) to all compounds, followed by charge calculation by the Electronegativity Equalization Method (EEM) [26], using OpenBabel software (v. 2.4.1) [27]. The three-dimensional structure of the protein, previously modeled, was prepared by the server PDB2PQR [28], in which hydrogens were added at pH 7.4 using AMBER force field [29] and pK_a prediction calculations were performed by ProPKa [30]. Gasteiger charges were assigned to the protein in Discovery Studio.

The software used for molecular docking calculations was Molegro Virtual Docker (MVD) [31]. After adding the structures of the reference ligands, decoys and protein, a cavity prediction analysis was performed, in which the cavity that best fit around the cocrystallized ligand was chosen, with a volume of 369 $Å^3$. Fifty-nine residues of the active site were configured as flexible and the MolDock Score GRID scoring function was used, with a resolution of 0.30 Å. The restriction sphere for calculations was defined around the binding site cavity, located at coordinates (x, y, z) = (136.5, 136.0, 174.0), with a radius of 13 units. The search algorithm used was MolDock SE, with 200 calculation rounds for each ligand and a total of 1500 iterations per round. At the end, the return of the 50 lowest energy poses for each step was configured.

Finished the docking processes, the energies were analyzed and stored in a spreadsheet. The lowest energy pose of the cocrystallized ligand was used for validation by redocking. The docking performance was evaluated by constructing a ROC curve, using the energies of the reference ligands and decoys. The total energy descriptor (Moldock Score) was used. The AUC (Area Under Curve) was calculated using OriginPro 2024 and the model was validated with an area > 0.7. All of the molecular dockings performed with the compounds found in the virtual screening process followed the same protocol validated by the AUC.

MLR based QSAR model building

A Quantitative Structure Activity Relationship (QSAR) model using descriptors encoding hydrophilic, electronic, structural and dimensional properties, selected by genetic algorithm (GA), was constructed, and these descriptors were regressed against pEC_{50} values using multiple linear regression (MLR) in Chemoface [32] software (v. 1.71). Thirty-nine compounds with EC_{50} values calculated by the [^{35}S]GTPγS method, extracted from the works of [12–15], were used to construct the model. The calculation of the MLR descriptors was performed in stages, with three computational tools. The average partial charges of the heavy atoms (other than hydrogen) were calculated by the OpenBabel software (v. 2.4.1). The method employed was the Electronegativity Equalization Method (EEM), rigorously parameterized with DFT calculations (B3LYP/6-31G*/MPA). The final file was saved in Sybil MOL2, the partial charge columns were transferred to a spreadsheet software and the average was calculated, excluding the hydrogen atoms. The rotational bond parameters, atomic mass of heavy atoms and number of valence electrons were calculated using ChemMaster software (v. 1.2) (Crescent Silico). Finally, the descriptors related to the number of hydrogen bond donors, solubility (ESOL - log S) and topological polar surface area (TPSA) were calculated using the SwissADME web server. The fitness function employed in GA selection was based on error/MAE parameters, and a total of 100 iterations, mutation probability of 0.3, and 30 selected equations in each generation were applied [33]. The quality of the MLR model was determined by the coefficient of determination of calibration (r^2), leave-one-out cross-

validation (q^2), and prediction (r^2_{pred}) [34]. The domain of applicability and, therefore, the presence of outliers was assessed using William's sample leverage plots *vs*. Student's residuals [35].

The matrix was then divided into training (75%) and test (25%) sets using Kennard-Stone sampling [36]. The training set was used for calibration tests [37], leave-one-out cross-validation (LOOCV), and y-randomization tests, while the test set was used for external validation. The quality and predictive capacity of the models were assessed through root mean square error (RMSE), as well as by the coefficients of determination in the calibration r^2 (values ≥ 0.6 are acceptable). Validation experiments were performed using y-randomization tests to ensure the absence of random correlation ${}^{c}r^2_p$ (values ≥ 0.5 are acceptable) [38], leave-one-out cross-validation q^2 (values ≥ 0.5 are acceptable) [39], and external validation r^2_{pred} (values ≥ 0.5 are acceptable). A measure of the proximity between the experimental activity values and the calculated data was obtained through the r^2_m metrics, as well as the mean r^2_m ($\bar{r}^2_m$) and Δr^2_m [40]. The mean absolute error (MAE) and the concordance correlation coefficient (CCC, which is a measure of precision and accuracy) were also calculated; values ≥ 0.80 are acceptable [41]. An additional bootstrapping procedure was performed for 10 randomly selected test sets to attest the stability of the models, and the parameters mentioned above were acquired for comparison [41].

Molecular similarity was assessed using the Tanimoto coefficient, based on ECFP6 fingerprints (radius 3, 2048 bits) generated from SMILES strings of the training set and proposals with the RDKit library. The fingerprints incorporated pharmacophoric features and chirality to better capture structural and stereochemical information. Pairwise Tanimoto similarity values were computed by comparing the binary vectors, resulting in a symmetric similarity matrix with values ranging from 0 (no similarity) to 1 (identical structures). The matrix was exported in CSV format for further analysis and converted to a heatmap. Details can be found in supplementary materials in **Script S1**.

Molecular dynamics simulations

To validate the dynamic stability of the virtual screening and MLR proposals, molecular dynamics simulations were performed with GROMACS 2021.4 [42]. First, the PDB file of the modeled protein was submitted to GROMACS, with the physiological protonation state preserved. The OPLS-AA force field [43] and the TIP3P water model [44] were applied to write the topology file of the protein. The ligands that underwent virtual screening, molecular docking and MLR-QSAR were prepared in identical ways, following the protocol: using the AnteChamber PYthon Parser interface (ACPYPE) software [45], present in the AmberTools23 package [46], the ligands were registered through PDB files for the construction of topologies, calculation of partial charges and energy minimization by SQM-AMBER, using a QM/MM level of theory, with the semiempirical AM1. After that, the MKTOP.pl [47] script was used to better identify the AtomTypes according to OPLS. Once the topologies of the protein and ligands were written, the complexes were constructed. The complexes were placed in a cubic solvation box, with distances of 1.5 nm beyond the edges of the receptor and subsequently solvated with 70,971 water molecules. To neutralize the charges of the system, 13 chloride ions were added at random positions in the solvent, ensuring that the total charge was zero. An APO system was also configured, where the protein was prepared by the same methodology, but without any ligand.

After the construction and solvation of the complexes, energy minimization steps were performed, using the Steepest Descent methods with and without position restrain, in addition to the application of the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm (**Figure S1**). Each minimization step occurred in 20,000 steps. Equilibrations were also performed. Using the canonical isothermal-isovolumetric (NVT) ensemble, in which the system temperature is equilibrated, a 100 ps simulation was performed at a temperature fixed at 310 K (**Figure S1**). In the isothermal-isobaric (NPT) ensemble, a reference pressure of 1 bar was fixed, with a simulation time of 100 ps

(**Figure S1**). At the end, the system temperature stabilized at 310 K, with a density of approximately 985 kg/m^3. After this, an extra water dynamics step was performed to ensure solvent relaxation, keeping the protein and ligands restricted to the equilibrium positions, for a time of 500 ps. After all these previous steps, the system was subjected, without position restrain, to a 200 ns simulation, with 100 million steps and a time increment of dt = 2 fs for temporal integration. The Verlet cutoff-scheme was used, at a temperature of 310 K, keeping only the bonds with hydrogen atoms restrained. At the end of the simulations, the calculations of RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuation) and hydrogen bonds were performed by GROMACS, saved in *.xvg files and plotted by the XMGRACE software.

Gibbs free energy landscapes

Large, flexible, and complex molecular systems such as proteins can generate a range of conformational states with different associated energy values. Thus, a Free Energy Landscape (FEL) study helped the recognition of stable conformations (or time intervals in which these conformations prevail most frequently) to guide deeper analyses, such as free energy calculations of receptor-ligand binding. To calculate the FEL, GROMACS was used in the same version previously reported to perform the diagonalization of the covariance matrices, using the gmx covar module, to also elucidte the eigenvectors and eigenvalues of the complexes and to construct the principal components (PC1 and PC2) in the time interval of 100-200 ns. Once the values of the first two principal components were obtained, a Python script was used to calculate the Gibbs energy (G), using the statistical-mechanical definition [48] and the Kernel Probability Density Function (KDE) [49] fed with the values of PC1 and PC2 (details are found in the supplementary material, in **Script S2**). Finally, the script generated a text file containing the values of the two principal components, the probability via Kernel density and *G*. The data were transferred to the OriginPro 2024 software and the three-dimensional graphs of the free energy landscapes were plotted, after data matrixing according to the coordinates (x, y, z) = ($PC1$, $PC2$, G).

MM-PBSA and QM/MM-GBSA calculations

To access thermodynamic potentials related to the binding energies between 5-HT$_{1A}$ and the proposed ligands, QM/MM-GBSA and MM-PBSA calculations were performed from the time intervals related to the most stable conformations, according to the FEL analyses. All MM-PBSA and QM/MM-GBSA calculations were performed with gmx_MMPBSA software (v. 1.6.4) [50]. All trajectories were centered and periodic boundary conditions (PBC) were removed. Translational and rotational movements were also excluded in order to improve the accuracy of configurational entropy calculations. The time interval used in the calculations covers the 150-200 ns interval, which corresponds to frames 7,500-10,000. The calculation temperature was set at 310 K and the entropy calculation method used was Interaction Entropy [51]. For MM-PBSA calculations, the polar solvation energy was obtained considering the implicit solvent model, by solving the Poisson-Boltzmann equation (PBE) in its linear form, and the potential energies of the complexes were obtained by the results of the van der Waals and electrostatic contributions, according to the FF14SB (protein) and GAFF2 (ligand) AMBER forcefields. For QM/MM-GBSA calculations, the polar solvation energy was determined by the generalized Born (GB) equation. The interacting amino acid residues of the active site Ala223, Ala333, Ala495, Ala513, Asn516, Asp246, Cys317, Gly519, Ile243, Ile319, Leu498, Lys321, Met507, Phe242, Phe491, Pro499, Ser329, Thr251, Thr330, Trp488, Tyr226, Tyr520, Val247, Phe492 and the ligands were quantum mechanically treated by means of the PM6-DH+ semiempirical Hamiltonian. The remainder of the complex was treated by molecular mechanics, using the energy parameters calculated by the classical

dynamics simulation. The nonpolar solvation energy for both cases was calculated by the Solvent Accessible Surface Area (SASA) model, with the dispersion term based on a surface integration model [52]. The dielectric constant of water was set to 80, while the internal dielectric constant was fixed to 3, according to optimization calculations from the reference ligand. An energy decomposition analysis was performed by MM-PBSA using the same interacting residues as in the QM/MM-GBSA step to elucidate the role of the amino acid residues in stabilizing the ligands.

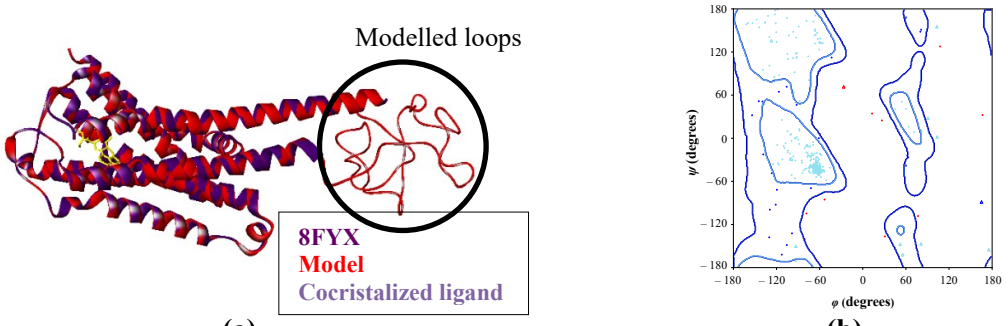
RESULTS

Homology Modeling

For the correction of the 5-HT_{1A} receptor structure complexed with a Buspirone molecule, homology modeling was performed to reconstruct the missing loops from the original crystal structure (PDB ID: 8FYX). The Swiss-Model results consisted of a total of four models, ranked according to their initial Global Model Quality Estimation (GMQE) scores, which were 0.66, 0.64, 0.46, and 0.43, respectively, for the templates with the following codes: UniProt Q6XXX9, PDB:8FYT, PDB:8DU3, and PDB:9CHU. The last two models (8DU3 and 9CHU) were discarded due to their low predicted quality scores.

The first template, UniProt Q6XXX9, was modeled from *Canis lupus familiaris* and showed a sequence identity of 93.20%. However, it lacked both a calculated value of QMEANDisCo Global (Qualitative Model Energy Analysis – Distance Constraints) and a co-crystallized ligand. The second template (PDB:8FYT) was selected for further analysis because it was based on a human 5-HT_{1A} structure, showed 100% sequence identity, had a QMEANDisCo Global score of 0.70, and included a co-crystallized ligand (LSD) that marked the binding site. The complete analysis per residue can be found in supplementary materials in the graph depicted in **Figure S2**.

Figure 2(a) shows the superposition of the generated model (in red) relative to the original receptor (5FYZ). It is evident that the structure and conformation of the active site were well conserved, particularly in the region surrounding the co-crystallized ligand, with a clear superposition throughout the domain containing the depicted α-helices. Additionally, the missing loops in the original structure, circled in black, demonstrate the model's capability to reconstruct regions of interest, thereby improving the quality of subsequent docking and molecular dynamics analyses. **Figure 2(b)** presents the Ramachandran plot, which correlates the torsion angles ψ and φ of the backbone relative to the α-carbons. The colors cyan, blue, and red (dots/triangles) represent torsion angles in favored, allowed, and disallowed regions, respectively; dots correspond to residues other than glycine, while triangles denote glycine. Visual inspection indicates that only 9 of the 381 residues in the generated model were located in disallowed regions in terms of torsion angles. The remaining residues displayed angles compatible with the expected conformations of the receptor's secondary structure.



(a) **(b)**

Figure 2: (a) Superimposed 3D structures of the original receptor 8FYX (purple) and model generated (red). The cocristalized LSD molecule (yellow) marks the active site; **(b)** The Ramachandran plot for the model [53].

Virtual screening and ADMET predictions

The structures of the Piperazine-(1,4R) (substructure A) and 1H-indole(3R) (substructure B) cores were drawn and submitted for a search on the ChEMBL Drugs server, as illustrated in **Table 1**. From substructure A, 12,113 compounds containing the same structural core were identified, while substructure B revealed an even more significant number, with 30,000 associated compounds. However, after the rigorous application of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) property-based filters, a significant reduction in the number of promising candidates was observed. Only 174 compounds derived from substructure A and 196 compounds from substructure B met the pharmacokinetic and toxicological acceptability criteria, making them viable for subsequent analyses in the rational drug screening process.

Table 1: Substructures used in CHEMBL Drugs search engine

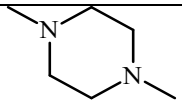
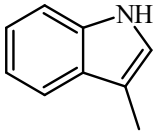
Substructure ID	Hit compounds	ADMET approved
		
A	12,113	174
		
B	30,000	196

Figure 3 presents a flowchart outlining the steps following the ChEMBL search. Out of a total of 42,208 compounds, 42,017 were classified by the server as preclinical (phase 0), while 191 were in advanced stages, such as phases 1, 2, 3, and 4 of drug development. For the ligands in later testing stages, identical compounds were first excluded, followed by an analysis of blood-brain barrier (BBB) penetration. Finally, predictions were made for oral bioavailability, cytochrome P450-3A4 inhibition potential, and hepatotoxicity. In the end, 16 compounds were selected with predicted ideal pharmacological properties. For the preclinical-stage ligands, the same methodological cascade was applied, with the additional step of screening for Lipinski's and Brenk's rule violations, as well as an assessment for the presence of PAINS (pan-assay interference compounds) structures. Ultimately, 227 compounds were subjected to molecular docking calculations to evaluate their binding modes with 5-HT_{1A}, compared to ligands with experimentally validated biological activity.

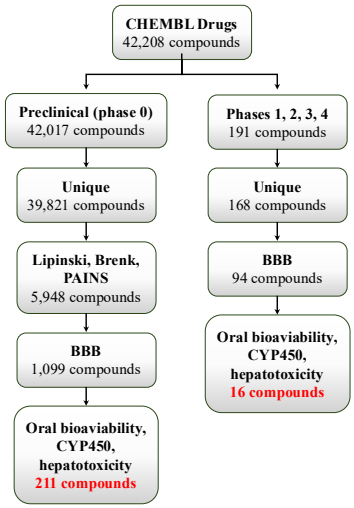


Figure 3: Schematic Representation of ADMET Screening Workflow Applied to ChEMBL Hits

Molecular docking

Prior to docking the compounds obtained from the ChEMBL database with the 5-HT_{1A} receptor, a docking methodology validation process was conducted through the construction of a Receiver Operating Characteristic (ROC) curve. For this purpose, 10 compounds with experimentally determined EC_{50} values from Zhu et al. (referred to as group A) were selected: 13M, 13G, 13H, 13O, 13F, 13N, 13P, 13E, 13K, and 13D, with EC_{50} values ranging between 7.00 and 9.00 nM. The binding energies obtained for these reference compounds, along with the energies of 50 decoys generated for each of them (totaling 500 inactive structures), were used to construct the ROC curve shown in **Figure 4**.

The area under the curve (AUC) value was 0.8230, indicating good discriminatory capability of the docking protocol in distinguishing active compounds from inactive ones (decoys) [54]. In addition to the compounds in group A, three other sets of molecules (groups B, C, and D, described in **Table 2**) were employed to subsequently develop a multiple linear regression (MLR) model. All additional compounds underwent docking calculations, exhibiting binding energies consistent with those of the active compounds in group A, further reinforcing the reliability of the applied methodology.

The experimental EC_{50} values were converted to pEC_{50} (-log EC_{50}), and the results are presented in **Table 2**, along with the corresponding MoldockScore values, which were used as the energy metric for validation.

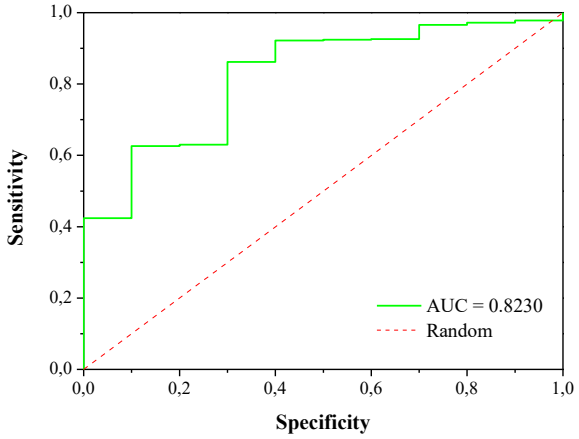


Figure 4: ROC curve for the molecular docking validation.

In addition to the ROC curve, a structural validation was performed through redocking, where the lowest-energy pose of the co-crystallized ligand was compared to its experimentally determined conformation in the three-dimensional crystal structure. The obtained RMSD (Root Mean Square Deviation) value was 0.8215 Å for LSD molecule (template's cocrystal ligand) and 1.1070 Å for buspirone (reference structure's cocrystal ligand), within the acceptable threshold (< 2.0 Å) [55], confirming the accuracy of the docking protocol in predicting reliable molecular poses. The superposition between the theoretical and experimental structures can be visualized in the supplementary material, **Figure S3**.

Although the compounds in **Table 2** show evidence of relevant biological activity, they all remain in the preclinical phase. Therefore, for comparative purposes, Buspirone docking was also performed. As a commercially available molecule with demonstrated in vitro and in vivo efficacy, Buspirone served as the standard reference ligand (cutoff molecule) in the virtual screening process. Only compounds with binding energies higher (more negative) than its established value of −148.51 kcal/mol were considered. The docking results for all molecules from CHEMBL and the structure of reference compounds can be found in **Table S1** and **Table S2** in the supplementary material.

Table 2: Reference compounds for validation of the molecular docking protocol by the ROC curve (group A, by Zhu et al) and other groups of potent 5-HT_{1A} agonist compounds extracted from the literature for construction of the MLR-QSAR model and external validation of the ROC curve (B by Lian et al, C by Wang et al and D by Xu et al).

Group	Number	Compound	EC_{50} (nM)	pEC50	MolDock (kcal/mol)
A	1	13A	167.7	6.78	-115.1
	2	13B	989.7	6.00	-160.01
	3	13C	197.2	6.71	-164.93
	4	13D	88.0	7.06	-159.61
	5	13E	49.2	7.31	-165.2
	6	13F	23.6	7.63	-187.49
	7	13G	5.35	8.27	-171.42
	8	13H	5.67	8.25	-210.65
	9	13I	400.5	6.40	-130.31
	10	13K	77.5	7.11	-180.27
	11	13L	523.9	6.28	-165.26
	12	13M	1.01	9.00	-177.31
	13	13N	31.6	7.50	-176.9
	14	13O	22.8	7.64	-188.45
	15	13P	48.2	7.37	-173.49
	16	13Q	174.7	6.76	-140.02
B	17	9A	106.8	6.97	-147.87
	18	9C	43.63	7.36	-171.36
	19	9D	6.53	8.19	-198.32
	20	9E	24.42	7.61	-186.74
	21	9F	287.11	6.54	-123.66
	22	9H	46.73	7.33	-166.8
	23	9I_2	11.95	7.92	-172.56
	24	9M	29.67	7.53	-188.91
	25	9O	15.81	7.80	-161.15
C	26	9_18	1.88	8.73	-223.47
	27	9_20	4.61	8.34	-200.18
	28	9_21	2.44	8.61	-187.63
	29	9_23	0.76	9.12	-231.71
	30	9_24	0.059	10.23	-205.13
D	31	FW01	7.0	8.15	-163.08
	32	FW02	77.0	7.11	-142.44
	33	FW03	404.0	6.39	-128.84
	34	FW04	128.0	6.89	-136.22
	35	FW05	534.0	6.27	-142.78
	36	FW06	434.0	6.36	-156.01
	37	FW07	340.0	6.47	-162.52
	38	FW08	572.0	6.24	-130.78
	39	FW09	597.0	6.22	-113.85
-	-	**Buspirone**	15.14	7.82	-148.51

Table 3 presents the compounds identified in the virtual screening of ChEMBL database entries, all exhibiting binding energies below the buspirone-derived threshold of −148.51 kcal/mol. Each compound satisfied rigorous ADMET criteria — including human oral bioavailability, blood-brain barrier (BBB) permeability, CYP450 3A4 interaction, half-life of drug, and hepatotoxicity potential (detailed in supplementary materials, in **Table S3**). Particular attention was given to BBB penetration due to 5-HT_{1A}'s CNS therapeutic relevance [56]. The compounds demonstrated energy profiles consistent with known bioactives (**Table 2**), supporting their candidate viability. From 37 approved molecules, 26 represent preclinical proposals (PP) and 11 are drug repurposing proposals (PR) — the latter involving clinical-phase compounds (phase 1+). All chemical structures are provided in supplementary materials – **Figure S4**.

Table 3: Molecular docking and ADMET results for the 37 proposals from CHEMBL screening.

Molecule	Moldock Score (kcal/mol)	CHEMBL-ID	Name	Phase test	Blood-Brain Barrier
PP1	-201.194	2031962	-	0	Penetrable (HC)
PP2	-188.313	2449412	-	0	Penetrable (HC)
PP3	-184.967	2326497	-	0	Penetrable (HC)
PP4	-181.632	2036648	-	0	Penetrable (HC)
PP5	-174.731	4446324	-	0	Penetrable (HC)
PP6	-174.331	4954080	-	0	Penetrable (HC)
PP7	-171.816	257743	-	0	Penetrable (HC)
PP8	-167.924	381781	-	0	Penetrable (HC)
PP9	-166.803	1963620	-	0	Penetrable (HC)
PP10	-165.596	1623490	-	0	Penetrable (HC)
PP11	-164.404	1624317	-	0	Penetrable (MC)
PP12	-163.279	98438	-	0	Penetrable (HC)
PP13	-163.065	270411	-	0	Penetrable (HC)
PP14	-160.091	5492590	-	0	Penetrable (HC)
PP15	-158.77	2220660	-	0	Penetrable (HC)
PP16	-156.296	4914762	-	0	Penetrable (HC)
PP17	-154.632	100891	-	0	Penetrable (HC)
PP18	-154.32	2426492	-	0	Penetrable (HC)
PP19	-154.167	494370	-	0	Penetrable (HC)
PP20	-153.298	4946422	-	0	Penetrable (MC)
PP21	-152.139	3492331	-	0	Penetrable (HC)
PP22	-151.911	1889146	-	0	Penetrable (HC)
PP23	-151.722	4966938	-	0	Penetrable (HC)
PP24	-150.291	1623304	-	0	Penetrable (HC)
PP25	-149.634	1542671	-	0	Penetrable (HC)
PP26	-148.602	5431131	-	0	Penetrable (HC)
PR1	-198.458	2110579	Anamorelin	3	Penetrable (MC)
PR2	-190.362	3301620	Temsavir	2	Penetrable (HC)
PR3	-177.598	4650316	Giredestrant	3	Penetrable (HC)
PR4	-176.84	4113131	Ensartinib	1	Penetrable (HC)
PR5	-171.796	4076837	Senexin b	1	Penetrable (HC)
PR6	-163.774	4113131	Ensartinib	3	Penetrable (HC)
PR7	-161.558	2111029	Oglufanide	3	Penetrable (LC)

Table 3 continuation…					
PR8	-161.129	4650351	Parpi	3	Penetrable (HC)
PR9	-161.03	4299940	Mitapivat	4	Penetrable (HC)
PR10	-160.166	1908397	Kw-2449	1	Penetrable (HC)
PR11	-159.432	1171829	Bitopertin	3	Penetrable (HC)

MLR-QSAR model building

Consolidated the compound proposals from the molecular docking and ADMET stages, a QSAR model was built using MLR to estimate pEC_{50} activity. Before exploring structure-activity relationships, an applicability domain test was performed to determine if the samples occupied the same chemical space, using leverages and studentized residuals (supplementary material, in **Figure S5**). Only one sample, cpd30 (referring to FW05), appeared as an outlier. Additional tests were conducted and it was kept in the dataset, thus the entire series of 39 compounds was considered (**Table 2**). Due to the large number of descriptors that was calculated for each sample, a genetic algorithm (GA) was employed to reduce descriptor dimensionality to a few explanatory variables, facilitating the analysis [57]. A MLR model was generated based on seven descriptors obtained by GA: heavy atoms mean partial charge (HA-MP/EEM), rotatable bonds (Rot Bonds), heavy atoms molecular weight (HA-MW), number of valence electrons (NV electrons), hydrogen bond donors (HB Donor), log *S* (ESOL), and TPSA.

Furthermore, a Kennard-Stone sampling approach was employed to split the dataset into 75% training samples and 25% test samples. The resulting equation (1) was as follows:

$$pEC_{50} = \sum_{i=1}^{7} a_i x_i + b + \varepsilon \qquad \frac{\partial pEC_{50}}{\partial x_i} = a_i \qquad (1)$$

where the coefficients can be consulted in **Table 5**. In total, 7 descriptors (x_i) were used, averaging approximately 6 molecules per descriptor in the training set, an acceptable amount [58]. The weight of each descriptor is expressed by the coefficient a_i, representing its influence on predicting biological activity (pEC_{50}). The linear coefficient and associated errors (deviations, ε) are also detailed in **Table 5**. The variable with the highest contribution was the mean partial charge of heavy atoms, while the lowest contribution came from the TPSA descriptor. Additionally, the Kennard-Stone model regression plot is shown in **Figure 5**, with model performance metrics as follows: $r^2 = 0.847$, RMSEC = 0.372, $r^2_{y\text{-}rand} = 0.204$, ${}^c r^2_p$ (y-rand) = 0.738, $RMSE_{y\text{-}rand} = 0.848$, $q^2 = 0.670$, $RMSE_{CV} = 0.569$, $r^2_{pred} = 0.751$, $RMSE_p = 0.562$, $r^2_m = 0.713$, avg. $r^2_m = 0.574$, $\Delta r^2_m = 0.215$, CCC = 0.789, MAE = 0.350.

Table 5: MLR-QSAR descriptors, weights and errors of the equation.

	Descriptor (x_i)	**Weight (a_i)**	**Deviation (ε)**
1	Mean charges EEM HA	-78.5561	± 15.6718
2	Rotatable bonds	0.2541	± 0.06160
3	Heavy Atoms Molecular Weight	-0.0519	± 0.01170
4	Num. Valence Electrons	0.0757	± 0.02200
5	H-bond donor	-0.6471	± 0.19840
6	log *S* (ESOL)	-0.7694	± 0.21540
7	TPSA	0.0388	± 0.01000
b	Linear coefficient	4.6929	± 1.11700

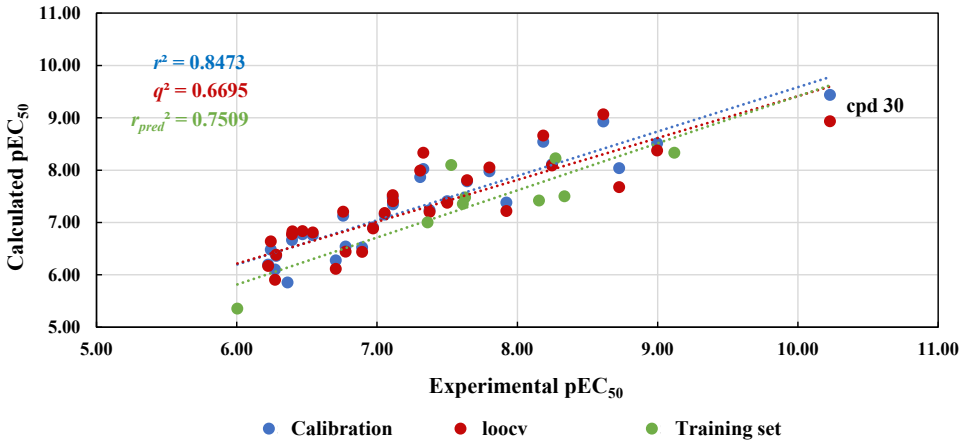


Figure 5: Regression lines of the MLR-QSAR model constructed using Kennard-Stone test sets.

These descriptors successfully generated robust models capable of making accurate predictions, as evidenced by an $r^2 > 0.6$, indicating a good fit to the data (and not merely a result of random correlation, given that ${}^{c}r^2_p$ of y-rand > 0.5). Furthermore, the models demonstrated strong predictive performance in both leave-one-out cross-validation ($q^2 > 0.5$) and external validation ($r^2_{pred} > 0.5$). The data exhibited appropriate dispersion, linearity, and satisfactory slope ($r^2_m > 0.5$ and CCC approaching 0.8), along with a low mean absolute error (MAE) [59].

To ensure model stability, a bootstrapping procedure was performed. In this process, 25% of the test set samples were randomly selected multiple times, totaling 10 repetitions [60]. The results, presented in **Table 6**, revealed highly consistent statistical behavior, as model quality remained stable across different training and test sets, with standard deviations of the evaluated statistical parameters below 10% - particularly for key metrics such as r^2, ${}^{c}r^2_p$, q^2, r^2_{pred}, average r^2_m, and CCC. This confirms that the GA-derived descriptor-based MLR models are reliable for predicting pEC_{50} values of the studied analogous compounds.

Table 6. Statistical parameters (10-cycle average) for MLR models built from bootstrapping experiments [a].

Parameters	Run 1	Run 2	Run3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	Average
RMSEC	0.37	0.37	0.39	0.38	0.40	0.39	0.39	0.40	0.40	0.41	0.39 ± 0.01
r^2	0.86	0.86	0.84	0.85	0.84	0.84	0.83	0.84	0.84	0.82	0.84 ± 0.01
$RMSE_{y\text{-}rand}$	0.87	0.83	0.81	0.83	0.88	0.85	0.77	0.86	0.88	0.83	0.84 ± 0.03
$r^2_{y\text{-}rand}$	0.21	0.27	0.31	0.27	0.21	0.25	0.31	0.25	0.21	0.25	0.25 ± 0.03
${}^{c}r^2_{p\ (y\text{-}rand)}$	0.75	0.71	0.67	0.70	0.73	0.70	0.66	0.70	0.73	0.68	0.70 ± 0.02
RMSECV	0.55	0.54	0.56	0.55	0.58	0.61	0.54	0.55	0.57	0.62	0.57 ± 0.02
q^2	0.71	0.71	0.70	0.69	0.68	0.65	0.68	0.70	0.68	0.62	0.68 ± 0.02
RMSEP	0.51	0.50	0.43	0.48	0.39	0.43	0.47	0.39	0.40	0.35	0.43 ± 0.04
r^2_{pred}	0.78	0.67	0.78	0.79	0.77	0.75	0.82	0.87	0.75	0.81	0.78 ± 0.04
$r^2_{m\ (test)}$	0.72	0.59	0.70	0.79	0.77	0.59	0.60	0.77	0.69	0.65	0.69 ± 0.06
Avg. r^2_m	0.70	0.57	0.62	0.72	0.70	0.67	0.76	0.80	0.67	0.73	0.70 ± 0.05
Δr^2_m	0.15	0.06	0.17	0.13	0.03	0.03	0.01	0.10	0.07	0.13	0.09 ± 0.05
CCC	0.85	0.82	0.82	0.84	0.85	0.85	0.87	0.90	0.86	0.89	0.85 ± 0.02
MAE	0.42	0.42	0.38	0.43	0.30	0.34	0.42	0.35	0.35	0.27	0.37 ± 0.05

Nº. of selected descriptors: 7. [a] Test sets: Run 1: 1; 8; 11; 14; 22; 24; 28; 31; 35; Run 2: 1; 5; 10; 15; 18; 19; 29; 31; 38; Run 3: 3; 4; 7; 9; 14; 21; 24; 33; 39; Run 4: 3; 9; 21; 22; 23; 28; 32; 33; 39; Run 5: 1; 6; 8; 10; 22; 32; 33; 35; 38; Run 6: 1; 5; 15; 20; 25; 27; 29; 34; 39; Run 7: 7; 9; 12; 13; 23; 31; 36; 37; 39; Run 8: 6; 8; 10; 17 ; 19; 25; 28; 36; 37; Run 9: 1; 4; 8; 12; 17; 19; 23; 32; 34; Run 10: 16; 17; 20; 23; 28; 34; 35; 38; 39.

Consequently, 37 candidate compounds from molecular docking and ADMET stages were evaluated using the QSAR models (**Table 4**), the calculation of the descriptors for the training set can

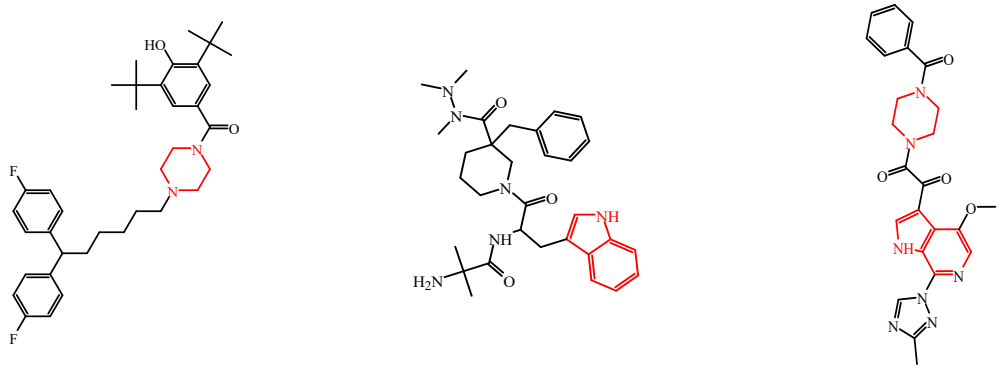
be found in supplementary materials (**Table S4**). Synergistic effects of substituents were observed for three proposals, with calculated pEC_{50} values higher than the experimental value obtained for most compounds in the training set. Notably, compounds PP1, PR1 (Anamorelin), and PR2 (Temsavir), whose structures are shown in **Figure 6**, exhibited pEC_{50} values above 8.9, indicating promising potential for further development.

Table 4: Calculated descriptors for all proposals and the predicted values of pEC_{50}.

Proposal	HA-MP Charge	Rot Bonds	HA-MW (g/mol)	NV electrons	HB Donor	log S (ESOL)	TPSA (Å^2)	Predicted pEC50
PP1	-0.0750	10	459.327	188	1	-5.00	70.08	9.686 ± 0.23
PP2	-0.0492	10	440.329	182	1	-4.83	70.08	7.897 ± 0.06
PP3	-0.0395	10	424.334	178	1	-4.99	71.57	7.778 ± 0.14
PP4	-0.0539	9	433.313	182	0	-4.85	53.09	8.460 ± 0.06
PP5	-0.0857	5	457.211	170	1	-5.94	48.71	7.922 ± 0.48
PP6	-0.0492	7	372.251	154	0	-3.68	68.31	8.346 ± 0.11
PP7	-0.0645	6	385.725	154	1	-4.54	44.81	7.698 ± 0.22
PP8	-0.0683	6	423.269	168	1	-3.71	74.49	7.659 ± 0.26
PP9	-0.0488	10	459.327	188	1	-5.00	70.08	7.458 ± 0.08
PP10	-0.0555	6	390.293	166	0	-4.32	45.25	8.142 ± 0.10
PP11	-0.0406	6	364.275	156	1	-4.54	53.01	7.233 ± 0.12
PP12	-0.0515	6	362.250	150	1	-4.24	43.78	7.241 ± 0.09
PP13	-0.0529	6	357.259	148	1	-3.81	44.81	7.198 ± 0.12
PP14	-0.0465	3	316.231	136	0	-3.79	42.01	7.674 ± 0.16
PP15	-0.0405	6	340.253	144	1	-3.86	53.01	7.082 ± 0.10
PP16	-0.0515	3	336.287	136	0	-3.23	66.07	7.595 ± 0.14
PP17	-0.0481	6	362.250	148	0	-4.30	40.62	7.439 ± 0.07
PP18	-0.0514	4	400.312	174	0	-4.19	47.10	7.309 ± 0.15
PP19	-0.0405	2	436.341	192	2	-5.80	64.01	5.739 ± 0.25
PP20	-0.0480	5	402.308	166	0	-4.10	61.68	7.103 ± 0.10
PP21	-0.0343	6	314.239	134	1	-2.95	48.71	6.316 ± 0.13
PP22	-0.0255	7	338.261	144	0	-4.07	45.67	6.749 ± 0.26
PP23	-0.0341	5	330.238	144	0	-3.17	45.25	6.683 ± 0.18
PP24	-0.0309	6	364.275	150	1	-4.41	53.01	5.877 ± 0.12
PP25	-0.0197	8	352.264	150	0	-4.34	42.01	6.302 ± 0.30
PP26	-0.0360	5	350.276	144	2	-4.18	64.26	5.879 ± 0.11
PR1	-0.0766	9	504.380	214	3	-4.50	114.77	9.089 ± 0.27
PR2	-0.0691	5	450.309	178	1	-3.81	126.31	8.902 ± 0.24
PR3	-0.0613	9	491.314	200	3	-5.83	54.53	6.067 ± 0.34
PR4	-0.0589	6	534.229	200	3	-5.39	122.47	5.189 ± 0.29
PR5	-0.0602	5	424.338	170	1	-5.22	85.15	8.329 ± 0.15
PR6	-0.0582	6	534.229	200	3	-5.39	122.47	5.123 ± 0.28
PR7	-0.0339	8	314.192	128	5	0.08	145.51	5.062 ± 0.18
PR8	-0.0457	4	466.318	182	1	-4.94	86.37	5.394 ± 0.10
PR9	-0.0569	6	424.356	166	1	-3.66	90.99	7.094 ± 0.14
PR10	-0.0314	3	312.247	126	2	-3.76	61.02	5.157 ± 0.14
PR11	-0.0802	5	523.301	198	0	-5.44	88.19	7.984 ± 0.37

A similarity analysis was also performed for the training set and the CHEMBL-derived proposed compounds that underwent docking. Based on the Tanimoto coefficient, the test set showed an average similarity of 0.431, while the proposed compounds, in relation to the training set, had an average similarity of 0.263. Focusing on the top three candidates – PP1, PR1, and PR2 – the average similarity values were 0.315, 0.325, and 0.264, respectively. The heatmap of the Tanimoto coefficients can be found in supplementary material, in **Figure S9**.

To deepen the study on the three most prominent compounds found in MLR, we made an interaction analysis of those compounds in the binding site of 5-HT_{1A} receptor (**Figure 7**) to compare with Buspirone (commercial reference) and 9_24C (highest activity reference). Compound PP1 (**Figure 7-a**) displays a π-anion interaction with residue Asp246, suggesting a relevant electrostatic contribution between the ligand's aromatic system and the amino acid's negative charge. Simultaneously, residue Asn516 forms a conventional hydrogen bond - a strong, directional interaction that contributes to precise ligand orientation in the binding site [61]. Beyond these core interactions, several π-alkyl contacts are observed with hydrophobic residues (Ala223 and Tyr226), which contribute to conformational fitting and complex stability through the piperazine moiety. Notably, a halogen bond with Phe242 is also present, which may have implications for molecular selectivity.



PP1 **PR1** **PR2**

Figure 6: Structures of the three most promising proposals

For compound PR1 (**Figure 7-b**), an analogous π-anion interaction with Asp246 is observed in the indolic portion, reinforcing the conservation of this electrostatic affinity with the negatively charged region of the active site, similar to PP1. The conventional hydrogen bond interaction with Asn516 is also maintained, demonstrating its importance for ligand recognition and anchoring. Additionally, interactions with residues Gly517 and Cys317 through hydrophobic contacts and van der Waals forces indicate good hydrophobic pocket occupancy. A significant number of π-alkyl interactions with aromatic residues (Tyr226 and Ile319) are particularly noteworthy, suggesting a diverse interaction profile.

Compound PR2 (**Figure 7-c**) maintains the π-anion interaction with Asp246, with a portion extremely similar to the indole group, with the presence of an aromatic heterocyclic nitrogen, confirming a recurring pattern among bioactive compounds targeting this receptor. The conventional hydrogen bond with Asn516 is also present, suggesting conservation of binding geometry relative to the previous compounds. Notable additional features include hydrophobic interactions with residues Ala333 and Val247, which enhance effective steric fitting of the ligand with the role of the piperazine group. Furthermore, the presence of π-stacked and π-sulfur interactions may also contribute to complex stability and increased binding affinity.

Regarding Buspirone (**Figure 7-d**), the commercial reference ligand, a non-conventional hydrogen bond with Asp246 is observed in the piperazine moiety. Residue Asn516 also participates through a conventional hydrogen bond, replicating key interactions seen in the candidate compounds. Additional stabilizing interactions include π-sigma and π-alkyl contacts with residues Trp488, Tyr520, and Phe491, though these are fewer in number compared to compounds PR1 and PR2. Finally, analyzing compound 9_24C (**Figure 7-e**), it is possible to visualize again the interaction of Asp246 with the indole

rings and the formation of unconventional hydrogen bonds with the piperazine nucleus, as occurred in buspirone. A conventional hydrogen bond can also be observed with Asn516.

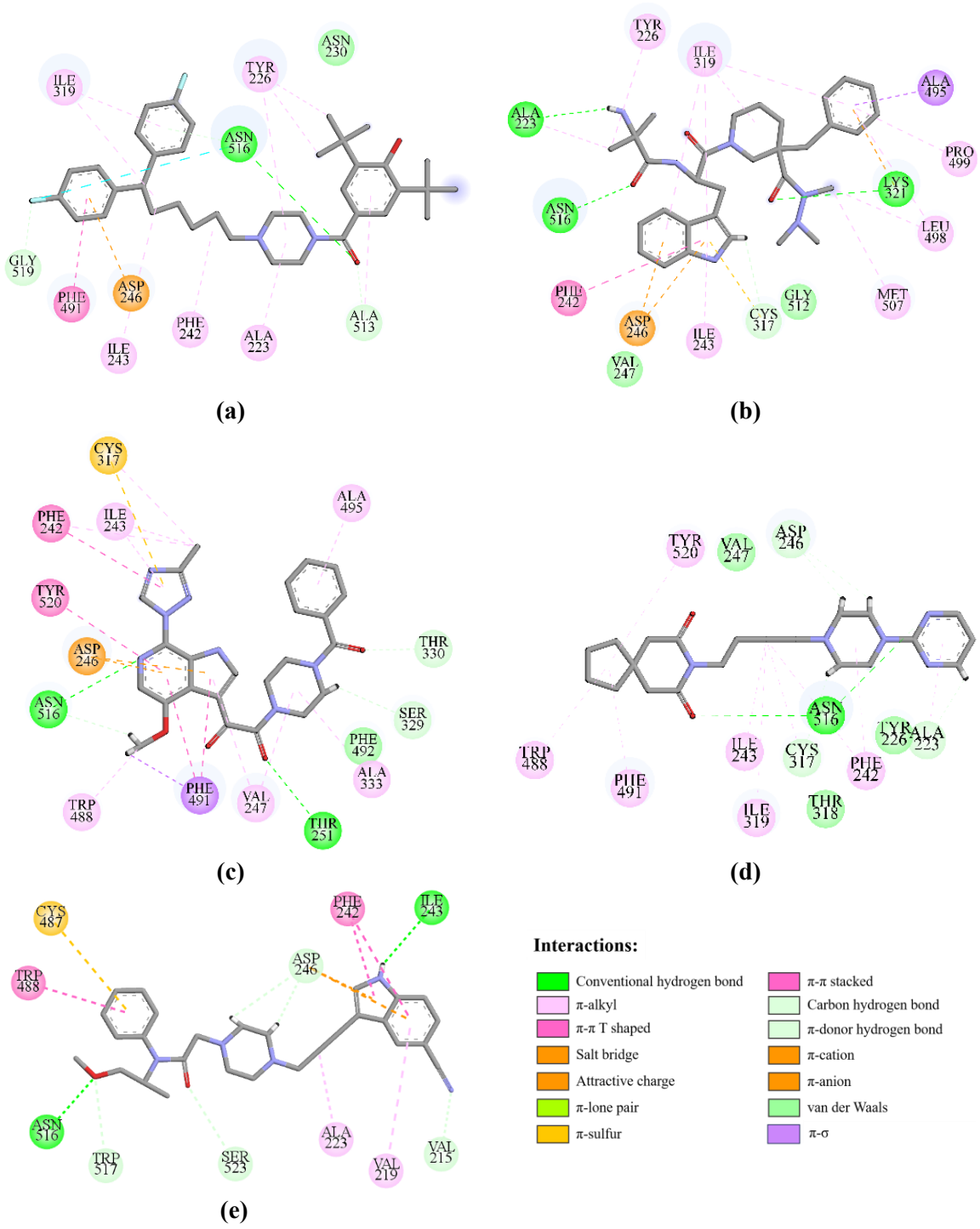


Figure 7: 2D interaction diagrams for the proposals and Buspirone. **(a)** PP1 interaction diagram; **(b)** PR1 interaction diagram; **(c)** PR2 interaction diagram; **(d)** Buspirone interaction diagram; **(e)** 9_24C interaction diagram.

Given the interaction profiles of the lowest-energy proposals with the best predicted pEC_{50} values from MLR-QSAR, a molecular dynamics study was conducted to evaluate the stability of these compounds in the 5-HT_{1A} binding site over time.

Molecular dynamics simulations

Four molecular dynamics simulations were conducted with the 5-HT_{1A} receptor complexed with the proposed compounds PP1, PR1, PR2, and with buspirone as reference. **Figure 8** presents the RMSD, RMSF and hydrogen bond analyses for all complexes. The protein RMSD values (**Figure 8-a**) ranged from 0 to 0.625 nm, with more pronounced fluctuations during the first 50 ns and a stabilization trend after 100 ns of simulation. The buspirone complex showed a brief fluctuation of approximately 0.25 nm between 138-150 ns before returning to an equilibrium RMSD around 0.375 nm. The other complexes stabilized after 100 ns with RMSDs between 0.375-0.625 nm. The RMSF analysis (also in **Figure 7-a**) revealed low fluctuations (0.125-0.375 nm) for residues 165-325 and 425-545, which comprise the ligand binding site. The region between residues 325-425 showed more significant variations, corresponding to flexible loop regions of the protein structure [62].

It is also important to highlight the stabilization of all bound complexes (HOLO), compared to the unbound form (APO). In terms of RMSD, all complexes reached lower values on average, indicating the efficiency of the ligands in the global stabilization of the protein conformation. There was also a reduction in RMSF values in general, but with a focus on residues 200-275, 350-375 and 490-520, the region of the binding site. These data corroborate the fact that those residues have acquired greater conformational freedom without the presence of ligands, which is expected.

Figure 8-(b) displays the RMSD variations for the bound ligands. PP1 and PR2 compounds remained highly stable throughout the simulation, with average RMSD values around 0.05 nm. The PR1 compound showed slightly more fluctuations compared to the other candidates, though still maintaining very low values [63] (never exceeding 0.1 nm). Histogram analysis of RMSD frequencies revealed stable distributions for all three proposed compounds, each showing a single predominant peak at 0.06 nm (PP1), 0.075 nm (PR1) and 0.06 nm (PR2). Buspirone exhibited greater conformational fluctuations, particularly between 125-175 ns, with two frequency peaks at 0.06 nm and 0.08 nm. However, the more pronounced peak at 0.06 nm indicates that despite some conformational variations, buspirone predominantly stabilized at RMSD values near 0.06 nm, suggesting reasonable binding stability. The comparative analysis demonstrates that the proposed compounds PP1, PR1 and PR2 maintain more stable binding modes than the reference ligand Buspirone throughout a first insight of the molecular dynamics simulations.

Finally, **Figure 8-(c)** shows the graphs of the number of hydrogen bonds (HB) formed during the simulation between the ligands and the protein. The PP1 candidate maintained a relatively constant number of hydrogen bonds throughout the simulation, predominantly forming a single hydrogen bond, with short intervals during which two bonds were observed. PR1 exhibited a peak of four hydrogen bonds within the first 50 ns, stabilizing at an average of one hydrogen bond throughout the remainder of the trajectory, with brief fluctuations to two HBs and occasional short intervals with no bonds. PR2, the compound showing the greatest potential for hydrogen bonding, reached the highest peak, forming a total of seven hydrogen bonds around the first 25 ns, and maintained up to five hydrogen bonds in the 25–80 ns interval. From the midpoint of the simulation onward, the number of hydrogen bonds decreased to 2–3, with minor fluctuations indicating the formation of up to four HBs. Buspirone, in comparison to the promising compounds, formed the fewest hydrogen bonds, displaying long intervals (from 30 to 120 ns) without any hydrogen bonding, with only slight peaks reaching two hydrogen bonds. In the final 50 ns of the simulation, Buspirone stabilized with fluctuations between one and two HBs.

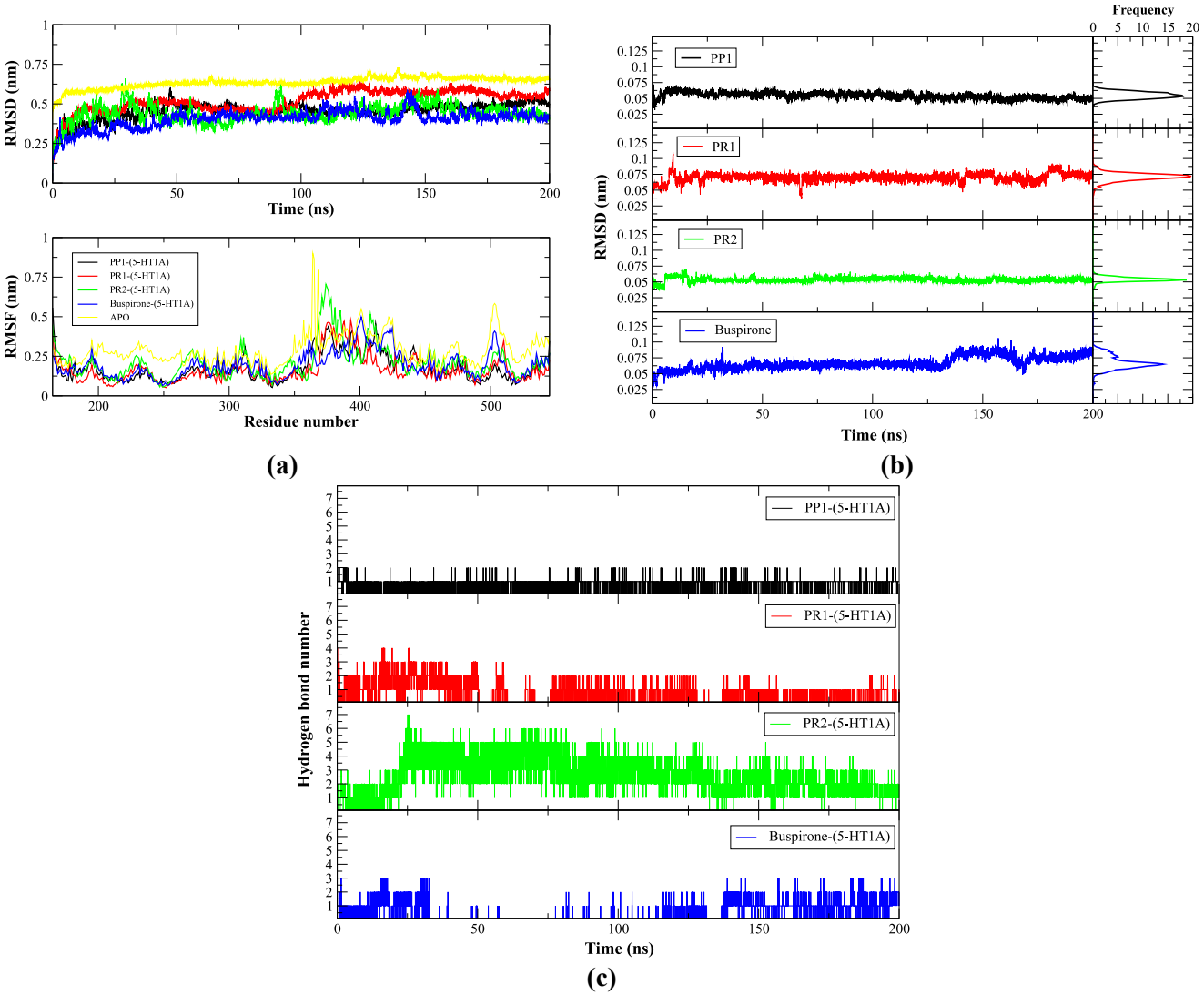


Figure 8: Trajectory calculations from molecular Dynamics. **(a)** RMSD and RMSF for the protein-ligand complexes; **(b)** RMSD values for the ligands complexed with Buspirone; **(c)** Hydrogen bonds formed between the protein and ligands.

The interaction analyses of the molecular docking poses revealed a consistent interaction pattern with two key residues, Asp246 and Asn516. **Figures 9-(a)**, **(c)**, **(e)**, and **(g)** illustrate the three-dimensional interactions of the candidate compounds and Buspirone with these residues, highlighting a strong π-anion interaction pattern between Asp246 and the indole moiety of candidates PR1 and PR2. Asp246 formed two π-anion interactions with both PR1 and PR2, and one with PP1 at the aromatic portion bound to a fluorine atom. Asn516 was responsible for hydrogen bonding in all proposed compounds. In PP1, the hydrogen bond occurred at the carbonyl of the (piperazin-1-yl)methanone core, whereas in PR2, the HB was established through the heterocyclic nitrogen of the indole-like group (pyrrolo-pyridine). In PR1, the hydrogen bond formed outside the axis of the Piperazine-(1,4R) and 1H-indole-(3R) moieties, yet still within a region near the carbonyl group of the piperazin-methanone.

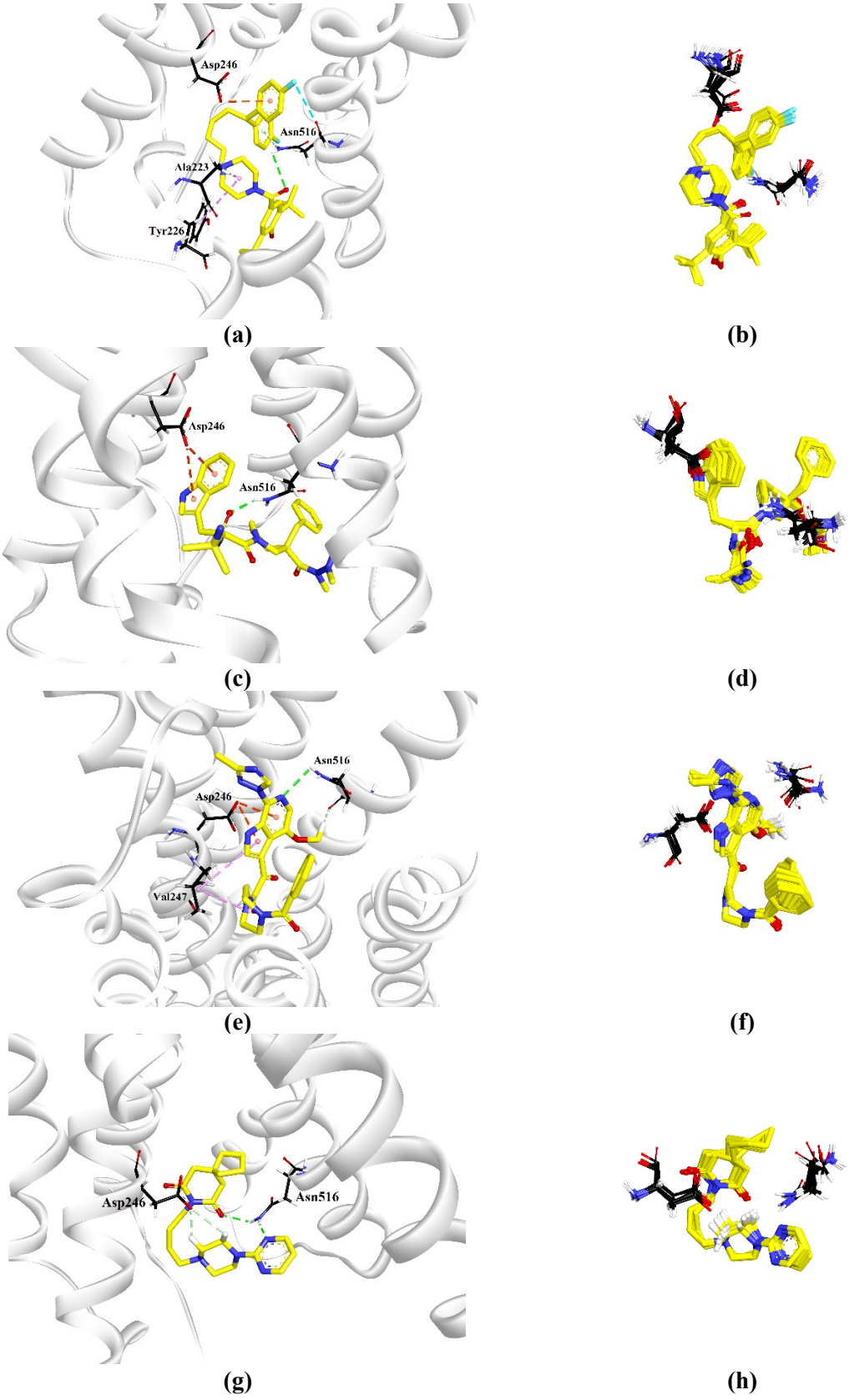


Figure 9: 3D interactions for the ligands with key amino acids and snapshots from MD simulations. **(a)** PP1 interactions; **(b)** Snapshots of PP1, Asp246 and Asn516; **(c)** PR1 interactions; **(d)** Snapshots of PR1, Asp246 and Asn516; **(e)** PR2 interactions; **(f)** Snapshots of P21, Asp246 and Asn516; **(g)** Buspirone interactions; **(h)** Snapshots of Buspirone, Asp246 and Asn516.

For the Buspirone molecule, the interactions with aspartate were weaker compared to the proposed compounds, involving a non-conventional hydrogen bond with the piperazine group. Asparagine, on the other hand, formed two hydrogen bonds: one with the carbonyl group of the piperidine-2,6-dione moiety and another with the aromatic nitrogen of the pyrimidine portion. Overall, in the proposed compounds, the indole group was crucial for the ionic stabilization of the Asp246 residue, engaging in π-stacking interactions involving the core's π-electrons. It also contributed to the formation of hydrogen bonds with Asn516. The piperidine portion of the phenyl-piperazinyl-methanone mainly engaged in hydrophobic interactions, as seen with buspirone, with a focus on the methanone carbonyl, which also exhibited potential for hydrogen bonding with asparagine.

Given the relevance of these two amino acid residues, a distance analysis was performed between the carbonyl oxygen of aspartate (OD2) and the indole and indole-like core carbons C26 and C8 (for PR1 and PR2, respectively), as well as the fluorine-bound aromatic carbon C20 of PP1. For Buspirone, the analysis considered the distance between the hydrogen H7 of the piperazine ring and the OD2 oxygen of the aspartate carbonyl group. For the analysis involving asparagine, the distances between the donor hydrogen (HD22) bonded to the nitrogen and the methanone oxygen atoms of PP1 and PR1 were measured, in addition to the nitrogen N1 of the indole group in PR2 and the O1 oxygen of the piperidine-2,6-dione moiety in Buspirone. Full details of this analysis, including the exact atomic locations, are available in the supplementary material (**Figure S6**).

As a result, for all compounds, interaction distances with Asp246 ranged between 0.3 and 0.6 nm, consistent with π-anion interactions [64], while Buspirone exhibited an average distance of approximately 0.25 nm, coherent with non-conventional hydrogen bonding [65]. As for Asn516, it maintained interaction distances between 0.2 and 0.3 nm with all ligands for most of the trajectory, supporting the formation of stable hydrogen bonds [65], as shown in the molecular docking results. **Figures 9-(b)**, **(d)**, **(f)**, and **(h)** show a total of eight frames extracted from the molecular dynamics simulations during the last 80 ns, illustrating the stability of ligand interactions with residues Asp246 and Asn516, thus supporting the discussion of their intermolecular interactions.

Gibbs free energy landscapes

The Free Energy Landscape (FEL) analysis was conducted using the final 100 ns (100-200 ns) of the molecular dynamics simulations, selected due to the enhanced conformational stability observed in this timeframe as evidenced by RMSD analyses (**Figure 7**). This analysis aimed to identify the most stable conformational states for subsequent free energy calculations. For PP1 (**Figure 10-a**), two distinct energy minima were identified - a deep minimum around (PC1, PC2) = (3.2, -3.5) with an energy of 0.52 kJ/mol indicating high stability, and a shallower minimum at (PC1, PC2) = (1.1, 0.8) with 18.51 kJ/mol representing a less frequently sampled conformation. PR1 (**Figure 10-b**) exhibited a broad energy well centered at (PC1, PC2) = (3.0, -1.0) with 0.88 kJ/mol, along with a less stable conformational state appearing as a funnel-shaped feature at (PC1, PC2) = (1.0, 1.0) with 15.5 kJ/mol. The PR2 system (**Figure 10-c**) displayed a single dominant well at (PC1, PC2) = (1.0, -1.5) with 0.77 kJ/mol, suggesting a predominant stable conformation. Buspirone (**Figure 10-d**) showed two minima - a deep one at (3.0, -1.5) with 0.54 kJ/mol and a shallower state at (-3.5, 3.8) with 19.86 kJ/mol. The APO system resulted in three different energy minimums (**Figure S7**), suggesting that the ligands were crucial for the stabilization of the conformation.

Notably, the most stable conformations across all systems predominantly occurred between 150-200 ns, validating this timeframe as representative of the dominant binding modes (the PCs versus time graphs can be found in supplementary materials, **Figure S7**). These stable conformational states will serve as crucial reference points for subsequent free energy calculations, enabling more accurate estimation of interaction energies and complex stability, considering the fundamental importance of

obtaining high-quality snapshots for PB/GBSA calculations [66,67]. The FEL analysis provides valuable insights into the dynamic behavior of these ligand-receptor complexes, with PR1 and PR2 showing particularly well-defined low-energy states compared to the reference compound buspirone.

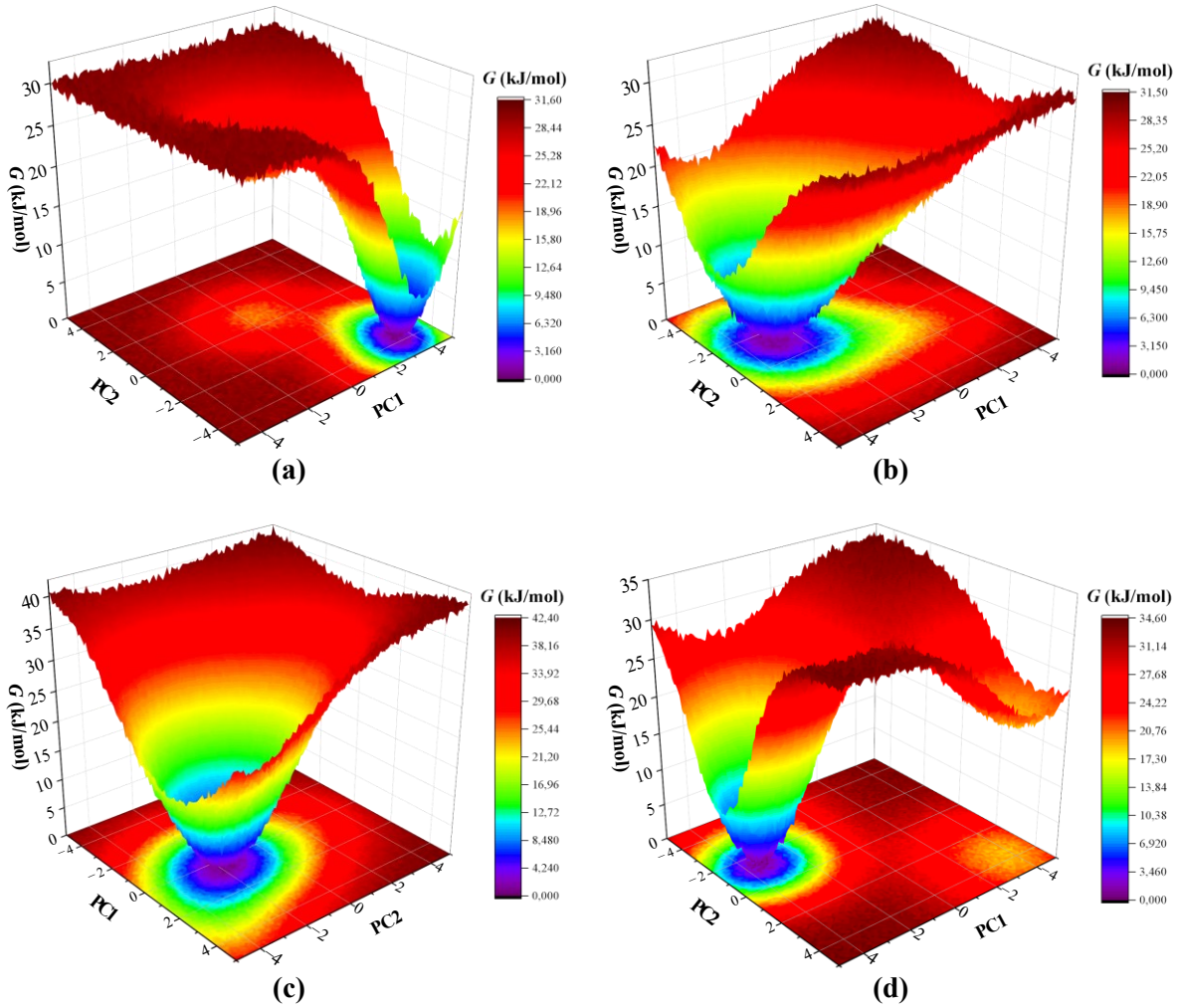


Figure 10: Free energy surfaces of the last 100 ns of simulation. (a) PP1 complex; (b) PR1 complex; (c) PR2 complex; (d) Buspirone complex.

MM-PBSA and QM/MM-GBSA calculations

The binding free energy calculations were performed using the QM/MM-GBSA approach with the PM6-DH^{+} method, considering the ligand in complex and key active site residues as the quantum region. The included residues were: Ala223, Ala333, Ala495, Ala513, Asn516, Asp246, Cys317, Gly519, Ile243, Ile319, Leu498, Lys321, Met507, Phe242, Phe491, Pro499, Ser329, Thr251, Thr330, Trp488, Tyr226, Tyr520, Val247, and Phe492. This selection was based on spatial proximity to the binding site, as verified in molecular docking studies. The results demonstrate clear variations in the energy profiles among the analyzed complexes, as shown in **Table 7**. The graphs for the delta energy terms are available in supplementary material (**Figure S8**).

The PR2 complex exhibited the most favorable thermodynamic profile, with $\Delta G_{binding}$ = -33.61 ± 4.73 kcal/mol. This value was supported by a highly stabilizing electronic contribution (ΔE_{SCF} = -22.77 ± 5.00 kcal/mol) and a strongly negative vacuum energy term (ΔG_{gas} = -50.88 ± 5.33 kcal/mol). Its solvation penalty (ΔG_{solv} = 7.65 ± 1.30 kcal/mol) was the lowest among all complexes, indicating excellent ligand compatibility with the binding site microenvironment. The PP1 system also demonstrated good stability ($\Delta G_{binding}$ = -29.06 ± 3.63 kcal/mol), with significant electronic stabilization

(ΔE_{SCF} = -19.91 ± 3.73 kcal/mol). However, it showed a slightly higher solvation penalty (ΔG_{solv} = 11.45 ± 1.10 kcal/mol), suggesting either greater exposure of polar groups to solvent or less effective binding site dehydration upon complex formation.

Table 7: Energies calculated through QM/MM-PBSA per complex.

Complex with	QM/MM-GBSA energies (kcal/mol)					
	ΔE_{SCF}	ΔG_{gas}	ΔG_{solv}	ΔG_{total}	$-T\Delta S$	$\Delta G_{binding}$
Buspirone	-13.11 ± 4.56	-30.85 ± 4.88	6.95 ± 1.67	-23.90 ± 3.76	9.85 ± 0.05	-14.05 ± 3.76
PP1	-19.91 ± 3.73	-48.00 ± 4.31	11.45 ± 1.10	-36.55 ± 3.63	7.49 ± 0.14	-29.06 ± 3.63
PR1	-7.87 ± 5.55	-37.07 ± 4.64	15.09 ± 1.69	-21.98 ± 4.78	8.01 ± 0.88	-13.97 ± 3.45
PR2	-22.77 ± 5.00	-50.88 ± 5.33	7.65 ± 1.30	-43.23 ± 4.63	9.62 ± 0.95	-33.61 ± 4.73

Buspirone displayed intermediate binding energy ($\Delta G_{binding}$ = -14.05 ± 3.76 kcal/mol), with ΔE_{SCF} = -13.11 ± 4.56 kcal/mol and a less favorable enthalpy term (ΔG_{gas} = -30.85 ± 4.88 kcal/mol). Its entropic energy ($-T\Delta S$ = 9.85 ± 0.05 kcal/mol) remained relatively constant, indicating lower conformational rigidity imposed on the system. The PR1 system showed the least favorable ($\Delta G_{binding}$ = -13.97 ± 3.45 kcal/mol), with the weakest electronic contribution (ΔE_{SCF} = -7.87 ± 5.55 kcal/mol) and highest solvation penalty (ΔG_{solv} = 15.09 ± 1.69 kcal/mol), suggesting suboptimal binding site compatibility. Entropic terms ($-T\Delta S$) showed minor variations between systems, with Buspirone having the highest penalty (9.85 ± 0.05 kcal/mol), followed by PR2 (9.62 ± 0.95 kcal/mol) and PR1 (8.01 ± 0.88 kcal/mol). PP1 showed the lowest entropic penalty (7.49 ± 0.14 kcal/mol). While entropically unfavorable contributions were present in all cases, their relative impact on binding free energy appeared less determinant compared to electronic and solvation effects.

Complementary MM-PBSA energy decomposition analysis (**Table 8**) identified key stabilizing residues: Asn516, Asp246, Phe491, Phe492, Tyr226 and Val247 showed consistently favorable contributions across all systems, with minimum ΔE_{total} values of -5.921 ± 0.510 kcal/mol (PR2) and -4.090 ± 0.281 kcal/mol (PR2), respectively. Phe491 and Phe492 contributed significantly, particularly in the Buspirone complex, while Tyr226 and Val247 showed stronger interactions in PP1, PR1 and PR2 systems, key residues that interacts with piperazine group. Ligand energy contributions varied, being most favorable for PP1 (-41.498 ± 1.645 kcal/mol). Complete decomposition data are available in supplementary materials – **Table S5**.

Table 8: Decomposition energies for the 6 residues with the greatest energetic contributions and ligands.

Residue	ΔE total (kcal/mol)			
	Buspirone	PP1	PR1	PR2
Asn516	-2.121 ± 0.551	-5.273 ± 0.542	-3.291 ± 0.538	-5.921 ± 0.510
Asp246	-2.070 ± 0.276	-3.278 ± 0.271	-1.573 ± 0.304	-4.090 ± 0.281
Phe491	-1.863 ± 0.282	-0.499 ± 0.281	-1.136 ± 0.301	-0.185 ± 0.283
Phe492	-2.046 ± 0.278	-1.157 ± 0.294	-1.100 ± 0.299	-0.034 ± 0.340
Tyr226	-0.264 ± 0.315	-2.973 ± 0.275	-0.039 ± 0.515	-0.804 ± 0.360
Val247	-2.149 ± 0.431	-0.991 ± 0.321	-1.086 ± 0.315	-2.032 ± 0.399
Ligand	-24.239 ± 1.546	-41.498 ± 1.645	-19.678 ± 1.935	-29.779 ± 1.611

DISCUSSION

Considering the importance of the 5-HT1A receptor in modulating neural pathways related to social behavior, anxiety, and cognition—domains often impaired in individuals with autism spectrum disorder (ASD)—as well as the lack of safe and effective pharmacological therapies targeting the core symptoms of this condition, the search for new selective modulators of this receptor holds significant clinical and scientific relevance. In this work, we conducted an extensive search for ligands derived from Piperazine-(1,4R) and 1H-Indole-(3R) with potent agonist potential using a range of computational methodologies. The goal was to guide the drug discovery process and reposition existing compounds, eliminating the initial stages of novel synthesis and clinical trials that take years, thereby accelerating the development of new treatments.

Initially, the crystallographic structure of the human 5-HT_{1A} receptor in complex with buspirone (PDB-ID: 8FYX) was selected as the structural template, as it features a complete, well-defined, and biologically relevant active site conformation, given that buspirone was used as the reference compound. Missing regions, limited to loops and peripheral segments of the binding site, were reconstructed using comparative modeling with SwissModel, employing the continuous chain of the structure itself and secondary homologous templates as references. The reconstruction was restricted to segments outside the active site to avoid modeling artifacts in the critical binding region. The final model was validated by stereochemical analysis (Ramachandran plot), with less important residues in sterically disallowed regions, and local quality assessment (QMEANDisCo = 0.70), deemed suitable for subsequent docking, QSAR, and molecular dynamics steps. Although the global topology of the model was restored, loop regions exhibit intrinsic flexibility that may not have been fully captured, which could indirectly affect aspects of global dynamics, such as solvation and stability of adjacent domains [68]. To minimize these effects, analyses and subsequent free energy calculations focused on the binding site and its immediate vicinity.

The accuracy of the docking protocol was confirmed by two key metrics: redocking (RMSD = 0.82 and 1.10 Å), as structural validation, and an AUC-ROC of 0.823, demonstrating high discriminative power between active and inactive compounds [69]. These results validate the use of the protocol for virtual screening of 5-HT_{1A} receptor ligands, considering also the correlation of the MolDock score with the experimental pEC_{50} values of active ligands. However, it is important to acknowledge that, despite its high performance, docking remains a conformationally fixed receptor-based approach [70] that does not fully account for protein flexibility, solvent rearrangement, or conformational entropy. Thus, the results should be interpreted as initial indicators of affinity rather than absolute predictions, requiring further validation through molecular dynamics and free energy methods, as performed in this work.

The MLR-QSAR model was built from descriptors selected by a genetic algorithm, prioritizing variables with physicochemical and pharmacological relevance. The mean partial charge of heavy atoms stood out for capturing electronic features associated with interactions at the 5-HT_{1A} receptor active site, such as electrostatic interactions and hydrogen bonds [71]. Other descriptors, such as log *S*, TPSA, and the number of hydrogen bond donors, reflect hydrophilic and permeability properties, which correlate with blood-brain barrier penetration [72], while geometric-electronic variables complement the structural encoding of the compounds. The model showed $q^2 = 0.723$, $r^2_{pred} = 0.751$, CCC = 0.789, and MAE = 0.491, indicating robust predictive performance. The initial data split using the Kennard-Stone method, combined with 10 bootstrap iterations, confirmed statistical stability across different training and test subsets [73]. Williams plot analysis allowed assessment of the model's applicability domain and identified a single compound as a potential outlier. Nevertheless, its removal did not significantly alter global statistical parameters, justifying its retention in the final dataset. However, despite the model's good correlation and internal stability, its applicability remains restricted to the chemical space

originally sampled (Piperazine-(1,4R) and 1H-Indole-(3R) derivatives) and does not replace additional validations for structures outside this scope.

ADMET analyses were employed to minimize the initial set generated by ChEMBL and screen molecules with optimal pharmacological properties according to the 5-HT_{1A} receptor's molecular environment. The BBB permeability descriptor was used as the primary screening criterion, given that 5-HT_{1A} receptors are predominantly located in central nervous system neurons. Efficient blood-brain barrier crossing is essential for candidate compounds to reach the binding site at therapeutic concentrations. Subsequent filters—such as oral bioavailability, cytochrome P450 3A4 inhibition (the most abundant in hepatocytes), hepatotoxicity (common in 5-HT receptor agonists [74]), Lipinski's rules (as a strong estimate of good pharmacokinetics [75]), Brenk filters (for good toxicological parameters [22]), and PAINS (to exclude potential promiscuous compounds [76]) – drastically reduced the number of compounds. Applying multiple filters may significantly narrow down candidate molecules, even eliminating some with potential biological activity that do not strictly meet the established criteria. However, this rigorous approach increases the likelihood of selecting compounds with better pharmacokinetic and toxicological profiles, greater specificity, and lower risk of off-target effects or undesirable properties [77]. Although the resulting set is quantitatively smaller, it is qualitatively superior, particularly in computational studies aimed at molecular and energetic understanding preceding experimental assays.

Molecular dynamics simulations revealed adequate conformational stability based on RMSD and RMSF values, with stable fluctuations for all analyzed ligands. A key finding from the MD simulations was the number of hydrogen bonds formed by PR2 (5-7 HBs) compared to buspirone (0-2 HBs), which was reflected in greater stability according to subsequent free energy calculations. OPLS-AA simulations were validated [78] and representative over a 200 ns duration, as previously done for receptors of the same family [79,80]. However, inherent limitations of classical force fields remain, such as the inability to capture quantum effects or subtle electronic changes during simulations. To address this, we extended our analyses to hybrid quantum-mechanical methods in free energy calculations. Subsequent free energy landscape (FEL) calculations were crucial in identifying time intervals corresponding to low-energy snapshots, rationalizing QM/MM-GB/PBSA calculations. For PP1 and buspirone complexes, where two free energy wells were clearly observed, the lowest-energy conformations were separated by energy barriers of 17.99 and 19.32 kJ/mol, indicating a predominance of conformations restricted to the deeper wells. Additionally, PR1 and PR2 exhibited single wells in the last 100 ns, corroborating RMSD results, which showed minimal fluctuations.

Free energy calculations were performed to support molecular dynamics findings, confirming structural stability. Active site residues were selected based on docking intermolecular interaction results, considering all residues in direct contact with ligands to compute possible immediate electronic interactions. The semi-empirical Hamiltonian PM6-DH+ was chosen for its parametrization focusing on dispersion corrections relevant to hydrogen bonding [81], extendable to π-stacking and π-ion interactions – essential for this system, given recurrent hydrogen bonding with Asn516, π-ion interactions with Asp246, and the presence of aromatic residues (phenylalanine, tryptophan, and tyrosine). The electronic interaction energy was crucial for PR2, the lowest free energy proposal (ΔG = -33.61 ± 4.73 kcal/mol), revealed by the self-consistent field energy (ΔE_{SCF}) of -22.77 kcal/mol. For PP1, binding free energy was strongly influenced by ΔG_{gas}, indicating complex formation stability with relatively low solvation penalty. In contrast, PR1, the lowest performer in QM/MM-GBSA calculations, exhibited a more pronounced solvation penalty, destabilizing the complex in solvent contact. Despite limitations of implicit solvent methods like GBSA, buspirone's inclusion helped validate energies, as its calculated value (-14.05 ± 3.76 kcal/mol, with an internal dielectric constant of 3) aligned with experimental K_i = 3.9 nM [82], converting energy ($\Delta G = RT \ln K_i$) to $\Delta G \approx$ -12.00 kcal/mol – remarkably close, considering methodological constraints.

Another key MD finding, confirmed by MM-PBSA energy decomposition, was the influence of Asn516 and Asp246. Conformational stabilization of these residues, assessed via interaction distance analysis, corresponded to prolonged hydrogen bonding and π-anion interactions. Interaction energies for these residues were evident across all analyzed complexes, ranging from -2.1 to -5.9 kcal/mol for Asn516 and -1.6 to -4.0 kcal/mol for Asp246 (the lowest values). These findings align with prior studies [83,84], initially reported in 1992 and 1993 but now focused on the active site. Computational evidence strongly underscores the importance of conserving these residues for future optimizations. As an immediate biological consequence, the phenyl-piperazin-methanone group in PP1 enabled an extremely stable hydrogen bond with Asn516, while the indole and pyrrolo-pyrimidine groups (structurally similar to indole, with an added nitrogen) in PR1 and PR2 were strongly stabilized by π-anion interactions with Asp246. These results may guide future designs of novel molecules with similar groups that effectively interact with these residues, enhancing agonist potential.

Additionally, PR2's superior interaction energy compared to buspirone may stem from replacing unconventional hydrogen bonds with Asp246 with stronger π-anion interactions involving the aromatic core, alongside conventional hydrogen bonding with Asn516. PR1 contains amides, ureas, and a fused aromatic ring (e.g., indole), favoring π-π stacking and hydrogen bonding in the active cavity. In contrast, buspirone's simpler structure and low functional density establish fewer simultaneous interactions, reducing its anchoring and stability. Thus, PR1 and PR2 exhibit greater receptor complementarity and potential for forming more stable, selective complexes.

Finally, Anamorelin (PP1) is a selective ghrelin receptor (GHS-R1a) agonist with potent orexigenic and anabolic activity. It promotes appetite stimulation, body weight gain, and lean mass increase, even indicated for cancer-associated cachexia [85]. It is currently in Phase 3 clinical evaluation. Temsavir, an HIV-1 entry inhibitor, binds viral envelope glycoprotein gp120, blocking its interaction with CD4 and CCR5/CXCR4 co-receptors, thus preventing host-cell fusion. It is used in combination with other antiretrovirals for multidrug-resistant HIV-1 infections [86] and is in Phase 2 clinical development. Repurposing these compounds for autism could revolutionize treatment for this condition, with high-impact medical implications.

Furthermore, subsequent experimental studies are essential to validate the biological efficacy of the proposals raised here, as bioactive compound discovery requires multidisciplinary teamwork. Theoretical studies like this shed light on molecular features underlying potent activity, identify critical residues, and provide energetic validation of key intermolecular interactions. Autism continues to affect millions worldwide, with no safe, efficient pharmacological alternatives to mitigate its effects.

CONCLUSIONS

This study investigated the applicability of integrated computational methodologies for identifying selective agonists of the serotoninergic 5-HT_{1A} receptor with therapeutic potential for autism spectrum disorder (ASD) treatment, as well as key intermolecular interactions at the binding site. Given the absence of specific pharmacological treatments for ASD core symptoms and the growing body of evidence indicating the modulatory role of the 5-HT_{1A} receptor in social and repetitive behaviors, this work proposed a rational drug repurposing approach based on Piperazine-(1,4R) and 1H-Indole-(3R) derived models.

Structural reconstruction of the human 5-HT_{1A} receptor from crystallographic data (PDB ID: 8FYX), followed by extensive stereochemical and functional validation steps, provided a reliable model for molecular docking, QSAR modeling, and molecular dynamics studies. The docking protocol demonstrated high accuracy (RMSD - redocking = 0.82 Å) and discriminative power (AUC-ROC = 0.823), while the QSAR model showed statistical robustness and adequate predictive performance (q^2 = 0.723; r^2_{pred} = 0.751). Candidate compound selection was further refined through rigorous

pharmacokinetic and toxicological filters (ADMET, Lipinski, Brenk, PAINS), resulting in molecules with properties compatible with bioavailability and central nervous system penetration.

Molecular dynamics simulations revealed ligand-receptor complex conformational stability, identifying three top candidates: PP1 (CHEMBL-ID 1963600), a preclinical-stage ligand; PP2 (Anamorelin), a ghrelin receptor agonist currently used for appetite stimulation and weight gain in cachexia treatment; and notably PR2 (Temsavir), an advanced-stage antiretroviral that demonstrated superior structural and energetic stability. This observation was corroborated by QM/MM-GBSA free energy analyses, where PR2 exhibited the lowest total free energy (ΔG = -33.61 ± 4.73 kcal/mol), emerging as the compound with optimal receptor interaction profile. Critical residue involvement (Asn516 and Asp246) was observed, along with the essential role of indole groups combined with piperazine moieties.

Thus, this work not only reinforces the 5-HT1A receptor's viability as a pharmacological target for ASD treatment but also demonstrates that integrated in silico approaches can significantly accelerate therapeutic candidate discovery and optimization. Temsavir (PR2) in particular emerges as a promising molecule for future experimental validation, consolidating molecular modeling's role as a strategic tool in rational neuropharmacology.

Author Contributions: Conceptualization, A. P. L. M.; formal analysis, E. F. F. C, M. P. F. and T. C. R.; writing-original draft preparation, A. P. L. M. and A. C. F.; writing—review and editing, A. P. L. M., A. C. F., E. F. F. C., M. P. F. and T. C. Ramalho; supervision, E. F. F. C and M. P. F.; project administration, T. C. Ramalho;

Funding: This work was funded by CAPES (001), CNPq (36-/2014-9), FAPEMIG (PPM-00831-15), and the University of Hradec Kralove (Faculty of Science, VT2019-2021).

Data Availability Statement: Data is contained within the article and supplementary material.

Ethics Statement: The authors declare that this study did not involve any experiments with human participants or animals. All research presented herein is entirely theoretical and computational in nature, conducted using in silico methods and simulations. Therefore, no ethical approval was required.

Statement of Usage of Artificial Intelligence: The authors declare that OpenAI's ChatGPT (version GPT-4, accessed via ChatGPT in 2025) was used solely to assist in translating the manuscript from Portuguese to English and for grammar refinement during the revision process. The content, scientific analysis, and interpretations are entirely the work of the authors.

Acknowledgments: The authors are thankful to CNPq, CAPES, FAPEMIG, Federal University of Lavras and University of Hradec Kralove.

Conflicts of Interest: The authors declare no conflicts of interest

REFERENCES

1. Albert, P.R.; Le François, B.; Millar, A.M. Transcriptional Dysregulation of 5-HT1A Autoreceptors in Mental Illness. *Mol Brain* **2011**, *4*, 21, doi:10.1186/1756-6606-4-21.
2. Newman-Tancredi, A.; Depoortère, R.Y.; Kleven, M.S.; Kołaczkowski, M.; Zimmer, L. Translating Biased Agonists from Molecules to Medications: Serotonin 5-HT1A Receptor Functional Selectivity for CNS Disorders. *Pharmacol Ther* **2022**, *229*, 107937, doi:https://doi.org/10.1016/j.pharmthera.2021.107937.

3. Polter, A.M.; Li, X. 5-HT1A Receptor-Regulated Signal Transduction Pathways in Brain. *Cell Signal* **2010**, *22*, 1406–1412, doi:https://doi.org/10.1016/j.cellsig.2010.03.019.
4. Garcia-Garcia, A.L.; Newman-Tancredi, A.; Leonardo, E.D. P5-HT1A Receptors in Mood and Anxiety: Recent Insights into Autoreceptor versus Heteroreceptor Function. *Psychopharmacology (Berl)* **2014**, *231*, 623–636, doi:10.1007/s00213-013-3389-x.
5. Kurahashi, H.; Kunisawa, K.; Tanaka, K.F.; Kubota, H.; Hasegawa, M.; Miyachi, M.; Moriya, Y.; Hasegawa, Y.; Nagai, T.; Saito, K.; et al. Autism Spectrum Disorder-like Behaviors Induced by Hyper-Glutamatergic NMDA Receptor Signaling through Hypo-Serotonergic 5-HT1A Receptor Signaling in the Prefrontal Cortex in Mice Exposed to Prenatal Valproic Acid. *Neuropsychopharmacology* **2025**, *50*, 739–750, doi:10.1038/s41386-024-02004-z.
6. Kondaurova, E.M.; Belokopytova, I.I.; Kulikova, E.A.; Khotskin, N. V; Ilchibaeva, T. V; Tsybko, A.S.; Popova, N.K.; Naumenko, V.S. On the Role of Serotonin 5-HT1A Receptor in Autistic-like Behavior: Cross Talk of 5-HT and BDNF Systems. *Behavioural Brain Research* **2023**, *438*, 114168, doi:https://doi.org/10.1016/j.bbr.2022.114168.
7. Saitow, F.; Takumi, T.; Suzuki, H. Upregulated 5-HT1A Receptor-Mediated Currents in the Prefrontal Cortex Layer 5 Neurons in the 15q11–13 Duplication Mouse Model of Autism. *Mol Brain* **2020**, *13*, 115, doi:10.1186/s13041-020-00655-9.
8. Nyrenius, J.; Waern, M.; Eberhard, J.; Ghaziuddin, M.; Gillberg, C.; Billstedt Background, E. Autism in Adult Psychiatric Out-Patients: Self-Reported Suicidal Ideation, Suicide Attempts and Non-Suicidal Self-Injury. *BJPsych Open* **2023**, *9*, e167, doi:10.1192/BJO.2023.553.
9. Hammond, R.K.; Hoffman, J.M. Adolescents With High-Functioning Autism: An Investigation of Comorbid Anxiety and Depression. *J Ment Health Res Intellect Disabil* **2014**, *7*, 246–263, doi:10.1080/19315864.2013.843223.
10. Lee, S.-T.; Park, J.-H.; Kim, M. Efficacy of the 5-HT1A Agonist, Buspirone Hydrochloride, in Migraineurs With Anxiety: A Randomized, Prospective, Parallel Group, Double-Blind, Placebo-Controlled Study. *Headache: The Journal of Head and Face Pain* **2005**, *45*, 1004–1011, doi:https://doi.org/10.1111/j.1526-4610.2005.05181.x.
11. Hrabak, M.; Ahmed, R.; Soriano, M.G.; Powell, A.; Thanos, P.K.; Steiner, H. Vilazodone, a Novel SSRI Antidepressant with 5-HT1A Partial Agonist Properties: Diminished Potentiation of Chronic Oral Methylphenidate-Induced Dynorphin Expression in the Striatum in Adolescent Male Rats. *Mol Neurobiol* **2025**, *62*, 4520–4532, doi:10.1007/s12035-024-04569-8.
12. Zhu, C.; Li, X.; Peng, W.; Fu, W. Discovery of Novel Indolealkylpiperazine Derivatives as Potent 5-HT1A Receptor Agonists for the Potential Future Treatment of Depression. *Molecules* **2020**, *25*, doi:10.3390/molecules25215078.
13. Lian, P.; Li, L.; Geng, C.; Zhen, X.; Fu, W. Higher-Affinity Agonists of 5-HT1AR Discovered through Tuning the Binding-Site Flexibility. *J Chem Inf Model* **2015**, *55*, 1616–1627, doi:10.1021/acs.jcim.5b00164.
14. Wang, W.; Zheng, L.; Li, W.; Zhu, C.; Peng, W.; Han, B.; Fu, W. Design, Synthesis, and Structure–Activity Relationship Studies of Novel Indolyalkylpiperazine Derivatives as Selective 5-HT1A Receptor Agonists. *J Chem Inf Model* **2020**, *60*, 235–248, doi:10.1021/acs.jcim.9b00926.
15. Xu, L.; Zhou, S.; Yu, K.; Gao, B.; Jiang, H.; Zhen, X.; Fu, W. Molecular Modeling of the 3D Structure of 5-HT1AR: Discovery of Novel 5-HT1AR Agonists via Dynamic Pharmacophore-Based Virtual Screening. *J Chem Inf Model* **2013**, *53*, 3202–3211, doi:10.1021/ci400481p.
16. Pinzi, L.; Bisi, N.; Rastelli, G. How Drug Repurposing Can Advance Drug Discovery: Challenges and Opportunities. *Frontiers in Drug Discovery* **2024**, *Volume 4-2024*, doi:10.3389/fddsv.2024.1460100.
17. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res* **2018**, *46*, W296–W303, doi:10.1093/nar/gky427.
18. Bienert, S.; Waterhouse, A.; De Beer, T.A.P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-New Features and Functionality. *Nucleic Acids Res* **2017**, *45*, D313–D319, doi:10.1093/NAR/GKW1132.
19. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J.P. ChEMBL Web Services: Streamlining Access to Drug Discovery Data and Utilities. *Nucleic Acids Res* **2015**, *43*, W612, doi:10.1093/NAR/GKV352.

20. Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E.J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Lopez, D.M.; Mosquera, J.F.; et al. The ChEMBL Database in 2023: A Drug Discovery Platform Spanning Multiple Bioactivity Data Types and Time Periods. *Nucleic Acids Res* **2024**, *52*, D1180–D1192, doi:10.1093/NAR/GKAD1004.
21. Myung, Y.; De Sá, A.G.C.; Ascher, D.B. Deep-PK: Deep Learning for Small Molecule Pharmacokinetic and Toxicity Prediction. *Nucleic Acids Res* **2024**, *52*, W469–W475, doi:10.1093/NAR/GKAE254.
22. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Scientific Reports 2017 7:1* **2017**, *7*, 1–13, doi:10.1038/srep42717.
23. Lipinski, C.A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov Today Technol* **2004**, *1*, 337–341, doi:10.1016/J.DDTEC.2004.11.007.
24. Baell, J.B.; Nissink, J.W.M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chem Biol* **2018**, *13*, 36–44, doi:10.1021/acschembio.7b00903.
25. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J Med Chem* **2012**, *55*, 6582–6594, doi:10.1021/JM300687E/SUPPL_FILE/JM300687E_SI_004.TXT.
26. Jiroušková, Z.; Vařeková, R.S.; Vaněk, J.; Koča, J. Electronegativity Equalization Method: Parameterization and Validation for Organic Molecules Using the Merz-Kollman-Singh Charge Distribution Scheme. *J Comput Chem* **2009**, *30*, 1174–1178, doi:10.1002/JCC.21142.
27. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An Open Chemical Toolbox. *J Cheminform* **2011**, *3*, 1–14, doi:10.1186/1758-2946-3-33/TABLES/2.
28. Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker, N.A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res* **2007**, *35*, W522–W525, doi:10.1093/nar/gkm276.
29. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25*, 1157–1174.
30. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J Chem Theory Comput* **2011**, *7*, 525–537, doi:10.1021/ct100578z.
31. Thomsen, R.; Christensen, M.H. MolDock: A New Technique for High-Accuracy Molecular Docking. *J Med Chem* **2006**, *49*, 3315–3321, doi:10.1021/JM051197E/SUPPL_FILE/JM051197ESI20060314_081922.PDF.
32. Nunes, C.A.; Freitas, M.P.; Pinheiro, A.C.M.; Bastos, S.C. Chemoface: A Novel Free User-Friendly Interface for Chemometrics. *J Braz Chem Soc* **2012**, *23*, 2003–2010, doi:10.1590/S0103-50532012005000073.
33. Ambure, P.; Aher, R.B.; Gajewicz, A.; Puzyn, T.; Roy, K. "NanoBRIDGES" Software: Open Access Tools to Perform QSAR and Nano-QSAR Modeling. *Chemometrics and Intelligent Laboratory Systems* **2015**, *147*, 1–13, doi:https://doi.org/10.1016/j.chemolab.2015.07.007.
34. Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *J Chem Inf Model* **2011**, *51*, 2320–2335, doi:10.1021/ci200211n.
35. Roy, K.; Kar, S.; Ambure, P. On a Simple Approach for Determining Applicability Domain of QSAR Models. *Chemometrics and Intelligent Laboratory Systems* **2015**, *145*, 22–29, doi:https://doi.org/10.1016/j.chemolab.2015.04.013.
36. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137–148, doi:10.1080/00401706.1969.10490666.
37. Geladi, P.; Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal Chim Acta* **1986**, *185*, 1–17, doi:https://doi.org/10.1016/0003-2670(86)80028-9.
38. Mitra, I.; Saha, A.; Roy, K. Exploring Quantitative Structure–Activity Relationship Studies of Antioxidant Phenolic Compounds Obtained from Traditional Chinese Medicinal Plants. *Mol Simul* **2010**, *36*, 1067–1079, doi:10.1080/08927022.2010.503326.
39. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *J Chem Inf Model* **2009**, *49*, 1669–1678, doi:10.1021/ci900115y.

40. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N. Some Case Studies on Application of "Rm2" Metrics for Judging Quality of Quantitative Structure–Activity Relationship Predictions: Emphasis on Scaling of Response Data. *J Comput Chem* **2013**, *34*, 1071–1082, doi:https://doi.org/10.1002/jcc.23231.
41. Shayanfar, S.; Shayanfar, A. Comparison of Various Methods for Validity Evaluation of QSAR Models. *BMC Chem* **2022**, *16*, 63, doi:10.1186/s13065-022-00856-4.
42. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindah, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25, doi:10.1016/J.SOFTX.2015.06.001.
43. Kaminski, G.A.; Friesner, R.A.; Tirado-Rives, J.; Jorgensen, W.L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J Phys Chem B* **2001**, *105*, 6474–6487, doi:10.1021/jp003919d.
44. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J Phys Chem A* **2001**, *105*, 9954–9960, doi:10.1021/jp003020w.
45. Sousa da Silva, A.W.; Vranken, W.F. ACPYPE - AnteChamber PYthon Parser InterfacE. *BMC Res Notes* **2012**, *5*, 367, doi:10.1186/1756-0500-5-367.
46. Case, D.A.; Aktulga, H.M.; Belfon, K.; Cerutti, D.S.; Cisneros, G.A.; Cruzeiro, V.W.D.; Forouzesh, N.; Giese, T.J.; Götz, A.W.; Gohlke, H.; et al. AmberTools. *J Chem Inf Model* **2023**, *63*, 6183–6191, doi:10.1021/ACS.JCIM.3C01153/ASSET/IMAGES/LARGE/CI3C01153_0002.JPEG.
47. Ribeiro, A.A.S.T.; Horta, B.A.C.; De Alencastro, R.B. MKTOP: A Program for Automatic Construction of Molecular Topologies. *J Braz Chem Soc* **2008**, *19*, 1433–1435, doi:10.1590/S0103-50532008000700031.
48. Nicolaï, A.; Petiot, N.; Grassein, P.; Delarue, P.; Neiers, F.; Senet, P. Free-Energy Landscape Analysis of Protein-Ligand Binding: The Case of Human Glutathione Transferase A1. *Applied Sciences* **2022**, *12*, doi:10.3390/app12168196.
49. Wahbah, M.; Mohandes, B.; EL-Fouly, T.H.M.; El Moursi, M.S. Unbiased Cross-Validation Kernel Density Estimation for Wind and PV Probabilistic Modelling. *Energy Convers Manag* **2022**, *266*, 115811, doi:https://doi.org/10.1016/j.enconman.2022.115811.
50. Valdés-Tresanco, M.S.; Valdés-Tresanco, M.E.; Valiente, P.A.; Moreno, E. Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J Chem Theory Comput* **2021**, *17*, 6281–6291, doi:10.1021/ACS.JCTC.1C00645/ASSET/IMAGES/LARGE/CT1C00645_0005.JPEG.
51. Duan, L.; Liu, X.; Zhang, J.Z.H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *J Am Chem Soc* **2016**, *138*, 5722–5728, doi:10.1021/JACS.6B02682/ASSET/IMAGES/JA-2016-026824_M008.GIF.
52. Tan, C.; Tan, Y.-H.; Luo, R. Implicit Nonpolar Solvent Models. *J Phys Chem B* **2007**, *111*, 12263–12274, doi:10.1021/jp073399n.
53. Kumar, M.; Rathore, R.S. RamPlot: A Webserver to Draw 2D, 3D and Assorted Ramachandran (, ψ) Maps. *J Appl Crystallogr* **2025**, *58*, 630–636, doi:10.1107/S1600576725001669.
54. Tang, Y.; Moretti, R.; Meiler, J. Recent Advances in Automated Structure-Based De Novo Drug Design. *J Chem Inf Model* **2024**, *64*, 1794–1805, doi:10.1021/ACS.JCIM.4C00247.
55. Sharma, A.; Yennamalli, R.M. Docking Strategies. *Basic Biotechniques for Bioprocess and Bioentrepreneurship* **2023**, 243–258, doi:10.1016/B978-0-12-816109-8.00016-7.
56. Kadry, H.; Noorani, B.; Cucullo, L. A Blood–Brain Barrier Overview on Structure, Function, Impairment, and Biomarkers of Integrity. *Fluids Barriers CNS* **2020**, *17*, 69, doi:10.1186/s12987-020-00230-3.
57. Abbod, M.; Safaie, N.; Gholivand, K. Genetic Algorithm Multiple Linear Regression and Machine Learning-Driven QSTR Modeling for the Acute Toxicity of Sterol Biosynthesis Inhibitor Fungicides. *Heliyon* **2024**, *10*, e36373, doi:https://doi.org/10.1016/j.heliyon.2024.e36373.
58. Gramatica, P. Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)* **2020**, *5*, 61–97, doi:10.4018/IJQSPR.20200701.oa1.
59. Golbraikh, A.; Tropsha, A. Beware of Q2! *J Mol Graph Model* **2002**, *20*, 269–276, doi:https://doi.org/10.1016/S1093-3263(01)00123-1.

60. Roy, K.; Mitra, I.; Kar, S.; Ojha, P.K.; Das, R.N.; Kabir, H. Comparative Studies on Some Metrics for External Validation of QSPR Models. *J Chem Inf Model* **2012**, *52*, 396–408, doi:10.1021/ci200520g.
61. Barillari, C.; Taylor, J.; Viner, R.; Essex, J.W. Classification of Water Molecules in Protein Binding Sites. *J Am Chem Soc* **2007**, *129*, 2577–2587, doi:10.1021/ja066980q.
62. Ramanathan, A.; Savol, A.; Burger, V.; Chennubhotla, C.S.; Agarwal, P.K. Protein Conformational Populations and Functionally Relevant Substates. *Acc Chem Res* **2014**, *47*, 149–156, doi:10.1021/ar400084s.
63. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin Drug Discov* **2015**, *10*, 449–461, doi:10.1517/17460441.2015.1032936.
64. Wang, D.-X.; Wang, M.-X. Exploring Anion−π Interactions and Their Applications in Supramolecular Chemistry. *Acc Chem Res* **2020**, *53*, 1364–1380, doi:10.1021/acs.accounts.0c00243.
65. Quyen, L.T.T.; Trung, N.T. Noticeable Characteristics of Conventional and Nonconventional Hydrogen Bonds in the Binary Systems of Chalcogenoaldehyde and Chalcogenocarboxylic Acid Derivatives††Electronic Supplementary Information (ESI) Available. See DOI: Https://Doi.Org/10.1039/D4ra07498j. *RSC Adv* **2024**, *14*, 40018–40030, doi:https://doi.org/10.1039/d4ra07498j.
66. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model* **2011**, *51*, 69–82, doi:10.1021/ci100275a.
67. Gundelach, L.; Fox, T.; Tautermann, C.S.; Skylaris, C.-K. BRD4: Quantum Mechanical Protein–Ligand Binding Free Energies Using the Full-Protein DFT-Based QM-PBSA Method. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25240–25249, doi:10.1039/D2CP03705J.
68. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem Rev* **2014**, *114*, 6589–6631, doi:10.1021/cr400525m.
69. Maeda-Gutiérrez, V.; Galván-Tejada, C.E.; Galván-Tejada, J.I.; Cruz, M.; Celaya-Padilla, J.M.; Gamboa-Rosales, H.; García-Hernández, A.; Luna-García, H.; Villalba-Condori, K.O. Evaluating Feature Selection Methods for Accurate Diagnosis of Diabetic Kidney Disease. *Biomedicines* **2024**, *12*, doi:10.3390/biomedicines12122858.
70. Agu, P.C.; Afiukwa, C.A.; Orji, O.U.; Ezeh, E.M.; Ofoke, I.H.; Ogbu, C.O.; Ugwuja, E.I.; Aja, P.M. Molecular Docking as a Tool for the Discovery of Molecular Targets of Nutraceuticals in Diseases Management. *Sci Rep* **2023**, *13*, 13398, doi:10.1038/s41598-023-40160-2.
71. Wang, B.; Truhlar, D.G. Screened Electrostatic Interactions in Molecular Mechanics. *J Chem Theory Comput* **2014**, *10*, 4480–4487, doi:10.1021/ct5005142.
72. Cornelissen, F.M.G.; Markert, G.; Deutsch, G.; Antonara, M.; Faaij, N.; Bartelink, I.; Noske, D.; Vandertop, W.P.; Bender, A.; Westerman, B.A. Explaining Blood–Brain Barrier Permeability of Small Molecules by Integrated Analysis of Different Transport Mechanisms. *J Med Chem* **2023**, *66*, 7253–7267, doi:10.1021/acs.jmedchem.2c01824.
73. Wood, D.J.; Carlsson, L.; Eklund, M.; Norinder, U.; Stålring, J. QSAR with Experimental and Predictive Distributions: An Information Theoretic Approach for Assessing Model Quality. *J Comput Aided Mol Des* **2013**, *27*, 203–219, doi:10.1007/s10822-013-9639-5.
74. Zeiss, R.; Schönfeldt-Lecuona, C.; Connemann, B.J.; Hafner, S.; Gahr, M. Hepatotoxicity of Antipsychotics: An Exploratory Pharmacoepidemiologic and Pharmacodynamic Study Integrating FAERS Data and in Vitro Receptor-Binding Affinities. *Front Psychiatry* **2024**, *Volume 15-2024*, doi:10.3389/fpsyt.2024.1479625.
75. Benet, L.Z.; Hosey, C.M.; Ursu, O.; Oprea, T.I. BDDCS, the Rule of 5 and Drugability. *Adv Drug Deliv Rev* **2016**, *101*, 89–98, doi:https://doi.org/10.1016/j.addr.2016.05.007.
76. Magalhães, P.R.; Reis, P.B.P.S.; Vila-Viçosa, D.; Machuqueiro, M.; Victor, B.L. Identification of Pan-Assay INterference CompoundS (PAINS) Using an MD-Based Protocol. In *Computational Design of Membrane Proteins*; Moreira, I.S., Machuqueiro, M., Mourão, J., Eds.; Springer US: New York, NY, 2021; pp. 263–271 ISBN 978-1-0716-1468-6.
77. Ferreira, L.L.G.; Andricopulo, A.D. ADMET Modeling Approaches in Drug Discovery. *Drug Discov Today* **2019**, *24*, 1157–1165, doi:https://doi.org/10.1016/j.drudis.2019.03.015.

78. Robertson, M.J.; Skiniotis, G. Development of OPLS-AA/M Parameters for Simulations of G Protein-Coupled Receptors and Other Membrane Proteins. *J Chem Theory Comput* **2022**, *18*, 4482–4489, doi:10.1021/acs.jctc.2c00015.
79. Seeber, M.; De Benedetti, P.G.; Fanelli, F. Molecular Dynamics Simulations of the Ligand-Induced Chemical Information Transfer in the 5-HT1A Receptor. *J Chem Inf Comput Sci* **2003**, *43*, 1520–1531, doi:10.1021/ci034069c.
80. Jha, P.; Shubhra, C.; Ruchika, B.; Nidhi, J.; and Mishra, A.K. Insights of Ligand Binding in Modeled H5-HT1A Receptor: Homology Modeling, Docking, MM-GBSA, Screening and Molecular Dynamics. *J Biomol Struct Dyn* **2022**, *40*, 11625–11637, doi:10.1080/07391102.2021.1961865.
81. Foster, M.E.; Sohlberg, K. Self-Consistent Addition of an Atomic Charge Dependent Hydrogen-Bonding Correction Function. *Comput Theor Chem* **2012**, *984*, 9–12, doi:https://doi.org/10.1016/j.comptc.2011.12.027.
82. Pokorny, T.; Preller, K.H.; Kraehenmann, R.; Vollenweider, F.X. Modulatory Effect of the 5-HT1A Agonist Buspirone and the Mixed Non-Hallucinogenic 5-HT1A/2A Agonist Ergotamine on Psilocybin-Induced Psychedelic Experience. *European Neuropsychopharmacology* **2016**, *26*, 756–766, doi:https://doi.org/10.1016/j.euroneuro.2016.01.005.
83. Ho, B.Y.; Karschin, A.; Branchek, T.; Davidson, N.; Lester, H.A. The Role of Conserved Aspartate and Serine Residues in Ligand Binding and in Function of the 5-HT1A Receptor: A Site-Directed Mutation Study. *FEBS Lett* **1992**, *312*, 259–262, doi:https://doi.org/10.1016/0014-5793(92)80948-G.
84. Chanda, P.K.; Minchin, M.C.; Davis, A.R.; Greenberg, L.; Reilly, Y.; McGregor, W.H.; Bhat, R.; Lubeck, M.D.; Mizutani, S.; Hung, P.P. Identification of Residues Important for Ligand Binding to the Human 5-Hydroxytryptamine1A Serotonin Receptor. *Mol Pharmacol* **1993**, *43*, 516–520, doi:https://doi.org/10.1016/S0026-895X(25)13406-8.
85. Taniguchi, J.; Mikura, S.; da Silva Lopes, K. The Efficacy and Safety of Anamorelin for Patients with Cancer-Related Anorexia/Cachexia Syndrome: A Systematic Review and Meta-Analysis. *Sci Rep* **2023**, *13*, 15257, doi:10.1038/s41598-023-42446-x.
86. Richard, J.; Prévost, J.; Bourassa, C.; Brassard, N.; Boutin, M.; Benlarbi, M.; Goyette, G.; Medjahed, H.; Gendron-Lepage, G.; Gaudette, F.; et al. Temsavir Blocks the Immunomodulatory Activities of HIV-1 Soluble Gp120. *Cell Chem Biol* **2023**, *30*, 540-552.e6, doi:https://doi.org/10.1016/j.chembiol.2023.03.003.

THIRD PART

3 FINAL CONSIDERATIONS

The two articles presented in Part 2 demonstrated the applications of virtual screening protocols in the identification of agonists for mGluR7/8 and 5-HT_{1A} receptors. Free energy calculations served as a crucial differential to support the computational activity explored. For mGluR7/8, among 27 ligands with relevant energetic and toxicological profiles, three underwent robust MM-PBSA free energy calculations, which revealed particularly promising potential for compound 412 – 3-[(3-cyanophenyl)methylsulfonylamino]-2-hydroxypropanoic acid. This compound displayed excellent ADMET-predicted properties, with improved toxicological and chiral characteristics when compared to the reference compound L-AP4, as well as a more favorable binding free energy. In the second article, three compounds with superior predicted activity compared to Buspirone – a widely known antidepressant and 5-HT_{1A} agonist—were identified. Special emphasis is given to Temsavir (PR2), which features a pyrimidine group and an indole-like scaffold substituted with a heterocyclic nitrogen, and demonstrated high biological activity according to QSAR models, along with the most negative free energy of binding among all compounds studied. It is noteworthy that Temsavir, an advanced antiretroviral currently in clinical trials, was proposed here as part of a drug repurposing strategy, exhibiting novel activity at the 5-HT_{1A} receptor.

The central objective of this dissertation was achieved, demonstrating the feasibility of classical and semiempirical methods for extracting thermodynamic potentials in implicit solvent, with strong correlation to the reference compounds used in both studies (L-AP4 and Buspirone). In the future, these identified compounds, once published in specialized journals, may serve as a basis for subsequent experimental testing, which could corroborate the theoretical data developed here. Furthermore, both articles incorporated structural investigations that analyzed the contribution of key amino acid residues, justifying the efficiency of the observed interactions. These findings offer a starting point for further studies aiming to optimize new compounds targeting the orthosteric sites of the two proteins investigated. Ultimately, this work aspires to contribute to the development of bioactive compounds with therapeutic potential for autism, a serious condition still lacking safe and effective pharmacological treatments. This is one step in a broader journey that will depend on multidisciplinary teams focused on solving this complex challenge, eventually encompassing compound optimization, synthesis, biological assays, in vivo testing, and finally, clinical trials.

REFERENCES

ABOUZIED, A. S. *et al.* Structural and free energy landscape analysis for the discovery of antiviral compounds targeting the cap-binding domain of influenza polymerase PB2. *Scientific Reports*, v. 14, n. 1, p. 25441, 2024. Disponível em: <https://doi.org/10.1038/s41598-024-69816-3>.

ALBERT, P. R.; VAHID-ANSARI, F. The 5-HT1A receptor: Signaling to behavior. *Biochimie*, v. 161, p. 34–45, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0300908418302979>.

AL-KHAFAJI, K.; TASKIN TOK, T. Molecular dynamics simulation, free energy landscape and binding free energy computations in exploration the anti-invasive activity of amygdalin against metastasis. *Computer Methods and Programs in Biomedicine*, v. 195, p. 105660, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169260720314930>.

ALYMOV, A. A.; KAPITSA, I. G.; VORONINA, T. A. Neurochemical Mechanisms of Pathogenesis and Pharmacological Correction of Autism Spectrum Disorders: Current Concepts and Prospects. *Neurochemical Journal 2021 15:2*, v. 15, n. 2, p. 129–138, 15 jul. 2021. Disponível em: <https://link-springer-com.ez26.periodicos.capes.gov.br/article/10.1134/S1819712421020033>. Acesso em: 7 ago. 2024.

AMBURE, P. *et al.* "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems*, v. 147, p. 1–13, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169743915001732>.

ANASHKINA, A. A.; ERLYKINA, E. I. Molecular mechanisms of aberrant neuroplasticity in autism spectrum disorders (Review). *Sovremennye Tehnologii v Medicine*, v. 13, n. 1, p. 78–92, 2021. Acesso em: 8 ago. 2024.

ARADI, B.; NIKLASSON, A. M. N.; FRAUENHEIM, T. Extended Lagrangian Density Functional Tight-Binding Molecular Dynamics for Molecules and Solids. *Journal of Chemical Theory and Computation*, v. 11, n. 7, p. 3357–3363, 14 jul. 2015. Disponível em: <https://doi.org/10.1021/acs.jctc.5b00324>.

ARNOLD, V. I. *Mathematical Methods of Classical Mechanics*. 2nd. ed. New York: Springer, 1989. v. 60. (Graduate Texts in Mathematics).

ATKINS, P.; DE PAULA, J.; KEELER, J. *Physical Chemistry*. 11th. ed. [S.l.]: Oxford University Press, 2018.

AUBERSON, Y. P. *et al.* EFMC: Trends in Medicinal Chemistry and Chemical Biology. *ChemBioChem*, v. 24, n. 7, 3 abr. 2023. Acesso em: 9 ago. 2024.

BACELLAR, C.; CHERGUI, M. Ultrafast X-ray Spectroscopy of Haem Proteins. *Chimia*, v. 76, n. 6, p. 538–545, 29 jun. 2022. Acesso em: 9 ago. 2024.

BANTICK, R. A. *et al.* Occupancy of Agonist Drugs at the 5-HT1A Receptor. *Neuropsychopharmacology*, v. 29, n. 5, p. 847–859, 2004. Disponível em: <https://doi.org/10.1038/sj.npp.1300390>.

BAUER, E. P. Serotonin in fear conditioning processes. *Behavioural Brain Research*, v. 277, p. 68–77, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S016643281400477X>.

BERTONI, M. *et al.* Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports 2017 7:1*, v. 7, n. 1, p. 1–15, 5 set. 2017. Disponível em: <https://www.nature.com/articles/s41598-017-09654-8>. Acesso em: 16 jul. 2024.

BHATI, A. P. *et al.* Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *Journal of Chemical Theory and Computation*, v. 13, n. 1, p. 210–222, 10 jan. 2017. Disponível em: <https://doi.org/10.1021/acs.jctc.6b00979>.

BLANES-MIRA, C. *et al.* Comprehensive Survey of Consensus Docking for High-Throughput Virtual Screening. *Molecules*, v. 28, n. 1, 1 jan. 2023. Acesso em: 9 ago. 2024.

BLOSSEY, R. *The Poisson–Boltzmann Equation: An Introduction*. Cham: Springer Nature, 2023. Disponível em: <https://doi.org/10.1007/978-3-031-24782-8>. (SpringerBriefs in Physics).

BOCCELLA, S. *et al.* The Modulation of Pain by Metabotropic Glutamate Receptors 7 and 8 in the Dorsal Striatum. *Current Neuropharmacology*, v. 18, n. 1, p. 34–50, 20 dez. 2019. Acesso em: 8 ago. 2024.

BOLNEO, E. *et al.* Investigating the Role of GABA in Neural Development and Disease Using Mice Lacking GAD67 or VGAT Genes. *International Journal of Molecular Sciences*, v. 23, n. 14, 1 jul. 2022. Acesso em: 8 ago. 2024.

BORJIAN BOROUJENI, M. *et al.* Computational driven molecular dynamics simulation of keratinocyte growth factor behavior at different pH conditions. *Informatics in Medicine Unlocked*, v. 23, p. 100514, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352914821000046>.

BORN, J. R. *et al.* The Impact of Assay Design on Medicinal Chemistry: Case Studies. *SLAS Discovery*, v. 26, n. 10, p. 1243–1255, 1 dez. 2021. Acesso em: 9 ago. 2024.

BORROTO-ESCUELA, D. O. *et al.* Receptor–Receptor Interactions in Multiple 5-HT1A Heteroreceptor Complexes in Raphe-Hippocampal 5-HT Transmission and Their Relevance for Depression and Its Treatment. *Molecules*, v. 23, n. 6, 2018. Disponível em: <https://www.mdpi.com/1420-3049/23/6/1341>.

BOSSE, J. L. *et al.* Efficient and practical Hamiltonian simulation from time-dependent product formulas. *Nature Communications*, v. 16, n. 1, p. 2673, 2025. Disponível em: <https://doi.org/10.1038/s41467-025-57580-5>.

BRAFA MUSICORO, V. *et al.* Gamma-Aminobutyric Acid Type A Receptor Genes and Their Related Epilepsies. *Journal of Pediatric Neurology*, v. 22, n. 1, p. 13–20, 24 maio 2024. Acesso em: 8 ago. 2024.

BRAŠIĆ, J. R. *et al.* Cerebral expression of metabotropic glutamate receptor subtype 5 in idiopathic autism spectrum disorder and fragile x syndrome: A pilot study. *International Journal of Molecular Sciences*, v. 22, n. 6, p. 1–16, 2 mar. 2021. Acesso em: 8 ago. 2024.

BUBAK, A. N. *et al.* The stalk-eyed fly as a model for aggression – is there a conserved role for 5-HT between vertebrates and invertebrates? *Journal of Experimental Biology*, v. 223, n. 1, p. jeb132159, 2 jan. 2020. Disponível em: <https://doi.org/10.1242/jeb.132159>.

BUFOLO, G. N.; SOBRAL, Y. D. Analysis of the leapfrog-Verlet method applied to the Kuwabara-Kono force model in discrete element method simulations of granular materials. *Advances in Computational Mathematics*, v. 50, n. 4, p. 82, 2024. Disponível em: <https://doi.org/10.1007/s10444-024-10162-3>.

BUKINA, E. S. *et al.* SLC6A1 and Neuropsychiatric Diseases: The Role of Mutations and Prospects for Treatment with Genome Editing Systems. *Neurochemical Journal*, v. 15, n. 4, p. 376–389, out. 2021. Acesso em: 8 ago. 2024.

CAIA, J. *et al.* Elevated plasma levels of glutamate in children with autism spectrum disorders. *NeuroReport*, v. 27, n. 4, p. 272–276, 2016. Disponível em: <https://journals.lww.com/neuroreport/fulltext/2016/03010/elevated_plasma_levels_of_glutamate_in_children.11.aspx>. Acesso em: 8 ago. 2024.

CALEBIRO, D. *et al.* G PROTEIN-COUPLED RECEPTOR-G PROTEIN INTERACTIONS: A SINGLE-MOLECULE PERSPECTIVE. *Physiological Reviews*, v. 101, n. 3, p. 857–906, 1 jul. 2021. Acesso em: 8 ago. 2024.

CAPELLI, R. *et al.* Chasing the Full Free Energy Landscape of Neuroreceptor/Ligand Unbinding by Metadynamics Simulations. *Journal of Chemical Theory and Computation*, v. 15, n. 5, p. 3354–3361, 14 maio 2019. Disponível em: <https://doi.org/10.1021/acs.jctc.9b00118>.

CASTELLAN, G. W. *Physical Chemistry*. 3rd. ed. [S.l.]: Addison-Wesley, 1983.

CHEN, J. *et al.* Roles of Accelerated Molecular Dynamics Simulations in Predictions of Binding Kinetic Parameters. *Mini-Reviews in Medicinal Chemistry*, v. 24, n. 14, p. 1323–1333, 24 jan. 2024. Acesso em: 9 ago. 2024.

CHIACCHIO, M. A. *et al.* Development of AMBER Parameters for Molecular Simulations of Selected Boron-Based Covalent Ligands. *Molecules*, v. 28, n. 6, 1 mar. 2023. Acesso em: 9 ago. 2024.

CHIEN, Y. L. *et al.* Prenatal and perinatal risk factors and the clinical implications on autism spectrum disorder. *Autism*, v. 23, n. 3, p. 783–791, 1 abr. 2019. Disponível em: <https://journals-sagepub-

com.ez26.periodicos.capes.gov.br/doi/10.1177/1362361318772813>. Acesso em: 8 ago. 2024.

CHIRICO, N.; GRAMATICA, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *Journal of Chemical Information and Modeling*, v. 51, n. 9, p. 2320–2335, 2011. Disponível em: <https://doi.org/10.1021/ci200211n>.

CHONG, S.-H.; HAM, S. Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins. *Scientific Reports*, v. 9, n. 1, p. 14927, 2019. Disponível em: <https://doi.org/10.1038/s41598-019-50825-6>.

CHRISTENSEN, A. S. *et al.* Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chemical Reviews*, v. 116, n. 9, p. 5301–5337, 11 maio 2016. Disponível em: <https://doi.org/10.1021/acs.chemrev.5b00584>.

CICCOTTI, G.; DECHERCHI, S.; MELONI, S. Foundations of molecular dynamics simulations: how and what. *La Rivista del Nuovo Cimento*, 2025. Disponível em: <https://doi.org/10.1007/s40766-025-00065-4>.

CONSONNI, V.; BALLABIO, D.; TODESCHINI, R. Comments on the Definition of the Q2 Parameter for QSAR Validation. *Journal of Chemical Information and Modeling*, v. 49, n. 7, p. 1669–1678, 2009. Disponível em: <https://doi.org/10.1021/ci900115y>.

COOPER, C. D.; BARBA, L. A. Poisson–Boltzmann model for protein–surface electrostatic interactions and grid-convergence study using the PyGBe code. *Computer Physics Communications*, v. 202, p. 23–32, 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0010465515004579>.

COURTNEY, N. A.; FORD, C. P. Mechanisms of 5-HT1A receptor-mediated transmission in dorsal raphe serotonin neurons. *The Journal of Physiology*, v. 594, n. 4, p. 953–965, 15 fev. 2016. Disponível em: <https://doi.org/10.1113/JP271716>.

CRABB, E. *et al.* Importance of Equilibration Method and Sampling for Ab Initio Molecular Dynamics Simulations of Solvent–Lithium-Salt Systems in Lithium-Oxygen Batteries. *Journal of Chemical Theory and Computation*, v. 16, n. 12, p. 7255–7266, 8 dez. 2020. Disponível em: <https://doi.org/10.1021/acs.jctc.0c00833>.

CRAMER, C. J. *Essentials of Computational Chemistry: Theories and Models*. 2nd. ed. [S.l.]: John Wiley & Sons, 2004.

DAINA, A.; MICHIELIN, O.; ZOETE, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports 2017 7:1*, v. 7, n. 1, p. 1–13, 3 mar. 2017. Disponível em: <https://www.nature.com/articles/srep42717>. Acesso em: 16 jul. 2024.

DE J. RÍOS-ROLDÁN, A. *et al.* Molecular Dynamics simulations and discrete perturbation theory for systems interacting via the parabolic-well pair potential. *Journal of Molecular Liquids*, v. 400, p. 124522, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167732224005786>.

DE RUITER, A.; BORESCH, S.; OOSTENBRINK, C. Comparison of thermodynamic integration and Bennett acceptance ratio for calculating relative protein-ligand binding free energies. *Journal of Computational Chemistry*, v. 34, n. 12, p. 1024–1034, 5 maio 2013. Disponível em: <https://doi.org/10.1002/jcc.23229>.

DE RUITER, A.; OOSTENBRINK, C. Free energy calculations of protein–ligand interactions. *Current Opinion in Chemical Biology*, v. 15, n. 4, p. 547–552, 2011. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1367593111000901>.

DELAIGLE, A.; GIJBELS, I. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, v. 56, n. 1, p. 19–47, 2004. Disponível em: <https://doi.org/10.1007/BF02530523>.

DEWAR, M. J. S. *et al.* Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, v. 107, n. 13, p. 3902–3909, 1 jun. 1985. Disponível em: <https://doi.org/10.1021/ja00299a024>.

DOHERTY, B. *et al.* Revisiting OPLS Force Field Parameters for Ionic Liquid Simulations. *Journal of Chemical Theory and Computation*, v. 13, n. 12, p. 6131–6135, 12 dez. 2017. Disponível em: <https://pubs.acs.org/doi/full/10.1021/acs.jctc.7b00520>. Acesso em: 9 ago. 2024.

DOMIN, H. Group III metabotropic glutamate receptors as promising targets for neuroprotective therapy: Particular emphasis on the role of mGlu4 and mGlu7 receptors. *Pharmacology Biochemistry and Behavior*, v. 219, 1 set. 2022. Acesso em: 8 ago. 2024.

DU, L. *et al.* Dockey: a modern integrated tool for large-scale molecular docking and virtual screening. *Briefings in Bioinformatics*, v. 24, n. 2, 1 mar. 2023. Acesso em: 9 ago. 2024.

DU, Q.; BEGLOV, D.; ROUX, B. Solvation Free Energy of Polar and Nonpolar Molecules in Water: An Extended Interaction Site Integral Equation Theory in Three Dimensions. *The Journal of Physical Chemistry B*, v. 104, n. 4, p. 796–805, 1 fev. 2000. Disponível em: <https://doi.org/10.1021/jp992712l>.

DUFOUR, B. D. *et al.* Distinct patterns of GABAergic interneuron pathology in autism are associated with intellectual impairment and stereotypic behaviors. *Autism*, v. 27, n. 6, p. 1730–1745, 1 ago. 2023. Disponível em: <https://journals-sagepub-com.ez26.periodicos.capes.gov.br/doi/10.1177/13623613231154053>. Acesso em: 8 ago. 2024.

DUPONT, A. C. *et al.* Metabotropic Glutamate Receptor Subtype 5 Positron-Emission-Tomography Radioligands as a Tool for Central Nervous System Drug Development: Between Progress and Setbacks. *Pharmaceuticals*, v. 16, n. 8, 1 ago. 2023. Acesso em: 8 ago. 2024.

EBRAHIMI-GHIRI, M.; KHAKPAI, F.; ZARRINDAST, M. R. Combined treatment of scopolamine and group III mGluR antagonist, CPPG, exerts antidepressant activity without

affecting anxiety-related behaviors. *Physiology & Behavior*, v. 224, p. 113034, 1 out. 2020. Acesso em: 8 ago. 2024.

EGBENYA, D. L.; AIDOO, E.; KYEI, G. Glutamate receptors in brain development. *Child's Nervous System*, v. 37, n. 9, p. 2753–2758, 1 set. 2021. Disponível em: <https://link-springer-com.ez26.periodicos.capes.gov.br/article/10.1007/s00381-021-05266-w>. Acesso em: 8 ago. 2024.

EKBERG, V.; RYDE, U. On the Use of Interaction Entropy and Related Methods to Estimate Binding Entropies. *Journal of Chemical Theory and Computation*, v. 17, n. 8, p. 5379–5391, 10 ago. 2021. Disponível em: <https://doi.org/10.1021/acs.jctc.1c00374>.

ELTOKHI, A. *et al.* Glutamatergic dysfunction and synaptic ultrastructural alterations in schizophrenia and autism spectrum disorder: Evidence from human and rodent studies. *International Journal of Molecular Sciences*, v. 22, n. 1, p. 1–26, 1 jan. 2021. Acesso em: 13 ago. 2024.

FANG, D.; LIU, D.; SARKAR, R. Time-Dependent Hamiltonian Simulation via Magnus Expansion: Algorithm and Superconvergence. *Communications in Mathematical Physics*, v. 406, n. 6, p. 128, 2025. Disponível em: <https://doi.org/10.1007/s00220-025-05314-5>.

FERREIRA, L. L. G.; ANDRICOPULO, A. D. ADMET modeling approaches in drug discovery. *Drug Discovery Today*, v. 24, n. 5, p. 1157–1165, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1359644618303301>.

FISCHER, F. P. *et al.* Molecular and Clinical Repercussions of GABA Transporter 1 Variants Gone Amiss: Links to Epilepsy and Developmental Spectrum Disorders. *Frontiers in Molecular Biosciences*, v. 9, 2 mar. 2022. Acesso em: 8 ago. 2024.

FLACHSENBERG, F. *et al.* Redocking the PDB. *Journal of Chemical Information and Modeling*, v. 64, n. 1, p. 219–237, 8 jan. 2024. Disponível em: <https://pubs.acs.org/doi/abs/10.1021/acs.jcim.3c01573>. Acesso em: 9 ago. 2024.

FREITAS, G. A.; NISWENDER, C. M. GRM7 gene mutations and consequences for neurodevelopment. *Pharmacology Biochemistry and Behavior*, v. 225, 1 abr. 2023. Acesso em: 8 ago. 2024.

GALINEAU, L. *et al.* Glutamatergic synapse in autism: a complex story for a complex disorder. *Molecular Psychiatry 2022 28:2*, v. 28, n. 2, p. 801–809, 25 nov. 2022. Disponível em: <https://www-nature-com.ez26.periodicos.capes.gov.br/articles/s41380-022-01860-9>. Acesso em: 7 ago. 2024.

GHOSH DASTIDAR, S. *et al.* Distinct regulation of bioenergetics and translation by group I mGluR and NMDAR. *EMBO reports*, v. 21, n. 6, 4 jun. 2020a. Acesso em: 8 ago. 2024.

GHOSH DASTIDAR, S. *et al.* Distinct regulation of bioenergetics and translation by group I mGluR and NMDAR. *EMBO reports*, v. 21, n. 6, 4 jun. 2020b. Acesso em: 8 ago. 2024.

GOEL, N.; PENG, K.; LU, Y. Neuromodulation by mGluRs in Sound Localization Circuits in the Auditory Brainstem. *Frontiers in Neural Circuits*, v. 14, 5 nov. 2020. Acesso em: 8 ago. 2024.

GRAMATICA, P. Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, v. 5, n. 3, p. 61–97, 2020. Disponível em: <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJQSPR.20200701.oa1>.

GRANIERI, J. E. *et al.* Atypical social communication is associated with positive initial impressions among peers with autism spectrum disorder. *https://doi-org.ez26.periodicos.capes.gov.br/10.1177/1362361320924906*, v. 24, n. 7, p. 1841–1848, 4 jun. 2020. Disponível em: <https://journals-sagepub-com.ez26.periodicos.capes.gov.br/doi/10.1177/1362361320924906>. Acesso em: 8 ago. 2024.

GROMACS DEVELOPMENT TEAM. *GROMACS User Manual*. . [S.l: s.n.]. Disponível em: <https://manual.gromacs.org/documentation/2024.1/>. , 2024

GRUBMÜLLER, H. *et al.* Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation*, v. 6, n. 1–3, p. 121–142, 1 mar. 1991. Disponível em: <https://doi.org/10.1080/08927029108022142>.

GRYGORENKO, O. O. Enamine Ltd.: The Science and Business of Organic Chemistry and Beyond. *European Journal of Organic Chemistry*, v. 2021, n. 47, p. 6474–6477, 21 dez. 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/ejoc.202101210>. Acesso em: 5 ago. 2024.

GUHA, P.; GHOSE-CHOUDHURY, A. Generalized Hamiltonian and Lagrangian aspects of a model for virus–tumor interaction in oncolytic virotherapy. *Mathematical Methods in the Applied Sciences*, v. 48, n. 4, p. 4173–4184, 15 mar. 2025. Disponível em: <https://doi.org/10.1002/mma.10538>.

HAIN, P. J. H.; MOSER, T. Optical measurement of glutamate release robustly reports short-term plasticity at a fast central synapse. *Frontiers in Molecular Neuroscience*, v. 17, p. 1351280, 28 fev. 2024. Acesso em: 8 ago. 2024.

HAMMOND, R. K.; HOFFMAN, J. M. Adolescents With High-Functioning Autism: An Investigation of Comorbid Anxiety and Depression. *Journal of Mental Health Research in Intellectual Disabilities*, v. 7, n. 3, p. 246–263, 2014. Acesso em: 8 ago. 2024.

HAPALA, P. *et al.* Mapping the electrostatic force field of single molecules from high-resolution scanning probe images. *Nature Communications 2016 7:1*, v. 7, n. 1, p. 1–8, 27 maio 2016. Disponível em: <https://www.nature.com/articles/ncomms11560>. Acesso em: 9 ago. 2024.

HENTER, I. D.; PARK, L. T.; ZARATE, C. A. Novel Glutamatergic Modulators for the Treatment of Mood Disorders: Current Status. *CNS Drugs*, v. 35, n. 5, p. 527–543, 1 maio 2021. Acesso em: 8 ago. 2024.

HERRERO, C.; JOLY, L. The Poisson–Boltzmann equation in micro- and nanofluidics: A formulary. *Physics of Fluids*, v. 36, n. 10, p. 101801, 25 out. 2024. Disponível em: <https://doi.org/10.1063/5.0238173>.

HODGES, H.; FEALKO, C.; SOARES, N. Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational Pediatrics*, v. 9, n. Suppl 1, p. S55, 1 fev. 2020. Disponível em: </pmc/articles/PMC7082249/>. Acesso em: 7 ago. 2024.

HOPKINS, C. W.; ROITBERG, A. E. Fitting of dihedral terms in classical force fields as an analytic linear least-squares problem. *Journal of Chemical Information and Modeling*, v. 54, n. 7, p. 1978–1986, 28 jul. 2014. Disponível em: <https://pubs.acs.org/doi/abs/10.1021/ci500112w>. Acesso em: 9 ago. 2024.

HUANG, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods*, v. 14, n. 1, p. 71, 29 dez. 2017. Disponível em: </pmc/articles/PMC5199616/>. Acesso em: 9 ago. 2024.

HUANG, M.; WANG, Y. GLOBAL AND TARGETED PROFILING OF GTP-BINDING PROTEINS IN BIOLOGICAL SAMPLES BY MASS SPECTROMETRY. *Mass Spectrometry Reviews*, v. 40, n. 3, p. 215–235, 1 maio 2021. Acesso em: 8 ago. 2024.

IZAT, N.; SAHIN, S. Hepatic transporter-mediated pharmacokinetic drug–drug interactions: Recent studies and regulatory recommendations. *Biopharmaceutics and Drug Disposition*, v. 42, n. 2–3, p. 45–77, 1 mar. 2021. Acesso em: 9 ago. 2024.

JENSEN, F. *Introduction to Computational Chemistry*. 2nd. ed. [S.l.]: John Wiley & Sons, 2007.

JHA, P. *et al.* Insights of ligand binding in modeled h5-HT1A receptor: homology modeling, docking, MM-GBSA, screening and molecular dynamics. *Journal of Biomolecular Structure and Dynamics*, v. 40, n. 22, p. 11625–11637, 12 dez. 2022. Disponível em: <https://doi.org/10.1080/07391102.2021.1961865>.

JIANG, M. *et al.* Autism spectrum disorder research: knowledge mapping of progress and focus between 2011 and 2022. *Frontiers in Psychiatry*, v. 14, 2023. Disponível em: </pmc/articles/PMC10168184/>. Acesso em: 7 ago. 2024.

JONG, Y. J. I. *et al.* Striatal mGlu5-mediated synaptic plasticity is independently regulated by location-specific receptor pools and divergent signaling pathways. *Journal of Biological Chemistry*, v. 299, n. 8, 1 ago. 2023. Acesso em: 8 ago. 2024.

KADRIU, B. *et al.* Positive AMPA receptor modulation in the treatment of neuropsychiatric disorders: A long and winding road. *Drug Discovery Today*, v. 26, n. 12, p. 2816–2838, 1 dez. 2021. Acesso em: 8 ago. 2024.

KAZMI, S. R. *et al.* In silico approaches and tools for the prediction of drug metabolism and fate: A review. *Computers in Biology and Medicine*, v. 106, p. 54–64, 1 mar. 2019. Acesso em: 9 ago. 2024.

KESSI, M. *et al.* Attention-deficit/hyperactive disorder updates. *Frontiers in Molecular Neuroscience*, v. 15, 21 set. 2022. Acesso em: 8 ago. 2024.

KIKKAWA, M.; YANAGISAWA, H. Identifying proteins in the cell by tagging techniques for cryo-electron microscopy. *Microscopy*, v. 71, p. I60–I65, 1 mar. 2022. Acesso em: 9 ago. 2024.

KILLAM, L. A. *et al.* Virtual screen-based clinical simulation: An integrative review of student experiences. *Teaching and Learning in Nursing*, v. 19, n. 2, p. e359–e368, 1 abr. 2024. Acesso em: 9 ago. 2024.

KISS, O. *et al.* Quantum neural networks force fields generation. *Machine Learning: Science and Technology*, v. 3, n. 3, 1 set. 2022. Acesso em: 9 ago. 2024.

KOEHL, A. *et al.* Structural Insights into Metabotropic Glutamate Receptor Activation. *Nature*, v. 566, n. 7742, p. 79, 7 fev. 2019. Disponível em: </pmc/articles/PMC6709600/>. Acesso em: 5 ago. 2024.

KOLODNY, T. *et al.* Concentrations of Cortical GABA and Glutamate in Young Adults With Autism Spectrum Disorder. *Autism Research*, v. 13, n. 7, p. 1111–1129, 1 jul. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/aur.2300>. Acesso em: 8 ago. 2024.

KONDAUROVA, E. M. *et al.* On the role of serotonin 5-HT1A receptor in autistic-like behavior: cross talk of 5-HT and BDNF systems. *Behavioural Brain Research*, v. 438, p. 114168, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166432822004375>.

KRAUSCH, N. *et al.* Monte Carlo Simulations for the Analysis of Non-linear Parameter Confidence Intervals in Optimal Experimental Design. *Frontiers in Bioengineering and Biotechnology*, v. Volume 7-2019, 2019. Disponível em: <https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2019.00122>.

KRIVOV, S. V. The Free Energy Landscape Analysis of Protein (FIP35) Folding Dynamics. *The Journal of Physical Chemistry B*, v. 115, n. 42, p. 12315–12324, 27 out. 2011. Disponível em: <https://doi.org/10.1021/jp208585r>.

KURAHASHI, H. *et al.* Autism spectrum disorder-like behaviors induced by hyper-glutamatergic NMDA receptor signaling through hypo-serotonergic 5-HT1A receptor signaling in the prefrontal cortex in mice exposed to prenatal valproic acid. *Neuropsychopharmacology*, v. 50, n. 5, p. 739–750, 2025. Disponível em: <https://doi.org/10.1038/s41386-024-02004-z>.

LAURENDEAU, N. C. *Statistical Thermodynamics: Fundamentals and Applications*. Cambridge: Cambridge University Press, 2005.

LEACH, A. R. *Molecular Modelling: Principles and Applications*. 2nd. ed. [S.l.]: Prentice Hall, 2001.

LEE, I. H. *et al.* Association between Neuroligin-1 polymorphism and plasma glutamine levels in individuals with autism spectrum disorder. *eBioMedicine*, v. 95, p. 104746, 1 set. 2023. Acesso em: 8 ago. 2024.

LEVINE, I. N. *Physical Chemistry*. 6th. ed. [S.l.]: McGraw-Hill Education, 2009.

LI, Q. *et al.* Group II metabotropic glutamate receptor agonist promotes retinal ganglion cell survival by reducing neuronal excitotoxicity in a rat chronic ocular hypertension model. *Neuropharmacology*, v. 170, p. 108016, 15 jun. 2020. Acesso em: 8 ago. 2024.

LIM, T. C. Mathematical relationships between bond-bending force fields. *Journal of Mathematical Chemistry*, v. 32, n. 3, p. 249–256, 2002. Acesso em: 9 ago. 2024.

LIMA, A. *et al.* Structure-Based Virtual Screening: Successes and Pitfalls. *Journal of the Brazilian Chemical Society*, 2024. Acesso em: 9 ago. 2024.

LIMON, A. *et al.* GABA and Glutamate Receptors of the Autistic Brain. *Autism - A Neurodevelopmental Journey from Genes to Behaviour*, 17 ago. 2011. Disponível em: <https://www.intechopen.com/chapters/18061>. Acesso em: 7 ago. 2024.

LIPINSKI, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies*, v. 1, n. 4, p. 337–341, 1 dez. 2004. Acesso em: 5 ago. 2024.

LORD, C. *et al.* Autism spectrum disorder. *Nature Reviews Disease Primers 2020 6:1*, v. 6, n. 1, p. 1–23, 16 jan. 2020. Disponível em: <https://www.nature.com/articles/s41572-019-0138-4>. Acesso em: 7 ago. 2024.

LUESSEN, D. J.; CONN, P. J. Allosteric Modulators of Metabotropic Glutamate Receptors as Novel Therapeutics for Neuropsychiatric Disease. *Pharmacological Reviews*, v. 74, n. 3, p. 630–661, 1 jul. 2022. Acesso em: 8 ago. 2024.

LV, D. *et al.* Enhanced NMDA receptor pathway and glutamate transmission in the hippocampal dentate gyrus mediate the spatial learning and memory impairment of obese rats. *Pflugers Archiv European Journal of Physiology*, v. 476, n. 5, p. 821–831, 1 maio 2024. Disponível em: <https://link-springer-com.ez26.periodicos.capes.gov.br/article/10.1007/s00424-024-02924-1>. Acesso em: 8 ago. 2024.

MARTIN, H. *et al.* Review: Dopamine, Serotonin, and the Translational Neuroscience of Aggression in Autism Spectrum Disorder. *JAACAP Open*, v. 3, n. 1, p. 29–41, 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2949732924000188>.

MARTYS, N. S.; MOUNTAIN, R. D. Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Physical Review E*, v. 59, n. 3, p. 3733–3736, 1 mar. 1999. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevE.59.3733>.

MATSUNAGA, H.; ARUGA, J. Trans-Synaptic Regulation of Metabotropic Glutamate Receptors by Elfn Proteins in Health and Disease. *Frontiers in Neural Circuits*, v. 15, 15 mar. 2021. Acesso em: 8 ago. 2024.

MAYES, S. D. *et al.* Suicide ideation and attempts in children with autism. *Research in Autism Spectrum Disorders*, v. 7, n. 1, p. 109–119, 1 jan. 2013. Acesso em: 8 ago. 2024.

MAZZITELLI, M. *et al.* Recent Advances in the Modulation of Pain by the Metabotropic Glutamate Receptors. *Cells*, v. 11, n. 16, 1 ago. 2022. Acesso em: 8 ago. 2024.

MCCULLOCK, T. W.; KAMMERMEIER, P. J. The evidence for and consequences of metabotropic glutamate receptor heterodimerization. *Neuropharmacology*, v. 199, p. 108801, 1 nov. 2021. Acesso em: 8 ago. 2024.

MCQUARRIE, D. A. *Statistical Mechanics*. [S.l.]: University Science Books, 2000.

MCQUARRIE, D. A.; SIMON, J. D. *Physical Chemistry: A Molecular Approach*. [S.l.]: University Science Books, 1997.

MELLER, A. *et al.* Toward physics-based precision medicine: Exploiting protein dynamics to design new therapeutics and interpret variants. *Protein Science*, v. 33, n. 3, 1 mar. 2024. Acesso em: 9 ago. 2024.

MENEZES, M. *et al.* Treatment of Depression in Individuals with Autism Spectrum Disorder: A Systematic Review. *Research in Autism Spectrum Disorders*, v. 78, p. 101639, 1 out. 2020. Acesso em: 7 ago. 2024.

MILLER, B. R. *et al.* MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, v. 8, n. 9, p. 3314–3321, 11 set. 2012. Disponível em: <https://pubs.acs.org/doi/full/10.1021/ct300418h>. Acesso em: 13 nov. 2024.

MITEVA, M. A.; VILLOUTREIX, B. O. Computational Biology and Chemistry in MTi: Emphasis on the Prediction of Some ADMET Properties. *Molecular Informatics*, v. 36, n. 10, 1 out. 2017. Acesso em: 9 ago. 2024.

MOLEXUS. *Molegro Virtual Docker*. Disponível em: <http://molexus.io/molegro-virtual-docker/>. Acesso em: 9 ago. 2024.

MORETTI, V. *Analytical Mechanics: Classical, Lagrangian and Hamiltonian Mechanics, Stability Theory, Special Relativity*. Tradução Simon G Chiossi. Cham, Switzerland: Springer Nature, 2023. v. 150. Disponível em: <https://doi.org/10.1007/978-3-031-27612-5>. (UNITEXT).

MORTIMER, R. G. *Physical Chemistry*. 3rd. ed. [S.l.]: Academic Press, 2008.

MOXON-EMRE, I. *et al.* Modulation of Dorsolateral Prefrontal Cortex Glutamate/Glutamine Levels Following Repetitive Transcranial Magnetic Stimulation in Young Adults With Autism. *Frontiers in Neuroscience*, v. 15, p. 711542, 6 out. 2021. Disponível em: <www.frontiersin.org>. Acesso em: 8 ago. 2024.

MUSKETT, A. *et al.* Repetitive behaviors in Autism Spectrum Disorder: Associations with depression and anxiety symptoms. *Research in Autism Spectrum Disorders*, v. 68, p. 101449, 1 dez. 2019. Acesso em: 7 ago. 2024.

MUTO, T. *et al.* Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 10, p. 3759–3764, 6 mar. 2007. Disponível em: <https://www.pnas.org/doi/abs/10.1073/pnas.0611577104>. Acesso em: 8 ago. 2024.

NIAZI, S. K.; MARIAM, Z. Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis. *Pharmaceuticals*, v. 17, n. 1, 1 jan. 2024. Acesso em: 9 ago. 2024.

NIESCIER, R. F.; LIN, Y. C. The Potential Role of AMPA Receptor Trafficking in Autism and Other Neurodevelopmental Conditions. *Neuroscience*, v. 479, p. 180–191, 15 dez. 2021. Acesso em: 8 ago. 2024.

NIKITIN, M. A. *et al.* Glutamate and GABA receptors in non-neural animals (Placozoa): Preadaptation to neural transmission. *Žurnal obŝej biologii*, v. 84, n. 3, p. 163–176, 1 maio 2023. Disponível em: <https://journals.rcsi.science/0044-4596/article/view/136440>. Acesso em: 8 ago. 2024.

NISWENDER, C. M.; CONN, P. J. Metabotropic Glutamate Receptors: Physiology, Pharmacology, and Disease. *Annual review of pharmacology and toxicology*, v. 50, p. 295, 10 fev. 2010. Disponível em: </pmc/articles/PMC2904507/>. Acesso em: 8 ago. 2024.

NOH, M.-H.; AND LEE, S.-Y. A bivariate Gaussian function approach for inverse cracks identification of forced-vibrating bridge decks. *Inverse Problems in Science and Engineering*, v. 21, n. 6, p. 1047–1073, 1 set. 2013. Disponível em: <https://doi.org/10.1080/17415977.2012.727086>.

NUSSENZVEIG, M. *Física, Volume 3: Eletromagnetismo*. 1ª ed. São Paulo: Editora Edgard Blücher, 2003.

NYRENIUS, J. *et al.* Autism in adult psychiatric out-patients: self-reported suicidal ideation, suicide attempts and non-suicidal self-injury. *BJPsych Open*, v. 9, n. 5, p. e167, set. 2023. Disponível em: <https://www-cambridge-org.ez26.periodicos.capes.gov.br/core/journals/bjpsych-open/article/autism-in-adult-psychiatric-outpatients-selfreported-suicidal-ideation-suicide-attempts-and-nonsuicidal-selfinjury/6293BD6BDB6B54DF388532D2B50E7139>. Acesso em: 8 ago. 2024.

OLMEDO, D. A. *et al.* Design and Diversity Analysis of Chemical Libraries in Drug Discovery. *Combinatorial Chemistry & High Throughput Screening*, v. 27, n. 4, p. 502–515, 6 jul. 2023. Acesso em: 9 ago. 2024.

OLSSON, M. H. M. *et al.* PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation*, v. 7, n. 2, p. 525–537, 2011. Disponível em: <https://doi.org/10.1021/ct100578z>.

ONG, A. W. W.; CAO, S. Y.; KWEK, L. C. An improved parameterization procedure for NDDO-descendant semi-empirical methods. *Journal of Molecular Modeling*, v. 29, n. 4, p. 118, 2023. Disponível em: <https://doi.org/10.1007/s00894-023-05499-3>.

ONUFRIEV, A.; BASHFORD, D.; CASE, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B*, v. 104, n. 15, p. 3712–3720, 1 abr. 2000. Disponível em: <https://doi.org/10.1021/jp994072s>.

ORMEÑO, F.; GENERAL, I. J. Convergence and equilibrium in molecular dynamics simulations. *Communications Chemistry*, v. 7, n. 1, p. 26, 2024. Disponível em: <https://doi.org/10.1038/s42004-024-01114-5>.

OYA, M. *et al.* Increased glutamate and glutamine levels and their relationship to astrocytes and dopaminergic transmissions in the brains of adults with autism. *Scientific Reports 2023 13:1*, v. 13, n. 1, p. 1–10, 19 jul. 2023. Disponível em: <https://www-nature-com.ez26.periodicos.capes.gov.br/articles/s41598-023-38306-3>. Acesso em: 8 ago. 2024.

OZTENEKECIOGLU, B. *et al.* Genetic and Epigenetic Alterations in Autism Spectrum Disorder. *Global Medical Genetics*, v. 08, n. 04, p. 144–148, dez. 2021. Acesso em: 8 ago. 2024.

PAPALEO, E. *et al.* Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *Journal of Molecular Graphics and Modelling*, v. 27, n. 8, p. 889–899, 2009. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1093326309000175>.

PARENT, H. H.; NISWENDER, C. M. Therapeutic Potential for Metabotropic Glutamate Receptor 7 Modulators in Cognitive Disorders. *Molecular Pharmacology*, v. 105, n. 5, p. 348–358, 1 maio 2024. Acesso em: 8 ago. 2024.

PARIMALA, K. Experimental and Computational Study on the Spectroscopic Approach, Hyperpolarizabilities, NBO Analysis, ADMET Studies, and In-Silico Ligand-Protein Docking of 2,4,6-Trifluoro-5-Chloro Pyrimidine. *Polycyclic Aromatic Compounds*, 2023. Acesso em: 9 ago. 2024.

PEREIRA, G. P.; CECCHINI, M. Multibasin Quasi-Harmonic Approach for the Calculation of the Configurational Entropy of Small Molecules in Solution. *Journal of Chemical Theory and Computation*, v. 17, n. 2, p. 1133–1142, 9 fev. 2021. Disponível em: <https://doi.org/10.1021/acs.jctc.0c00978>.

PICKLES, A. *et al.* Predictors of language regression and its association with subsequent communication development in children with autism. *Journal of Child Psychology and Psychiatry*, v. 63, n. 11, p. 1243–1251, 1 nov. 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/jcpp.13565>. Acesso em: 8 ago. 2024.

PLESTED, A. J. R.; POULSEN, M. H. Crosslinking glutamate receptor ion channels. *Methods in Enzymology*, v. 652, p. 161–192, 1 jan. 2021. Acesso em: 8 ago. 2024.

POLTER, A. M.; LI, X. 5-HT1A receptor-regulated signal transduction pathways in brain. *Cellular Signalling*, v. 22, n. 10, p. 1406–1412, 2010. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0898656810000938>.

RABEH, N. *et al.* Targeting mGluR group III for the treatment of neurodegenerative diseases. *Biomedicine & Pharmacotherapy*, v. 168, p. 115733, 1 dez. 2023. Acesso em: 8 ago. 2024.

RÁCZ, A. *et al.* Molecular Dynamics Simulations and Diversity Selection by Extended Continuous Similarity Indices. *Journal of Chemical Information and Modeling*, v. 62, n. 14, p. 3415–3425, 25 jul. 2022. Disponível em: <https://doi.org/10.1021/acs.jcim.2c00433>.

ŘEZÁČ, J. *et al.* Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *Journal of Chemical Theory and Computation*, v. 5, n. 7, p. 1749–1760, 14 jul. 2009. Disponível em: <https://doi.org/10.1021/ct9000922>.

RINGE, S. *et al.* Implicit Solvation Methods for Catalysis at Electrified Interfaces. *Chemical Reviews*, v. 122, n. 12, p. 10777–10820, 22 jun. 2022. Disponível em: <https://doi.org/10.1021/acs.chemrev.1c00675>.

RONDARD, P. *et al.* Coupling of agonist binding to effector domain activation in metabotropic glutamate-like receptors. *The Journal of biological chemistry*, v. 281, n. 34, p. 24653–24661, 25 ago. 2006. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/16787923/>. Acesso em: 8 ago. 2024.

ROY, K.; KAR, S.; AMBURE, P. On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, v. 145, p. 22–29, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169743915000969>.

RUSAKOV, D. A.; STEWART, M. G. Synaptic environment and extrasynaptic glutamate signals: The quest continues. *Neuropharmacology*, v. 195, p. 108688, 1 set. 2021. Acesso em: 8 ago. 2024.

SAFARIPOUR, S.; ANAND, G.; SNOEYINK, C. Thermodynamic Study of the Electric Field Effect on Liquid–Vapor Mixture at Equilibrium: An Analysis on a Water–Ethanol Mixture. *The Journal of Physical Chemistry B*, v. 127, n. 29, p. 6500–6506, 27 jul. 2023. Disponível em: <https://doi.org/10.1021/acs.jpcb.3c01578>.

SAITOW, F.; TAKUMI, T.; SUZUKI, H. Upregulated 5-HT1A receptor-mediated currents in the prefrontal cortex layer 5 neurons in the 15q11–13 duplication mouse model of autism. *Molecular Brain*, v. 13, n. 1, p. 115, 2020. Disponível em: <https://doi.org/10.1186/s13041-020-00655-9>.

SALGADO, N. D. M. *et al.* Autism Spectrum Disorder in Children: A Systematic Review of the Increasing Incidence and Diagnosis. *Research, Society and Development*, v. 11, n. 13, p. e512111335748–e512111335748, 16 out. 2022. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/35748>. Acesso em: 7 ago. 2024.

SALINAS, S. R. A. *Introdução à Física Estatística*. São Paulo: Editora da Universidade de São Paulo (EdUSP), 1997. v. 9.

SAVA, A. D. IN SILICO SCREENING OF SOME OLD DRUGS FOR NEW APPROACHES IN HEART FAILURE. *Revista Medico-Chirurgicala*, v. 127, n. 3, p. 488–496, 30 set. 2023. Acesso em: 9 ago. 2024.

SCHAUPERL, M. *et al.* Data-driven analysis of the number of Lennard–Jones types needed in a force field. *Communications Chemistry 2020 3:1*, v. 3, n. 1, p. 1–12, 13 nov. 2020. Disponível em: <https://www.nature.com/articles/s42004-020-00395-w>. Acesso em: 9 ago. 2024.

SCHKERYANTZ, J. *et al.* Determination of L-AP4-bound human mGlu8 receptor amino terminal domain structure and the molecular basis for L-AP4's group III mGlu receptor functional. *ElsevierJM Schkeryantz, Q Chen, JD Ho, S Atwell, A Zhang, MC Vargas, J Wang, JA Monn, J HaoBioorganic & medicinal chemistry letters, 2018•Elsevier*, [S.d.]. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0960894X18300477>. Acesso em: 16 abr. 2024.

SCHMID, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal : EBJ*, v. 40, n. 7, p. 843–856, jul. 2011. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21533652/>. Acesso em: 9 ago. 2024.

SEARS, S. M. S.; HEWETT, S. J. Influence of glutamate and GABA transport on brain excitatory/inhibitory balance. *https://doi-org.ez26.periodicos.capes.gov.br/10.1177/1535370221989263*, v. 246, n. 9, p. 1069–1083, 7 fev. 2021. Disponível em: <https://journals-sagepub-com.ez26.periodicos.capes.gov.br/doi/10.1177/1535370221989263>. Acesso em: 8 ago. 2024.

SERAJEE, F. J. *et al.* The metabotropic glutamate receptor 8 gene at 7q31: partial duplication and possible association with autism. *Journal of medical genetics*, v. 40, n. 4, 2003. Acesso em: 13 ago. 2024.

SHAHRAJABIAN, M. H.; SUN, W. Characterization of Intrinsically Disordered Proteins in Healthy and Diseased States by Nuclear Magnetic Resonance. *Reviews on Recent Clinical Trials*, v. 19, n. 3, p. 176–188, 27 fev. 2024. Acesso em: 9 ago. 2024.

SHANNON, R. J. *et al.* Anharmonic Molecular Mechanics: Ab Initio Based Morse Parametrizations for the Popular MM3 Force Field. *The journal of physical chemistry. A*, v. 123, n. 13, p. 2991–2999, 4 abr. 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30793911/>. Acesso em: 9 ago. 2024.

SHARMA, A.; YENNAMALLI, R. M. Docking strategies. *Basic Biotechniques for Bioprocess and Bioentrepreneurship*, p. 243–258, 1 jan. 2023. Acesso em: 5 ago. 2024.

SHAYANFAR, S.; SHAYANFAR, A. Comparison of various methods for validity evaluation of QSAR models. *BMC Chemistry*, v. 16, n. 1, p. 63, 2022. Disponível em: <https://doi.org/10.1186/s13065-022-00856-4>.

SHENG, L.; CHEN, Z. Molecular dynamics study of dispersion and fluidity of porous liquids with different pore sizes. *Journal of Molecular Liquids*, v. 333, p. 115890, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167732221006176>.

SILVA, G. M. *et al.* Improving the Born equation: Origin of the Born radius and introducing dielectric saturation effects. *Fluid Phase Equilibria*, v. 576, p. 113955, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378381223002352>.

SOMMA, R. D. *et al.* Quantum simulations of physics problems. 4 ago. 2003, [S.l: s.n.], 4 ago. 2003. p. 96–103. Disponível em: <https://doi.org/10.1117/12.487249>.

SPAIN, D.; ZIVRALI YARAR, E.; HAPPÉ, F. Social anxiety in adults with autism: a qualitative study. *International Journal of Qualitative Studies on Health and Well-being*, v. 15, n. 1, 1 jan. 2020. Disponível em: <https://www-tandfonline-com.ez26.periodicos.capes.gov.br/doi/abs/10.1080/17482631.2020.1803669>. Acesso em: 8 ago. 2024.

SPREITER, Q.; WALTER, M. Classical Molecular Dynamics Simulation with the Velocity Verlet Algorithm at Strong External Magnetic Fields. *Journal of Computational Physics*, v. 152, n. 1, p. 102–119, 1999. Disponível em: <https://www.sciencedirect.com/science/article/pii/S002199919996237X>.

SPRENGER, K. G.; JAEGER, V. W.; PFAENDTNER, J. The general AMBER force field (GAFF) can accurately predict thermodynamic and transport properties of many ionic liquids. *Journal of Physical Chemistry B*, v. 119, n. 18, p. 5882–5895, 7 maio 2015. Disponível em: <https://pubs.acs.org/doi/abs/10.1021/acs.jpcb.5b00689>. Acesso em: 9 ago. 2024.

STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, v. 19, n. 1, p. 1–32, 2013. Disponível em: <https://doi.org/10.1007/s00894-012-1667-x>.

STOPPEL, D. C. *et al.* mGluR5 Negative Modulators for Fragile X: Treatment Resistance and Persistence. *Frontiers in Psychiatry*, v. 12, 29 set. 2021. Acesso em: 8 ago. 2024.

STUDER, G. *et al.* ProMod3—A versatile homology modelling toolbox. *PLOS Computational Biology*, v. 17, n. 1, p. e1008667, 28 jan. 2021. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008667>. Acesso em: 16 jul. 2024.

TANG, D. *et al.* An Evolutionary Frog Leaping Algorithm for Global Optimization Problems and Applications. *Computational Intelligence and Neuroscience*, v. 2021, n. 1, p. 8928182, 1 jan. 2021. Disponível em: <https://doi.org/10.1155/2021/8928182>.

TANG, Y.; MORETTI, R.; MEILER, J. Recent Advances in Automated Structure-Based De Novo Drug Design. *Journal of Chemical Information and Modeling*, v. 64, n. 6, p. 1794–1805, 25 mar. 2024. Acesso em: 9 ago. 2024.

TEIXEIRA, J. P.; RAMALHO, T. C. Regulation of Protein Synthesis: An Approach to Treat Autism Spectrum Disorder (ASD). *Current Medicinal Chemistry*, v. 28, n. 34, p. 7141–7156, 20 abr. 2021. Disponível em: <https://www.eurekaselect.com/article/115384>. Acesso em: 8 ago. 2024.

THE AMBER DEVELOPERS. *AMBER 2024 Reference Manual*. . [S.l: s.n.]. Disponível em: <https://ambermd.org/doc12/Amber24.pdf>. , 2024

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, v. 49, n. 11, p. 3315–3321, 1 jun. 2006. Disponível em: <https://pubs.acs.org/doi/abs/10.1021/jm051197e>. Acesso em: 5 ago. 2024.

TINGLE, B. I. *et al.* ZINC-22─A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *Journal of Chemical Information and Modeling*, v. 63, n. 4, p. 1166–1176, 27 fev. 2023. Disponível em: <https://pubs.acs.org/doi/full/10.1021/acs.jcim.2c01253>. Acesso em: 5 ago. 2024.

TOSCO, P.; STIEFL, N.; LANDRUM, G. Bringing the MMFF force field to the RDKit: Implementation and validation. *Journal of Cheminformatics*, v. 6, n. 1, p. 1–4, 12 jul. 2014. Disponível em: <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-014-0037-3>. Acesso em: 9 ago. 2024.

TRAN, T. T. VAN; TAYARA, H.; CHONG, K. T. Artificial Intelligence in Drug Metabolism and Excretion Prediction: Recent Advances, Challenges, and Future Perspectives. *Pharmaceutics*, v. 15, n. 4, 1 abr. 2023. Acesso em: 9 ago. 2024.

TROPICALE, C. *et al.* Determinant Factors of Sensory in Creating Autism Learning Environment. *Environment-Behaviour Proceedings Journal*, v. 6, n. 16, p. 113–118, 27 mar. 2021. Disponível em: <https://ebpj.e-iph.co.uk/index.php/EBProceedings/article/view/2696>. Acesso em: 8 ago. 2024.

TUCKERMAN, M. E.; MARTYNA, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *The Journal of Physical Chemistry B*, v. 104, n. 2, p. 159–178, 1 jan. 2000. Disponível em: <https://doi.org/10.1021/jp992433y>.

UPTHEGROVE, R. *et al.* Autism and psychosis: Clinical implications for depression and suicide. *Schizophrenia Research*, v. 195, p. 80–85, 1 maio 2018. Acesso em: 8 ago. 2024.

URQUIZA-CARVALHO, G. A.; FRAGOSO, W. D.; ROCHA, G. B. Assessment of semiempirical enthalpy of formation in solution as an effective energy function to discriminate native-like structures in protein decoy sets. *Journal of Computational Chemistry*, v. 37, n. 21, p. 1962–1972, 5 ago. 2016. Disponível em: <https://doi.org/10.1002/jcc.24415>.

VALDÉS-TRESANCO, M. S. *et al.* Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of Chemical Theory and Computation*, v. 17, n. 10, p. 6281–6291, 12 out. 2021. Disponível em: <https://pubs.acs.org/doi/full/10.1021/acs.jctc.1c00645>. Acesso em: 13 nov. 2024.

VIDAL-LIMON, A.; AGUILAR-TOALÁ, J. E.; LICEAGA, A. M. Integration of Molecular Docking Analysis and Molecular Dynamics Simulations for Studying Food Proteins and Bioactive Peptides. *Journal of Agricultural and Food Chemistry*, v. 70, n. 4, p. 934–943, 2 fev. 2022. Acesso em: 9 ago. 2024.

VOSS, J. H.; MÜLLER, C. E. Heterotrimeric G Protein α-Subunits - Structures, Peptide-Derived Inhibitors, and Mechanisms. *Current Medicinal Chemistry*, v. 29, n. 42, p. 6359–6378, 14 mar. 2022. Acesso em: 8 ago. 2024.

WADE, A. D. *et al.* Alchemical Free Energy Estimators and Molecular Dynamics Engines: Accuracy, Precision, and Reproducibility. *Journal of Chemical Theory and Computation*, v. 18, n. 6, p. 3972–3987, 14 jun. 2022. Disponível em: <https://doi.org/10.1021/acs.jctc.2c00114>.

WANG, C.-C. *et al.* 5-HT1A-receptor agonist modified amygdala activity and amygdala-associated social behavior in a valproate-induced rat autism model. *International Journal of Neuropsychopharmacology*, v. 16, n. 9, p. 2027–2039, 1 out. 2013. Disponível em: <https://doi.org/10.1017/S1461145713000473>.

WANG, J. *et al.* Development and Testing of a General Amber Force Field. *J Comput Chem*, v. 25, p. 1157–1174, 2004. Acesso em: 9 ago. 2024.

WANG, X. *et al.* The GluN2B-Trp373 NMDA Receptor Variant is Associated with Autism-, Epilepsy-Related Phenotypes and Reduces NMDA Receptor Currents in Rats. *Neurochemical Research*, v. 47, n. 6, p. 1588–1597, 1 jun. 2022. Disponível em: <https://link-springer-com.ez26.periodicos.capes.gov.br/article/10.1007/s11064-022-03554-8>. Acesso em: 8 ago. 2024.

WARSHEL, A. *et al.* Electrostatic Basis for Enzyme Catalysis. *Chemical Reviews*, v. 106, n. 8, p. 3210–3235, 2006. Disponível em: <https://doi.org/10.1021/cr0503106>.

WATERHOUSE, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, v. 46, n. W1, p. W296–W303, abr. 2018. Disponível em: <https://doi.org/10.1093/nar/gky427>.

WOOD, D. J. *et al.* QSAR with experimental and predictive distributions: an information theoretic approach for assessing model quality. *Journal of Computer-Aided Molecular Design*, v. 27, n. 3, p. 203–219, 2013. Disponível em: <https://doi.org/10.1007/s10822-013-9639-5>.

WU, P. Y. *et al.* Loss of synaptopodin impairs mGluR5 and protein synthesis–dependent mGluR-LTD at CA3-CA1 synapses. *PNAS Nexus*, v. 3, n. 2, 1 fev. 2024. Disponível em: <https://dx.doi.org/10.1093/pnasnexus/pgae062>. Acesso em: 8 ago. 2024.

XU, P. *et al.* Structural insights into the lipid and ligand regulation of serotonin receptors. *Nature*, v. 592, n. 7854, p. 469–473, 2021. Disponível em: <https://doi.org/10.1038/s41586-021-03376-8>.

XU, Y.; VIRGIN, L. N. Probing the force field to identify potential energy. *Journal of Applied Mechanics, Transactions ASME*, v. 86, n. 10, 1 out. 2019. Acesso em: 9 ago. 2024.

YANG, Y. *et al.* Photochemical transformation of liquid crystal monomers in simulated environmental media: Kinetics, mechanism, toxicity variation and QSAR modeling. *Water Research*, v. 261, 1 set. 2024. Acesso em: 9 ago. 2024.

YANG, Z.-Q. Agonists and antagonists for group III metabotropic glutamate receptors 6, 7 and 8. *Current topics in medicinal chemistry*, v. 5, n. 9, p. 913–918, 15 ago. 2005. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/16178735/>. Acesso em: 5 ago. 2024.

YAU, M. Q. *et al.* Evaluating the performance of MM/PBSA for binding affinity prediction using class A GPCR crystal structures. *Journal of Computer-Aided Molecular Design*, v. 33, n. 5, p. 487–496, 2019. Disponível em: <https://doi.org/10.1007/s10822-019-00201-3>.

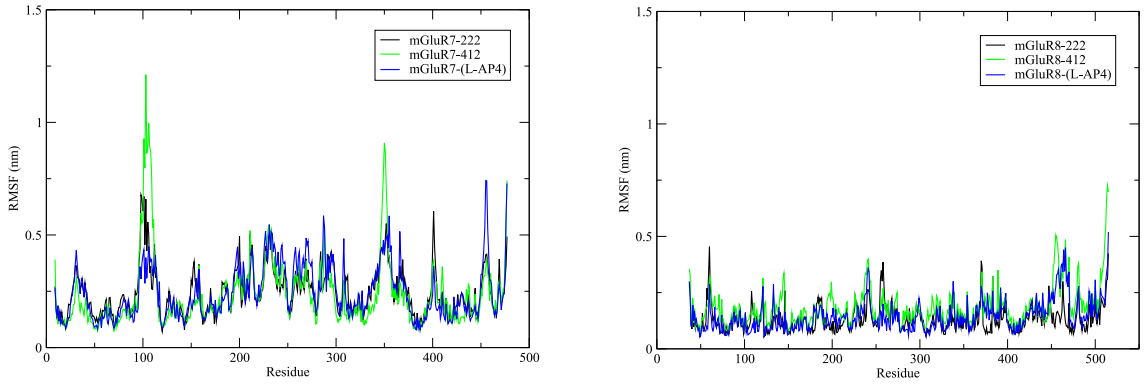
ZHAI, H.; ALEXANDROVA, A. N. Ensemble-Average Representation of Pt Clusters in Conditions of Catalysis Accessed through GPU Accelerated Deep Neural Network Fitting Global Optimization. *Journal of Chemical Theory and Computation*, v. 12, n. 12, p. 6213–6226, 13 dez. 2016. Disponível em: <https://doi.org/10.1021/acs.jctc.6b00994>.

ZHANG, J. *et al.* Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *Journal of Chemical Theory and Computation*, v. 13, n. 3, p. 1034–1043, 14 mar. 2017. Disponível em: <https://doi.org/10.1021/acs.jctc.7b00169>.

ZWECKSTETTER, M.; DITYATEV, A.; PONIMASKIN, E. Structure of serotonin receptors: molecular underpinning of receptor activation and modulation. *Signal Transduction and Targeted Therapy*, v. 6, n. 1, p. 243, 2021. Disponível em: <https://doi.org/10.1038/s41392-021-00668-3>.

APPENDIX A – SUPPLEMENTARY DATA OF PAPER 1

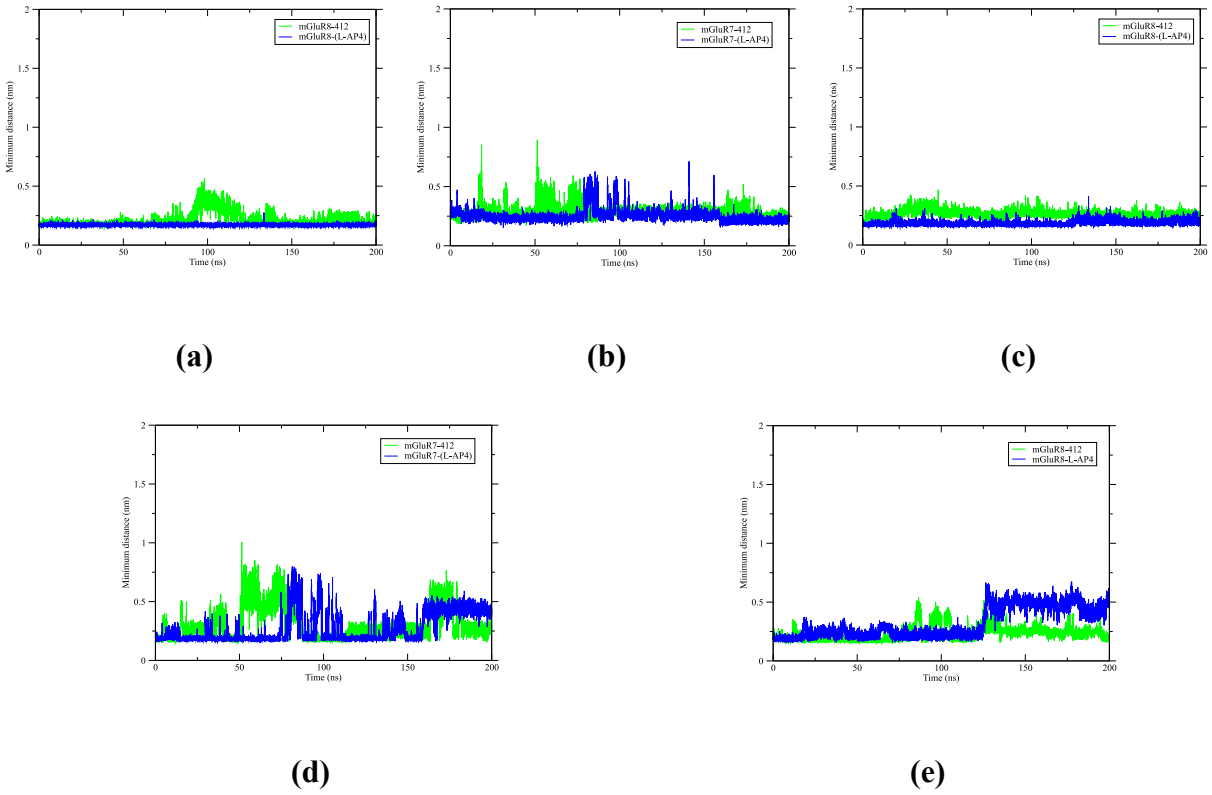
A.1 RMSF Analysis



(a) **(b)**

Figure A.1: RMSF for the protein complexes. **(a)** mGlurR7; **(b)** mGluR8.

A.2 Minimum distance analysis



(a) **(b)** **(c)**

(d) **(e)**

Figure A.2. Minimum distance plots for the best energy residues from molecular docking. **(a)** Minimum distance between Ser156 and complexed ligands with mGluR8; **(b)** Minimum distance between Gly126 and complexed ligands with mGluR7; **(c)** Minimum distance between Ser127 and complexed ligands with mGluR7; **(d)** Minimum distance between Ala177 and

complexed ligands with mGluR8; **(e)** Minimum distance between Thr179 and complexed ligands with mGluR8;

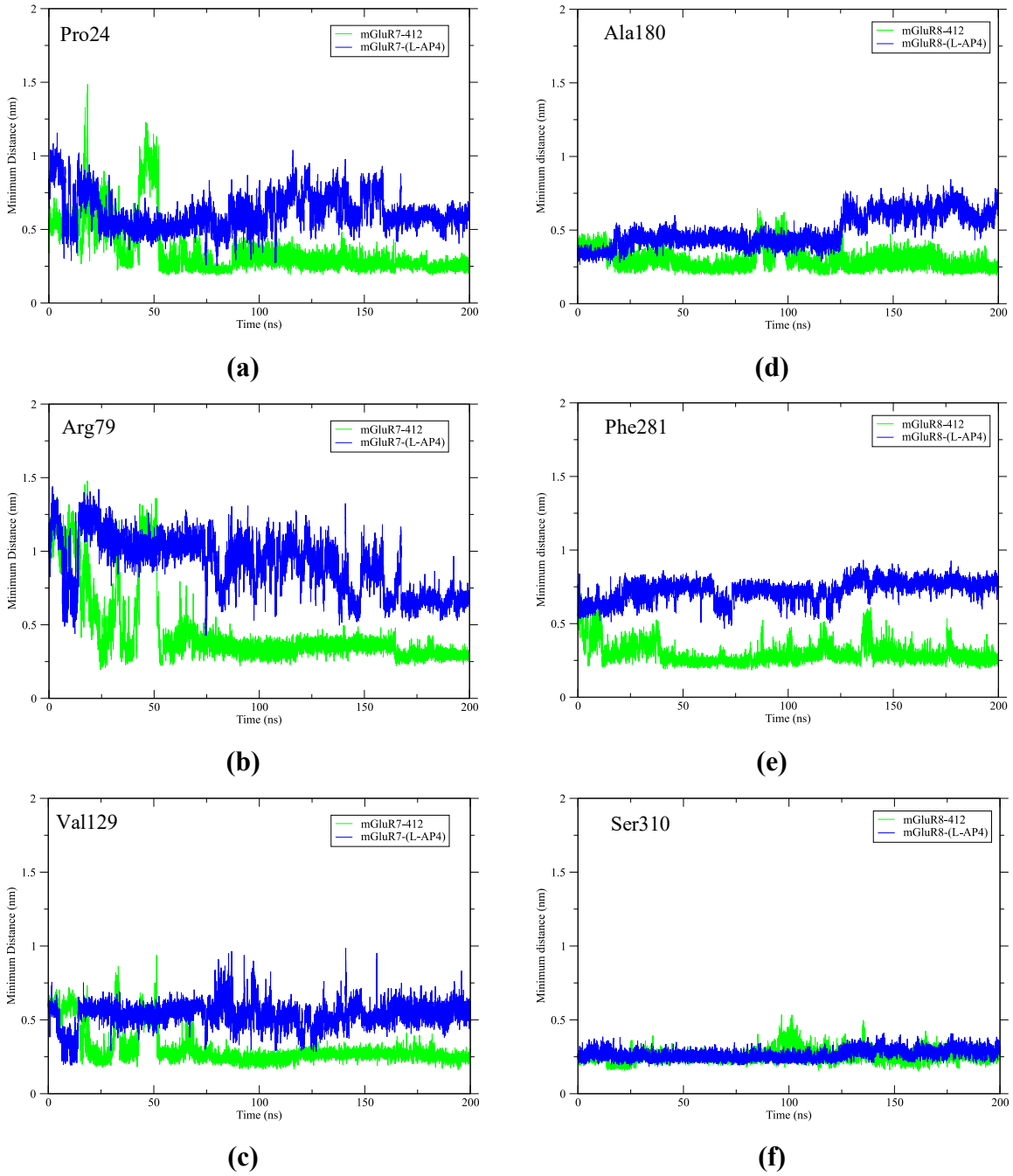
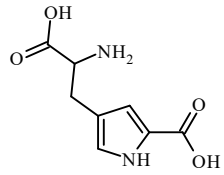
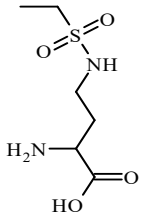


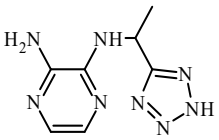
Figure A.3. Minimum distance plots for the lowest RMSF residues in the active site. **(a)** Minimum distance between Pro24 and complexed ligands with mGluR7; **(b)** Minimum distance between Arg79 and complexed ligands with mGluR7; **(c)** Minimum distance between Val129 and complexed ligands with mGluR7; **(d)** Minimum distance between Ala180 and complexed ligands with mGluR8; **(e)** Minimum distance between Phe281 and complexed ligands with mGluR8; **(f)** Minimum distance between Ser310 and complexed ligands with mGluR8;



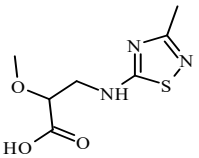
338



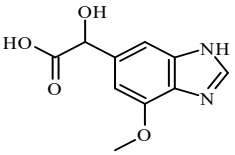
343



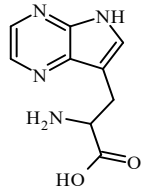
417



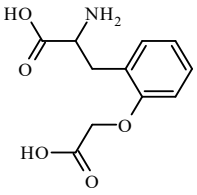
413



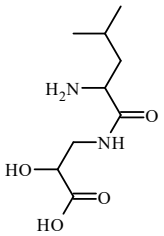
341



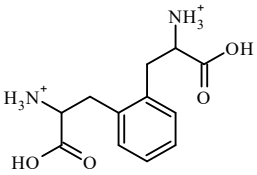
222



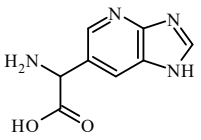
268



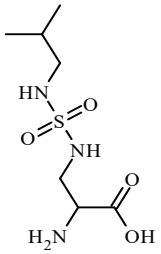
278



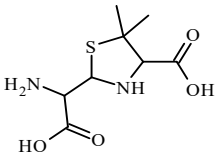
203



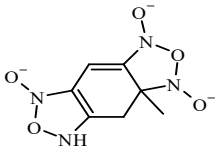
318



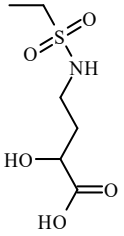
379



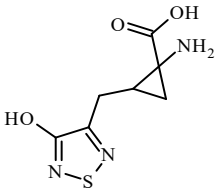
366



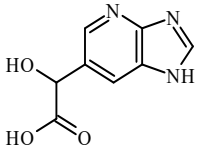
427



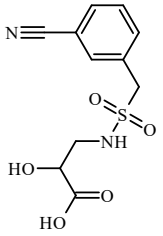
374



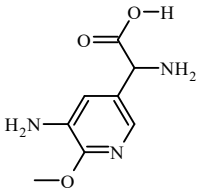
1



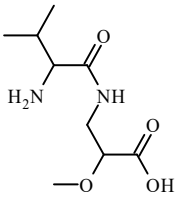
362



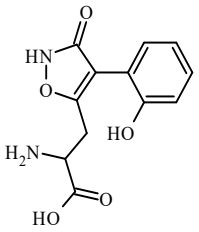
412



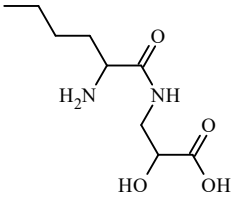
292



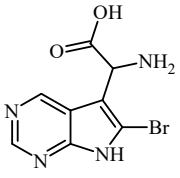
421



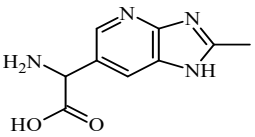
400



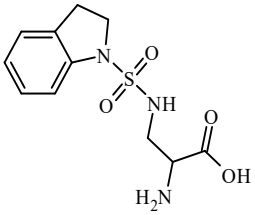
398



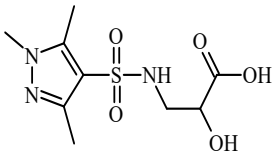
416



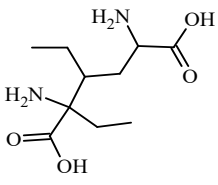
252



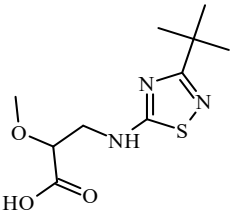
259



326



244



428

Figure A.4. 27 dual ligands structures.

A.3 Molecular dynamics simulations

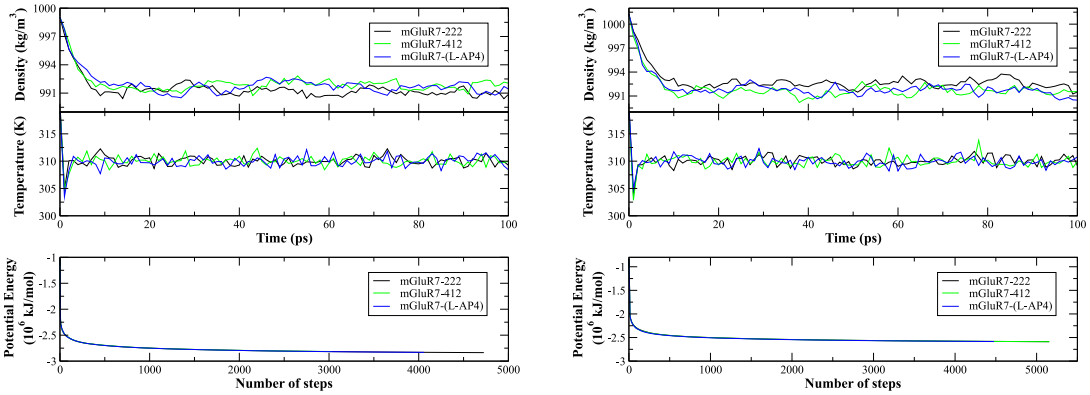


Figure A.5. Equilibration and minimization plots from molecular dynamics. The density and temperature plots are related to NPT and NVT equilibrations, respectively.
The potential energy was plotted after the last minimization step (Steepest descent without position restrain) before the 200 ns molecular dynamics.

A.4 Protein internal dielectric constant

Table A.1: Optimization of the protein's internal dielectric constant. Internal dielectric constant values were assigned to the complex with the reference ligand until the free energy matched the experimental value, as a form of model calibration.

Complex	Internal dielectric constant	$\Delta G_{binding}$
mGluR7-(L-AP4)	1	-1.77 ± 2.21
	2	-3.16 ± 1.21
	3	-8.10 ± 3.79
	4	-10.44 ± 2.41
mGluR8-(L-AP4)	1	2.35 ± 2.61
	2	-0.25 ± 1.94
	3	-2.07 ± 2.88
	4	-4.11 ± 2.47
	5	-8.97 ± 5.25
	6	-13.33 ± 4.27
	7	-16.92 ± 3.31
	8	-22.62 ± 4.45

APPENDIX B – SUPPLEMENTARY DATA OF PAPER 2

B.1 Molecular dynamics simulations

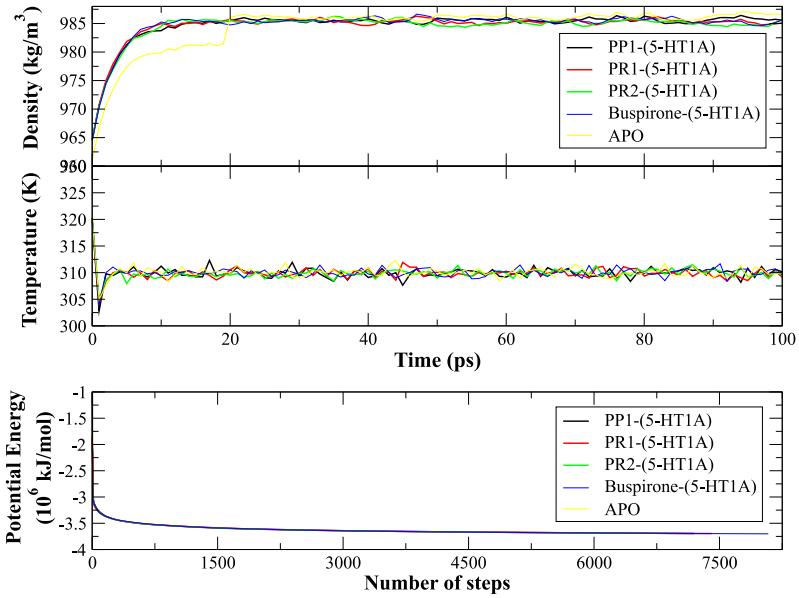


Figure B.1: The NPT equilibration converged to a density of approximately 985 kg/m³, while the NVT phase stabilized at a temperature of 310 K. Additionally, the energy minimization step using the LBFGS algorithm converged in approximately 7,500 steps for all systems.

Script B.1 (a): This Python script calculates a Tanimoto similarity matrix for a set of molecules represented by their SMILES strings. It begins by reading the SMILES from a text file, then converts each into an RDKit molecular object. Using the Extended-Connectivity Fingerprint (ECFP4) with a radius of 3 and 2048 bits, it generates molecular fingerprints that consider pharmacophoric features and chirality. A pairwise Tanimoto similarity is then computed for all molecules, resulting in a symmetrical matrix where each cell reflects the structural similarity between two compounds. Finally, the matrix is saved as a CSV file, with SMILES strings as row and column headers. This allows for easy visualization and analysis of chemical similarity across a dataset.

```
# -*- coding: utf-8 -*-
import csv
from rdkit import Chem
from rdkit.Chem import AllChem
from rdkit.DataStructs import TanimotoSimilarity
import sys

def read_smiles_from_txt(file_path):
    with open(file_path, 'r') as file:
        smiles_list = [line.strip() for line in file if line.strip()]
    return smiles_list

def generate_fingerprints(smiles_list):
    mols = [Chem.MolFromSmiles(smi) for smi in smiles_list]
    fps = []
    for mol in mols:
        if mol is None:
            fps.append(None)
        else:
            fp = AllChem.GetMorganFingerprintAsBitVect(
                mol,
                radius=3,           # ECFP4
                nBits=2048,         # Tamanho do vetor
```

```
                useFeatures=True,    # Grupos farmacofóricos
                useChirality=True    # Quiralidade
            )
            fps.append(fp)
    return fps

def compute_tanimoto_matrix(fps):
    n = len(fps)
    matrix = []
    for i in range(n):
        row = []
        for j in range(n):
            if fps[i] is None or fps[j] is None:
                row.append('')
            else:
                sim = TanimotoSimilarity(fps[i], fps[j])
                row.append(round(sim, 4))
        matrix.append(row)
    return matrix

def save_matrix_to_csv(matrix, smiles_list, output_path):
    with open(output_path, 'w', newline='') as f:
        writer = csv.writer(f)
        header = [''] + smiles_list
        writer.writerow(header)
        for smi, row in zip(smiles_list, matrix):
            writer.writerow([smi] + row)

if __name__ == '__main__':
    if len(sys.argv) != 3:
        print("Uso: python tanimoto_matrix.py entrada.txt saida.csv")
        sys.exit(1)

    input_file = sys.argv[1]
    output_file = sys.argv[2]

    smiles = read_smiles_from_txt(input_file)
    fps = generate_fingerprints(smiles)
    matrix = compute_tanimoto_matrix(fps)
    save_matrix_to_csv(matrix, smiles, output_file)

    print(f"Tanimoto matrix saved in: {output_file}")
```

Script B.1 (b): This Python script generates a heatmap of the Tanimoto similarity matrix to visualize pairwise molecular similarities based on SMILES representations. It starts by loading the matrix from a tab-delimited .txt file, where each row and column corresponds to a molecule, and the values indicate the Tanimoto similarity between them. The heatmap is then created using the Seaborn library with the "viridis" color palette to represent similarity values. To improve readability, axis labels are shown every five units, with customized tick positions and rotations. The plot includes a title indicating the use of a Morgan fingerprint with a radius of 3. Finally, the heatmap is saved as a high-resolution image file (tanimoto_heatmap.png) and displayed on the screen.

```
# -*- coding: utf-8 -*-
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np


# Carrega o arquivo
file_path = "tanimoto_matrix.txt"
df = pd.read_csv(file_path, sep='\t', index_col=0)

# Cria o heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(df, cmap='viridis', square=True, cbar_kws={'label': 'Tanimoto Index'},
            xticklabels=10, yticklabels=10)  # mostra 1 label a cada 10

# Eixos: marcações de 5 em 5 unidades
x_positions = np.arange(4, df.shape[1], 5)
y_positions = np.arange(4, df.shape[0], 5)

# Rótulos: 5, 10, 15, ...
x_labels = [str(i+1) for i in x_positions]
y_labels = [str(i+1) for i in y_positions]
```

```
plt.xticks(ticks=x_positions, labels=x_labels, rotation=90)
plt.yticks(ticks=y_positions, labels=y_labels, rotation=0)

plt.title("Tanimoto Similarity Heatmap (Radius = 3)", fontsize=14)
plt.savefig("tanimoto_heatmap.png", dpi=300)
plt.show()
```

Script B.2: This script performs a two-dimensional kernel density estimation (KDE) based on two principal components (PC1 and PC2) extracted from molecular data. Using the gaussian_kde function from the SciPy library, it calculates the probability density over a 2D grid defined in the principal component space. From this density, the Gibbs free energy (G) is computed using the thermodynamic relation $G = -k_B T \ln P$, where k_B is the Boltzmann constant and T is the temperature (310 K). The resulting data — PC1 and PC2 coordinates, probability density (P), and Gibbs free energy (G) – are saved in a tab-delimited CSV file for future analysis and plotting with software such as Origin Pro 2024. Finally, the script generates a 3D plot of the Gibbs free energy landscape using Matplotlib, displaying the surface with a gradient color map to facilitate visualization of regions with higher and lower energy.

```
# -*- coding: utf-8 -*-
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import gaussian_kde
from mpl_toolkits.mplot3d import Axes3D

k_B = 0.008314  # kJ/mol·K
T = 310         # Temperature (Kelvin)

dados = pd.read_csv('pc12.txt', sep='\t')
pc1 = dados['PC1'].values
pc2 = dados['PC2'].values

xy = np.vstack([pc1, pc2])  # shape (2, N)
kde = gaussian_kde(xy)

n_grid = 100
x_min, x_max = pc1.min(), pc1.max()
y_min, y_max = pc2.min(), pc2.max()

x = np.linspace(x_min, x_max, n_grid)
y = np.linspace(y_min, y_max, n_grid)
X, Y = np.meshgrid(x, y)

positions = np.vstack([X.ravel(), Y.ravel()])
P = kde(positions).reshape(X.shape)

P[P == 0] = 1e-12

#G(npT)
G = -k_B * T * np.log(P)

output_data = pd.DataFrame({
    'PC1': X.ravel(),
    'PC2': Y.ravel(),
    'P': P.ravel(),
    'G': G.ravel()
})
output_data.to_csv('fel_kde_output.txt', sep='\t', index=False)

#plots
fig = plt.figure(figsize=(10, 7))
ax = fig.add_subplot(111, projection='3d')
surf = ax.plot_surface(X, Y, G, cmap='viridis')

ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
ax.set_zlabel('Energia Livre de Gibbs (kJ/mol)')
ax.set_title('Paisagem de Energia Livre de Gibbs (KDE)')

plt.colorbar(surf, label='G (kJ/mol)')
plt.tight_layout()
plt.show()
```

B.2 Homology modelling validation

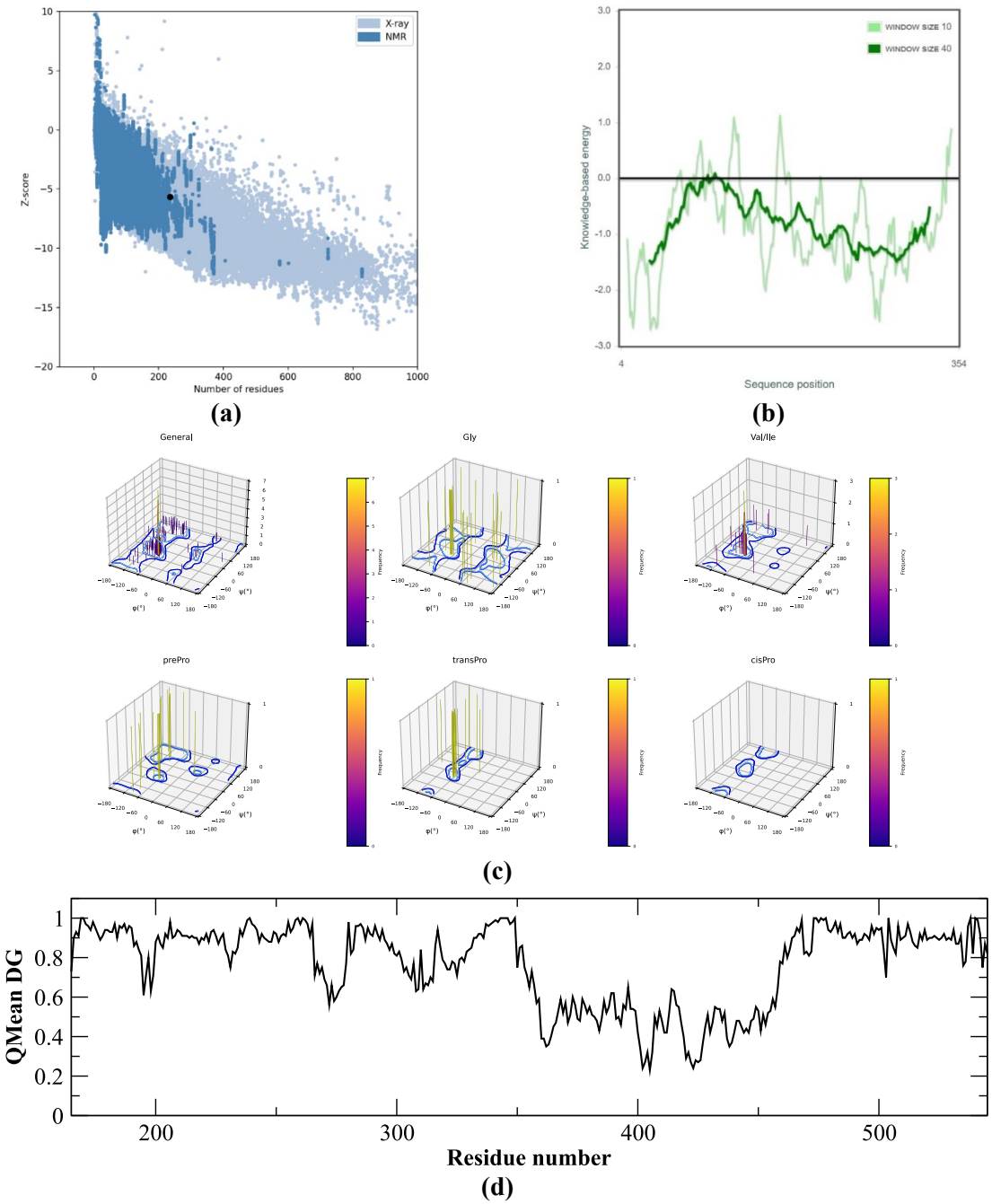


Figure B.2: In **(a)**, the graph shows the Z-score data of -5.18, indicating that the model falls within an acceptable range compared to experimental X-ray and NMR structures. In **(b)**, the figure presents the energy values of each residue in the model, with the majority exhibiting negative energies. Panel **(c)** displays the three-dimensional Ramachandran plot divided into general residues, glycines, valine/isoleucine, pre-proline, trans-proline, and cis-proline categories. All residues located in disallowed regions are few and correspond to low frequencies of approximately 1. Finally, **(d)** shows the per-residue global QMEANDisCo score graph generated by the Swiss-Model.

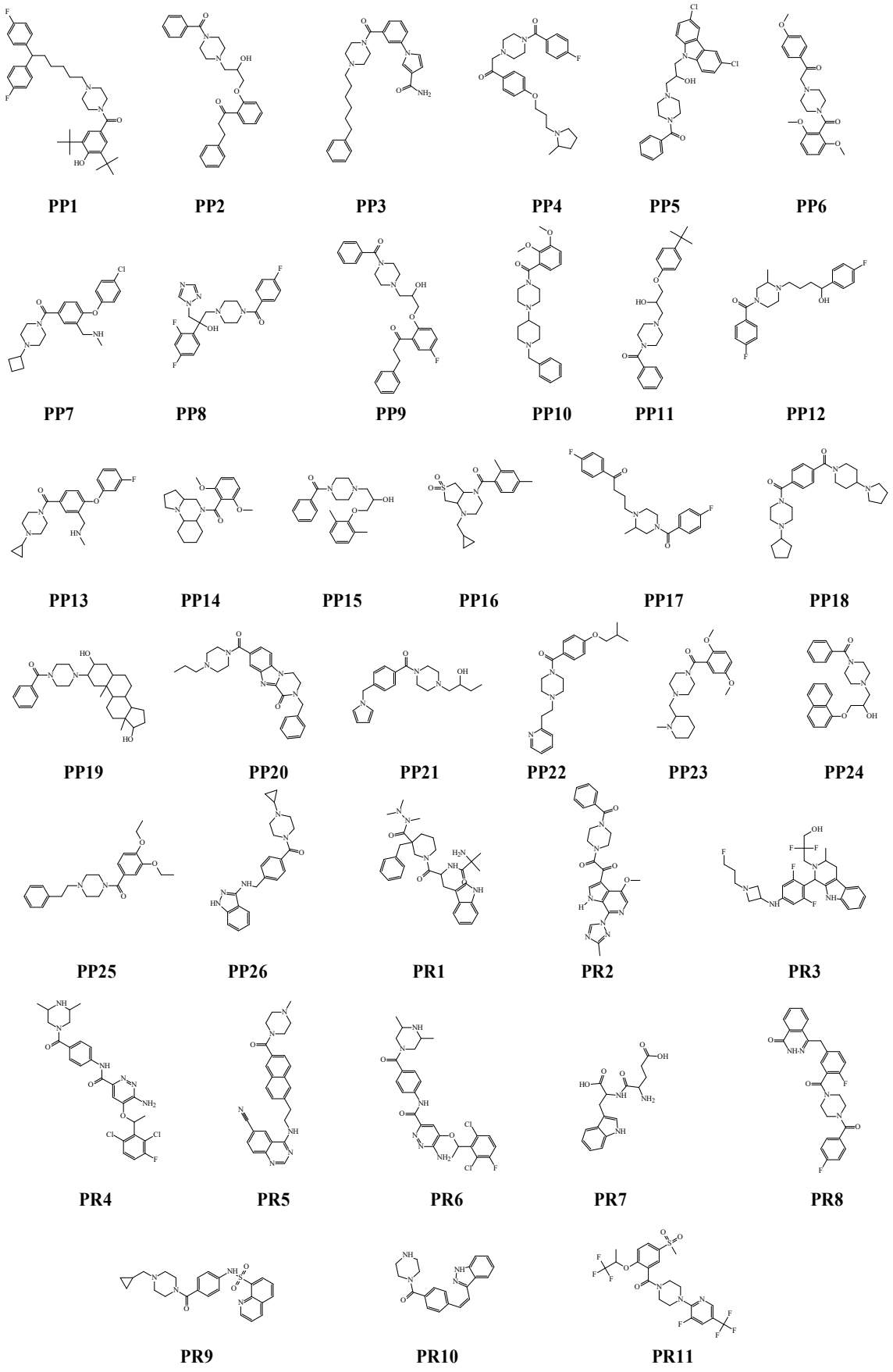


Figure B.3: Chemical structures of the proposals from MLR-QSAR.

B.3 Interatomic distances

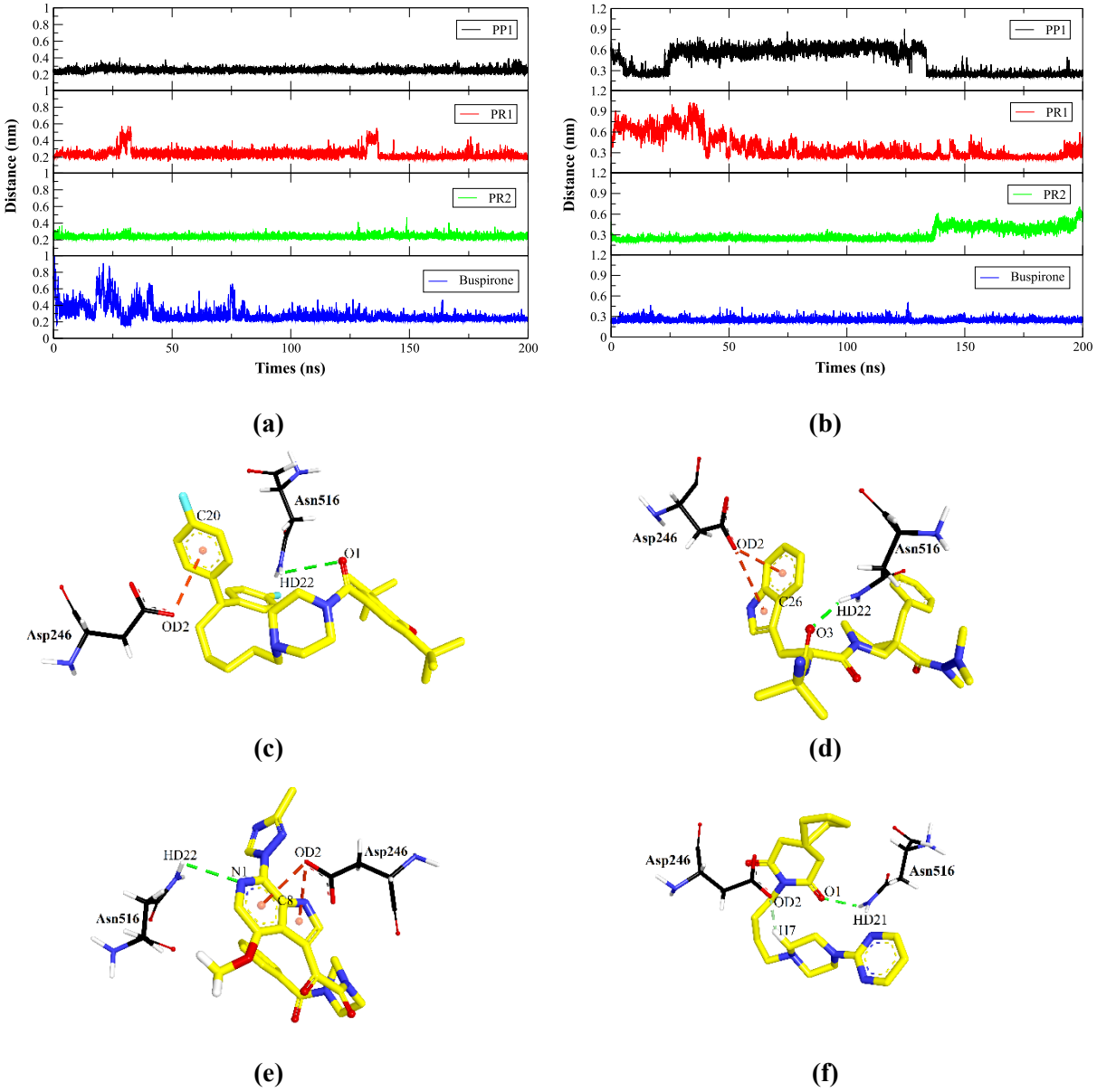


Figure B.4: (a) Distance plots for atoms HD22 and O1 for PP1, HD22 and O3 for PR1, HD22 and N1 for PR2 and HD21 and O1 for Buspirone. These distances are related to the hydrogen bond formation between those atoms; (b) Distance plots for atoms OD2 and C20 for PP1, OD2 and C25 for PR1, OD2 and C8 for PR2 and OD2 and H7 for Buspirone. These distances are related to the π-anion and non-conventional hydrogen bonds formation; (c) 3D distance scheme used for PP1; (d) 3D distance scheme used for PR1; (e) 3D distance scheme used for PR2; (f) 3D distance scheme used for Buspirone.

B.4 PCA analysis

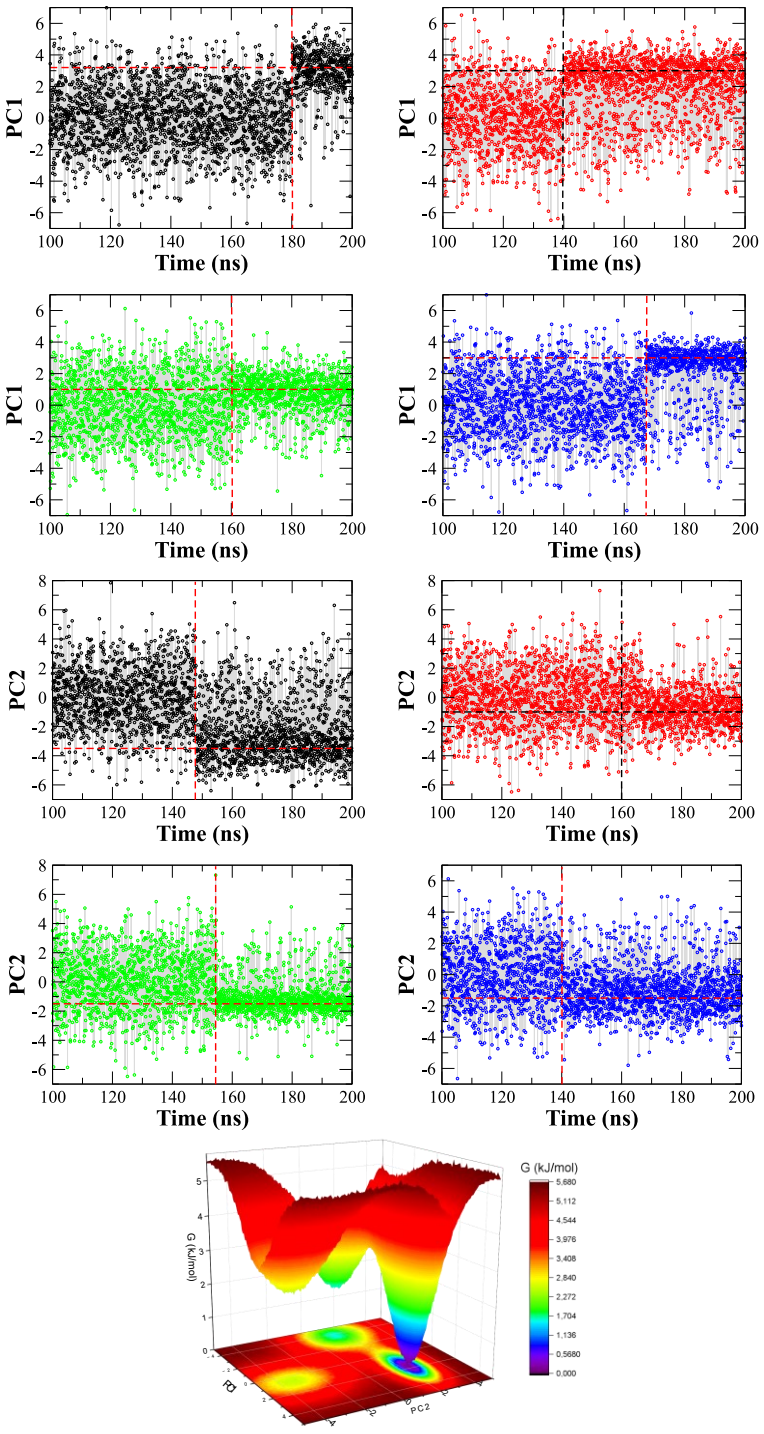


Figure B.5: The figure shows the plots of the principal components PC1 and PC2 over time. Black, red, green, and blue correspond respectively to PC1 and PC2 for PP1, PR1, PR2, and Buspirone. Overall, the complexes exhibited more condensed PC1 and PC2 values during the last 50 ns, which correspond to the principal components associated with the lowest energy wells in the free energy landscapes (FELs). Additionally, at the end, there is an image of the 3D/2D FEL for the APO system, showing three distinct wells, indicating the formation of more than one set of low-energy conformations in the absence of ligands.

APPENDIX C – FORMALISMS AND DERIVATIONS

C.1 Lagrangean and Hamiltonian mechanics of conservative systems

According to Arnold (1989) and Moretti (2023) (ARNOLD, 1989; MORETTI, 2023), given a system composed of N particles of mass m_i and generalized coordinates ($\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$,…, $\mathbf{q}_i$), whose velocities are given by the time derivatives of the position ($\dot{\mathbf{q}}_1$, $\dot{\mathbf{q}}_2$, $\dot{\mathbf{q}}_3$,…, $\dot{\mathbf{q}}_i$), with momenta ($\dot{\mathbf{p}}_1$, $\dot{\mathbf{p}}_2$, $\dot{\mathbf{p}}_3$,…, $\dot{\mathbf{p}}_i$), the Lagrangian as a function of positions, velocities and time, given the time dependence implicit in the kinetic energy term, can be defined as

$$\mathcal{L}(\mathbf{q}_i, \dot{\mathbf{q}}_i) = \mathcal{K}(\dot{\mathbf{q}}_i) - \mathcal{V}(\mathbf{q}_i) \tag{C.1.1}$$

considering that all components of the system are subjected to a conservative potential energy field, without external potentials, with a quadratic kinetic energy function, the sum of the energies of each particle by the Lagrangian function will be

$$\mathcal{L}(\mathbf{q}_i, \dot{\mathbf{q}}_i) = \sum_i^N \frac{1}{2} m_i \mathbf{q}_i^2 - \sum_i^N \mathcal{V}(\mathbf{q}_i) \tag{C.1.2}$$

the total differential of (C.1.1), given the Legendre transform, will be

$$d\mathcal{L} = \sum_i^N \left(\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} d\mathbf{q}_i + \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i} d\dot{\mathbf{q}}_i \right) + \frac{\partial \mathcal{L}}{\partial t} dt \tag{C.1.3}$$

Extracting the derivatives of the position and velocities of (C.1.2), according to the total differential, we will have

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} = -\sum_I^N \frac{\partial \mathcal{V}}{\partial \mathbf{q}_i} = -\nabla \mathcal{V} = \mathbf{F} \tag{C.1.4}$$

where the force is calculated by the negative gradient of the potential energy. The force can be written in terms of the momentum. Given the definition:

$$\mathbf{p}_i = m_i \dot{\mathbf{q}}_i = m_i \frac{d\mathbf{q}_i}{dt}$$

differentiating with respect to time,

$$\frac{d\mathbf{p}_i}{dt} = \dot{\mathbf{p}} = \frac{d^2 \mathbf{q}_i}{dt^2} = m_i \ddot{\mathbf{p}}_i = \mathbf{F} \tag{C.1.5}$$

equating (C.1.4) and (C.1.5), we obtain an important quantity extracted from the Lagrangian.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} = \dot{\mathbf{p}} \tag{C.1.6}$$

Deriving the Lagrangian (4) with respect to the velocities ($\dot{\mathbf{q}}_i$), we obtain

$$\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i} = \frac{\partial}{\partial \dot{\mathbf{q}}_i} \sum_i^N \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 = \sum_i^N m_i \dot{\mathbf{q}}_i$$

since the momentum is defined by $\mathbf{p}_i = m_i\dot{\mathbf{q}}$, we obtain another important quantity:

$$\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i} = \mathbf{p}_i \tag{C.1.7}$$

using the quantities in (C.1.6) and (C.1.7) in the total differential (C.1.3), we obtain

$$d\mathcal{L} = \sum_i^N (\dot{\mathbf{p}}\, d\mathbf{q}_i + \mathbf{p}_i\, d\dot{\mathbf{q}}_i) + \frac{\partial \mathcal{L}}{\partial t}\, dt \tag{C.1.8}$$

Having defined the Lagrangian function, we can write it in terms of positions and momenta, thus defining the Hamiltonian function (ARNOLD, 1989; SALINAS, 1997). Starting from (C.1.3):

$$\begin{aligned}
d\mathcal{L} &= \sum_i^N \left(\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}_i}\, d\dot{\mathbf{q}}_i \right) + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
&= \sum_i^N \left(\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + \mathbf{p}_i\, d\dot{\mathbf{q}}_i \right) + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
&= \sum_i^N \left[\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + d(\mathbf{p}_i\, \dot{\mathbf{q}}_i) \right] + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
&= \sum_i^N \left(\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + \mathbf{p}_i\, d\dot{\mathbf{q}}_i + \dot{\mathbf{q}}_i\, d\mathbf{p}_i - \dot{\mathbf{q}}_i\, d\mathbf{p}_i \right) + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
&= \sum_i^N \left(\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + d(\mathbf{p}_i\, \dot{\mathbf{q}}_i) - \dot{\mathbf{q}}_i\, d\mathbf{p}_i \right) + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
&= \sum_i^N \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + \sum_i^N d(\mathbf{p}_i\, \dot{\mathbf{q}}_i) - \sum_i^N \dot{\mathbf{q}}_i\, d\mathbf{p}_i + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
d\mathcal{L} &= \sum_i^N \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + d\left(\sum_i^N \mathbf{p}_i\, \dot{\mathbf{q}}_i \right) - \sum_i^N \dot{\mathbf{q}}_i\, d\mathbf{p}_i + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
d\mathcal{L} - d\left(\sum_i^N \mathbf{p}_i\, \dot{\mathbf{q}}_i \right) &= \sum_i^N \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i - \sum_i^N \dot{\mathbf{q}}_i\, d\mathbf{p}_i + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
-d\left(\sum_i^N \mathbf{p}_i\, \dot{\mathbf{q}}_i - \mathcal{L} \right) &= \sum_i^N \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i - \sum_i^N \dot{\mathbf{q}}_i\, d\mathbf{p}_i + \frac{\partial \mathcal{L}}{\partial t}\, dt \\
d\left(\sum_i^N \mathbf{p}_i\, \dot{\mathbf{q}}_i - \mathcal{L} \right) &= -\sum_i^N \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i}\, d\mathbf{q}_i + \sum_i^N \dot{\mathbf{q}}_i\, d\mathbf{p}_i - \frac{\partial \mathcal{L}}{\partial t}\, dt
\end{aligned}$$

thus, we define the Hamiltonian function as

$$\mathcal{H}(\mathbf{p}, \mathbf{q}, t) = \sum_i^N \mathbf{p}_i\, \dot{\mathbf{q}}_i - \mathcal{L} \tag{C.1.9}$$

so,

$$d\mathcal{H} = \sum_i^N \left(-\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} d\mathbf{q}_i + \dot{\mathbf{q}}_i d\mathbf{p}_i \right) - \frac{\partial \mathcal{L}}{\partial t} dt \tag{C.1.10}$$

The total differential of the Hamiltonian can now be written in terms of its original variables, that is

$$d\mathcal{H} = \sum_i^N \left(\frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} d\mathbf{p}_i + \frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} d\mathbf{q}_i \right) + \frac{\partial \mathcal{H}}{\partial t} dt \tag{C.1.11}$$

taking the equality of (C.1.10) and (C.1.11), we arrive at Hamilton's equations:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{p}_i} = \dot{\mathbf{q}}_i \tag{C.1.12}$$

using equation (C.1.6)

$$\frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} = -\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} \quad \therefore \quad \frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} = -\dot{\mathbf{p}}_i \tag{C.1.13}$$

$$\frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t} \tag{C.1.14}$$

finally, substituting Hamilton's equations into the total differential of the Hamiltonian, given the implicit time dependence of the kinetic energy function, the Hamiltonian becomes

$$d\mathcal{H}(\mathbf{p}_i, \mathbf{q}_i) = \sum_i^N (\dot{\mathbf{q}}_i d\mathbf{p}_i - \dot{\mathbf{p}}_i d\mathbf{q}_i) \tag{C.1.15}$$

Once the functions are defined, it is easy to show, from $\mathcal{H}$ and $\mathcal{L}$ that the total energy of a classical conservative system is the Hamiltonian of the system itself. Combining (C.1.9) and (C.1.2):

$$\begin{aligned}
\mathcal{H}(\mathbf{p}, \mathbf{q}, t) &= \sum_i^N \mathbf{p}_i \dot{\mathbf{q}}_i - \sum_i^N \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 + \sum_i^N \mathcal{V}(\mathbf{q}_i) \\
&= \sum_i^N m_i \dot{\mathbf{q}}_i \dot{\mathbf{q}}_i - \sum_i^N \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 + \sum_i^N \mathcal{V}(\mathbf{q}_i) \\
&= \sum_i^N m_i \dot{\mathbf{q}}_i^2 - \sum_i^N \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 + \sum_i^N \mathcal{V}(\mathbf{q}_i) \\
&= \sum_i^N \frac{1}{2} m_i \dot{\mathbf{q}}_i^2 + \sum_i^N \mathcal{V}(\mathbf{q}_i) \\
\mathcal{H}(\dot{\mathbf{q}}, \mathbf{q}, t) &= \sum_i^N \mathcal{K}(\dot{\mathbf{q}}) + \sum_i^N \mathcal{V}(\mathbf{q}_i)
\end{aligned} \tag{C.1.16}$$

However, since kinetic energy can be written in terms of momenta,

$$\mathcal{K}(\mathbf{p}_i) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} \tag{C.1.17}$$

the Hamiltonian function can be written in terms of momenta and position only, this being the definition of the total energy of a conservative classical system.

$$\mathcal{H}(\mathbf{p}_i, \mathbf{q}_i) = \mathcal{K}(\mathbf{p}_i) + \mathcal{V}(\mathbf{q}_i) \tag{C.1.18}$$

C.2 Velocity integration schemes

The algorithms for integrating the equations of motion are defined by expansions of the Taylor series of the type

$$\frac{d\mathbf{q}(t+\delta t)}{dt} = \frac{d\mathbf{q}(t)}{dt} + \sum_{k=1}^{\infty} \frac{1}{k!} \delta t^k \frac{d^{k+1}\mathbf{q}(t)}{dt^{k+1}} \tag{2.1.24}$$

with this, the equations for velocity, acceleration and higher derivatives with respect to the generalized position $\mathbf{q}_i$ will be

$$\begin{aligned} \mathbf{v} = \dot{\mathbf{q}}_i(t+\delta t) &= \frac{d\mathbf{q}_i(t)}{dt} + \delta t \frac{d^2\mathbf{q}_i(t)}{dt^2} + \frac{1}{2}\delta t^2 \frac{d^3\mathbf{q}_i(t)}{dt^3} + \frac{1}{6}\delta t^3 \frac{d^4\mathbf{q}_i(t)}{dt^4} + \cdots \\ \mathbf{a} = \ddot{q}_i(t+\delta t) &= \frac{d^2\mathbf{q}_i(t)}{dt^2} + \delta t \frac{d^3\mathbf{q}_i(t)}{dt^3} + \frac{1}{2}\delta t^2 \frac{d^4\mathbf{q}_i(t)}{dt^4} + \cdots \\ \mathbf{b}(t+\delta t) &= \frac{d^3\mathbf{q}_i(t)}{dt^3} + \delta t \frac{d^4\mathbf{q}_i(t)}{dt^4} + \cdots \end{aligned} \tag{2.1.25}$$

For the Verlet algorithm, the equation is derived by expanding the position $\mathbf{q}_i$ forward and backward in time:

$$\begin{aligned} \mathbf{q}_i(t+\delta t) &= \mathbf{q}_i(t) + \delta t \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t^2 \frac{d^2\mathbf{q}_i}{dt^2} + \cdots \\ \mathbf{q}_i(t-\delta t) &= \mathbf{q}_i(t) - \delta t \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t^2 \frac{d^2\mathbf{q}_i}{dt^2} - \cdots \end{aligned} \tag{C.2.1}$$

by adding these two expressions, the terms involving the first derivative cancel out, leading to the final update equation:

$$\mathbf{q}_i(t+\delta t) = 2\mathbf{q}_i(t) - \mathbf{q}_i(t-\delta t) + \delta t^2 \frac{d^2\mathbf{q}_i}{dt^2} \tag{C.2.2}$$

For the Leap-Frog algorithm (1970), velocities are evaluated at smaller integer steps. The velocity at time *t*, and a correction term to the force (acceleration) upon the particle (TANG, D. *et al.*, 2021) is given by:

$$\mathbf{q}_i(t+\delta t) = \mathbf{q}_i(t) + \delta t \frac{d}{dt}\mathbf{q}_i(t) + \frac{1}{2}\delta t \frac{d^2\mathbf{q}_i}{dt^2} \tag{C.2.3}$$

in the second expression, the velocity at the intermediate time $t + \delta t/2$ is defined in terms of the velocity at time *t* and the acceleration at *t*, including the change in velocity due to the force:

$$\dot{\mathbf{q}}_i(t+\frac{1}{2}\delta t) = \frac{d}{dt}\mathbf{q}_i(t) - \frac{1}{2}\delta t\frac{d^2\mathbf{q}_i}{dt^2} + \delta t\frac{d^2\mathbf{q}_i(t)}{dt^2} \tag{C.2.4}$$

finally, the third equation shows that the velocity at time *t* can be reconstructed as the average of its forward and backward staggered values:

$$\dot{\mathbf{q}}_i(t) = \frac{1}{2}\left(\frac{d}{dt}\mathbf{q}(t) + \frac{1}{2}\delta t\frac{d^2\mathbf{q}_i}{dt^2} + \frac{d}{dt}\mathbf{q}(t) - \frac{1}{2}\delta t\frac{d^2\mathbf{q}_i}{dt^2}\right) \tag{C.2.5}$$

For Velocity-Verlet algorithm, the position is updated using both the velocity and the acceleration at time *t*:

$$\mathbf{q}_i(t+\delta t) = \mathbf{q}_i(t) + \delta t\frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t^2\frac{d^2\mathbf{q}_i(t)}{dt} \tag{C.2.6}$$

then, the velocity at the next time step is calculated as the initial velocity plus the average of the acceleration at time *t* and at time $t + \delta t/2$, which requires the force to be evaluated again at the new position:

$$\dot{\mathbf{q}}_i(t+\delta t) = \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t\left(\frac{d^2\mathbf{q}_i(t)}{dt^2} + \frac{d^2\mathbf{q}_i(t+\delta t)}{dt^2}\right) \tag{C.2.7}$$

An auxiliary expression shows the intermediate form of the velocity update, useful in practice for force evaluation.

$$\dot{\mathbf{q}}_i(t+\frac{1}{2}\delta t) = \frac{d\mathbf{q}_i(t)}{dt} + \frac{1}{2}\delta t\frac{d^2\mathbf{q}_i}{dt^2} \tag{C.2.8}$$

C.3 Electrostatic potential for spherical geometries

Gauss's law (WARSHEL *et al.*, 2006) (C.3.1) defines the total electric flux (Φ_E) in terms of the closed surface integral of the inner product of the electric field vector (**E**) by the area vector ($d\mathbf{A}$) of the Gaussian surface. Furthermore, the total electric flux is also equivalent to the quotient of the internal charge of the system (q_{int}) by the vacuum permittivity constant (ε_0):

$$\Phi_E = \oint_S \mathbf{E}\cdot d\mathbf{A} = \frac{q_{int}}{\varepsilon_0} \tag{C.3.1}$$

The vector $d\mathbf{A}$ is perpendicular at each point on the Gaussian surface, so it will always have the same direction as the electric field vector. If the surface charge is positive, the electric field will have the same direction and sense as the area vector. In the case of negative charges, the electric field will have the same direction, but opposite sense to the area vector. Therefore, if $q_i = q_-$, the angle between the vectors will be zero and the inner product

$$\begin{aligned}\mathbf{E}\cdot d\mathbf{A} &= |\mathbf{E}||d\mathbf{A}|\cos\theta = E\,dA\cos\theta \\ &= E\,dA\cos 0\end{aligned}$$

$$= E\,dA$$

and if $q_i = q_+$, the angle between the vectors will be π and the inner product

$$\begin{aligned}\mathbf{E}\cdot d\mathbf{A} &= E\,dA\cos\theta\\ &= E\,dA\cos\pi\\ &= -E\,dA\end{aligned}$$

Generally, we always apply the electric field module, that is,

$$E = |\mathbf{E}| = \begin{cases} E, & \theta > 0 \\ -E, & \theta < 0 \end{cases} \tag{C.3.2}$$

Another important consideration is that the electric field will always be a function of the radius of the Gaussian surface, $\mathbf{E} = f(\mathbf{r})$. Therefore, applying Gauss's law under these conditions, we have

$$\begin{aligned}\Phi_E &= \oint_S \mathbf{E}\cdot d\mathbf{A}\\ &= \oint_S E(\mathbf{r})\,dA\\ &= E(\mathbf{r})\oint_S dA\end{aligned}$$

where the line integral is the surface area of the sphere, which can be solved given the appropriate transformation into spherical coordinates:

$$\oint_S dA = \int_0^{2\pi}\int_0^{\pi} r^2\sin\theta\,d\theta d\phi = 4\pi r^2 \tag{C.3.3}$$

with this, Gauss's law becomes

$$\Phi_E = 4\pi\mathbf{r}^2 E(\mathbf{r}) = \frac{q_{int}}{\varepsilon_0} \tag{C.3.4}$$

Since the internal charge of the Gaussian surface will be the same charge q_i as the conducting surface of the sphere, the electric field will be

$$E(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}^2} \tag{C.3.5}$$

where $\mathbf{r}$ is the radius of the sphere, given the vector's module $\mathbf{r} = |\mathbf{r}_i - \mathbf{r}_0|$. Therefore, the electric field can also be written in the form

$$E(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}_i - \mathbf{r_0}|^2} \tag{C.3.6}$$

In terms now of the electric potential of a charged sphere, let us consider the strength of a conservative field of potential energy ($\mathcal{V}$) to be, according to equation (2.13)

$$\mathbf{F}(\mathbf{r}) = -\vec{\nabla}\mathcal{V}(\mathbf{r}) = -\left(\frac{\partial \mathcal{V}}{\partial \mathbf{r}_1}, \frac{\partial \mathcal{V}}{\partial \mathbf{r}_2}, \frac{\partial \mathcal{V}}{\partial \mathbf{r}_3}, \ldots\right)$$

The work done to move a body along a path C, from position $\mathbf{r}_0$ to $\mathbf{r}_i$ is (MORTIMER, 2008)

$$w = \int_{\mathbf{r}_0}^{\mathbf{r}_i} \mathbf{F}(\mathbf{r})\, d\mathbf{r} \tag{C.3.7}$$

Multiplying the equation (2.13) by $d\mathbf{r}$ and integrating, we will have an important relationship between work (C.3.7) and potential energy:

$$w = \int_{\mathbf{r}_0}^{\mathbf{r}_i} \mathbf{F}(\mathbf{r})\, d\mathbf{r} = -\int_{\mathbf{r}_0}^{\mathbf{r}_i} \vec{\nabla}\mathcal{V}(\mathbf{r})\, d\mathbf{r}$$

$$w = -\left[\mathcal{V}(\mathbf{r}_i) - \mathcal{V}(\mathbf{r}_0)\right] = -\Delta U \tag{C.3.8}$$

Following the definitions of electric potential $\varphi(\mathbf{r})$ (NUSSENZVEIG, 2003)

$$\varphi(\mathbf{r}) = \frac{\mathcal{V}(\mathbf{r})}{q} \tag{C.3.9}$$

and electric field,

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{q} \tag{C.3.10}$$

Given definition (C.3.9), the difference between two electric potentials $\varphi_2 - \varphi_1$ will be

$$\varphi_2 - \varphi_1 = \frac{1}{q}(\mathcal{V}_2 - \mathcal{V}_1) = \frac{\Delta \mathcal{V}}{q} \tag{C.3.11}$$

Equation (C.3.8) allows us to write (C.3.11) as

$$\begin{aligned} \varphi_2 - \varphi_1 &= -\frac{w}{q} \\ &= \frac{1}{q}\int_{\mathbf{r}_0}^{\mathbf{r}_i} \mathbf{F}(\mathbf{r})\, d\mathbf{r} \\ &= \int_{\mathbf{r}_0}^{\mathbf{r}_i} \frac{\mathbf{F}(\mathbf{r})}{q}\, d\mathbf{r} \end{aligned}$$

but by definition (C.3.10),

$$\varphi_2 - \varphi_1 = \int_{\mathbf{r}_0}^{\mathbf{r}_i} \mathbf{E}(\mathbf{r})\, d\mathbf{r}$$

The electric potential at infinite distances is zero, since one is no longer under the influence of the body's electric field. Thus, taking the modulus of the vector $\mathbf{E}$ and setting $\varphi_2(\infty) = 0$ and $\varphi_1 = \varphi(\mathbf{r})$, we will have

$$\varphi(\mathbf{r}) = \int_{\mathbf{r}_0}^{\infty} E(\mathbf{r})\, d\mathbf{r} \tag{C.3.12}$$

the final expression for the electric potential, also known as the electrostatic potential. The higher the value of the electric potential, the greater the magnitude of the work done by a body at a given point in an electric field.

For a spherical geometry, we can substitute the function $\mathbf{E}(\mathbf{r})$ from (34). Let $\mathbf{r}_0$ be the distance of the vector $\mathbf{r}$, we have:

$$\begin{aligned}\varphi(\mathbf{r}) &= \int_{\mathbf{r}}^{\infty} \frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}^2}\,d\mathbf{r} \\ &= -\frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}}\Big|_{\mathbf{r}}^{\infty} \\ &= -\lim_{\mathbf{r}\to\infty}\frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}} - \left(-\lim_{\mathbf{r}\to\mathbf{r}}\frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}}\right)\end{aligned}$$

therefore, the final expression for the electric potential of a charged conducting sphere is

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}} \tag{C.3.13}$$

or, analogously to what was done with the final expression for the electric field,

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{|\mathbf{r}_i - \mathbf{r_0}|} \tag{C.3.14}$$

C.4 The Poisson-Boltzmann equation

The Poisson equation can be obtained from the differential form of Gauss's law (NUSSENZVEIG, 2003) (C.3.1), given the application of the divergence theorem in its integral form:

$$\nabla \cdot \mathbf{E} = \frac{\rho(\mathbf{r})}{\varepsilon} \tag{C.4.1}$$

Given the electric field relation as being the negative gradient of the electric potential $\mathbf{E} = -\nabla^2\varphi(\mathbf{r})$, this application of Gauss's law leads directly to Poisson's equation:

$$\nabla^2\varphi(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\varepsilon} \tag{C.4.2}$$

Boltzmann's contribution is given from the distribution of the electrostatic potential in terms of ion concentrations. The total chemical potential μ can be written as under the influence of an electric field is:

$$\begin{aligned}\mu &= \mu_{ideal} + \mathcal{V}_{el} \\ \mu &= \mu^0 + k_B T \ln\frac{c_i}{c_i^0} + \mathcal{V}_{el}\end{aligned} \tag{C.4.3}$$

Previously, electric potential energy was defined as the product of the charge and the electric potential $\mathcal{V}_{el} = q_i\,\varphi(\mathbf{r})$ (C.3.9). However, for a system of ions, the total charge corresponds to its valence z_i and the elementary charge $\mathbf{e}$. With this, $q_i = z_i\mathbf{e}$ and electric potential energy becomes, in the chemical potential

$$\mu = \mu^0 + k_B T \ln \frac{c_i}{c_i^0} + z_i \mathbf{e} \varphi(\mathbf{r}) \tag{C.4.4}$$

The equilibrium condition of this system requires equality of chemical potentials, that is $\mu = \mu^0$, leading us to

$$k_B T \ln \frac{c_i}{c_i^0} + z_i \mathbf{e} \varphi(\mathbf{r}) = 0$$

Solving the equation for the concentration c_i:

$$k_B T \ln \frac{c_i}{c_i^0} = -z_i \mathbf{e} \varphi(\mathbf{r})$$

$$\ln \frac{c_i}{c_i^0} = -\frac{z_i \mathbf{e} \varphi(\mathbf{r})}{k_B T}$$

$$\frac{c_i}{c_i^0} = e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T}$$

finally,

$$c_i = c_i^0 e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} \tag{C.4.5}$$

the final expression that corresponds to the Boltzmann distribution of the electric potential of an electrolyte solution. To combine the Poisson (C.4.2) and Boltzmann (C.4.5) equations, we start from the definition of the charge density, expressed by

$$\rho(\mathbf{r}) = \sum_i^N z_i \mathbf{e} c_i \tag{C.4.6}$$

which depends on the concentration of electrolytes in solution, given by the Boltzmann distribution. Combining (C.4.5) and (C.4.6),

$$\rho(\mathbf{r}) = \sum_i^N z_i \mathbf{e} c_i^0 e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} \tag{C.4.7}$$

which substituted into the original Poisson equation (C.4.2), returns the Poisson-Boltzmann equation (BLOSSEY, 2023):

$$\nabla^2 \varphi(\mathbf{r}) = \sum_i^N \frac{z_i \mathbf{e} c_i^0}{\varepsilon} e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} \tag{C.4.8}$$

Another way of expressing the PB equation is in cases of systems composed of binary electrolytes and symmetrically opposite charges, such as NaCl, for example, the concentrations of the Boltzmann equation must be modeled in terms of the charges of these two components, with this the equation becomes

$$\nabla^2 \varphi(\mathbf{r}) = \sum_i^N \frac{\mathbf{e} c_i^0}{\varepsilon} \left[z_i e^{z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} - z_i e^{-z_i \mathbf{e} \varphi(\mathbf{r})/k_B T} \right] \tag{C.4.9}$$

by definition, the trigonometric identity of the hyperbolic sine is given by

$$\sinh(x) = \frac{1}{2}(e^{x} - e^{-x}) \tag{C.4.10}$$

therefore, the PB equation can also be written in its hyperbolic form, as

$$\nabla^2\varphi(\mathbf{r}) = -\frac{2\mathbf{e}c_i^0}{\varepsilon}\sinh\left[\frac{\mathbf{e}\varphi(\mathbf{r})}{k_BT}\right] \tag{C.4.11}$$

Debye and Hückel approximated the entire exponential term of the PB equation in terms of a Taylor series for low electric potentials, which states

$$e^{-z_i\mathbf{e}\varphi(\mathbf{r})/k_BT} \approx 1 - \frac{-z_i\mathbf{e}\varphi(\mathbf{r})}{k_BT} \tag{C.4.12}$$

substituting in the exponential PB (C.4.8),

$$\nabla^2\varphi(\mathbf{r}) = \sum_i^N \frac{z_i\mathbf{e}c_i^0}{\varepsilon}\left(1 + \frac{z_i\mathbf{e}\varphi(\mathbf{r})}{k_BT}\right)$$

$$\nabla^2\varphi(\mathbf{r}) = \sum_i^N \frac{z_i\mathbf{e}c_i^0}{\varepsilon} + \sum_i^N \frac{z_i^2\mathbf{e}^2c_i^0}{\varepsilon k_BT}\varphi(\mathbf{r}) \tag{C.4.13}$$

Considering the given definition of charge density, in neutral solutions with no external electric field, the density is always zero, which implies that the sum of $z_i\mathbf{e}c_i^0$ is also zero. Therefore,

$$\nabla^2\varphi(\mathbf{r}) = \sum_i^N \frac{z_i^2\mathbf{e}^2c_i^0}{\varepsilon k_BT}\varphi(\mathbf{r}) \tag{C.4.14}$$

Now defining the constant term of the Debye length (ATKINS; DE PAULA; KEELER, 2018) as being

$$\kappa^2 = \sum_i^N \frac{z_i^2\mathbf{e}^2c_i^0}{\varepsilon k_BT} \tag{C.4.15}$$

the equation is summarized in a compact, linear form (C.4.16):

$$\nabla^2\varphi(\mathbf{r}) = \kappa^2\varphi(\mathbf{r}) \tag{C.4.16}$$

For a spherical charge, with radial symmetry, we write the Laplacian in spherical coordinates and isolate the differentials:

$$\nabla^2 = \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}\right)$$

$$\nabla^2\varphi(\mathbf{r}) = \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right) = \kappa^2\varphi(\mathbf{r})$$

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} - \kappa^2\varphi = 0$$

this is an ordinary differential equation of the second degree, whose solutions are well known, being

$$\varphi(\mathbf{r}) = \frac{Ae^{-\kappa \mathbf{r}} + Be^{\kappa \mathbf{r}}}{\mathbf{r}} \tag{C.4.17}$$

As a boundary condition, the electric potential must be zero when the radius of the sphere tends to infinity, which is only satisfied with the negative exponential of the solution (BLOSSEY, 2023), therefore,

$$\varphi(\mathbf{r}) = \frac{Ae^{-\kappa \mathbf{r}}}{\mathbf{r}} \tag{C.4.18}$$

with this, we can rewrite that

$$\mathbf{r}\varphi(\mathbf{r}) = Ae^{-\kappa r} \tag{C.4.19}$$

by defining the electric potential of a sphere (C.3.14), as seen previously

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i}{\mathbf{r}}$$

$$\mathbf{r}\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} q_i$$

so, the constant A will be

$$A = \frac{\mathbf{r}\varphi(\mathbf{r})}{e^{-\kappa r}}$$

$$A = \frac{1}{e^{-\kappa r}}\frac{1}{4\pi\varepsilon_0} q_i$$

For $\mathbf{r}$ tending to zero, $e^{-\kappa \mathbf{r}} = 1$ and

$$A = \frac{1}{4\pi\varepsilon_0} q_i \tag{C.4.20}$$

thus, the final Debye-Hückel solution leads us to

$$\varphi(\mathbf{r}) = \frac{q_i}{4\pi\varepsilon_0}\frac{e^{-\kappa \mathbf{r}}}{\mathbf{r}} \tag{C.4.21}$$

The linearized PB equation can also be solved in cylindrical coordinates, in which the resulting differential equation is, in cylindrical coordinates

$$\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr} - \kappa^2\varphi = 0 \tag{C.4.22}$$

resulting in an electrical potential of

$$\varphi(\mathbf{r}) = \varphi(a)\frac{\sqrt{\frac{\pi}{2\kappa \mathbf{r}}}e^{-\kappa \mathbf{r}}}{\sqrt{\frac{\pi}{2\kappa a}}e^{-\kappa a}} \tag{C.4.23}$$

where $\varphi(a)$ is the electric potential of a cylinder of radius a. The solution can also be obtained more trivially by considering a charged plane surface. The differential equation becomes

$$\frac{d^2\varphi}{dr^2} = \kappa^2\phi \tag{C.4.24}$$

with solution

$$\varphi(\mathbf{r}) = \varphi(0)e^{-\kappa \mathbf{r}} \tag{C.4.25}$$

For the nonlinear case, using the numeric solution, given in equation (2.52) it takes the form of a power series written in terms of the hyperbolic portion:

$$\sinh \varphi(\mathbf{r}) = 1 + \frac{\varphi^2(\mathbf{r})}{3!} + \frac{\varphi^4(\mathbf{r})}{5!} + \frac{\varphi^6(\mathbf{r})}{7!} \cdots \sum_{n=0}^{\infty} \frac{\varphi^n(\mathbf{r})}{(n+1)!} \tag{C.4.26}$$

C.5 *NVT* and *NpT* ensembles formalism

Given a Hamiltonian function that represents a continuous and conservative system of particles $\mathcal{H}(\mathbf{p},\mathbf{q})$, the probability density function $\rho(\mathbf{p},\mathbf{q})$ for a generic ensemble is (C.5.1):

$$\iint \rho(\mathbf{p},\mathbf{q})\, d\mathbf{p}^3 d\mathbf{q}^3 = 1 \tag{C.5.1}$$

Following the formalism shown by McQuarrie (MCQUARRIE, 2000), under these conditions, for the discrete quantum case, in order to obtain greater algebraic simplicity, the molecular partition function z_{NVT}, which corresponds to the totality of possible microstates of energy ε_i for each molecule, is defined from the Boltzmann factor (66):

$$z_{NVT} = \sum_{i}^{N} e^{-\varepsilon_i / k_B T} \tag{C.5.2}$$

For classical and continuous systems, in which all energies are likely to occur, the summation can be replaced by an integral over the entire phase space, as occurred with the probability density function (C.5.1).

$$z_{NVT} \approx \iint e^{-\varepsilon_i / k_B T}\, d\mathbf{p}^3 d\mathbf{q}^3 \tag{C.5.3}$$

By definition, the total partition function is $\mathcal{Z} = z^N/N!$, where the term $N!$ is allocated to ensure the indistinguishability of each particle, i.e., to prevent identical particles from being counted more than once, so

$$\mathcal{Z}_{NVT} \approx \frac{1}{N!} \iint (e^{-\varepsilon_i / k_B T})^N\, d\mathbf{p}^{3N} d\mathbf{q}^{3N} = \frac{1}{N!} \iint e^{-N\varepsilon_i / k_B T}\, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.4}$$

the term $N\varepsilon_i$ corresponds to the total energy of each microstate of the N components of the ensemble. As seen previously in the classical conservative case, this quantity can be expressed by the Hamiltonian $\mathcal{H}(\mathbf{p},\mathbf{q})$, that is

$$\mathcal{Z}_{NVT} \approx \frac{1}{N!} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q}) / k_B T}\, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.5}$$

Applying Plank's correction factor, we have the classical partition function for the canonical ensemble.

$$\mathcal{Z}_{NVT} = \frac{1}{N!h^{3N}} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.6}$$

Furthermore, according to Laurendeau (LAURENDEAU, 2005), considering all six dimensions of the phase space, the complete expression for the canonical partition function will be

$$\mathcal{Z}_{NVT} = \frac{1}{N!h^{3N}} \int_{\mathbf{p}_x^N} \int_{\mathbf{p}_y^N} \int_{\mathbf{p}_z^N} \int_{\mathbf{q}_x^N} \int_{\mathbf{q}_y^N} \int_{\mathbf{q}_z^N} e^{-\mathcal{H}(\mathbf{p}_x,\mathbf{p}_y,\mathbf{p}_z,\mathbf{q}_x,\mathbf{q}_y,\mathbf{q}_z)/k_B T} \, d\mathbf{p}_x^N d\mathbf{p}_y^N d\mathbf{p}_z^N d\mathbf{q}_x^N d\mathbf{q}_y^N d\mathbf{q}_z^N \tag{C.5.7}$$

Mathematically, an observable $\langle A \rangle$ for the *NVT* ensemble can be obtained by equation (C.5.8):

$$\langle A \rangle = \frac{1}{N!h^{3N}} \frac{1}{Z} \iint A e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.8}$$

To obtain the main observables of the canonical ensemble (pressure, internal energy and entropy), we will start from the classical thermodynamic definition of an adiabatic system (q = 0). The first law becomes $dE = \mathcal{H} = -pdV$. Taking the derivative with respect to volume, keeping the composition and temperature constant, we will have

$$-\left(\frac{\partial \mathcal{H}}{\partial V}\right)_{N,T} = p \tag{C.5.9}$$

By the definition of observable,

$$p = \langle p \rangle = \frac{1}{N!h^{3N}} \frac{1}{\mathcal{Z}_{NVT}} \iint p e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.10}$$

substituting by pressure in terms of the derivative of energy:

$$p = \langle p \rangle = \frac{1}{\mathcal{Z}_{NVT}} \frac{1}{N!h^{3N}} \iint -\left(\frac{\partial \mathcal{H}}{\partial V}\right)_{N,T} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.11}$$

Now deriving the partition function (C.5.6) with respect to volume, by the chain rule and considering the implicit dependence of the total energy on volume

$$\begin{aligned}
\left(\frac{\partial \mathcal{Z}_{NVT}}{\partial V}\right)_{T\,N} &= \frac{1}{N!h^{3N}} \iint \frac{\partial}{\partial V} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{N!h^{3N}} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \frac{1}{k_B T} \frac{\partial}{\partial V} [-\mathcal{H}(\mathbf{p},\mathbf{q})] \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{N!h^{3N}} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \left[-\frac{1}{k_B T} \left(\frac{\partial \mathcal{H}}{\partial V}\right)_{T,V} \right] d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{k_B T} \left[-\frac{1}{N!h^{3N}} \iint e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \left(\frac{\partial \mathcal{H}}{\partial V}\right)_{T,V} d\mathbf{p}^{3N} d\mathbf{q}^{3N} \right] \\
&= \frac{1}{k_B T} \left[\frac{1}{N!h^{3N}} \iint -\left(\frac{\partial \mathcal{H}}{\partial V}\right)_{T,V} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \right]
\end{aligned}$$

so,

$$k_B T \left(\frac{\partial \mathcal{Z}_{NVT}}{\partial V} \right)_{T,N} = \frac{1}{N! h^{3N}} \iint - \left(\frac{\partial \mathcal{H}}{\partial V} \right)_{T,V} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.12}$$

Substituting the derivative of the partition function (C.5.12) into the definition of the observable p (C.5.11), we have

$$p = k_B T \frac{1}{\mathcal{Z}_{NVT}} \left(\frac{\partial \mathcal{Z}_{NVT}}{\partial V} \right)_{T,N} \tag{C.5.13}$$

simplifying

$$p = k_B T \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial V} \right)_{T,N} \tag{C.5.14}$$

Now, for the observable of the internal energy (U) in terms of the Hamiltonian,

$$U = \langle \mathcal{H} \rangle = \frac{1}{N! h^{3N}} \frac{1}{\mathcal{Z}_{NVT}} \iint \mathcal{H} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.15}$$

Deriving the canonical partition function with respect to temperature, keeping composition and volume constant:

$$\begin{aligned}
\left(\frac{\partial \mathcal{Z}_{NVT}}{\partial T} \right)_{N,V} &= \frac{1}{N! h^{3N}} \iint \frac{\partial}{\partial T} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{N! h^{3N}} \iint \mathcal{H} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \left[\frac{\partial}{\partial T} \left(-\frac{1}{k_B T} \right) \right] d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{N! h^{3N}} \iint \mathcal{H} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \left[\frac{1}{k_B T^2} \right] d\mathbf{p}^{3N} d\mathbf{q}^{3N} \\
&= \frac{1}{k_B T^2} \left[\frac{1}{N! h^{3N}} \iint \mathcal{H} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \right]
\end{aligned}$$

therefore,

$$k_B T^2 \left(\frac{\partial \mathcal{Z}_{NVT}}{\partial T} \right)_{N,V} = \frac{1}{N! h^{3N}} \iint \mathcal{H} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} \, d\mathbf{p}^{3N} d\mathbf{q}^{3N} \tag{C.5.16}$$

Substituting (C.5.16) in the observable of energy (C.5.15),

$$U = k_B T^2 \frac{1}{\mathcal{Z}_{NVT}} \left(\frac{\partial \mathcal{Z}_{NVT}}{\partial T} \right)_{N,V}$$

or

$$U = k_B T^2 \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T} \right)_{N,V} \tag{C.5.17}$$

To determine entropy, we start from the thermodynamic definition $dS = dU/T + p/TdV$. That is,

$$dS = \frac{k_B}{T} d \left[T^2 \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T} \right)_{N,V} \right] + k_B \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial V} \right)_{T,N} dV$$

Integrating from zero to a final state f, different from zero, considering that $S(0) = 0$ in the thermodynamic limit, we have

$$S = \frac{k_B}{T}\int_0^f d\left[T^2\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V}\right] + k_B\int_0^f \left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial V}\right)_{T,N} dV$$

$$S = k_B T\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V} + k_B \ln \mathcal{Z}_{NVT} \tag{C.5.18}$$

Once the observables of pressure, enthalpy and entropy are determined, the Helmholtz energy can be written by the definition $A = U - TS$. Using (C.5.17) and (C.5.18):

$$A = k_B T^2\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V} - T\left[k_B T\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V} + k_B \ln \mathcal{Z}_{NVT}\right]$$

$$= k_B T^2\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V} - k_B T^2\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial T}\right)_{N,V} + k_B T \ln \mathcal{Z}_{NVT}$$

therefore,

$$A = k_B T \ln \mathcal{Z}_{NVT} \tag{C.5.19}$$

Finally, to determine the Gibbs energy, we start from the definition $G = A + pV$, joining equations (C.5.19) and (C.5.14):

$$G = k_B T \ln \mathcal{Z}_{NVT} + k_B T V\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial V}\right)_{T,N}$$

or

$$G = k_B T \ln \mathcal{Z}_{NVT} + k_B T\left(\frac{\partial \ln \mathcal{Z}_{NVT}}{\partial \ln V}\right)_{T,N} \tag{C.5.20}$$

The partition function NpT, to compute a microstate with energy $\Delta E + p\Delta V$, the volumetric fluctuations are proportional to a Boltzmann distribution based on the pressure and volume, integrating over every possible differential change dV:

$$\mathcal{Z}_{NpT} \propto \int e^{pV/k_B T}\, dV\, \mathcal{Z}_{NVT} \tag{C.5.21}$$

The particle distinguishability and dimensionlessness correction terms are also applied here, explicitly defining the partition function NpT (C.5.22) as:

$$\mathcal{Z}_{NpT} = \frac{1}{N! h^{3N}}\int_V \iint_{\mathbf{p},\mathbf{q}} e^{-\mathcal{H}(\mathbf{p},\mathbf{q})/k_B T} e^{pV/k_B T}\, d\mathbf{p}^{3N}\, d\mathbf{q}^{3N}\, dV$$

$$\mathcal{Z}_{NpT} = \frac{1}{N! h^{3N}}\int_V \iint_{\mathbf{p},\mathbf{q}} e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_B T}\, d\mathbf{p}^{3N}\, d\mathbf{q}^{3N}\, dV \tag{C.5.22}$$

an observable for this ensemble can be obtained from an expression analogous to the canonical ensemble, by equation:

$$\langle A \rangle = \frac{1}{N! h^{3N}}\frac{1}{\mathcal{Z}_{NpT}}\int_V \iint_{\mathbf{p},\mathbf{q}} A e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_B T}\, d\mathbf{p}^{3N}\, d\mathbf{q}^{3N}\, dV \tag{C.5.23}$$

To obtain the expression for the Gibbs energy, a shorter sequence of steps is sufficient, which consists of defining the volume observable, followed by the derivation of the partition function with respect to pressure. The observable will be

$$V = \langle V \rangle = \frac{1}{\mathcal{Z}_{NpT}} \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} V e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \tag{C.5.24}$$

The first derivative of the partition function with respect to pressure is:

$$\begin{aligned}
\left(\frac{\partial \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T} &= \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} \frac{\partial}{\partial p} e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \\
&= \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} -\frac{\partial}{\partial p} \left[\frac{\mathcal{H}(\mathbf{p},\mathbf{q}) + pV}{k_BT} \right] e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \\
&= \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} -\frac{\partial}{\partial p} \left[\frac{\mathcal{H}(\mathbf{p},\mathbf{q})}{k_BT} + \frac{pV}{k_BT} \right] e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \\
&= \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} -\frac{\partial}{\partial p} \left[\frac{pV}{k_BT} \right] e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \\
&= \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} -\frac{V}{k_BT} e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV
\end{aligned}$$

so,

$$\left(\frac{\partial \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T} = -\frac{1}{k_BT} \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} V e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \tag{C.5.25}$$

or, rearranging

$$-k_BT \left(\frac{\partial \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T} = \frac{1}{N!h^{3N}} \int_V \iint_{\mathbf{p},\mathbf{q}} V e^{-[\mathcal{H}(\mathbf{p},\mathbf{q})+pV]/k_BT} \, d\mathbf{p}^{3N} \, d\mathbf{q}^{3N} \, dV \tag{C.5.26}$$

Joining the equality (C.5.26) to the equation of the volumetric observable (C.5.24), we have

$$V = -k_BT \frac{1}{\mathcal{Z}_{NpT}} \left(\frac{\partial \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T}$$

or

$$V = -k_BT \left(\frac{\partial \ln \mathcal{Z}_{NpT}}{\partial p} \right)_{N,T} \tag{C.5.27}$$

which is the final expression for the observable volume of the isothermal-isobaric ensemble. The connection with macroscopic thermodynamics can be established by the Gibbs energy itself, $G = f(N, p, T)$:

$$dG = \left(\frac{\partial G}{\partial T}\right)_{p,N} dT + \left(\frac{\partial G}{\partial p}\right)_{T,N} dp + \sum_i \left(\frac{\partial G}{\partial N_i}\right)_{p,T,N_{j\neq i}} dN_i \tag{C.5.28}$$

Gibbs energy is also defined in terms of entropy and volume, therefore

$$dG = -S\,dT + V\,dp + \sum_i \mu_i\, dN_i$$

Fixing temperature and number of particles, we have the following identity:

$$\left(\frac{\partial G}{\partial p}\right)_{T,N} = V \tag{C.5.29}$$

Therefore, integrating (C.5.29) from an initial state of zero to a final state f

$$\left(\frac{\partial G}{\partial p}\right)_{T,N} = -k_B T \left(\frac{\partial \ln \mathcal{Z}_{NpT}}{\partial p}\right)_{N,T}$$

$$\int_0^f \left(\frac{\partial G}{\partial p}\right)_{T,N} dp = -k_B T \int_0^f \left(\frac{\partial \ln \mathcal{Z}_{NpT}}{\partial p}\right)_{N,T} dp$$

Finally,

$$G = -k_B T \ln \mathcal{Z}_{NpT} \tag{C.5.30}$$

C.6 PM6 semiempirical Hamiltonian

The Hamiltonian of the PM6-DH+ model consists of the Hamiltonian of the PM6 method, based on the parameterized Fock operator, with a correction term for the dispersion and hydrogen bonding energy:

$$\hat{H}^{PM6-DH+} = \hat{F}_{\mu\nu} + E_{disp} + E_{H-bond} \tag{C.6.1}$$

in which the exchange integrals $(\mu\sigma|\lambda\mu)$, and the orbital interactions of different atomic centers, such as $(\mu\nu|\lambda\sigma)$ are parameterized, the overlap integrals between different atoms $\langle\phi_\mu|\phi_\nu\rangle$ are considered zero and the integrals involving four or more orbitals, of different atomic centers are totally neglected (STEWART, 2013). The Coulomb intergals for the same atom are maintained

$$(\mu\mu|\nu\nu) = \iint \phi_\mu^2(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}\phi_\nu^2(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \tag{C.6.2}$$

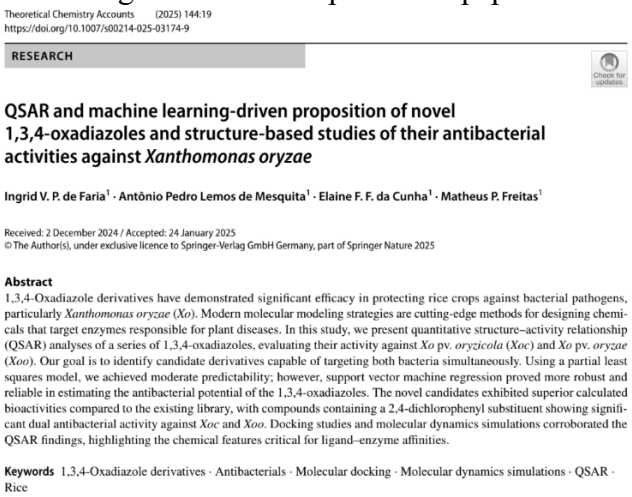
Just like diagonal monatomic integrals

$$(\mu\nu|\mu\nu) = \iint \phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}\phi_\mu(\mathbf{r}_2)\phi_\nu(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \tag{C.6.3}$$

ATTACHMENT A – ADDITIONAL ACTIVITIES

During my master's program, I had the opportunity to participate in two external research projects that resulted in two publications (Figs .A.1 and A.2). The first, in collaboration with Ingrid V. P. F., Elaine F. F. C., and Matheus P. F., is titled *"QSAR and machine learning-driven propositions of novel 1,3,4-oxadiazoles and structure-based studies of their antibacterial activities against Xanthomonas oryzae"* (https://doi.org/10.1007/s00214-025-03174-9) (Fig. A1), and was published in *Theoretical Chemistry Accounts*.

Figure A.1 – First published paper.



Theoretical Chemistry Accounts (2025) 144:19
https://doi.org/10.1007/s00214-025-03174-9

RESEARCH

QSAR and machine learning-driven proposition of novel 1,3,4-oxadiazoles and structure-based studies of their antibacterial activities against *Xanthomonas oryzae*

Ingrid V. P. de Faria[1] · Antônio Pedro Lemos de Mesquita[1] · Elaine F. F. da Cunha[1] · Matheus P. Freitas[1]

Received: 2 December 2024 / Accepted: 24 January 2025

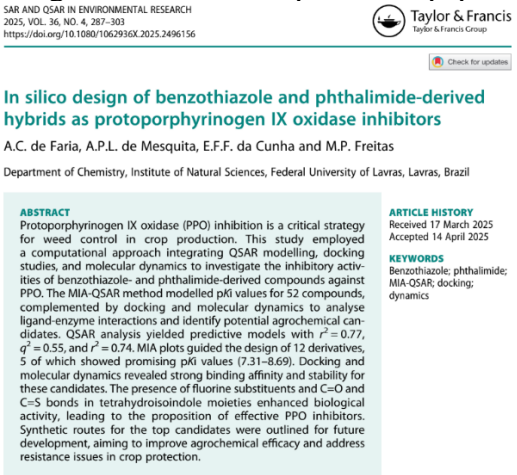
Abstract
1,3,4-Oxadiazole derivatives have demonstrated significant efficacy in protecting rice crops against bacterial pathogens, particularly *Xanthomonas oryzae* (*Xo*). Modern molecular modeling strategies are cutting-edge methods for designing chemicals that target enzymes responsible for plant diseases. In this study, we present quantitative structure–activity relationship (QSAR) analyses of a series of 1,3,4-oxadiazoles, evaluating their activity against *Xo* pv. *oryzicola* (*Xoc*) and *Xo* pv. *oryzae* (*Xoo*). Our goal is to identify candidate derivatives capable of targeting both bacteria simultaneously. Using a partial least squares model, we achieved moderate predictability; however, support vector machine regression proved more robust and reliable in estimating the antibacterial potential of the 1,3,4-oxadiazoles. The novel candidates exhibited superior calculated bioactivities compared to the existing library, with compounds containing a 2,4-dichlorophenyl substituent showing significant dual antibacterial activity against *Xoc* and *Xoo*. Docking studies and molecular dynamics simulations corroborated the QSAR findings, highlighting the chemical features critical for ligand–enzyme affinities.

Keywords 1,3,4-Oxadiazole derivatives · Antibacterials · Molecular docking · Molecular dynamics simulations · QSAR · Rice

Source: author (2025)

The second article, (Fig.A.2), in collaboration with Adriana C. F., Elaine F. F. C., and Matheus P. F., is titled *"In silico design of benzothiazole and phthalimide-derived hybrids as protoporphyrinogen IX oxidase inhibitors"*, and was published in *SAR and QSAR in Environmental Research* (https://doi.org/10.1080/1062936X.2025.2496156).

Figure A.2 – Second published paper



SAR AND QSAR IN ENVIRONMENTAL RESEARCH
2025, VOL. 36, NO. 4, 287–303
https://doi.org/10.1080/1062936X.2025.2496156

Taylor & Francis
Taylor & Francis Group

Check for updates

In silico design of benzothiazole and phthalimide-derived hybrids as protoporphyrinogen IX oxidase inhibitors

A.C. de Faria, A.P.L. de Mesquita, E.F.F. da Cunha and M.P. Freitas

Department of Chemistry, Institute of Natural Sciences, Federal University of Lavras, Lavras, Brazil

ABSTRACT
Protoporphyrinogen IX oxidase (PPO) inhibition is a critical strategy for weed control in crop production. This study employed a computational approach integrating QSAR modelling, docking studies, and molecular dynamics to investigate the inhibitory activities of benzothiazole- and phthalimide-derived compounds against PPO. The MIA-QSAR method modelled pKi values for 52 compounds, complemented by docking and molecular dynamics to analyse ligand-enzyme interactions and identify potential agrochemical candidates. QSAR analysis yielded predictive models with $r^2 = 0.77$, $q^2 = 0.55$, and $r^2 = 0.74$. MIA plots guided the design of 12 derivatives, 5 of which showed promising pKi values (7.31–8.69). Docking and molecular dynamics revealed strong binding affinity and stability for these candidates. The presence of fluorine substituents and C=O and C=S bonds in tetrahydroisoindole moieties enhanced biological activity, leading to the proposition of effective PPO inhibitors. Synthetic routes for the top candidates were outlined for future development, aiming to improve agrochemical efficacy and address resistance issues in crop protection.

ARTICLE HISTORY
Received 17 March 2025
Accepted 14 April 2025

KEYWORDS
Benzothiazole; phthalimide; MIA-QSAR; docking; dynamics

Source: author (2025)