erCHARLES UNIVERSITY
FACULTY OF PHARMACY IN HRADEC KRÁLOVÉ

Department of Biophysics and Physical Chemistry

STRUCTURE-BASED DRUG DISCOVERY: COMPUTATIONAL INSIGHTS INTO TGR5 AGONISTS AND SUBSTITUTED CATECHOL VASODILATORS

Objevování léčiv založené na strukturních metodách: výpočetní pohled na agonisty receptoru TGR5 a substituované katecholové vazodilátory

Doctoral Dissertation

(Commentary on Published Articles)

Study program: Bioorganic chemistry Rudy Salam, M.Biomed

Hradec Králové 2025

STATEMENT OF AUTHORSHIP

I hereby declare that I am the sole author of this thesis. I have written this dissertation independently under the supervision of Assoc. Prof. Dipl.-Math. Erik Jurjen Duintjer Tebbens, PhD. and co-supervision by Assoc. Prof. PharmDr. Jan Zitko, PhD., using only the mentioned and duly cited sources and literature, and that the work has not been used in another university study program or to obtain the same or another academic title.

In Hradec Králové

Rudy Salam, M.Biomed

Date: __April 22, 2025___

_____

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Assoc. Prof. doc. Dipl.-Math. Erik Jurjen Duintjer Tebbens, Ph.D. for his invaluable guidance, endless patience, and unwavering support in the completion of this dissertation. I greatly appreciate his tremendous support and tolerance especially during the stressful phases of this research work.

Special appreciation should also be given to my consultant, Assoc. Prof. doc. PharmDr. Jan Zitko, Ph.D. who always provided feedback in advancing our scientific work.

I would also like to thank Prof. PharmDr. Přemysl Mladěnka, Ph.D. for believing in me to be involved in his research project.

Furthermore, I would like to thank Ing. Martin Drastík, Ph.D. for guiding me in the early days of this research to reach this point.

On a personal note, I would like to thank Oktavia Eka Puspita, my beloved wife, who has sacrificed our time together for almost 5 years for me to complete this study. She always motivated me and believed in me to be even better. I would also like to thank my father, mother and father-in-law and mother-in-law, who always support me.

Lastly, my heartfelt thanks to my son Kaizen Emir Arsyanendra, whom I was unable to be with at the time of his birth.

Each of these individuals have played a unique and invaluable role in shaping the outcome of this dissertation, and for that, I am deeply grateful.

This dissertation was made possible by the support of the Specialized University Research of Charles University (Grant No. SVV 260 666) and Martina Roeselová Memorial Fellowship awarded by the IOCB Tech Foundation to Alžbeta Štefela.

The author thanks the Ministry of Education, Youth, and Sport of the Czech Republic for computational resources supplied by the project “e-Infrastruktura CZ” (e-INFRA CZ ID: 90140 and e-INFRA CZ LM2018140). Computations were implemented in the MetaCentrum and IT4I supercomputing facilities.

ABSTRAKT

Univerzita Karlova, Farmaceutická fakulta v Hradci Králové

Katedra:	Katedra biofyziky a fyzikální chemie
Kandidát:	Rudy Salam, M.Biomed
Školitel:	Assoc. Prof. Dipl.-Math. Erik Jurjen Duintjer Tebbens, Ph.D.
Konzultant:	Assoc. Prof. PharmDr. Jan Zitko, Ph.D.
Název disertační práce:	Objevování léčiv založené na strukturních metodách: výpočetní pohled na agonisty receptoru TGR5 a substituované katecholové vazodilátory

Počítačem podporovaný návrh léčiv (CADD) urychluje a zefektivňuje vývoj léčiv prostřednictvím integrace výpočetních přístupů. Objevování léčiv založené na strukturních metodách (SBDD), základní kámen racionálního návrhu léčiv, identifikuje vůdčí sloučeniny na základě jejich interakcí ve vazebném místě receptoru. V teoretické části této disertační práce popisuji základní principy a metody SBDD, jako jsou molekulární dokování, farmakoforové vyhledávání a molekulové dynamiky (MD).

Tato interdisciplinární disertační práce komentuje publikace doktoranda a ukazuje jako příklad implementaci výpočetních metod na příkladu dvou projektů: 1. Objevování nových agonistů Takeda G protein-coupled receptoru 5 (TGR5) kombinací farmakoforového vyhledávání, molekulárního dokování a MD simulací. 2. Využití molekulárního dokování k objasnění molekulárního cíle substituovaných katecholů s vazodilatačním účinkem.

V první publikaci jsme použili vysokokapacitní virtuální screening (HTVS) k identifikaci nových ligandů TGR5. Virtuální screening, který zahrnoval farmakoforové vyhledávání a dvě kola molekulárního dokování, přinesl pět hitů sloučenin splňujících kritéria výběru: vyhovují farmakoforovému modelu, docking skóre pod −9,2 kcal/mol (práh definovaný dle skóre referenčního ligandu) a vykazovaly klíčové interakce podobné referenčnímu ligandu. Těchto pět sloučenin (hitů) bylo podrobeno biologickým testům, které potvrdily, že pouze Hit-3 (CSC089939231) byl schopen aktivovat TGR5. Tyto výsledky byly podpořeny MD simulacemi, které odlišily Hit-3 od ostatních hitů zdůrazněním role vodíkové vazby s Tyr240 při aktivaci TGR5.

Druhá část výzkumu se zaměřila na objasnění mechanismu účinku odpovědného za vazodilatační efekt substituovaných katecholů, konkrétně 4-methylkatecholu (4-MC) a 3-methoxykatecholu (3-MOC). Reverzní molekulární dokování bylo využito jako doplnění in vitro studií na izolované aortě potkana. Obě sloučeniny vykazovaly výrazný vazodilatační účinek, přičemž 3-MOC byl účinnější. Z hlediska mechanismu účinku nebyly závislé na L-typových Ca^{2+} kanálech, ale působily prostřednictvím aktivace napěťově řízených K^+ (K_V) kanálů, konkrétně izoformy $K_V7.4$, jak bylo potvrzeno selektivními inhibičními studiemi. Molekulární dokování dále podpořilo tento mechanismus, když odhalilo klíčové vodíkové a hydrofobní interakce s aminokyselinovými zbytky Trp242 a Phe246 v kanálu $K_V7.4$. Tyto poznatky řadí 3-MOC a 4-MC mezi perspektivní kandidáty pro vývoj antihypertenziv, která neovlivňují koronární oběh.

ABSTRACT

Charles University, Faculty of Pharmacy in Hradec Králové

Department: Department of Biophysics and Physical Chemistry

Candidate: Rudy Salam, M.Biomed

Supervisor: Assoc. Prof. Dipl.-Math. Erik Jurjen Duintjer Tebbens, Ph.D.

Consultant: Assoc. Prof. PharmDr. Jan Zitko, Ph.D.

Title of dissertation: Structure-Based Drug Discovery: Computational Insights into TGR5 Agonists and Substituted Catechol Vasodilators

Computer-aided drug design (CADD) accelerates and enhances drug development by integrating computational approaches. Structure-based drug discovery (SBDD), a cornerstone of rational drug design, identifies lead compounds by focusing on their interactions within the target binding site. In the theoretical part of this thesis, I have described the fundamental principles and selected SBDD methods, such as molecular docking, pharmacophore modeling, and molecular dynamics (MD) simulations.

This interdisciplinary dissertation thesis comments on the published work of the candidate and aims to provide an example of the implementation of computational approaches in two projects: 1. Discovering new Takeda G protein-coupled receptor 5 (TGR5) agonists by combining pharmacophore modelling, molecular docking, and MD simulations, and 2. Utilization of molecular docking in unravelling the molecular target of substituted catechols acting as vasodilators.

In the first publication, we used a high-throughput virtual screening (HTVS) to identify new ligands of TGR5. Virtual screening, using pharmacophore searching and two rounds of molecular docking, yielded five hit compounds that met the selection criteria: fulfilled pharmacophore features, obtained a docking score below -9.2 kcal/mol (threshold defined by the score of the reference ligand), and showed key interactions pattern similar to the reference ligand. Five hits were advanced to biological assays, which confirmed only Hit-3 (CSC089939231) was able to activate TGR5. These findings were supported by MD simulations, which distinguished Hit-3 from the other hit compounds by emphasizing the role of hydrogen bonding with Tyr240 in TGR5 activation.

The second part of the research focused on elucidating the mechanism of action responsible for the vasodilatory effects of substituted catechols, specifically 4-methylcatechol (4-MC) and 3-methoxycatechol (3-MOC). Reverse molecular docking was used to complement in vitro studies on isolated rat aorta. Both compounds demonstrated significant vasodilatory activity, with 3-MOC showing greater potency. Mechanistically, their action was independent of L-type Ca^{2+} channels and was instead mediated by the activation of voltage-gated K^+ (KV) channels, particularly the KV7.4 isoform, as confirmed through selective inhibition studies. Molecular docking further supported this mechanism, revealing key hydrogen bonding and hydrophobic interactions with residues Trp242 and Phe246 within the KV7.4 channel. These findings position 3-MOC and 4-MC as promising leads for the development of antihypertensive agents that do not compromise coronary circulation.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Name
3-MOC	3-methoxycatechol
4-MC	4-methylcatechol
ADMET	absorption, distribution, metabolism, elimination and toxicity
AUC	area under the curve
CADD	Computer-aided drug discovery
cAMP	cyclic adenosine monophosphate
cGMP	cyclic guanosine monophosphate
CGenFF	CHARMM General Force Field
cyro-EM	cryo-electron microscopy
EF	enrichment factor
FEP	free energy perturbation
GA	Genetic algorithm
GAFF	General AMBER Force Field
GH	Güner-Henry
GPU	graphical processing unit
HTS	high-throughput screening
HTVS	high-throughput virtual screening
IFD	Induced fit docking
LBDD	ligand-based drug discovery
LCA	lithocholic acid
LGA	Lamarckian Genetic Algorithm
MOA	mechanism of action
MD	molecular dynamics
MC	Monte Carlo
MC/MD	Monte Carlo/molecular dynamics
MM	molecular mechanics
NMR	nuclear magnetic resonance
OPLS-AA	Optimized Potentials for Liquid Simulations in all atoms
PDB	Protein Data Bank
PMF	Potential of Mean Force
PSO	particle swarm optimization
QM	quantum mechanics
QSAR	quantitative structure-activity relationship
Rg	radius of gyration
RMSD	root mean square deviation
RMSF	root mean square fluctuation
ROC	Receiver Operating Characteristic
SBDD	structure-based drug discovery
SAR	structure-activity relationship
SA	simulated annealing
TGR5	Takeda G protein-coupled receptor 5
TS	Tabu Search
vdW	van der Waals
VS	virtual screening

1 INTRODUCTION

In recent years, computer-aided drug discovery (CADD) simplifies the process by using computational techniques to design and optimize bioactive compounds, thereby reducing cost and time (1, 2). CADD methods are categorized based on the level of theoretical approximation into molecular mechanics (MM) and quantum mechanics (QM) approaches (3). MM relies on classical physics and applies empirical force fields to model molecular interactions, calculating the potential energy of a system based on parameters such as bond lengths, angles, torsions, and non-bonded interactions. MM-based methods are computationally efficient, making them particularly suitable for high-throughput screening and early-stage drug discovery. In contrast, quantum mechanics methods, which are based on the principles of quantum theory and involve solving the Schrödinger equation, provide a more detailed description of electronic structure but are computationally more intensive and often reserved for later-stage analysis, such as refining binding interactions or modelling reactive processes (3). CADD can be categorized as well into structure-based and ligand-based approaches (4). Structure-based drug discovery (SBDD) and ligand-based drug discovery (LBDD) are complementary approaches in early drug discovery, each utilizing different information (4). SBDD, a widely used MM strategy, designs ligands based on the 3D structure of the target protein, using techniques such as molecular docking and molecular dynamics (MD) simulations to predict binding affinity and optimize lead compounds (5). This method is effective with high-resolution protein structure data (5). LBDD is employed in the absence of a target protein structure, relying on known ligands to identify common features to predict new compounds. It is useful for target identification and lead optimization in the absence of protein structure data (5). An example of LBDD is quantitative structure-activity relationship (QSAR) modelling, which employs statistical and machine learning techniques to correlate molecular descriptors with biological activity, allowing activity predictions for new compounds (6).

SBDD has arisen as a cornerstone in the rational design of new drugs, allowing the identification of lead compounds with high specificity and potency (7). Utilising detailed three-dimensional structures of biological targets such as proteins, SBDD allows precise modelling of molecular interactions, facilitating the prediction of binding affinities and the design of compounds with improved pharmacokinetic and pharmacodynamic properties (8). SBDD encompasses a range of computational and experimental techniques, including pharmacophore modelling, molecular docking, MD simulations, and structure-activity relationship (SAR) studies. These approaches are supported by advances in X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM), which provide high-resolution protein structural data. The integration of these methods has led to numerous successes in drug discovery, including the development of

inhibitors that target kinases, proteases, and other important enzymes involved in various diseases (5). Many therapeutically relevant proteins lack experimentally resolved structures, hindering structure-based drug design (SBDD). Homology modelling has been used to predict 3D structures based on sequence similarity to known structures, by artificial intelligence (AI). Particularly AlphaFold has revolutionized structure prediction (9). AlphaFold's highly accurate predictions, accessible through public databases, have expanded the proteome's structural coverage, enabling drug discovery for previously intractable targets. Homology modelling and AlphaFold are now powerful tools in SBDD, facilitating ligand design and virtual screening (9).

Pharmacophore modelling is a concept of capturing the essential features responsible for biological activity of a molecule. The model represents the spatial arrangement of functional groups that interact with a specific biological target, thus allowing researchers to identify and design new compounds with similar activity profiles. Pharmacophore models can be derived from the structural information of known active compounds (ligand-based approach) or the three-dimensional structure of the target protein (structure-based approach). Structure-based pharmacophore modelling leverages high-resolution ligand-receptor complex structures (e.g., from X-ray crystallography or molecular binding) to map specific ligand-receptor interactions (hydrogen bonding, π-π stacking, salt bridges, hydrophobic contacts). This precisely defines the spatial and chemical requirements for receptor binding (10, 11). Structure-based pharmacophore modelling is direct and more precise than ligand-based modelling as it incorporates the important interactions occurring between the ligand and the receptor in eliciting a biological response (10, 12). These models facilitate virtual screening of large compound libraries, thus prioritising candidates for further experimental validation. By simplifying the identification of potential drug candidates, pharmacophore modelling significantly reduces the time and resources required for the preliminary screening step (13).

On the other hand, molecular docking predicts the binding orientation and affinity of small molecules toward their target proteins (14). By simulating the interaction between ligands and their receptors, docking studies provide insight into modelling the interaction between ligands and proteins at the atomic level, enabling characterization of ligand behaviour within the ligand-binding domain to elucidate fundamental biochemical processes (15). Apart from predicting binding modes, molecular docking provides important information on the thermodynamics and kinetics of ligand-receptor interactions. Coupled with a scoring function, docking studies allow to determine the ranking of potential drug candidates based on the predicted binding affinity. This information is invaluable for prioritization of compounds for synthesis and biological assay. Moreover, in combination with MD simulations, docking can reveal dynamic interaction patterns that are essential for understanding

ligand binding stability and specificity (16). The integration of pharmacophore modelling and molecular docking has been successfully applied in various drug discovery projects. An important advantage of combining these methods is the ability to cross-validate the findings. Pharmacophore models can be employed to filter compound libraries prior to docking, thereby reducing the computational workload and increasing the efficiency of the screening process as well as gaining a more comprehensive understanding of ligand-receptor interactions and improving the accuracy of virtual screening workflows.

Computational modelling is essential in unravelling the mechanism of action of bioactive compounds, particularly in cases of limited experimental data. The molecular docking process involves not only predicting the binding affinity but also unlocking the mechanism of action (MOA) of the ligand. This involves understanding how the ligand induces specific conformational changes in the protein, which can affect the biological function of the protein. Ligand binding can inhibit or activate target proteins, and docking simulations provide a detailed view of interactions at the atomic level, such as hydrogen bonding, hydrophobic interactions, and electrostatic forces (17). By analysing these interactions, docking studies can elucidate the molecular mechanisms at the atomic level controlling the activity of ligands and their potential therapeutic effects (15).

This dissertation applies molecular docking approaches to discover bioactive small molecules with potential pharmacological properties. The main study in this dissertation emphasizes molecular docking as an essential tool in discovering drugs modulating the Takeda G protein-coupled receptor 5 (TGR5) protein and combines it with pharmacophore modelling for preliminary screening of a vast compound library and with MD simulations to rationalize the observed biological activity of candidate compounds. The other two studies explore the potential vasodilatory mechanisms of catechol derivatives, 3-methoxycatechol (3-MOC) and 4-methylcatechol (4-MC). All studies utilize molecular docking to elucidate ligand-receptor interactions, identify key binding residues and are supported by experimental validation.

The dissertation is structured as follows: the first and second chapter consist of the introduction and of the aims of the thesis, respectively. In the third chapter, we will focus on the theoretical background of SBDD with molecular docking as the main part along with pharmacophore theory and MD simulation as part of SBDD. The fourth chapter, central component of the dissertation, will highlight the application of in silico techniques in both the identification of novel TGR5 agonists and the mechanistic investigation of substituted catechol compounds with vasodilatory potential. Chapter 5, as conclusion of this work, synthesizes the main findings, reflecting on their significance within the

broader landscape of drug discovery. This is followed by a discussion on the Prospects for Further Research (chapter 6), which outlines future directions and potential expansions of the work. Finally, research outputs (chapter 7) summarize the scientific contributions made throughout the dissertation, including peer-reviewed publications and other relevant outputs such as oral presentation. Together, these sections provide a coherent and structured presentation of the research process and its outcomes.

2 AIMS OF THE DISSERTATION THESIS

The aim of this dissertation is to explore and implement molecular docking as a core computational approach for the identification, characterization, and mechanistic prediction of small-molecule bioactivity in the context of drug discovery. In addition to molecular docking, this thesis addresses other computational approaches including pharmacophore modelling for preliminary database screening and MD simulations for the rationalization of biological activity of candidate compounds. This methodology complements experimental methods, offering insights into ligand-target interactions, stability, and potency.

The specific applications this thesis focuses on, are:

I. Identifying new TGR5 agonists through computational screening using pharmacophore-based screening, molecular docking, and MD simulation
II. Utilizing molecular docking to unravel the mechanical role of catechol derivatives in vasodilation

By leveraging molecular docking to simulate and predict ligand–receptor interactions at the atomic level, this dissertation aims to contribute to the rational design of lead compounds with therapeutic potential, while demonstrating the versatility and value of docking methodologies in mechanism of action studies.

3 THEORETICAL BACKGROUNDS

CADD has emerged as a transformative approach in pharmaceutical research, integrating computational techniques to streamline the drug development process (1, 2). The fundamental goal of CADD is to design and optimize biologically active compounds while minimizing the costs, time, and labour associated with traditional drug discovery methods. The origins of CADD trace back to the 1970s, gaining public recognition in 1981 when Fortune magazine highlighted its potential in a cover article on Merck's computational drug design efforts (18-20). Despite initially intriguing, CADD faced competition from high-throughput screening (HTS), an automated, robust approach to identifying bioactive compounds. However, HTS was associated with low success rates and high costs, restricting its accessibility to only well-funded research programmes (20). Over time, CADD has re-emerged as a complementary strategy, allowing researchers to efficiently screen virtual compound libraries, prioritise active candidates, and reduce the need for labour-intensive experimental screening (2). A case study from Pharmacia (now part of Pfizer) demonstrated the effectiveness of CADD: virtual screening identified 365 potential inhibitors for tyrosine phosphatase-1B, with 127 showing significant activity (hit rate of 35%), whereas traditional HTS identified only 81 active compounds from 400,000 tested molecules (hit rate of 0.021%) (21). This example highlights how computational screening can drastically improve the efficiency of drug discovery by reducing both time and cost while maintaining high-quality lead identification.

CADD plays a crucial role in multiple stages of drug discovery:

1. Target Identification and Validation

   Computational tools assist the identification of bioactive molecules and their molecular targets through structure- or ligand-based approaches.
2. Hit Discovery and Lead Optimization

   Virtual screening techniques, such as molecular docking, pharmacophore screening, and quantitative structure-activity relationship (QSAR) modelling, are applied to predict potential drug candidates with beneficial pharmacokinetic aspects (absorption, distribution, metabolism, excretion and toxicity/ADMET) and pharmacodynamic properties.
3. Preclinical and Clinical Development

   CADD supports the refinement of drug candidates before proceeding to experimental assays and costly clinical trials (20).

The pipeline in CADD typically has three main objectives: (1) screening of a large compound database to obtain a small number of hit compounds for experimental assay and subsequent lead

compounds; (2) lead compound optimisation, either to improve its affinity or the pharmacokinetic properties of the drug including ADMET; (3) synthesis of derivative compounds from the obtained lead compounds in order to obtain other drug candidates with increased potency (20). In general, CADD can be categorised into two main approaches: structure-based approaches with the discovery of drug candidates focused on the orientation in the target binding cavity and ligand-based approaches with drug discovery centred on the ligand (4).

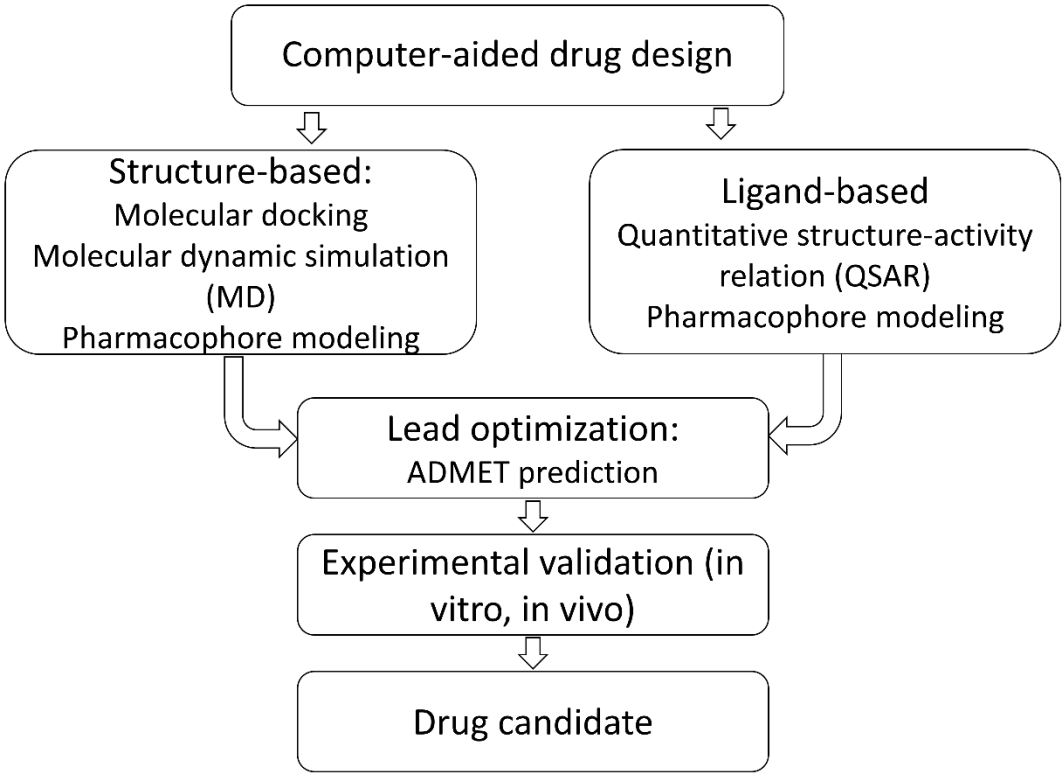

Figure 1. Drug discovery pipeline using computer-aided drug design approach (adapted from (2, 20).

SBDD relies on technologies such as X-ray crystallography and NMR to determine the 3D structure of target proteins (2, 22, 23). Knowing the 3D structure of the target protein provides insight into ligand binding, conformation, and the mechanism of action of the active compound (2, 24). Therefore, identifying the drug target protein is crucial for SBDD as highlighted in Figure 2. However, many therapeutically relevant proteins lack experimentally resolved structures due to limitations in techniques such as X-ray crystallography or cryo-electron microscopy. To address this gap, homology modelling has traditionally been employed to predict the 3D structure of a target protein based on its sequence similarity to proteins with known structures (templates). This method assumes that homologous proteins adopt similar folds, and it involves sequence alignment, model building, and refinement to generate a plausible structural model suitable for SBDD applications such as molecular docking, pharmacophore modelling, and virtual screening (25).

Recent advances in artificial intelligence (AI) have revolutionized structure prediction with the development of AlphaFold, a deep learning-based system by DeepMind. AlphaFold has demonstrated unprecedented accuracy in predicting protein structures directly from amino acid sequences, achieving near-experimental resolution in many cases (9). Its predictions are not only significantly more reliable than traditional homology models, especially for proteins with low sequence identity to known structures, but also widely accessible through public databases like the AlphaFold Protein Structure Database. As a result, AlphaFold has become a game-changer in SBDD by vastly expanding the structural coverage of the proteome and enabling drug discovery efforts against previously intractable targets. Together, homology modelling and AlphaFold serve as powerful tools to fill the structural knowledge gap in SBDD with highly accurate predictions across a broader range of proteins, facilitating ligand design and virtual screening workflows with greater confidence and speed.

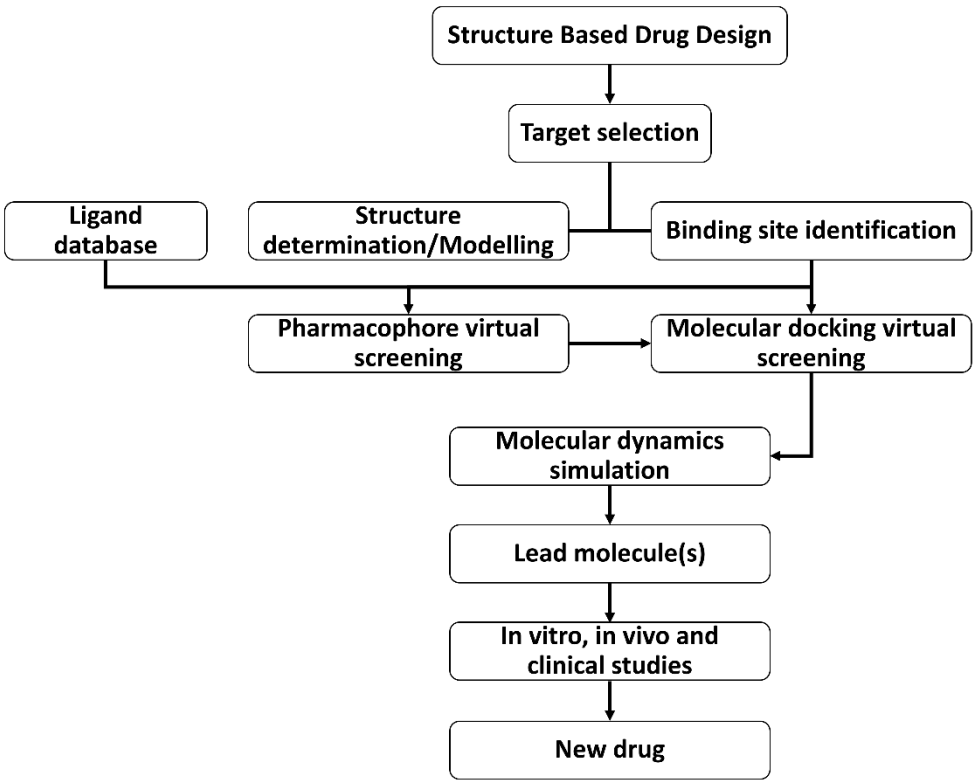

Figure 2. Structure-based drug design pipeline for the identification of lead molecules (adapted from (2)).

Once the target protein is known, detailed information about its 3D structure and binding cavity is crucial. This information aids in predicting optimal ligand conformation for protein inhibition or activation(2). In case a reference ligand exists within the crystallographic structure of the protein, the essential binding site amino acid residues can be determined. Otherwise, van der Waals forces,

hydrogen bond interactions, and binding energy calculations can map these residues. (2). With the target protein and key binding site residues identified, virtual screening can discover lead compounds for in vitro/in vivo testing. Virtual screening computationally identifies potential drug candidates from vast databases, using either a pharmacophore approach followed by molecular docking, or direct molecular docking. (26).

3.1 Molecular Docking

Molecular docking, one of CADD methods, predicts the position of the ligand in the protein binding cavity (called pose) and estimates binding affinity (27). The predicted ligand poses, its position and arrangement, are defined by translation (location), orientation (rotation), and conformation (internal geometry). Accurate pose is crucial for predicting binding affinity and biological activity, as changes in these parameters impact molecular interactions. Binding cavity predictions reveal insights into protein functionality, while binding affinity estimates reveal insights into the strength of ligand-protein binding (28, 29). Docking studies expose molecular interactions, such as hydrogen bond formation, that indicate the stability of the complex (28). Characterization of ligand-binding behaviour is essential for rational drug design and understanding biochemical processes (17).

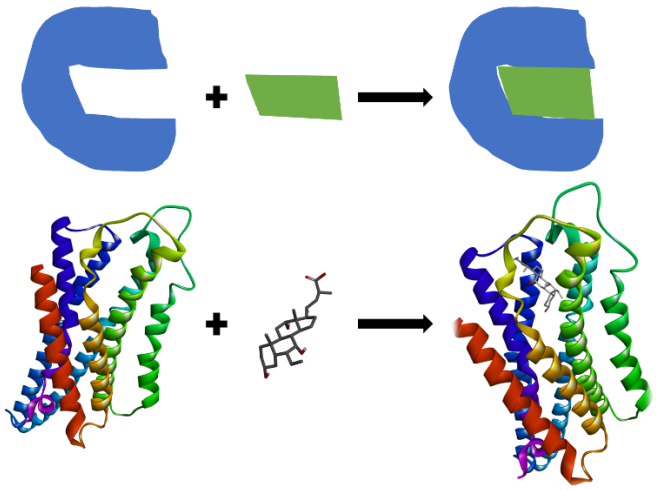

Figure 3. Schematic illustration of docking a small molecule ligand (green) to a protein target (blue) producing a stable complex.

In contrast to conventional docking, which seeks optimal ligands for a known target, reverse docking reverses this strategy by screening a single compound against a library of protein structures to identify potential binding targets. This method is particularly useful in target fishing, mechanism of action studies, and drug repurposing, especially when the

biological target of a compound is unknown (30). Reverse docking involves systematically docking the compound of interest into the binding sites of multiple proteins, often retrieved from databases such as the Protein Data Bank (PDB) (31). Scoring functions then rank the predicted protein-ligand interactions to prioritize targets for further validation such as experimental validation. Together, molecular docking and reverse docking offer a powerful in silico toolkit for both ligand-based screening and target discovery, streamlining early-stage drug development and guiding experimental validation.

In the docking process, the ligand is docked in the protein binding cavity based on the applied docking algorithm. For instance, a ligand with a relatively simple structure could possess various degrees of freedom corresponding to possible conformations in the protein binding cavity, making it a challenge to obtain the most suitable conformation in the protein binding cavity with considerable accuracy in a short period of time (17). As an evaluation of the conformation or docking pose of the ligand, a scoring function is added to the docking algorithm. The scoring function is set up to predict the biological activity based on ligand-protein interaction (17).

3.1.1 Docking Methodology

Docking methodologies are generally classified into rigid docking and flexible docking, each with varying levels of complexity and accuracy. Rigid docking simplifies the docking process by treating both the ligand and protein as static objects, which enhances computational efficiency but reduces accuracy, particularly for highly flexible biomolecules (32). This approach is appropriate for high-throughput virtual screening (HTVS) due to the rapidity of it. Flexible docking, on the other hand, takes into account conformational changes in either the ligand, protein, or both, thus improving accuracy but increasing computational cost (33). Flexible docking, also known as induced fit docking (IFD), allows for a more accurate prediction of binding poses, especially for targets with flexible or partially unresolved active sites (34). Docking methodologies also vary in their approach to identifying optimal ligand conformations. Systematic search methods exhaustively explore all possible conformations but are computationally demanding. In contrast, stochastic search algorithms, such as GA, Simulated Annealing (SA), and Monte Carlo (MC) methods, introduce an element of randomness to efficiently sample the conformational space (35, 36). Furthermore, ensemble docking, which considers multiple receptor conformations, improves predictions for targets exhibiting significant structural flexibility (37).

3.1.1.1 Rigid Docking

Rigid docking is a molecular docking approach in which both the ligand and receptor are treated as fixed, disallowing any conformational flexibility during the docking process. This methodology assumes that the protein maintains a static structure, while different orientations and positions of the ligand are assessed within the binding cavity (33). Rigid docking is computationally efficient and widely used in HTVS, where screening thousands to millions of compounds against a target protein within a reasonable timeframe (17). However, this method has limitations, especially when applied to flexible molecules or dynamic proteins that undergo conformational changes on ligand binding. Since rigid docking fails to consider the effects of induced fit, it can lead to inaccurate binding predictions, especially for highly flexible ligands (32). However, this method can be used on both rigid and highly flexible ligands depending on the preparation of favourable ligand conformations for the docking input. Flexible ligands, however, require pre-generated conformer sets by more sophisticated methods such as molecular dynamics or Monte Carlo sampling to represent their diverse conformations, enabling rigid docking to evaluate multiple plausible shapes without real-time generation of ligand conformers (38).

Rigid docking is frequently used as starting point of screening before employing more computationally intensive flexible docking methods. Numerous docking programs, including AutoDock, DOCK, and Glide, offer rigid docking as a default option for rapid ligand binding assessment (35). The scoring functions in rigid docking primarily evaluate non-bonded interactions, such as van der Waals forces, electrostatics, and hydrogen bond, to rank potential binding poses. However, rigid docking neglects protein flexibility and its results may require post-processing techniques, such as MD simulations, to refine and validate the predicted binding conformation (33). Despite these limitations, rigid docking remains a valuable tool in structure-based drug discovery, especially when high-resolution experimental protein structures are available and ligand flexibility is limited.

3.1.1.2 Semi Flexible Docking

Semi-flexible docking is a molecular docking approach in that the ligand is treated as flexible, allowing it to change its shape, while the receptor is kept rigid. This method offers a practical balance between computational efficiency and biological relevance, especially when dealing with small, conformationally flexible ligands and well-defined receptor structures. During semi-flexible docking, the ligand is permitted to explore different conformations by

rotating its torsional bonds, enabling the identification of optimal binding poses within the static receptor binding site (39, 40). This approach is widely used in virtual screening workflows because it can efficiently handle large compound libraries while capturing essential aspects of molecular recognition by considering how well the ligand's shape matches the receptor's and how well they can form hydrogen bonds. While semi-flexible docking simplifies the process by ignoring the receptor's flexibility (which can be important for certain binding mechanisms like induced fit), it can still provide reliable predictions when the binding site of receptor is relatively rigid, or when a detailed structure of the receptor is available (e.g., from X-ray crystallography). Tools like AutoDock Vina, Glide, and GOLD commonly implement semi-flexible docking algorithms, making this method a cornerstone in structure-based drug discovery (35, 41). By allowing conformational sampling of the ligand while maintaining a fixed receptor geometry, semi-flexible docking serves as an effective compromise between speed and accuracy, particularly for early-stage hit identification and lead optimization.

3.1.1.3 Flexible Docking

Flexible docking is a molecular docking approach designed to overcome the limitations of rigid docking by accounting for the conformational flexibility of the ligand and the receptor during the binding process. Unlike rigid docking, which treats both molecules as static, flexible docking allows for structural changes, leading to more biologically relevant and precise predictions of ligand-protein interactions (33). This strategy provides a more realistic simulation of molecular recognition, particularly when dealing with flexible ligands or dynamic protein binding sites. The conformations of the ligands are treated as described above. There are several strategies to address the flexibility of the protein receptor. The first strategy accounts for side-chain flexibility, particularly of residues in the binding cavity to move and adapt during ligand binding. This can be implemented through precomputed rotamer libraries of specific amino acids, or by allowing full conformational sampling of selected side chains (treating them in a similar way as the ligand, used for example in AutoDock and AutoDock Vina. Although it requires more computing power, this approach provides better predictions, especially when the receptor's shape changes significantly when binding occurs (35, 38). Alternatively, ensemble docking provides another solution for receptor flexibility, docking ligands into multiple static conformations of the protein derived from experimental structures or molecular dynamics simulations, enabling the capture of a broader range of potential interactions (37). Flexible docking methodologies are expected to further enhance drug discovery by providing more accurate and physiologically relevant binding predictions.

3.1.2 Molecular Docking Algorithms

Several algorithms have been applied in the development of molecular docking programs (42). This part focusses on an overview of the most used algorithms along with the molecular docking software samples using the algorithms. Molecular docking software can use one or multiple algorithms in the search for specific ligand conformations in the protein binding cavity.

3.1.2.1 Monte Carlo Algorithms

The Monte Carlo (MC) algorithm is a stochastic simulation method applied in several docking software packages such as AutoDock, MCDOCK, AutoDock Vina, QXP, and ROSETTALIGAND (35, 43-45). MC does not adopt a deterministic fixed-rule approach but is based on the MC algorithm of Metropolis, which provides acceptance criteria in the evolution of docking search (46, 47). The MC algorithm utilises random sampling to examine the possible states of the system, making it valuable for studying molecular conformations, ligand-receptor interactions, and thermodynamic properties. By generating random configurations of molecules and evaluating their energies using statistical mechanics, these simulations assist in identifying the most favourable conformations with high binding affinity (48). The efficiency of Monte Carlo sampling can be further enhanced through such techniques as simulated annealing and the Metropolis-Hastings algorithm, which enable the system to move out of the local minimum energy and explore a wider conformational space (49). Monte Carlo methods have been successfully applied in different aspects of structure-based drug design, including flexible molecular docking (explained earlier in the section 3.1.1.3), solvation modelling, and protein folding simulations. For example, the hybrid Monte Carlo/molecular dynamics (MC/MD) approach combines the strengths of both methods, allowing efficient sampling of molecular movement while incorporating physics-based force fields (50). One of the limitations of MC algorithms is that it is not suitable for time-dependent simulations such as molecular dynamics simulations (39). However, Monte Carlo can play an essential role in the estimation of binding free energies in molecular systems and can complement other computational approaches such as MD (51).

3.1.2.2 Genetic Algorithms

A genetic algorithm (GA) is an evolutionary optimisation method inspired by Darwin's theory of natural selection and genetic variation, which is widely used in molecular docking to explore binding interactions between ligands and target proteins (52). GA are valuable tools

in molecular docking, especially when the precise geometry or conformational flexibility of the binding cavity of the protein is unclear or dynamic (36). In molecular docking, GA is highly effective in searching a wide conformational space of ligand poses to identify the optimal binding mode with minimal energy (36, 53). The algorithm runs through an iterative process of selection, crossover, and mutation, mimicking biological evolution to fine-tune ligand conformation and orientation (54). Initially, a population of ligand poses is generated, and each pose is evaluated using a scoring function that estimates binding affinity. The best-performing poses (i.e., those with the lowest binding energy) are selected for reproduction, allowing crosses and mutations to introduce diversity and explore new conformations (36). This iterative refinement allows the GA to efficiently converge to the most favourable ligand-binding mode, making it a powerful approach for flexible molecular docking. One well-known application of GA in molecular docking is in the AutoDock software suite, which applies the Lamarckian Genetic Algorithm to optimise ligand docking while incorporating local minimisation to improve accuracy (32). In comparison with traditional exhaustive searching methods, GA-based docking provides a balance between efficiency and accuracy of computation, enabling fast screening of large chemical libraries (55). As a result, GA-based docking remains a key strategy in modern structure-based drug discovery, enabling the identification of novel therapeutics with improved precision.

3.1.2.3 Simulated Annealing Algorithms

In molecular docking, simulated annealing (SA) is employed to explore the conformational space of ligand-receptor interactions by mimicking the physical annealing process in metallurgy that involves the factor of temperature. The term temperature refers to a control parameter that governs the probability of accepting a higher-energy conformation during the exploration of optimal ligand-receptor binding poses. The algorithm starts by generating an initial ligand conformation and iteratively perturbs it to explore alternative binding poses. At each iteration, alterations with lower energy of the system are favoured, while higher-energy conformations are also favoured with a certain probability of passing the minimum local energy. As the 'temperature' gradually decreases, the system stabilises into an optimised docking conformation with minimised binding energy (49). The ability of SA to efficiently navigate complex energy landscapes makes it particularly effective for flexible molecular docking, where ligands and proteins can undergo conformational changes during binding (49). SA has been successfully integrated into various molecular docking programmes, such as AutoDock, as an addition to genetic algorithms (GA) for flexible docking (explained

earlier in the section 3.1.1.3) (32). Studies have shown that SA-based docking methods can improve binding pose accuracy and enhance sampling efficiency, especially when dealing with highly flexible ligands or challenging docking scenarios (56).

3.1.2.4 Tabu Search Algorithms

Tabu Search (TS) is an advanced metaheuristic optimisation algorithm to guide the search for an optimal result by maintaining a memory-based adaptive search mechanism. In complex molecular docking, standard local search methods often become trapped in local energy minima. TS addresses this limitation by employing a memory-based strategy to explore both local and global regions of the energy landscape. TS utilizes a "tabu list" to prevent revisiting recently evaluated molecular configurations, allowing it to escape local minima and explore more promising areas. Aspiration criteria and intensification/ diversification techniques further refine TS's ability to navigate these complex energy landscapes. Consequently, TS bridges local and global search by avoiding cycling and promoting broader exploration, thereby increasing the probability of identifying the optimal molecular docking conformations (57). Furthermore, systematically moving from non-optimal binding poses and continuously refining ligand orientation, TS improves docking accuracy and efficiency in SBDD (58). Recently, TS has been integrated into docking software such as PRO_LEADS and AutoDock to improve ligand binding predictions (59, 60). Studies have shown that TS outperforms traditional stochastic search algorithms in certain docking scenarios by quickly converging to the global minimum energy conformation while maintaining computational efficiency (61).

The integration of various docking algorithm methods plays a pivotal role in enhancing the accuracy and reliability of molecular docking studies. Different algorithms such as GA, SA, and TS offer unique advantages in exploring the conformational space of ligand–receptor interactions. By combining these approaches, docking tools can more effectively balance thorough conformational sampling with efficient energy minimization. For instance, hybrid algorithms like the Lamarckian Genetic Algorithm used in AutoDock integrate evolutionary search strategies with local refinement techniques to improve docking precision. Additionally, incorporating flexible docking protocols alongside rigid docking can better account for the dynamic nature of molecular interactions, particularly in cases involving flexible ligands or partially flexible protein binding sites. This methodological integration not only increases the likelihood of identifying accurate binding poses but also reduces false

positives, ultimately improving the predictive power of docking in structure-based drug design.

3.1.3 The Scoring Function

The success of docking an entire molecule on a binding cavity of a target protein or screening a vast virtual combinatorial library in the end depends on the accuracy of the scoring function that ranks the compounds (62). Scoring functions play an important role in molecular docking by estimating binding affinity or interaction energy between a ligand and its target protein. The ligand conformation is assessed instantly while the ligand is posed inside the binding cavity, or all the generated poses can be scored at the end (62, 63). Scoring functions are broadly classified into three categories: empirical, knowledge-based, and force field-based scoring functions (64). Empirical scoring functions rely on derived mathematical models from experimental data on binding properties, whereas knowledge-based functions use statistical analyses of known protein-ligand complexes to estimate binding affinity (65, 66). Force field-based scoring functions, on the other hand, use basic principles of molecular mechanics and thermodynamics to simulate the energy contribution of each molecular interaction (67).

3.1.3.1 Empirical Scoring Methods

Empirical scoring methods in molecular docking approximate ligand-binding affinity through mathematical functions derived from experimental data of binding properties. These methods employ a linear regression model that assigns weighted coefficients to various molecular interaction terms, including van der Waals forces, hydrogen bonds, electrostatics, and desolvation effects, based on established protein-ligand complexes (68). The ability of empirical scoring functions to provide rapid binding affinity calculations is one of their key advantages, making them beneficial for high-throughput virtual screening in drug discovery. In contrast to physics-based methods, which require computationally intensive simulations, empirical methods estimate binding free energy based on statistical correlations, enabling accelerated evaluation of vast compound libraries (69).

Established empirical scoring functions like ChemScore, GlideScore, and X-Score incorporate experimentally derived parameters to improve docking accuracy (70). Although efficient, these methods have limitations, primarily their dependence on predefined weighting parameters that may not generalize well to different molecular systems. Their accuracy depends heavily on the quality and diversity of the training set and may not generalize well across different protein families or ligand types (69). Consensus scoring

approaches, which integrate several scoring methods, are frequently employed to address this problem, and enhance reliability (71).

3.1.3.2 Knowledge-based Methods

Knowledge-based assessment methods in molecular docking estimate binding affinity by obtaining statistical probabilities from vast datasets of experimentally determined protein-ligand complexes. These methodologies operate under the premise that the frequency of observed molecular interactions within known structures is indicative of favourable binding affinity, allowing them to construct statistical energy functions based on the spatial distribution of atoms within the binding site (72). In contrast to empirical scoring functions, which are parameterized using experimentally determined binding affinities, knowledge-based methods leverage interaction patterns extracted from structural databases, such as the Protein Data Bank (PDB), to generate pairwise interaction potentials. These potentials consider van der Waals forces, hydrogen bonds. hydrophobic and electrostatic interactions, providing a data-driven approach to ranking anchored poses (73). In summary, empirical methods quantitatively predict binding strength using experimental data, while knowledge-based methods assess binding likelihood by recognizing favorable interaction patterns in existing structural data.

Knowledge-based scoring functions offer a significant advantage by efficiently improving docking predictions through the utilization of extensive structural information. Popular knowledge-based scoring functions include DrugScore, PMF (Potential of Mean Force), and ITScore, all of which derive parameters from statistical analyses of protein-ligand complexes (74). However, these methods are limited by their reliance on the quality and diversity of training datasets, potentially introducing biases. To overcome these issues, hybrid approaches have emerged, integrating knowledge-based methods with empirical or machine learning-driven scoring functions to improve prediction accuracy (75). With the continued growth of structural databases and the refinement of interaction potentials by machine learning models, knowledge-based scoring functions are expected to become even more crucial in molecular docking and structure-based drug discovery.

3.1.3.3 Force Field-based Methods

Force field-based scoring methods work by calculating the interaction energy between the ligand and target protein, utilizing classical physics-based force fields. These methods rely on molecular mechanics (MM) principles, wherein the potential energy of the

system is determined as the sum of bonded (bond stretching, angle bending, and torsional rotation) and non-bonded (van der Waals and electrostatic) interactions (76). In contrast to empirical and knowledge-based scoring functions, force field-based methods explicitly model atomic interactions through established energy functions such as CHARMM, AMBER, and OPLS, thereby enhancing physical interpretability (77). Due to their basis in fundamental physical laws, force field-based scoring functions can provide accurate predictions, contingent upon proper parameterization and integration with solvent models to account for desolvation effects.

Force field-based scoring methods come with a significant advantage in structure-based drug design through their capacity to model molecular flexibility and dynamic interactions. Nevertheless, these methods are computationally demanding, require extensive energy minimization and MD simulations for ligand pose refinement (78). Hybrid approaches that integrate force field-based approaches with empirical or knowledge-based scoring functions are employed to enhance efficiency, achieving a balance between accuracy and computational expense (71). In addition, advances in GPU-accelerated computing and enhanced sampling techniques have increased the feasibility of force field-based scoring for large-scale virtual screening (79). As computation resources continue to grow, force field-based methods are expected to play an increasingly prominent role in molecular docking, especially when combined with free energy perturbation calculations for precise binding affinity estimations.

Scoring functions, despite their importance in molecular docking, have limitations that impact the accuracy of binding affinity predictions. Conventional scoring functions simplify calculations, sacrificing the complexity of molecular interactions by inadequately addressing conformational flexibility, solvation, and entropy (68, 80). This can lead to incorrect pose ranking and inaccurate virtual screening results. Scoring functions also lack transferability and struggle with absolute binding energy predictions. To address these challenges, consensus scoring, combining multiple scoring functions to enhance predictive accuracy, has become a widely adopted approach (69). Furthermore, advances in artificial intelligence and machine learning have facilitated the generation of data-driven scoring functions that incorporate deep learning models to improve binding affinity predictions (71).

3.2 AI-docking

Artificial intelligence (AI) is rapidly transforming the field of drug discovery, particularly in molecular docking, significantly improving the accuracy and efficiency of ligand-protein interaction predictions. While traditional docking methods are constrained by predefined scoring functions, heuristic algorithms, rigid receptor models, oversimplified energy functions, and limited generalization, AI-based docking approaches leverage machine learning (ML) and deep learning (DL) to learn complex, non-linear relationships directly from structural and experimental data, overcoming these limitations. A significant advancement in AI docking is the emergence of learned scoring functions. These functions replace traditional physics-based or empirical scoring methods with models trained on extensive datasets of known protein–ligand complexes and their corresponding binding affinities. Machine learning techniques, including random forests, support vector machines (SVMs), and deep neural networks (DNNs), have proven effective in enhancing both pose prediction and binding affinity estimation (75). By learning from data, these models can capture uncommon interactions, such as binding site flexibility, water-mediated interactions, and ligand strain, which are often overlooked by classical scoring functions.

Graph neural networks (GNNs) and 3D convolutional neural networks (3D-CNNs) have become prominent in docking due to their ability to process complex spatial information (81). Tools like AtomNet, DeepDock, and GNINA have shown improved virtual screening performance by learning to predict binding poses and affinities from 3D atomic environments (81-83). Furthermore, AI enhances the sampling stage of docking. Reinforcement learning and generative models improve exploration of ligand conformational space and can propose novel poses or *de novo* ligands. These AI-driven approaches increase hit rates, reduce false positives, and accelerate early drug design. AI-based docking is transforming computational drug discovery, enabling more intelligent, data-driven decision-making and expanding the possibilities for accurate prediction of ligand–target interactions. The synergy between evolving AI models and structural biology tools like AlphaFold is expected to significantly enhance docking and lead optimization, accelerating the drug discovery process.

3.3 Molecular Dynamics Simulation

Molecular dynamics (MD) simulation is a computational method employed to investigate the physical movements of atoms and molecules over time (84). These simulations provide detailed insight into the structural and dynamic behavior of biomolecular systems, such as proteins, nucleic acids, and drug molecules, under physiological conditions (85). Unlike

static molecular docking, which predicts a single ligand-receptor binding pose, MD simulation allows researchers to observe how molecules interact dynamically, accounting for solvent effects, temperature fluctuations, and conformational changes (86). This method is frequently employed to identify equilibrium distributions across various conformations, determine novel conformations, and characterize shifts in structural conformational distribution results from ligand binding or structural mutation (28). MD simulation uses Newton's motion laws regarding velocities and position of atoms to simulate atomic trajectories, with interactions defined by molecular mechanics force fields such as AMBER, CHARMM, OPLS, or GROMOS (77, 78, 87, 88). In addition to structural insights, molecular dynamics MD simulations enable accurate estimation of thermodynamic properties through free energy perturbation (FEP) calculations. FEP calculates the free energy difference between two molecular states, for example different ligands bound to the same receptor, by alchemically transforming one molecule into the other. This transformation is achieved through a series of simulations across multiple intermediate states (lambda windows), and the free energy difference is derived from statistical averaging. By accounting for solvent interactions, conformational flexibility, and entropic effects, FEP is particularly useful for ranking ligand affinities with high precision during lead optimization phases in drug discovery (89, 90).

MD simulations generate trajectory files with time-specific atomic coordinates (91), enabling dynamic analysis of ligand-protein behaviour, unlike molecular docking in which no time perspective is involved, therefore obtaining fixed coordinates. MD trajectories are further analysed depending on the objectives such as information on stability, flexibility, and ligand-protein interactions over time (7). Some typical evaluation parameters in MD trajectories include root mean square deviation (RMSD) to measure structural stability, root mean square fluctuation (RMSF) to assess residue flexibility, and radius of gyration (Rg) to evaluate biomolecular compactness (84). RMSD can provide information on the stability of the protein or ligand complex and conformation-related data of the ligand in the protein binding cavity in minimized global energy conditions (84). Moreover, RMSF provides insight into the fluctuations of protein residues to determine residues with flexible tendencies such as the N and C terminals and the loops of the protein (92, 93). Rg provides further information related to protein compactness or ligand-protein complexes to understand the nature of conformational changes or variations in proteins (94).

Further study objects relating to more insight into molecular interactions and conformational changes are hydrogen bonds. Hydrogen bonds contribute to molecular

recognition, stability, and strength of ligand-protein complexes as well as the specificity of ligand-protein interactions (95, 96). The number, geometry and fluctuations of hydrogen bonds formed during MD simulations can provide a view of the persistence of key interactions that contribute to the binding affinity and stability of the ligand. For example, high occupancy of hydrogen bonds of a ligand in the protein binding cavity indicates strong and stable interactions that can be linked to the biological activity of the protein (93). Another analysis related to molecular interactions is ionic interaction. Ionic interactions (electrostatic attractions between oppositely charged groups) are crucial for the stability, specificity, and function of biomolecular systems. In molecular dynamics (MD) simulations, analysing these interactions reveals how salt bridges, ion pairs, and charged residues impact binding affinity, stability, and dynamics. MD monitors ionic interactions by measuring the distance and persistence between oppositely charged atoms (e.g., aspartate/glutamate and lysine/arginine), often using distance-based criteria and frequency analysis. MD force fields (e.g., AMBER, CHARMM, GROMOS) model electrostatics with Coulomb's law and Particle Mesh Ewald (PME) for long-range interactions, allowing observation of ionic interaction changes during simulations, which is vital for studying conformational changes, ligand binding, and mutational effects.

3.3.1 Force Fields for Ligands

In molecular dynamics (MD) simulations, accurate representation of both the protein and the ligand is essential for obtaining meaningful results. While protein force fields such as CHARMM36 or AMBER ff14SB are well-established, ligands often require specialized parameterization due to their structural diversity and chemical complexity. Unlike proteins, ligands are not covered comprehensively by standard biomolecular force fields, necessitating the use of general-purpose force fields specifically developed for small molecules. To ensure compatibility with the protein force field and accurate modelling of ligand behavior, small molecules are typically parameterized using general force fields such as CHARMM General Force Field (CGenFF), General AMBER Force Field (GAFF), or Optimized Potentials for Liquid Simulations in all atom (OPLS-AA) compatible tools (87, 97).

One widely used solution is the CGenFF, compatible with the CHARMM family of force fields and designed to parameterize a wide range of drug-like compounds. Based on a structure of ligand, parameters such as bond lengths, angles, dihedrals, and partial charges can be automatically generated using the CGenFF web server or ParamChem tools (97). Similarly, the GAFF in the AMBER ecosystem provides parameters for small molecules. Tools

like Antechamber facilitate the conversion of ligand structures into topology files compatible with protein simulations (87). OPLS-AA force fields offer equivalent parameterization schemes for use in compatible MD engines. The accuracy of ligand force field parameters is critical for reliable protein-ligand simulations because they directly impact binding energy calculations, interaction stability, and conformational sampling. For example, in this work (refers to P1 article on section 4.1), ligand topologies were generated using CGenFF to ensure compatibility with the CHARMM36 force field used for the protein targets. These parameters facilitated accurate modelling of ligand dynamics within the receptor complex, enhancing the reliability of the molecular dynamics simulation results (98).

3.3.2 Force Fields for Proteins

Force fields are fundamental components of molecular dynamics (MD) simulations, providing the mathematical framework to describe the potential energy of a molecular system. For proteins, force fields define how atoms interact through bonded (bond lengths, angles, torsions) and non-bonded (electrostatics, van der Waals) terms. These parameters are carefully derived from experimental data and quantum mechanical calculations to ensure the accurate modelling of protein structure, dynamics, and interactions (98). These interactions govern atomic motion according to Newton's laws, enabling the simulation of diverse molecular behaviors. The combination of bonded and non-bonded terms determines the total potential energy, which is then used to compute the forces acting on individual atoms. Widely used biomolecular force fields like AMBER, CHARMM, GROMOS, and OPLS have been rigorously validated for simulating proteins, lipids, and nucleic acids (98).

The accuracy of a MD simulation is highly dependent on the quality of its force field parameters, which are derived from experimental data and high-level quantum mechanical calculations. Force fields are generally empirical, meaning they are parameterized to reproduce known molecular geometries, vibrational frequencies, solvation properties, and thermodynamic measurements. Several well-established all-atom force fields are widely used in protein simulations, including CHARMM36, AMBER ff14SB, and OPLS-AA. Each has been extensively validated and optimized to reproduce key biophysical properties such as protein folding, backbone flexibility, and secondary structure stability (98). For complex or novel molecules, such as drug-like ligands, additional parameterization using tools like CGenFF (CHARMM General Force Field) or Antechamber (AMBER) may be necessary to maintain consistency with the protein force field model. In essence, force fields are essential for MD simulations, providing the energy functions that govern atomic movements and dictate

molecular behavior. Ongoing advancements in simulation techniques and computational power are driving the evolution of force fields to include polarizable models and machine learning-based parameterization, thereby enhancing the realism and predictive power of MD simulations.

3.3.3 MD Methodology

In MD simulations, the methodological setup profoundly impacts the accuracy of biomolecular system modelling. The choice of thermodynamic ensemble, solvent representation, and system boundaries are key considerations, directly influencing the realism and relevance of simulation outcomes. In term of thermodynamic ensembles, MD simulations are performed under specific statistical ensembles, each defining the conserved thermodynamic properties of the system. The choice of ensemble is crucial and depends on the biological context being modeled. The most common ensembles are: NVE (constant number of particles, volume, and energy), NVT (constant number of particles, volume, and temperature), and NPT (constant number of particles, pressure, and temperature). For example, the NPT ensemble is widely used to mimic physiological conditions where pressure and temperature are maintained (99).

Solvation profoundly affects biomolecule behavior, making solvent representation a critical consideration. While implicit solvent models offer a computationally efficient approach by treating the solvent as a continuous dielectric medium, this simplification can limit the accuracy of simulating solvent effects on solute dynamics. Conversely, explicit solvent models, which represent individual water molecules using models like TIP3P or SPC, provide a more accurate representation of hydration effects and hydrogen bonding, but at a significantly higher computational cost (100). Defining the system boundary is a further crucial consideration for ensuring accurate MD simulations. Without proper handling, the artificial edges of a finite-sized simulation can introduce inaccuracies. Boundary conditions are implemented to address these "edge effects." Periodic boundary conditions (PBC) are frequently used to overcome this limitation. PBC effectively creates a repeating, infinite system by duplicating the simulation box in all directions. This allows particles to pass through one boundary and reappear on the opposite side, thus minimizing surface artifacts and better representing the behavior of a bulk material (101).

The most widely adopted and biologically relevant setup for simulating ligand-protein complexes in MD involves using the NPT ensemble, explicit solvent models, and PBC. This

combination closely mimics physiological conditions and ensures accurate modelling of molecular behavior in a solvated and pressure-regulated environment. MD Simulations are typically carried out in the NPT ensemble, maintaining a constant number of particles, pressure, and temperature parameters that reflect in vivo environments. A Berendsen or Nose–Hoover thermostat and a Parrinello–Rahman barostat are often used to control temperature and pressure, respectively (99). For solvation, explicit solvent models such as TIP3P or SPC water are used to represent water molecules individually, providing detailed and realistic hydration dynamics. These models allow the accurate simulation of hydrogen bonding, electrostatic screening, and solute-solvent interactions, which are especially critical in ligand-protein binding studies (100). To eliminate edge effects and mimic an infinite system, PBC is applied, whereby the simulation box is replicated in all directions. This ensures that molecules leaving one side of the box re-enter from the opposite side, maintaining the system's continuity, and avoiding artifacts associated with finite box size (101). In summary, NPT ensemble conditions, explicit water solvation, and periodic boundary conditions are the most common and recommended setup for simulating ligand-protein complexes offering a good balance between computational accuracy and biological realism.

3.4 Pharmacophores

The pharmacophore concept, which was first presented in 1970, is employed in CADD to assist in fully understanding small molecule interactions without requiring in-depth structural knowledge (18, 102). The term pharmacophore refers to an abstract representation of the essential molecular features required for a ligand to interact with a specific biological target and elicit a desired biological or pharmacological function of a small molecule (ligand) (10). A pharmacophore is defined as *"an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to modulate its biological response"* (103). Pharmacophoric features typically include hydrogen bond accepting and donating groups (HBA/HBD), aromatic rings, hydrophobic centers, and positive or negative charges, along with their precise spatial arrangement. These features are derived either from the properties of known active ligands (ligand-based pharmacophore modelling) or through analysis of ligand-target binding interactions (structure-based pharmacophore modelling) as represented in Figure 4 (104).

By utilizing methods like X-ray crystallography or molecular docking, structure-based pharmacophore modelling obtains pharmacophore features directly from the three-dimensional structure of the target protein binding cavity (104). In contrast, ligand-based

pharmacophore modelling does not require knowledge of the target structure in order to identify similar features among a set of known active compounds (10). Ligand-based pharmacophore modelling identifies shared structural features in active molecules to enable binding to a protein target without known 3D structure. By aligning known active ligands and identifying common chemical features, a pharmacophore model is generated (10, 102). Pharmacophore model construction is a critical step in the pharmacophore-based drug discovery workflow. The model acts as a template, defining spatial arrangement of features and inter-feature distances, enabling virtual screening and lead optimization to find molecules mimicking essential target-binding interactions (12, 105).

Pharmacophore-based screening offers significant efficiency and scalability advantages over molecular docking, enabling the rapid identification of bioactive compounds from vast chemical libraries. Their application extends to lead optimization, scaffold hopping, and hit-to-lead development, significantly accelerating the drug discovery process and reducing costs (106). In the context of scaffold hopping, pharmacophore models are used to identify structurally diverse compounds that retain the key interaction features required for biological activity. By preserving important features of the ligands while replacing the unimportant part (scaffold), this strategy enables the discovery of novel compounds with potentially improved pharmacokinetic properties or reduced off-target effects (107). Instead of calculating the complex interactions between receptors and ligands, pharmacophore screening identifies ligand conformations that align with crucial interaction features. Molecular docking, however, demands a more thorough approach, requiring the simulation of various ligand positions within the receptor's binding site, followed by an assessment and ranking based on predicted binding strength (108). This makes molecular docking a computationally extensive method, particularly when analysing vast compound libraries. However, a common challenge in pharmacophore screening is the preparation of ligand conformations. Ligands in the database must be converted into 3D structures and pre-generated in multiple low-energy conformations to effectively match a pharmacophore model. This process can be time-consuming and resource-intensive, especially when screening millions of compounds (108).

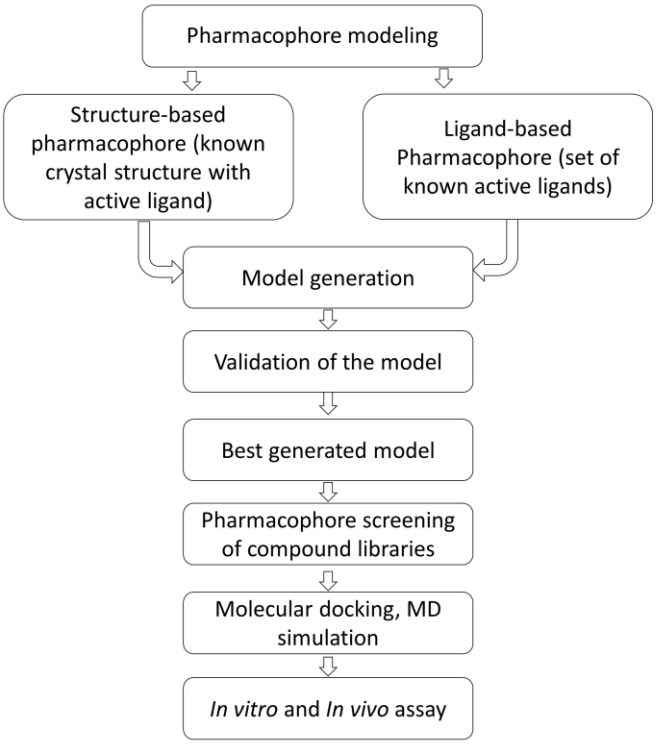

Figure 4. Flowchart illustrating pharmacophore modelling process in a CADD workflow (adapted from (104)).

Structure-based pharmacophore begins by analysing the binding site of the target protein, often using crystallographic or modelled structures of ligand–receptor complexes. Critical interactions such as hydrogen bonds, ionic interactions, hydrophobic contacts, and π-π stacking are identified between the ligand and residues in the binding cavity. These interactions are then abstracted into pharmacophore features with specific 3D spatial coordinates that collectively define the essential chemical characteristics for ligand binding (109). Furthermore, pharmacophore models are constructed by analysing the spatial arrangement of key interaction points in the receptor binding pocket, such as hydrogen bond donors and acceptors, hydrophobic regions, and charged residues (104). Notably, structure-based pharmacophore modelling is more applicable when high-resolution structural data (e.g., X-ray crystallography or molecular docking results) are available. This allows for scaffold hopping, even identifying novel chemotypes that bind in a biologically meaningful way based on the protein structure (106).

Pharmacophore model validation is an important step in ensuring the reliability and predictive strength prior to utilization for virtual screening or drug design. A well-validated pharmacophore model should accurately identify active compounds while minimizing false

positives and false negatives (104). The model could require to be further enhanced through validation data if the results of pharmacophore-based screening and experimental validation differ (110). Validation is commonly performed using a combination of internal validation, external validation, and statistical evaluation (109). Internal validation refers to theoretical validation in which a training dataset of known active and inactive compounds is applied to validate a pharmacophore model. A decoy set is introduced once there are lacking inactive compounds (111). The pharmacophore model is tested against a database composed of two sets: a small group of known active compounds and a large group of decoy compounds, which are physicochemically similar to actives (in terms of properties like molecular weight, logP, and hydrogen bond count) but are presumed inactive against the target. The ability of the model to distinguish between these compounds provides an initial measure of its effectiveness (104). Screening hits are classified as true positives (correctly detected active), false positives (inactive compound misidentified as active), true negatives (correctly rejected inactive), and false negatives (active compound misidentified as inactive). Performance is measured using sensitivity (percentage of correctly identified active substances), specificity (percentage of correctly identified inactive substances), yield of active, and enrichment factor (EF) (93).

External validation is performed by screening an independent test set containing compounds not used during model construction, to ensure the models can generalize to new molecules (10). Furthermore, model performance is assessed using statistical evaluation metrics like the Güner-Henry (GH) scoring method, the EF and the receiver operating characteristic (ROC) curve (93). The Enrichment Factor (EF) quantifies a model's ability to identify active compounds compared to random selection. It is the ratio of the proportion of active compounds retrieved by the model to the proportion expected by chance. An EF exceeding 1 indicates performance better than random screening, with higher values signifying greater screening efficiency (112). The Güner–Henry (GH) score comprehensively evaluates pharmacophore model performance based on the number of identified active compounds, the total number of retrieved hits, and the database size. GH scores range from 0 to 1, with scores above 0.7 indicating a highly effective model. This metric is particularly valuable for simultaneously assessing selectivity and efficiency (102). Good model robustness and efficiency in retrieving active compounds are indicated by high EF and GH values. The ROC curve assesses the efficiency of the model by plotting the rate of true positives versus false positives. A well-performing model is indicated by a steep curve with a high area under the curve (AUC) value close to 1, while random selection is indicated by an AUC of 0.5 (Figure 5)

(110). The pharmacophore model is set up for virtual screening once it has been validated (93).

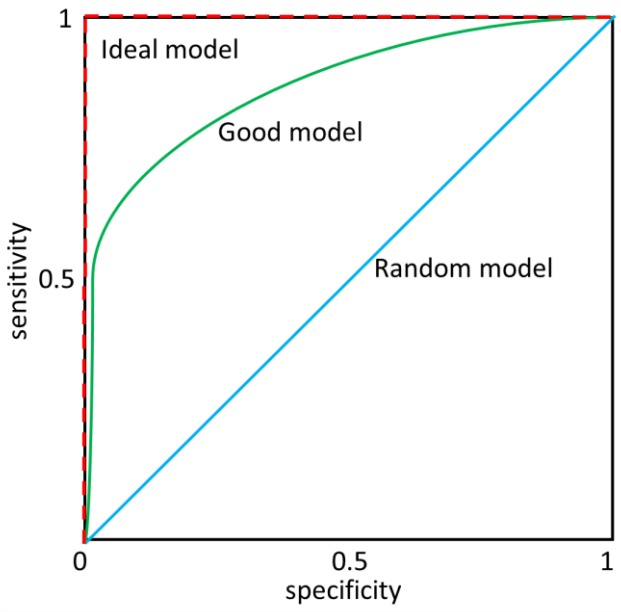

Figure 5. Ideal ROC curve (shown with red dashed line); Good performance ROC curve (shown with green line); Random curve on blue diagonal with specificity and sensitivity values of 0.5 (113)

4 RESULTS AND DISCUSSION

Molecular docking is essential in drug discovery to predict ligand-receptor interactions and design bioactive compounds. In my research, molecular docking serves as a core technique across three studies aimed at discovering novel TGR5 agonists and elucidation of the vasodilatory mechanism of substituted catechol compounds. The following publications highlight how molecular docking was integrated with pharmacophore modelling, molecular dynamics simulations, and experimental validation to advance our understanding of small molecule–target interactions and support early-stage drug development for substituted catechol compounds.

4.1. The Discovery of a New Non-Bile Acid Modulator of Takeda G Protein-Coupled Receptor 5: an Integrated Computational Approach (P1)

Salam R, Bakker M, Krutáková M, Štefela A, Pávek P, Duintjer Tebbens J, Zitko J. The discovery of a new non-bile acid modulator of Takeda G protein-coupled receptor 5: An integrated computational approach. Arch Pharm (Weinheim). 2025 Jan;358(1):e2400423. https://doi.org/10.1002/ardp.202400423.

In this study, HTVS scheme was effectively implemented as part of an integrated computational strategy to discover new non-bile acid agonists of the Takeda G protein-coupled receptor 5 (TGR5) through an SBDD approach involving pharmacophore screening, molecular docking and MD simulation as seen in Figure 6. TGR5 (GPBAR1) is a widely expressed G protein-coupled bile acid receptor that regulates metabolism and impacts various pathophysiological processes (114). TGR5 activation affects bile acid metabolism, blood sugar, energy intake, and macrophage activation (115). Primarily, TGR5 functions in blood glucose homeostasis and energy expenditure by promoting GLP-1 release, amplifying insulin secretion, and improving blood glucose homeostasis, making it a potential target for type 2 diabetes therapy (116). The discovery of the electron microscopy structure of TGR5 provided insights into ligand interactions in the TGR5 binding cavity that facilitated the construction of structure-based pharmacophore models in the search for new TGR5 agonists.

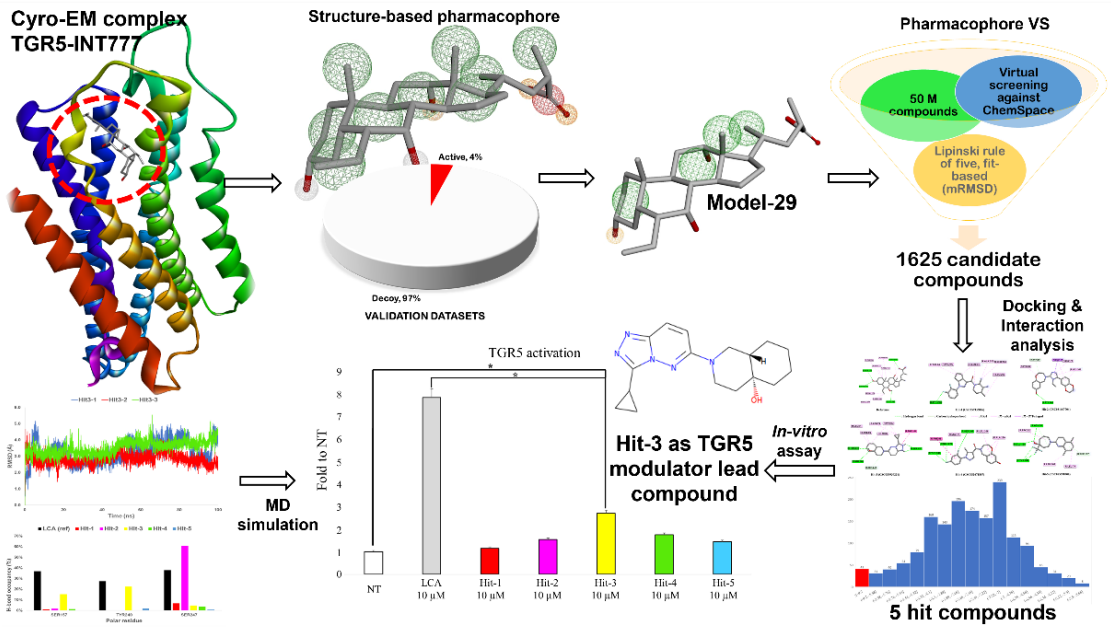

Figure 6. HTVS scheme of discovering new non-bile acid agonists of TGR5 (taken from P1 under CC-BY 3.0)(93).

The pharmacophore models for TGR5 were constructed based on the interaction of INT-777 with TGR5 (PDB ID: 7CFN). INT-777 is a semi synthesized bile acid derivative that has proven to activate the TGR5 receptor (117). Seven TGR5 residues were identified to directly interact with INT-777, emphasizing the importance of hydrogen bonding and hydrophobic interactions. The main interactions involved hydrogen bonds with Tyr240 and Ser247, as well as hydrophobic interactions enhancing binding stability. The hydroxyl group at the 3α position significantly affected the selectivity of the ligand towards TGR5. A pharmacophore model with six features was selected to optimize selectivity and reduce false positives. Of the 40 generated models, model-29 was selected due to having the highest selectivity and specificity (Figure 7). Model-29 was applied in a pharmacophore-based virtual screening using the ChemSpace database (50 million compounds), with additional filter conditions including Lipinski's rules, PAINS filtering, and RMSD values. The screening process narrowed down the candidate compounds to 1625 for further docking-based virtual screening applying semi-flexible docking method using AutoDock Vina with default setting. The method was validated by redocking the reference ligand INT-777, achieving a RMSD of 1.78 Å, within the acceptable range to confirm docking reliability.

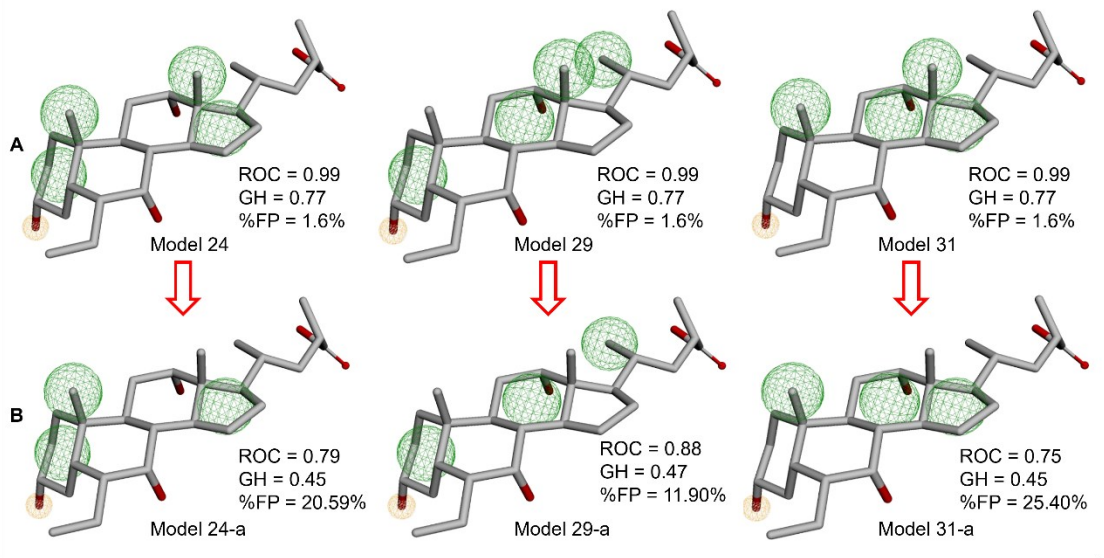

Figure 7. The three-dimensional (3D) arrangements of the selected pharmacophore models: model-24, model-29, and model-31. (b) The deletion of hydrophobic features in C13 of the three models (taken from P1 under CC-BY 3.0) (93).

For all candidate compounds, docking scores ranged from -10.2 to -5.2, of which forty-one had scores comparable to INT-777 (-9.2 kcal/mol). Subsequently, 41 candidate compounds were subjected to a second docking round with exhaustiveness increased to 32 (8 is default setting) to refine pose prediction and interaction profiling.

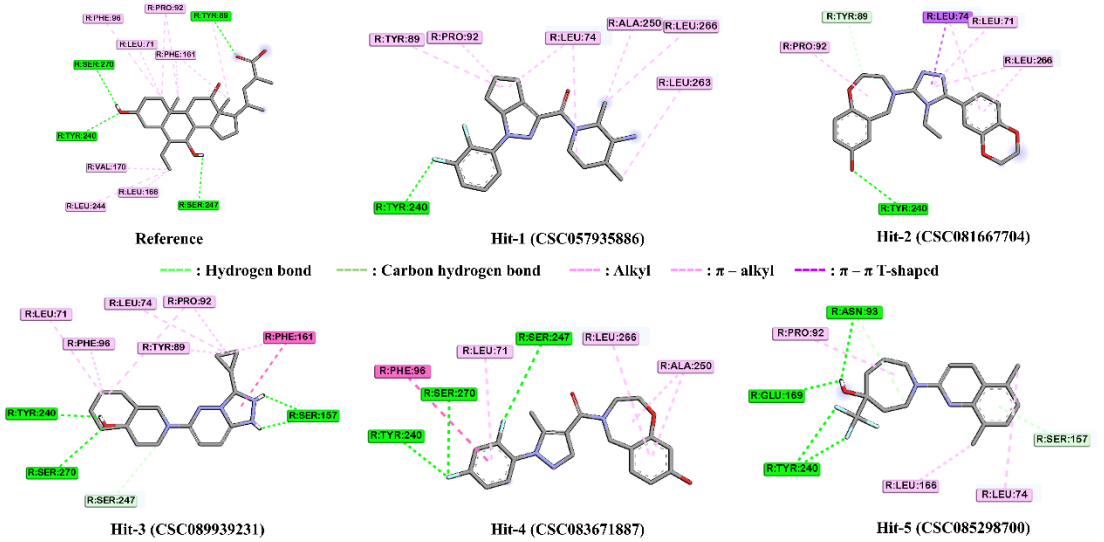

Figure 8. Two-dimensional (2D) interaction diagrams of five hit compounds in the ligand-binding domain of TGR5. (Dashed green line shows a hydrogen bond interaction; dashed purple line indicates a hydrophobic interaction) (taken from P1 under CC-BY 3.0) (93).

Of the 41, five hit compounds were selected for biological assays based on their ability to reproduce key interactions with TGR5 residues, particularly the hydrogen bond with Tyr240, a critical residue for receptor activation as shown in Figure 8. These hits also demonstrated supportive hydrophobic and π-π interactions that mirrored those of INT-777.

The biological validation of the computationally identified hit compounds revealed Hit-3 (CSC089939231) as the most promising non-bile acid TGR5 agonist among five candidates selected through pharmacophore and docking-based virtual screening. In cell-based assays using HepG2 cells, Hit-3 demonstrated statistically significant activation of TGR5 at 10 µM, though its efficacy was lower compared to lithocholic acid (LCA), the bile acid control (Figure 9). Notably, Hit-3 displayed concentration-dependent activation, indicating its potential as a genuine modulator of TGR5 activity. Due to solubility limitations, the exact EC_{50} of Hit-3 could not be determined, but it is expected to be higher than that of LCA. The reduced potency compared to LCA was attributed to sample composition, a mixture of stereoisomers of Hit-3, with only one isomer (4aR,8aR) predicted to form the critical hydrogen bond with Tyr240, a residue essential for TGR5 activation.

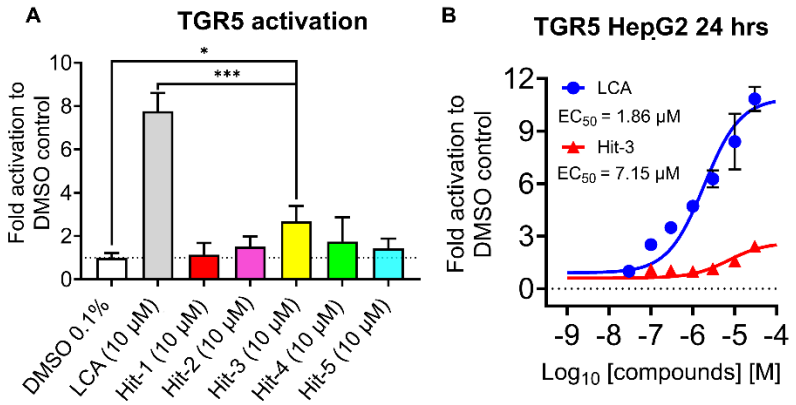

Figure 9. Takeda G protein-coupled receptor 5 (TGR5) activation in transfected HepG2 cells. (A) All hit compounds at 10 µM. (B) Concentration-response curves for the activation by lithocholic acid (LCA) and Hit-3. EC50 values were calculated using nonlinear fitting of concentration–response curves. Values are presented as means ± SD from three independent experiments performed in triplicates. *p < 0.05 Hit-3 versus NT; ***p < 0.001 LCA versus Hit-3 (taken from P1 under CC-BY 3.0) (93).

To further support and rationalize the observed bioactivity, MD simulations were performed for all hit compounds and the LCA control to analysis the stability and ligand-receptor interactions. Simulations were performed using GROMACS 2020, with 100 ns production runs for each complex (118). Protein topology files were generated using the CHARMM36 force

field, a robust and validated method for modelling biomolecules (119). CHARMM36 is particularly well-suited for simulating G protein-coupled receptors (GPCRs) such as TGR5, due to its compatibility with membrane proteins and its accuracy in representing structure and ligand interactions . CGenFF, the CHARMM General Force Field, was used to generate ligand topology parameters, complementing the CHARMM36 force field (97). This integrated use of the CHARMM36 and CGenFF force fields ensured consistent and reliable representation of interatomic forces throughout the simulation process.

TGR5 is known as a transmembrane protein but in this MD simulation, the transmembrane protein model was not applied. This MD simulation of TGR5 focused on ligand binding stability and interaction profiles rather than conformational changes or receptor activation mechanisms. This approach, consistent with previous studies (reported by Cheng et al., 2017 (68, 120)) showing soluble domain models yielding relevant insights for early drug discovery, is appropriate because many small molecule ligands primarily interact with the extracellular ligand-binding domains of GPCRs such as TGR5, and key binding interactions can be effectively captured in soluble domain models (120, 121). Aqueous phase simulations using soluble ligand-receptor complexes were chosen as a practical and widely acceptable alternative for early-stage screening and hit validation studies. This approach allowed us to assess ligand stability, binding interactions, and dynamics within the binding cavity with sufficient accuracy for the purposes of mechanistic insight and hit evaluation. The structural stability of TGR5 was not an issue in this study, as becomes evident from monitoring the RSMD and Rg of the protein.

Among the five candidates, Hit-3 (CSC089939231) demonstrated the most favorable binding characteristics. RMSD analysis revealed that Hit-3 had the lowest average RMSD value (3.20 ± 0.34 Å), indicating a highly stable binding pose throughout the simulation (Figure 10). Additionally, root mean square fluctuation (RMSF) analysis confirmed minimal fluctuations in the binding site residues, ensuring stable interactions between the ligand and TGR5 over time . Crucially, H-bonds occupancy analysis indicated that Hit-3 retained consistent H-bonds with key residues Tyr240 and Ser157, which both are critical for TGR5 activation (Figure 9). The ability of Hit-3 to form and retain these interactions during the trajectory distinguishes it from other hits that either lacked interaction with Tyr240 or exhibited greater structural instability. These findings from MD simulations corroborate the biological assay results, positioning Hit-3 as the most promising non-bile acid TGR5 agonist identified in this study, with the potential for further optimization in drug development pipelines.

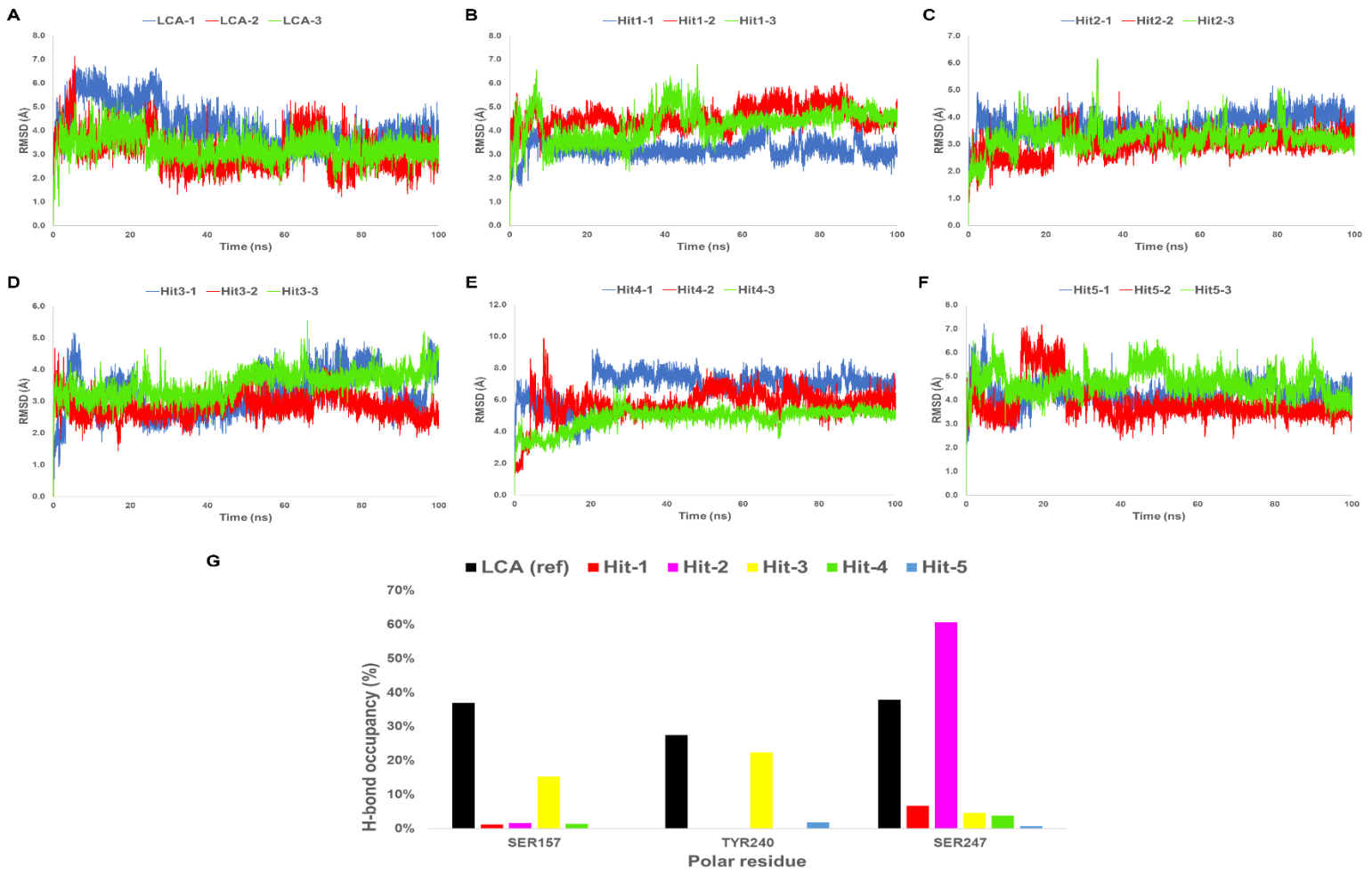

Figure 10. (A-F) RMSD graph of the protein backbone complexed with LCA (reference) and Hits 1–5 across three independent replicas. (G) Hydrogen bond occupancies (%) to TGR5 amino acid residues important for the LCA binding (occupancy > 10%) (adapted from P1 under CC-BY 3.0) (93).

4.2. The Quercetin Metabolite 4-Methylcatechol Causes Vasodilation via Voltage-Gated Potassium (K_V) Channels (P2)

Dias P, **Salam R**, Pourová J, Vopršálová M, Konečný L, Jirkovský E, Duintjer Tebbens J, Mladěnka P. The quercetin metabolite 4-methylcatechol causes vasodilation via voltage-gated potassium (K_v) channels. Food & Function. 2024;15(22):11047-59. https://doi.org/10.1039/D3FO04672A

In the experimental part of this study, it was observed that 4-methylcatechol (4-MC) induces relaxation in 4-MC-induced rat aorta and mesenteric arteries in vitro. It leads to a mean arterial pressure reduction of 10% in spontaneously hypertensive rats after a single iv dose of 2.5 mg/kg. Interestingly, the observed effects are not mediated through the heart, and most likely rely on a decrease in vascular resistance. Ex-vivo assay resulted in 4-MC enhanced cGMP and cAMP-mediated vasodilation in rat aorta, but not through direct activation of soluble guanylyl cyclase (sGC)/protein kinase G (PKG) or by affecting cAMP levels (Figure 10). Figure 3 (Panels A and B) suggests 4-MC enhances vasorelaxation through cGMP and cAMP pathways, as it boosts the effects of SNP and forskolin. However, inhibiting sGC and PKG with ODQ, DT3, and Rp-8-pCPT-cGMPS didn't reduce 4-MC's effect (Panels C and D), indicating that it doesn't directly target these enzymes. 4-MC likely influences cyclic nucleotides through a different mechanism.

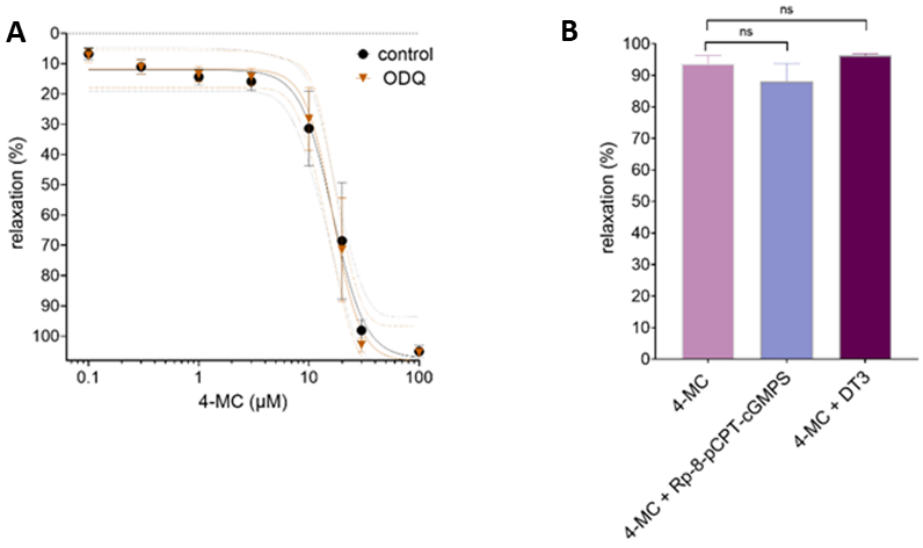

Figure 10. 4-Methylcatechol (4-MC)-induced vasorelaxation is not mediated via the (A) soluble guanylate cyclase/sGC (B) protein kinase G/PKG pathway (adapted from P2 under CC-BY 3.0) (122).

While L-type Ca2+ channels appear uninvolved, this study demonstrates that 4-MC promotes vasodilation primarily by activating K_v7 potassium channels (Figure 11). Blocking other channels with their respective inhibitors such as iberiotoxin (BK_{Ca}), glibenclamide (K_{ATP}), and BaCl2 (K_{IR}) had no impact on the effects of 4-MC, however, blocking K_v channels with 4-aminopiridine significantly reduced vasodilation, highlighting K_v channel importance. Further experiments using linopirdine to target K_v7 channels showed a partial reduction in vasodilation, suggesting a specific role for K_v7, particularly $K_v7.4$. Based on how well it binds, fits, and interacts with key molecules, $K_v7.4$ is the most likely channel responsible for the effects of 4-MC.

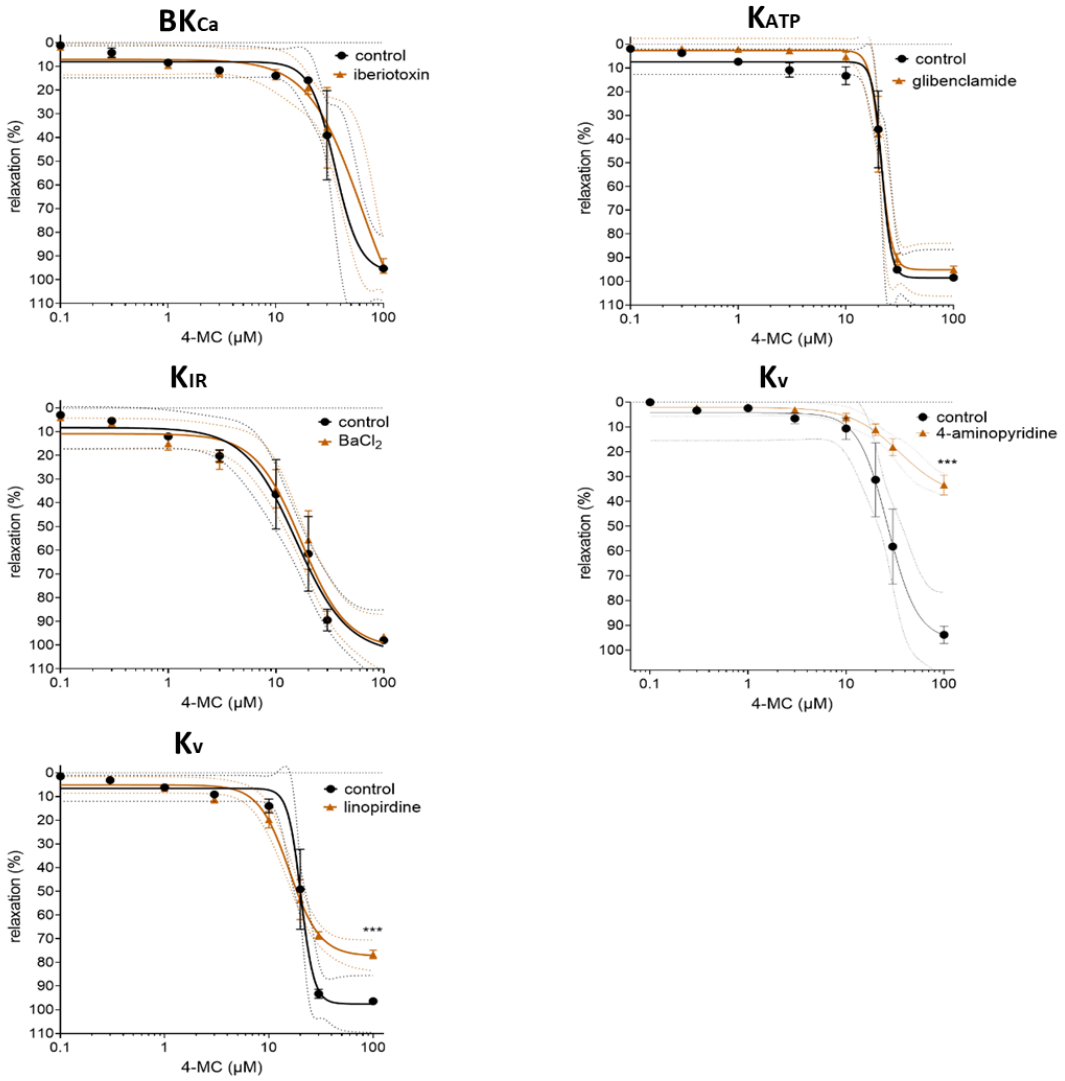

Figure 11. 4-Methylcatechol (4-MC)-induced vasorelaxation involves voltage-gated potassium (K_v) channel but not large conductance calcium-activated K^+ channels (BK_{Ca}), ATP-sensitive K^+ channels (K_{ATP}), inwardly-rectifier K^+ channels (K_{IR}) (adapted from P2 under CC-BY 3.0) (122).

In this study, molecular docking was employed as part of a reverse docking strategy (explained in section 3.1) to identify potential protein targets responsible for the vasorelaxant effects of 4-MC and to confirm the results of in vitro experiments on isolated rat aorta. For molecular docking studies, the crystal structures of all selected target proteins were retrieved from the RCSB Protein Data Bank (PDB) as listed in Table 1. High-resolution structures from *Homo sapiens* were used where available to enhance biological relevance. These targets were strategically selected to cover both upstream signaling enzymes and direct ion channel effectors involved in smooth muscle relaxation, allowing for a comprehensive evaluation of the vasodilatory mechanism of 4-MC using computational tools. Although some of the crystal structures of the target proteins have low resolution, by using experimentally validated human protein structures, the study ensured that docking predictions would closely reflect potential in vivo interactions of 4-MC within the human vascular system.

The ligand structures of 4-MC and its sulfate conjugates (4-methylcatechol-1-sulfate/4-MC-1-S and 4-methylcatechol-2-sulfate/4-MC-2-S) and the reference ligands of each target protein were retrieved from the PubChem database, energy-minimized using the MM2 force field in Chem3D, and subsequently converted into PDB format which was further converted to PDBQT format for docking. The crystal structures of target proteins were obtained from the RCSB Protein Data Bank and prepared using UCSF Chimera built-in Dock Prep v. 1.15. Binding sites of each protein, along with grid box and center box settings, were determined based on the co-crystal structure of the reference ligand bound to the protein. Docking simulations were performed using AutoDock Vina with semi-flexible method (explained in section 3.1.1.2), in which grid boxes configured to cover the active or binding regions of each protein. The docked poses were evaluated based on binding affinity scores and ligand efficiency (LE) (binding energy per heavy atom), and the best poses were further analyzed for molecular interactions using UCSF Chimera and Discovery Studio Visualizer.

The ligand structures of 4-MC and its sulfate conjugates (4-methylcatechol-1-sulfate/4-MC-1-S) and 4-methylcatechol-2-sulfate/4-MC-2-S) and the reference ligands of each target protein were retrieved from the PubChem database, energy-minimized using the MM2 force field in Chem3D, and subsequently converted into PDB format which was further converted to PDBQT format for docking. The crystal structures of target proteins were obtained from the RCSB Protein Data Bank and prepared using UCSF Chimera built-in Dock Prep v. 1.15.

Table 1. Target proteins investigated in the mechanistic studies of 4-methylcatechol (4-MC) used in the molecular docking studies.

Protein name	PDB ID	Organism	Uniprot	Method	Resolution (Å)	Reference ligand (Role of the ligand)
adenylate cyclase (AC)	1CJK	*Canis lupus familiaris**	P30803	X-ray diffraction	3.00	FOK (activator)
soluble guanylate cyclase (sGC)	7D9T	*Homo sapiens*	Q02108	cryogenic electron microscopy (cryo-EM)	4.10	Z90 (activator)
protein kinase G isoform Iα (PKG-Iα)	6C0T	*Homo sapiens*	Q13976	X-ray diffraction	1.98	EE4 (inhibitor)
protein kinase G isoforms Iβ (PKG-Iβ)	5JAX	*Homo sapiens*	Q13976	X-ray diffraction	1.49	6J7 (inhibitor)
protein kinase A (PKA)	3POO	*Homo sapiens*	P17612	X-ray diffraction	1.60	S69 (inhibitor)
sarco/endoplasmic reticulum calcium ATPase (SERCA)	6JJU	*Homo sapiens*	P16615	X-ray diffraction	3.20	BHQ (inhibitor)
voltage-gated K^+ channels $K_V7.1$ ($K_V7.1$)	7TCI	*Xenopus laevis**	P70057	cryo-EM	3.90	I0S (inhibitor)
voltage-gated K^+ channels $K_V7.2$ ($K_V7.2$)	7CR1	*Homo sapiens*	O43526	cryo-EM	3.40	GB9 (inhibitor)
voltage-gated K^+ channels $K_V7.4$ ($K_V7.4$)	7BYM	*Homo sapiens*	P56696	cryo-EM	3.10	FBX (inhibitor)

* No human structure available

Binding sites of each protein, along with grid box and center box settings, were determined based on the co-crystal structure of the reference ligand bound to the protein. Docking simulations were performed using AutoDock Vina with semi-flexible method (explained in section 3.1.1.2), in which grid boxes configured to cover the active or binding regions of each protein. The docked poses were evaluated based on binding affinity scores and ligand efficiency (LE) (binding energy per heavy atom), and the best poses were further analyzed for molecular interactions using UCSF Chimera and Discovery Studio Visualizer. LE, a critical parameter to assess binding quality, defined as the binding energy per non-hydrogen (heavy) atom, provides a size-normalized measure of ligand–protein interaction strength.

An important physiological modification of 4-MC was the sulfonation of one of its hydroxyl groups, which occurred in vivo and affected its interaction with target proteins. 4-MC, a phenolic metabolite derived from dietary polyphenols, undergoes sulfonation in the gut modulated by the protonated state of one of the hydroxyl groups, resulting in the formation of 4-MC-1-S and 4-MC-2-S. Molecular docking analysis was performed to evaluate the binding affinities emphasized on LE of 4-MC and its sulfate conjugates across several protein targets. Despite lower absolute docking scores compared to reference ligands, 4-MC showed higher LE values as shown in Table 2, particularly when docked to K_v7.1, K_v7.2, K_v7.4, and PKGIα, suggesting its small structure promotes efficient binding. In case of PKGIα, both in vitro experiment and molecular docking results provided strong evidence to rule it out as a relevant target. Additionally, molecular docking showed that although 4-MC exhibited a relatively favorable ligand efficiency score for PKG Iα, its binding pose and interaction profile were less compelling compared to confirmed targets such as K_v7.4.

Table 2. Docking scores and ligand efficiency (LE) of reference, 4-methylcatechol (4-MC), 4-methylcatechol 1-sulfate (4-MC-1S) and 4-methylcatechol 2-sulfate (4-MC-2S) for each target protein.

PDB ID	Target protein	Docking score (kcal/mol)				LE (docking score/heavy atom)			
		Ref	4-MC	4-MC-1-S	4-MC-2-S	Ref	4-MC	4-MC-1-S	4-MC-2-S
7D9T	sGC	-10.7	-5.3	-5.9	-5.9	-0.25	-0.59	-0.45	-0.45
6C0T	PKGIα	-8.2	-6.5	-6.6	-6.5	-0.31	-0.72	-0.51	-0.50
5JAX	PKGIβ	-11.6	-5.8	-5.7	-7.0	-0.48	-0.64	-0.44	-0.54
3POO	PKA	-9.3	-5.9	-6.8	-6.8	-0.40	-0.66	-0.52	-0.52
7TCI	K_V7.1	-8.8	-6.0	-6.0	-6.4	-0.28	-0.67	-0.46	-0.49
7CR1	K_V7.2	-8.1	-6.1	-5.9	-6.3	-0.37	-0.68	-0.45	-0.48
7BYM	K_V7.4	-7.6	-6.0	-5.8	-6.1	-0.35	-0.67	-0.45	-0.47
1CJK	AC	-9.7	-5.6	-6.2	-6.3	-0.33	-0.62	-0.48	-0.48
6JJU	SERCA	-9.7	-5.5	-6.0	-6.5	-0.31	-0.61	-0.46	-0.50

PDB ID, Protein Data Bank Identifiers; Ref, reference; sGC, soluble guanylate cyclase; PKGIα, protein kinase G isoform Iα; PKGIβ, protein kinase G isoform Iβ; PKA, protein kinase A; K_V, voltage-gated potassium channel; AC, adenylate cyclase; SERCA, sarco/endoplasmic reticulum calcium-ATPase.

K_v7.4 emerged as a key candidate, with docking simulations revealing shared binding residues and spatial overlap between 4-MC and the reference ligand retigabine. The high ligand efficiency value (>0.65) for K_v7.4, along with stable interactions involving critical residues Trp242 and Phe246, highlights the selective and energetically favorable binding of 4-MC to this channel. Docking results of its conjugates also revealed that sulfonation did not affect the binding poses of 4-MC, which affected the interaction with key residues (Figure 12). In the case of K_v7.4, the sulfonated derivatives retained the ability to engage in hydrogen bonds and hydrophobic interactions, particularly with the crucial residues on K_v7.4. This showed that although sulfonation could affect binding efficiencies, it did not eliminate the potential of 4-MC to interact with K_v7.4, reinforcing its role as a key target for vasodilation. These findings, consistent with in vitro data, strengthen the hypothesis that K_v7 channels, especially K_v7.4, are major targets contributing to the vasodilatory effects of 4-MC, positioning it as a potential novel antihypertensive lead compound.

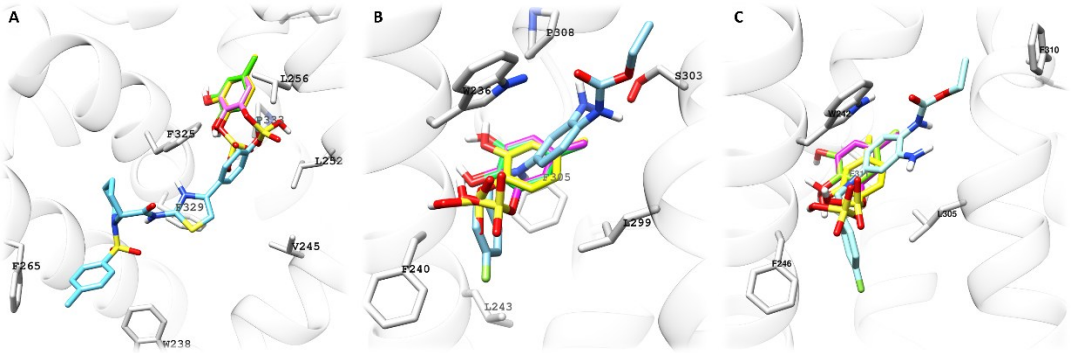

Figure 12. Docking pose of 4-methylcatechol (green), 4-methylcatechol-1-sulfate (yellow), 4-methylcatechol-2-sulfate (purple), and retigabine (turquoise) in the ligand binding domain of the human voltage-gated potassium channel: (A) K_v7.1; (B) K_v7.2; (C) K_v7.4 (taken from P2 under CC-BY 3.0) (122).

In conclusion, the overlapping binding poses of 4-MC and sulfonated derivatives with retigabine further support its potential role as a K_v7.4 agonist. These computational findings are in line with ex-vivo studies, where K_v channel inhibition significantly attenuated 4-MC-induced vasorelaxation, and with in-vivo data showing blood pressure lowering effects. Collectively, these results highlight K_v7.4 as a potential target of 4-MC, reinforcing its role in vascular smooth muscle relaxation and suggesting potential therapeutic relevance in cardiovascular modulation.

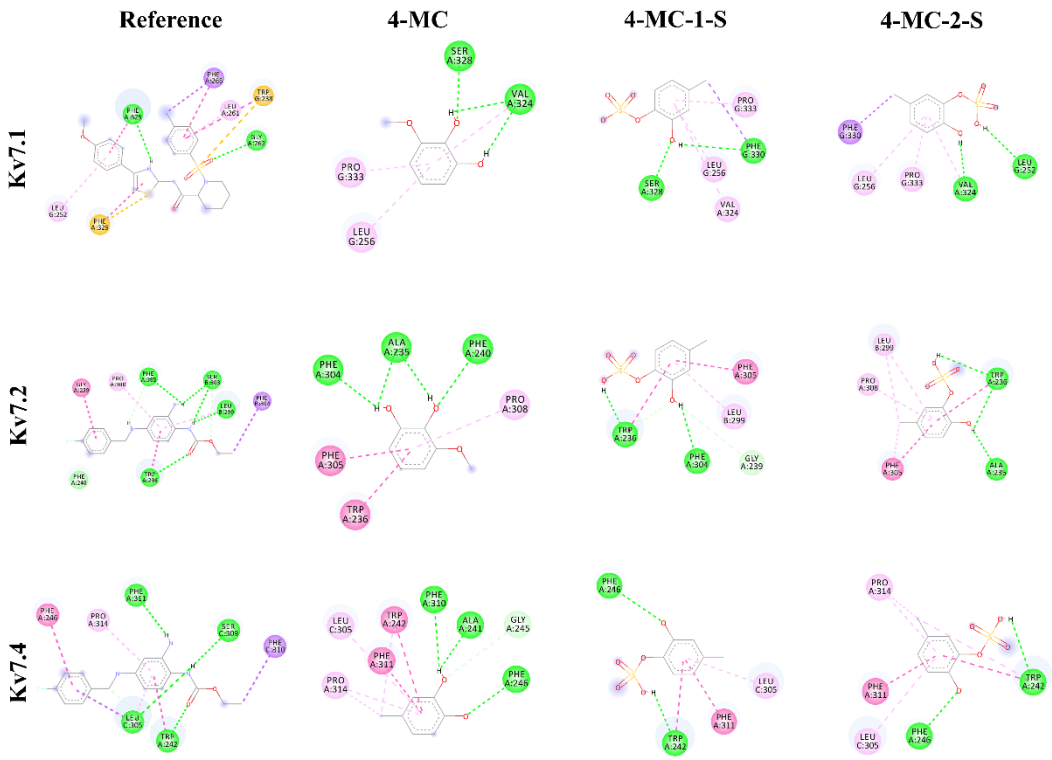

Figure 13. 2D representation of molecular interactions between the reference ligand (ML277 for K_v7.1 and retigabine for K_v7.2; 7.4), 4-methylcatechol (MC), 4-methylcatechol-1-sulfate (4-MC-1-S), and 4-methylcatechol-2-sulfate (4-MC-2-S). Dashed green lines show the hydrogen bond interactions; dashed purple lines indicate the hydrophobic interactions (taken from P2 under CC-BY 3.0) (122).

4.3. 3-Methoxycatechol Causes Vasodilation Likely via K_V Channels: Ex Vivo, In Silico Docking and In Vivo Study (P3)

Dias P, **Salam R**, Moravcová M, Saadat S, Pourová J, Vopršalová M, Jirkovský E, Duintjer Tebbens J, Mladěnka P. 3-methoxycatechol causes vasodilation likely via KV channels: ex vivo, in silico docking and in vivo study. Vascular Pharmacology. 2024 Sep 1;156:107418. https://doi.org/10.1016/j.vph.2024.107418

The objective of this study was to systematically evaluate the vasodilatory potential of 3-methoxycatechol (3-MOC) and structurally related catechol derivatives, identify the most potent compounds, and elucidate the molecular mechanisms underlying its vascular effects. By integrating ex vivo, in silico, and in vivo approaches, this study aimed to establish a

mechanistic basis for the vasorelaxation properties of 3-MOC and assess its potential as a novel antihypertensive agent.

Evaluation of 22 structurally related catechol derivatives for vasodilatory activity revealed that 3-methoxycatechol (3-MOC) was the most potent, displaying an EC_{50} of 9.64 µM in rat aortic rings. This was considerably lower than the EC_{50} values of other active catechols, including 4-ethylcatechol, 3,5-dichlorocatechol, and 4-*tert*-butylcatechol, which ranged from approximately 10 to 24 µM. Figure 14 demonstrates the involvement of different potassium (K^+) channels in mediating the vasodilatory effects of 3-MOC on rat aortic rings. Panels A–C show that the blockade of BK_{Ca} (with iberiotoxin), K_{ATP} (with glibenclamide), and K_{IR} (with Ba^{2+}) channels did not significantly affect the vasorelaxation induced by 3-MOC, indicating these channels are not primarily involved. In contrast, panel D shows that inhibition of voltage-gated K^+ (K_V) channels using 4-aminopyridine substantially reduced 3-MOC-induced relaxation, and panel E further reveals that selective inhibition of the K_V7 channel subtype with linopirdine also attenuated the effect.

These findings suggest that K_V channels, especially the K_V7 family, play a key role in the vasodilatory mechanism of 3-MOC. Furthermore, 3-MOC induced vasodilation in mesenteric resistance arteries but showed limited activity in porcine coronary arteries. These findings suggest that 3-MOC selectively modulates systemic vascular tone with minimal effect on coronary circulation. To further investigate the mechanism of action, molecular docking studies were performed targeting several vascular proteins (refer to Table 1 in section 4.2 for details), including sGC, PKGIα, PKGIβ, PKA, AC, SERCA, and K_V channels. These targets were strategically chosen to encompass both upstream signaling enzymes and direct ion channel effectors involved in smooth muscle relaxation, enabling a comprehensive evaluation of 3-MOC's vasodilatory mechanism through docking studies.

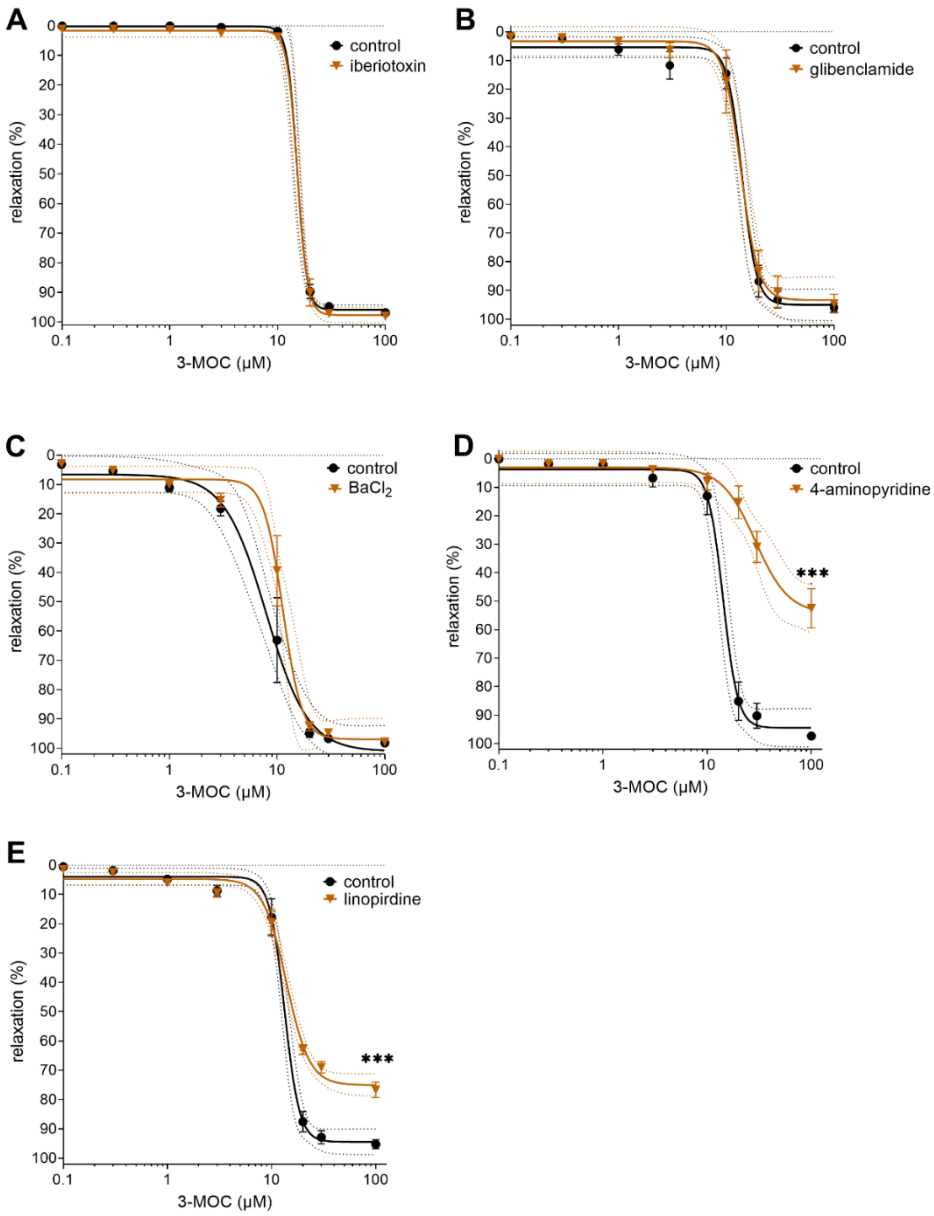

Figure 14. *Vasorelaxant effects of 3-methoxycatechol (3-MOC) on isolated rat aortic rings in the presence of selective K^+ channel inhibitors.* Concentration–response curves were generated for 3-MOC-induced relaxation in the absence (control, black) and presence (orange) of selective potassium channel blockers: (A) iberiotoxin (large conductance calcium-activated K^+/BK_{Ca} channel blocker), (B) glibenclamide (ATP-sensitive K^+/ K_{ATP} channel blocker), (C) $BaCl_2$ (inwardly-rectifier K^+/K_{IR} channel blocker), (D) 4-aminopyridine (voltage-gated K^+/K_v channel blocker), and (E) linopirdine (K_v7 channel blocker (taken from P3) (123).

3-MOC exhibited relatively weak docking affinities for sGC (–5.3 kcal/mol), AC (–5.5 kcal/mol), and PKG isoforms (–5.6 to –6.2 kcal/mol), suggesting a low probability of direct involvement of these pathways in the vasodilatory effect. Similarly, docking results for $K_v7.1$ and $K_v7.2$ did

not indicate strong binding interactions, further reducing the likelihood of these channels being primary targets. Although the binding affinity of 3-MOC to K_v7.4 (−5.4 kcal/mol) was lower than that of the known K_v7.4 agonist, retigabine (−7.6 kcal/mol), the docking pose of 3-MOC significantly overlapped with that of the reference ligand, suggesting potential functional relevance (Figure 15). The superimposition of the binding poses of 3-MOC with retigabine further suggested its potential as a K_v7.4 modulator. Interaction analysis confirmed that 3-MOC binds within the K_v7.4 channel binding pocket, forming key hydrogen bonds with residues Ala241, Phe246, and Phe310, and π-π interactions with Trp242 and Phe246. Notably, these residues are critical for K_v7.4 channel activation (124). To confirm the involvement of K_v7.4 in 3-MOC-induced vasodilation and explore the specificity of molecular interactions, the study employed docking experiments using both wild-type and site-directed mutated K_v7.4 channel models.

Three mutant variants of K_v7.4 were generated using UCSF Chimera's Rotamers tool based on Dunbrack's method: (1) Leu305 mutated to Trp, (2) Phe246 mutated to Ala, and (3) a double mutant combining both substitutions. These mutations targeted residues were shown to play crucial roles in ligand binding through hydrogen bonding and π-π stacking interactions in the wild-type structure (125). Docking analysis revealed that while 3-MOC retained its overall binding pose in the mutated channels, the interaction patterns were notably altered (Figure 16). Specifically, in the Leu305 mutated to Trp mutant, interactions with Phe246 and Trp305 were lost. In the Phe246 mutated to Ala, the interaction with the altered Ala246 residue was abolished, though interactions with Leu305 and Trp242 persisted. In the double mutant, 3-MOC was only able to maintain interactions with Ala246 and Trp242. These findings indicate that Phe246 and Leu305 are essential for stabilizing the ligand–channel interaction, and their alteration significantly disrupts binding affinity. This mutation-based docking approach provided mechanistic insights supporting the functional importance of K_v7.4 in mediating the vasodilatory effects of 3-MOC, aligning with pharmacological inhibition data.

Computational findings align with ex-vivo studies showing that K_v channel inhibitors, particularly the K_v7-selective inhibitor linopirdine, significantly reduced the vasodilatory effect of 3-MOC. Moreover, in-vivo experiments in spontaneously hypertensive rats confirmed that intravenous 3-MOC effectively lowered arterial blood pressure without altering heart rate. Taken together, these results identify K_v7.4 as a key molecular target of 3-MOC, suggesting its potential as a selective vasodilator for systemic vascular conditions.

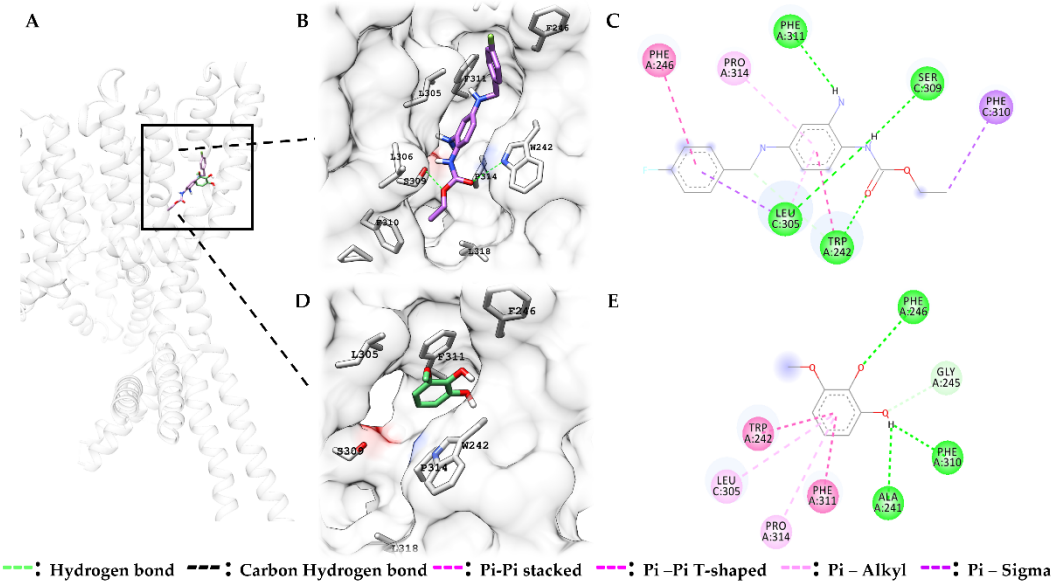

Figure 15. Reverse docking of 3-MOC with K_v7.4 channels. (A) The K_v7.4 channel activator, retigabine (purple) and 3-MOC (green) in the binding cavity of K_v7.4. (B) Detailed presentation of retigabine in the binding cavity of K_v7.4. (C) 2D molecular interaction between retigabine and K_v7.4. (D) Detailed presentation of 3-MOC in the binding cavity of K_v7.4. (E) 2D molecular interaction between 3-MOC and K_v7.4 (taken from P3) (123).

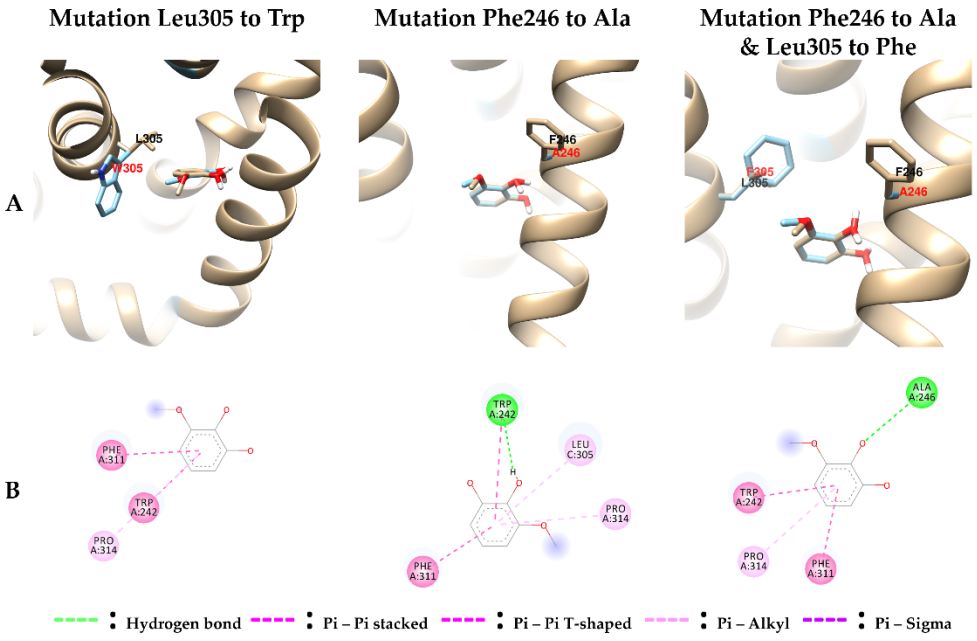

Figure 16. (A) The best pose of 3-MOC on the ligand binding domain of wild-type (brown) and mutated (blue) K_v7.4. (B). 2D molecular interaction between 3-MOC and mutated K_v7.4 (taken from P3) (123).

5 CONCLUSIONS

This dissertation demonstrated the strategic application to computational drug discovery methods, using methods such as molecular docking, pharmacophore modelling, and molecular dynamics simulations, to identify novel bioactive compounds and elucidate their potential mechanisms of action. The results of the three publications presented in this work lead to the following conclusions:

1. An integrated computational approach combining pharmacophore modelling, molecular docking and molecular dynamics simulations successfully identified a new TGR5 agonist, Hit-3 (CSC089939231). Pharmacophore screening of 50 million compounds yielded 41 candidates, with five candidates chosen based on an interaction pattern similar to that of the reference ligand (INT-777). Biological assays confirmed the potential of Hit-3 to activate TGR5 in a concentration-dependent manner, albeit with lower potency than lithocholic acid as control. MD simulations supported this activity, demonstrated that Hit-3 retained essential hydrogen bonds with Tyr240, a crucial residue for TGR5 activation, while exhibiting stable binding dynamics. ADMET predictions highlight the favorable drug-like properties of Hit-3, although solubility limitations and isomeric purity require optimization. This study positioned Hit-3 as a promising nonsteroidal lead for further development, emphasizing the utility of computational strategies in finding selective TGR5 agonists with potential therapeutic advantages over bile acid derivatives, such as reduced off-target effects and improved safety profiles.
2. The application of molecular docking allowed us to further strengthen evidence of the potential mechanism of action of substituted catechol compounds associated with vasodilatory effects. 3-methoxycatechol (3-MOC) and 4-methylcatechol (4-MC) emerge as potential vasodilators as shown in isolated rat aorta ex vivo. Both compounds were shown to effectively lower arterial blood pressure in hypertensive rats without altering heart rate and to enhance vasodilation through cAMP and cGMP pathways without influencing L-type Ca^{2+} channels. Instead, their mechanism involved activating voltage-gated K^+ (K_v) channels, particularly the $K_v7.4$ Molecular docking of both compounds suggested high probability of interaction with $K_v7.4$ residues Trp242 and Phe246, which are crucial for channel activation. These findings highlight the potential of 3-MOC and 4-MC as candidates for the development of novel antihypertensives targeting $K_v7.4$ channels.

6 PROSPECTS FOR FUTURE RESEARCH

Hit-3 (CSC089939231), identified as a promising nonsteroidal TGR5 agonist via virtual screening and molecular docking, was initially evaluated as a racemic mixture due to availability constraints. Docking studies indicated that only the (4aR,8aR) stereoisomer forms a key hydrogen bond with Tyr240, essential for TGR5 activation. This suggests that the observed moderate activity may result from dilution by inactive isomers. Therefore, future research should prioritize separation and synthesis of pure Hit-3 enantiomers to determine individual activity profiles and validate the (4a*R*,8a*R*) isomer as the active pharmacophore for SAR studies and lead optimization. Integrating enantioselective docking, molecular dynamics simulations, and chiral resolution techniques could enhance potency, reduce off-target effects, and accelerate Hit-3's preclinical development as a selective and safe TGR5 modulator.

The identification of 3-MOC and 4-MC as potential vasodilators targeting $K_v7.4$ channels provides a strong foundation for developing novel antihypertensive agents. Future research should focus on optimizing these catechol derivatives to improve their potency, selectivity, and pharmacokinetic properties through computational and medicinal chemistry approaches. SBDD strategies with pharmacophore and molecular docking approaches as a virtual screening of vast compound libraries, can identify potential new scaffolds that interact with the key residues Trp242 and Phe246, which are important for $K_v7.4$ activation to obtain compounds with improved $K_v7.4$ binding affinity. Currently, preliminary studies with pharmacophore and molecular docking have obtained 8 candidate compounds to proceed to the MD simulation step. MD simulation can provide deeper insights into ligand stability and binding dynamics, guiding rational modifications to improve efficacy. In addition to computational studies, experimental validation in ex vivo and in vivo models is essential. Functional testing using patch-clamp electrophysiology can confirm $K_v7.4$ activation directly, while advanced pharmacokinetic studies will assess bioavailability, metabolic stability and potential off-target effects. Furthermore, investigating the therapeutic effects of optimized $K_v7.4$ activators in models of hypertension and vascular dysfunction may pave the way for preclinical development.

7 RESEARCH OUTPUTS

7.1 Articles related to the topic of the dissertation

SALAM R, BAKKER M, KRUTÁKOVÁ M, ŠTEFELA A, PÁVEK P, DUINTJER TEBBENS J, ZITKO J.:
The discovery of a new non-bile acid modulator of Takeda G protein-coupled receptor 5: An integrated computational approach. Arch Pharm (Weinheim). 2025 Jan;358(1):e2400423. https://doi.org/10.1002/ardp.202400423, ISSN: 1521-4184, IF_{2023} 4.3, Q_{AIS} 2
Candidate's contribution:
First author, investigation & methodology, data analysis, interpretation of the results, visualization, writing the article, reviewing, and editing for submission.

DIAS P*, **SALAM R***, POUROVÁ J, VOPRŠALOVÁ M, KONEČNÝ L, JIRKOVSKÝ E, DUINTJER TEBBENS J, MLADĚNKA P.:
The quercetin metabolite 4-methylcatechol causes vasodilation via voltage-gated potassium (K_v) channels. Food & Function. 2024;15(22):11047-59.
https://doi.org/10.1039/D3FO04672A, ISSN: 2042-6496, IF_{2023} 5.1, Q_{AIS} 1
Candidate's contribution:
Shared First author, investigation & methodology, data analysis, interpretation of the results, visualization, writing the article, reviewing, and editing for submission.
** The authors contributed equally to this work*

DIAS P, **SALAM R**, MORAVCOVÁ M, SAADAT S, POUROVÁ J, VOPRŠALOVÁ M, JIRKOVSKÝ E, DUINTJER TEBBENS J, MLADĚNKA P.:
3-methoxycatechol causes vasodilation likely via K_v channels: ex vivo, in silico docking and in vivo study. Vascular Pharmacology. 2024 Sep 1;156:107418. https://doi.org/10.1016/j.vph.2024.107418, ISSN: 1537-1891, IF_{2023} 3.5, Q_{AIS} 2

Candidate's contribution:
Investigation & methodology, data analysis, visualization, writing the article

7.2 Oral presentations

SALAM R, DUINTJER TEBBENS J: Agonist and Antagonist Affect Conformational Changes in The TGR5 Ligand Binding Domain. 6th World Chemistry Conference and Exhibition. Barcelona, Spain, 11 – 12.09.2023

SALAM R, DIAS P, POUROVÁ J, VOPRŠALOVÁ M, KONEČNÝ L, JIRKOVSKÝ E, DUINTJER TEBBENS J, MLADĚNKA P: Insights into the vasodilatory mechanism of flavonoid metabolite 4-methylcatechol via Kv voltage channel by reverse molecular docking approach. International Graduate Student Conference on Pharmaceutical Science 2024. Surabaya, Indonesia, 10.08.2024

SALAM R, DUINTJER TEBBENS J, PAVEK P: Agonist and Antagonist Affect Conformational Changes in The TGR5 Binding Cavity; 13th Postgraduate and Postdoc Conference Hradec Králové, Czech Republic, 1. - 2. February 2023

SALAM R, DRASTIK M, BAKKER M, DUINTJER TEBBENS J, PAVEK P: An Integrated Computational Approach to The Discovery of a New TGR5 Agonist; 12th Postgraduate and Postdoc Conference Hradec Králové, Czech Republic, 1. - 2. February 2022

SALAM R, DRASTIK M, BAKKER M, DUINTJER TEBBENS J: Virtual Screening of TGR5 Small-Molecule: Agonist or Antagonist? 11th Postgraduate and Postdoc Conference, Hradec Králové, Czechia, 27–28 January 2021.

7.3 Scientific experience abroad

Traineeship on molecular dynamic (MD) simulation of complex biological systems (transmembrane proteins) at Molecular Dynamic Group of Faculty of Chemical Science, Universidad de Salamanca supervised by Prof. Pablo Garcia Jambrina for 3 months from September 1 - December 2, 2023.

8 REFERENCES:

1. Veselovsky A, Ivanov A. Strategy of computer-aided drug design. Current Drug Targets-Infectious Disorders. 2003;3(1):33-40.
2. Tiwari A, Singh S. Chapter 13 - Computational approaches in drug designing. In: Singh DB, Pathak RK, editors. Bioinformatics: Academic Press; 2022. p. 207-17.
3. Jorgensen WL. The many roles of computation in drug discovery. Science. 2004;303(5665):1813-8.
4. Gupta P, Chanana A, Kulkarni YR, Narayan A, Shanker O, Mahato B, et al. Computer-aided Drug Design: Innovation and its Application in Reshaping Modern Medicine. Current Artificial Intelligence. 2024.
5. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014;14(16):1923-38.
6. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR Modeling: Where Have You Been? Where Are You Going To? Journal of Medicinal Chemistry. 2014;57(12):4977-5010.
7. Pant S, Verma S, Pathak RK, Singh DB. Chapter 14 - Structure-based drug designing. In: Singh DB, Pathak RK, editors. Bioinformatics: Academic Press; 2022. p. 219-31.
8. Schneider G, Böhm H-J. Virtual screening and fast automated docking methods. Drug Discovery Today. 2002;7(1):64-70.
9. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. nature. 2021;596(7873):583-9.
10. Langer T. Pharmacophores in Drug Research. Molecular Informatics. 2010;29(6-7):470-5.
11. Leach AR, Gillet VJ, Lewis RA, Taylor R. Three-dimensional pharmacophore methods in drug discovery. Journal of medicinal chemistry. 2010;53(2):539-58.
12. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure and Applied Chemistry. 1998;70(5):1129-43.
13. Schaller D, Šribar D, Noonan T, Deng L, Nguyen TN, Pach S, et al. Next generation 3D pharmacophore modeling. WIREs Computational Molecular Science. 2020;10(4):e1468.
14. Stanzione F, Giangreco I, Cole JC. Chapter Four - Use of molecular docking computational tools in drug discovery. In: Witty DR, Cox B, editors. Progress in Medicinal Chemistry. 60: Elsevier; 2021. p. 273-343.
15. McConkey BJ, Sobolev V, Edelman M. The performance of current methods in ligand-protein docking. Current Science. 2002;83:845-56.
16. Xuan-Yu M, Hong-Xing Z, Mihaly M, Meng C. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. Current Computer-Aided Drug Design. 2011;7(2):146-57.
17. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery. 2004;3(11):935-49.
18. Van Drie JH. Computer-aided drug design: the next 20 years. Journal of computer-aided molecular design. 2007;21(10):591-601.
19. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. Nature. 2023;616(7958):673-85.
20. Sliwoski G, Kothiwale S, Meiler J, Lowe EW, Jr. Computational methods in drug discovery. Pharmacol Rev. 2014;66(1):334-95.
21. Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, et al. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. Journal of medicinal chemistry. 2002;45(11):2213-21.
22. Middleton D. Solid-state NMR spectroscopy as a tool for drug design: from membrane-embedded targets to amyloid fibrils. Biochemical Society Transactions. 2007;35(5):985-90.

23. Hosfield D, Palan J, Hilgers M, Scheibe D, McRee DE, Stevens RC. A fully integrated protein crystallization platform for small-molecule drug discovery. Journal of structural biology. 2003;142(1):207-17.
24. Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. Current opinion in structural biology. 2018;48:93-102.
25. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative protein structure modeling of genes and genomes. Annual review of biophysics and biomolecular structure. 2000;29(1):291-325.
26. Maia EHB, Assis LC, de Oliveira TA, da Silva AM, Taranto AG. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. Frontiers in Chemistry. 2020;8.
27. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015;20(7):13384-421.
28. Singh S, Chaturvedi N, Rai G. De novo modeling and structural characterization of IL9-IL9 receptor complex: a potential drug target for hematopoietic stem cell therapy. Network Modeling Analysis in Health Informatics and Bioinformatics. 2020;9:1-10.
29. Agoni C, Olotu FA, Ramharack P, Soliman ME. Druggability and drug-likeness concepts in drug design: are biomodelling and predictive tools having their say? Journal of molecular modeling. 2020;26:1-11.
30. Chen Y, Zhi D. Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins: Structure, Function, and Bioinformatics. 2001;43(2):217-26.
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic acids research. 2000;28(1):235-42.
32. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of computational chemistry. 1998;19(14):1639-62.
33. Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: current status and future challenges. Proteins: structure, function, and bioinformatics. 2006;65(1):15-26.
34. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. Journal of medicinal chemistry. 2006;49(2):534-53.
35. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010;31(2):455-61.
36. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of molecular biology. 1997;267(3):727-48.
37. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins: Structure, Function, and Bioinformatics. 2007;66(2):399-421.
38. Morris GM, Lim-Wilby M. Molecular docking. Molecular modeling of proteins. 2008:365-82.
39. Singh S, Bani Baker Q, Singh DB. Chapter 18 - Molecular docking and molecular dynamics simulation. In: Singh DB, Pathak RK, editors. Bioinformatics: Academic Press; 2022. p. 291-304.
40. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design. 2011;7(2):146-57.
41. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009;30(16):2785-91.
42. Dias R, Macedo Timmers LFS, Caceres RA, de Azevedo J, Walter F. Evaluation of molecular docking using polynomial empirical scoring functions. Current drug targets. 2008;9(12):1062-70.

43. Meiler J, Baker D. ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. Proteins: Structure, Function, and Bioinformatics. 2006;65(3):538-48.
44. Liu M, Wang S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. Journal of computer-aided molecular design. 1999;13:435-51.
45. Mcmartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. Journal of computer-aided molecular design. 1997;11:333-44.
46. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. The journal of chemical physics. 1953;21(6):1087-92.
47. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. 1970.
48. Hansmann UH. Parallel tempering algorithm for conformational studies of biological molecules. Chemical Physics Letters. 1997;281(1-3):140-50.
49. Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. science. 1983;220(4598):671-80.
50. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. The Journal of chemical physics. 2008;129(12).
51. Chipot C, Pohorille A. Free energy calculations: Springer; 2007.
52. Li Z, Gu J, Zhuang H, Kang L, Zhao X, Guo Q. Adaptive molecular docking method based on information entropy genetic algorithm. Applied Soft Computing. 2015;26:299-302.
53. Goldberg DE. Genetic Algorithm in Search, Optimization and Machine Learning, Addison. W esley Publishing Company, R eading, MA. 1989;1(98):9.
54. Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. Multimedia tools and applications. 2021;80:8091-126.
55. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. Journal of computational chemistry. 2007;28(6):1145-52.
56. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins: Structure, Function, and Bioinformatics. 1990;8(3):195-202.
57. Glover F. Tabu search—part I. ORSA Journal on computing. 1989;1(3):190-206.
58. Battiti R, Tecchiolli G. The reactive tabu search. ORSA journal on computing. 1994;6(2):126-40.
59. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins: Structure, Function, and Bioinformatics. 1998;33(3):367-82.
60. Fogel GB, Cheung M, Pittman E, Hecht D. Modeling the inhibition of quadruple mutant Plasmodium falciparum dihydrofolate reductase by pyrimethamine derivatives. Journal of Computer-Aided Molecular Design. 2008;22:29-38.
61. Westhead DR, Clark DE, Murray CW. A comparison of heuristic search algorithms for molecular docking. Journal of Computer-Aided Molecular Design. 1997;11:209-28.
62. Tripathi A, Bankaitis VA. Molecular Docking: From Lock and Key to Combination Lock. J Mol Med Clin Appl. 2017;2(1).
63. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. Journal of medicinal chemistry. 2006;49(23):6789-801.
64. Böhm HJ, Stahl M. The use of scoring functions in drug discovery applications. Reviews in computational chemistry. 2002;18:41-87.
65. Guedes IA, Pereira FS, Dardenne LE. Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. Frontiers in pharmacology. 2018;9:1089.
66. Schulz-Gasch T, Stahl M. Scoring functions for protein–ligand interactions: a critical perspective. Drug Discovery Today: Technologies. 2004;1(3):231-9.

67. Vanommeslaeghe K, Guvench O. Molecular mechanics. Current pharmaceutical design. 2014;20(20):3281-92.
68. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-based virtual screening for drug discovery: a problem-centric review. The AAPS journal. 2012;14:133-41.
69. Wang R, Lu Y, Fang X, Wang S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein− ligand complexes. Journal of chemical information and computer sciences. 2004;44(6):2114-25.
70. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. Proteins: Structure, Function, and Bioinformatics. 2003;52(4):609-23.
71. Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2015;5(6):405-24.
72. Muegge I, Martin YC. A general and fast scoring function for protein− ligand interactions: a simplified potential approach. Journal of medicinal chemistry. 1999;42(5):791-804.
73. Gohlke H, Klebe G. Statistical potentials and scoring functions applied to protein–ligand binding. Current opinion in structural biology. 2001;11(2):231-5.
74. Huang SY, Zou X. An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials. Journal of computational chemistry. 2006;27(15):1866-75.
75. Ballester PJ, Mitchell JB. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;26(9):1169-75.
76. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of computational chemistry. 1996;17(5-6):490-519.
77. Jorgensen WL, Tirado-Rives J. Monte Carlo vs molecular dynamics for conformational sampling. The Journal of Physical Chemistry. 1996;100(34):14508-13.
78. Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. Journal of computational chemistry. 2009;30(10):1545-614.
79. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. Journal of Chemical Theory and Computation. 2013;9(9):3878-88.
80. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. Journal of medicinal chemistry. 2006;49(20):5912-31.
81. Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv:151002855. 2015.
82. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. Journal of cheminformatics. 2021;13(1):43.
83. Liao Z, You R, Huang X, Yao X, Huang T, Zhu S, editors. DeepDock: Enhancing Ligand-protein Interaction Prediction by a Combination of Ligand and Structure Information. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2019 18-21 Nov. 2019.
84. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;99(6):1129-43.
85. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature structural biology. 2002;9(9):646-52.
86. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chemical reviews. 2006;106(5):1589-615.
87. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of computational chemistry. 2004;25(9):1157-74.

88. van Gunsteren WF, Daura X, Mark AE. GROMOS Force Field. Encyclopedia of Computational Chemistry1998.
89. Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. The Journal of Chemical Physics. 1954;22(8):1420-6.
90. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. Journal of the American Chemical Society. 2015;137(7):2695-703.
91. Likhachev IV, Balabaev NK, Galzitskaya OV. Available Instruments for Analyzing Molecular Dynamics Trajectories. Open Biochem J. 2016;10:1-11.
92. Fraccalvieri D, Pandini A, Stella F, Bonati L. Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. BMC bioinformatics. 2011;12:1-19.
93. Salam R, Bakker M, Krutáková M, Štefela A, Pávek P, Duintjer Tebbens J, et al. The discovery of a new nonbile acid modulator of Takeda G protein-coupled receptor 5: An integrated computational approach. Archiv der Pharmazie. 2025;358(1):e2400423.
94. Ivankov DN, Bogatyreva NS, Lobanov MY, Galzitskaya OV. Coupling between properties of the protein shape and the rate of protein folding. PloS one. 2009;4(8):e6476.
95. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science. 2011;334(6055):517-20.
96. Ippolito JA, Alexander RS, Christianson DW. Hydrogen bond stereochemistry in protein structure and function. Journal of molecular biology. 1990;215(3):457-71.
97. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of computational chemistry. 2010;31(4):671-90.
98. Ponder JW, Case DA. Force fields for protein simulations. Advances in protein chemistry. 2003;66:27-85.
99. Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications: Elsevier; 2023.
100. Im W, Beglov D, Roux B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer physics communications. 1998;111(1-3):59-75.
101. Allen MP, Tildesley DJ. Computer simulation of liquids: Oxford university press; 2017.
102. Guner O. History and evolution of the pharmacophore concept in computer-aided drug design. Current topics in medicinal chemistry. 2002;2(12):1321-32.
103. Chalk S. IUPAC, Compendium of Chemical Terminology (the “Gold Book”). Book IUPAC Compendium of Chemical Terminology—The Gold Book. 2019.
104. Tyagi R, Singh A, Chaudhary KK, Yadav MK. Chapter 17 - Pharmacophore modeling and its applications. In: Singh DB, Pathak RK, editors. Bioinformatics: Academic Press; 2022. p. 269-89.
105. Schwab CH. Conformations and 3D pharmacophore searching. Drug Discovery Today: Technologies. 2010;7(4):e245-e53.
106. Steindl TM, Schuster D, Laggner C, Langer T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. Journal of chemical information and modeling. 2006;46(5):2146-57.
107. Schuster D, Laggner C, Langer T. Why drugs fail-a study on side effects in new chemical entities. Current pharmaceutical design. 2005;11(27):3545-59.
108. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. Journal of chemical information and modeling. 2005;45(1):160-9.
109. Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug discovery today. 2010;15(11-12):444-50.

110. Akram M, Waratchareeyakul W, Haupenthal J, Hartmann RW, Schuster D. Pharmacophore modeling and in silico/in vitro screening for human cytochrome P450 11B1 and cytochrome P450 11B2 inhibitors. Frontiers in Chemistry. 2017;5:104.
111. John S, Thangapandian S, Arooj M, Hong JC, Kim KD, Lee KW, editors. Development, evaluation and application of 3D QSAR Pharmacophore model in the discovery of potential human renin inhibitors. BMC bioinformatics; 2011: Springer.
112. Truchon J-F, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. Journal of chemical information and modeling. 2007;47(2):488-508.
113. Murgueitio MS, Rakers C, Frank A, Wolber G. Balancing inflammation: Computational design of small-molecule toll-like receptor modulators. Trends in pharmacological sciences. 2017;38(2):155-68.
114. Di Leva FS, Di Marino D, Limongelli V. Structural insight into the binding mode of FXR and GPBAR1 modulators. Bile acids and their receptors. 2019:111-36.
115. Sepe V, Distrutti E, Limongelli V, Fiorucci S, Zampella A. Steroidal scaffolds as FXR and GPBAR1 ligands: from chemistry to therapeutical application. Future medicinal chemistry. 2015;7(9):1109-35.
116. Kumar DP, Asgharpour A, Mirshahi F, Park SH, Liu S, Imai Y, et al. Activation of transmembrane bile acid receptor TGR5 modulates pancreatic islet α cells to promote glucose homeostasis. Journal of Biological Chemistry. 2016;291(13):6626-40.
117. Yang F, Mao C, Guo L, Lin J, Ming Q, Xiao P, et al. Structural basis of GPBAR activation and bile acid recognition. Nature. 2020;587(7834):499-504.
118. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. Journal of computational chemistry. 2005;26(16):1701-18.
119. Huang J, MacKerell Jr AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. Journal of computational chemistry. 2013;34(25):2135-45.
120. Cheng RK, Segala E, Robertson N, Deflorian F, Doré AS, Errey JC, et al. Structures of human A1 and A2A adenosine receptors with xanthines reveal determinants of selectivity. Structure. 2017;25(8):1275-85. e4.
121. Latorraca NR, Venkatakrishnan A, Dror RO. GPCR dynamics: structures in motion. Chemical reviews. 2017;117(1):139-55.
122. Dias P, Salam R, Pourová J, Vopršalová M, Konečný L, Jirkovský E, et al. The quercetin metabolite 4-methylcatechol causes vasodilation via voltage-gated potassium (KV) channels. Food & Function. 2024;15(22):11047-59.
123. Dias P, Salam R, Moravcová M, Saadat S, Pourová J, Vopršalová M, et al. 3-methoxycatechol causes vasodilation likely via KV channels: ex vivo, in silico docking and in vivo study. Vascular Pharmacology. 2024;156:107418.
124. Schenzer A, Friedrich T, Pusch M, Saftig P, Jentsch TJ, Grötzinger J, et al. Molecular determinants of KCNQ (Kv7) K+ channel sensitivity to the anticonvulsant retigabine. Journal of Neuroscience. 2005;25(20):5051-60.
125. Li T, Wu K, Yue Z, Wang Y, Zhang F, Shen H. Structural basis for the modulation of human KCNQ4 by small-molecule drugs. Molecular Cell. 2021;81(1):25-37. e4.