Universidade do Minho

Escola de Ciências

Maria Francisca Araújo

In Silico Identification of Protein Targets Associated to the Insecticide Activity of Natural Products

In Silico Identification of Protein Targets Associated to the Insecticide Activity of Natural Products

Maria Francisca Araújo

UMinho |

Universidade do Minho

Escola de Ciências

Maria Francisca Araújo

In Silico Identification of Protein Targets Associated to the Insecticide Activity of Natural Products

Master Thesis

Master in Biophysics and Bionanosystems

Work developed under the
supervision of **Dr. Sérgio Sousa** and
Prof. Dr. Elisabete Coutinho

march 2023

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Atribuição-NãoComercial-SemDerivações
CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

“Para ser grande, sê inteiro: nada

Teu exagera ou exclui.

Sê todo em cada coisa. Põe quanto és

No mínimo que fazes.

Assim em cada lago a lua toda

Brilha, porque alta vive.”

- Ricardo Reis

Acknowledgments/Agradecimentos

Esta dissertação representa não só a conclusão do meu percurso na Academia, mas também um dos maiores desafios que superei até então.

Começo por expressar a minha sincera gratidão aos meus pais, Conceição e Carlos, às minhas irmãs, Inês e Madalena, e ao meu sobrinho Tomás, pelo amor, apoio e incentivo inabaláveis ao longo, não só desta etapa, mas de toda a minha vida. Dedico-vos este trabalho pois sei o quanto vos significa também.

Aos meus amigos, sem os quais nada disto seria possível: à Bruna e à Maria, que deram os primeiros passos na vida académica de mão dada comigo e que não a largaram até hoje. Ao Albano, à Madalena, ao Daniel, à Beatriz, e à Francisca por estarem sempre presentes, por me ouvirem, aconselharem e por me encorajarem a cada passo do caminho. Obrigada por tanto, e por tudo.

Ao Joel, por acreditar sempre em mim, principalmente (e com mais força) nos dias em que eu não fui capaz. Obrigada pelo companheirismo, pelo amor, pelo carinho e por nunca me teres deixado desistir.

A todos com quem cruzei caminho no BioSIM, em especial à Tatiana, à Rita e ao Rodrigo, agradeço o tempo, o cuidado e a amizade.

Ao meu orientador Doutor Sérgio Sousa, agradeço as oportunidades, a sabedoria, a partilha de conhecimento e também a paciência.

À Professora Doutora Elisabete Coutinho, agradeço a disponibilidade, generosidade e apoio que me tem vindo a dar desde a primeira vez que contactamos, ainda na Licenciatura.

Obrigada a todos, sou imensamente grata por tudo o que fazem por mim.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

In Silico Identification of Protein Targets Associated to the Insecticide Activity of Natural Products

Resumo

O crescimento rápido e exponencial da população mundial, bem como a necessidade de uma rede alimentar segura, conduziu inevitavelmente a um acréscimo da utilização de inseticidas, dado o papel crucial que estes desempenham na proteção das culturas agrícolas, protegendo-as contra os danos provocados por insetos e assim, garantindo uma produção segura de alimento. Apesar de todos os seus benefícios, os inseticidas convencionais estão associados a vários efeitos adversos, entre os quais a poluição ambiental, os danos provocados a espécies não-alvo e a resistência, que tem aumentado e afetado extensamente o controlo efetivo das populações de insetos.

A identificação de alvos proteicos associados à atividade inseticida de compostos naturais é determinante para o desenvolvimento de novos inseticidas viáveis e capazes de ultrapassar as problemáticas atuais. Nesta dissertação, desafiámo-nos a identificar tais alvos proteicos recorrendo a métodos in silico. Numa primeira seção, forneceu-se uma visão abrangente dos inseticidas - a sua importância e necessidade crescente, os inseticidas mais utilizados e os seus efeitos adversos, quais os principais alvos conhecidos para a atividade inseticida (moléculas que perturbam o sistema nervoso do inseto, reguladores de crescimento, endotoxinas, entre outros), a problemática da resistência a inseticidas e ainda algumas soluções alternativas para os inseticidas convencionais.

O estudo recorreu a métodos de inverted virtual screening para a identificação de possíveis alvos inseticidas para o eugenol e 11 dos seus derivados. Os resultados mostraram que estes compostos têm uma maior afinidade de ligação para as odorant binding proteins e acetilcolinesterases. Em paralelo, este estudo permitiu o desenvolvimento de uma base de dados de estruturas tridimensionais de alvos mais comuns para moléculas com atividade insecticida, que poderá vir a ser usada no futuro para a identificação e otimização de novas moléculas com atividade insecticida.

Palavras-Chave: inseticidas; controlo de pragas; sustentabilidade ambiental; carbamatos; organofosfatos; neonicotinóides; organoclorados; controle biológico; resistência; biopesticidas; molecular docking; virtual screening; inverted virtual screening.

In Silico Identification of Protein Targets Associated to the Insecticide Activity of Natural Products

Abstract

The increasing demand for food security and the rapidly growing population have led to an increased demand for insecticides. Insecticides play a crucial role in protecting crops from insect damage and ensuring food security. However, conventional insecticides have several adverse effects, including environmental pollution and harm to non-target species. Moreover, resistance to insecticides has become a major problem, making it difficult to control insect populations effectively. In this context, there is a need for new and innovative insecticides that are effective and environmentally friendly.

The identification of protein targets associated with the insecticide activity of natural compounds is crucial in the development of new insecticides. In this study, we aimed to identify such protein targets using in silico methods. The first section of this thesis provided a comprehensive overview of insecticides, including their significance and growing demand, the most commonly used insecticides, and their associated adverse effects. The section also delved into the major known targets for insecticidal activity, including molecules that disrupt the insect's nervous system, growth regulators, endotoxins, and others. The issue of insecticide resistance and alternative solutions for conventional insecticides was also highlighted.

The study leverages inverted virtual screening techniques to identify potential insecticidal targets for eugenol and 11 of its derivatives. The results show that eugenol derivatives have a higher binding affinity for odorant binding proteins and acetylcholinesterases. This project presents a straightforward approach for the application of in silico methods in identifying possible targets for new insecticides. In addition, this project enabled the development of a database of tridimensional structures of protein targets of molecules with insecticide activity that can be used in the future for the development or optimization of new insecticide compounds.

Keywords: insecticides; pest management; environmental sustainability; carbamates; organophosphates; neonicotinoids; organochlorines; biological control; resistance; biopesticides; molecular docking; virtual screening; inverted virtual screening.

Table of Contents

ACKNOWLEDGMENTS/AGRADECIMENTOS 6

INDEX OF FIGURES 13

INDEX OF EQUATIONS 14

INDEX OF TABLES 15

LIST OF ABBREVIATIONS 16

A. INTRODUCTION 19

1. INSECTICIDES: AN OVERVIEW 20

2. INSECTICIDES: IMPORTANCE AND INCREASING DEMAND 23

3. PROMINENT INSECTICIDES AND THEIR ADVERSE EFFECTS 24

4. MAJOR KNOWN TARGETS FOR INSECTICIDAL ACTIVITY 27

4.1. Molecules disrupting insect's nervous system 27

4.2. Metabolic targets 31

4.3. Growth regulators and others 32

5. THE PROBLEM OF RESISTANCE 33

6. ALTERNATIVES FOR CONVENTIONAL INSECTICIDES 37

7. CONCLUSIONS 40

B. COMPUTATIONAL METHODS 41

1. COMPUTER-AIDED DRUG DESIGN 42

2. PROTEIN-LIGAND INTERACTIONS 45

2.1. Physicochemical Mechanisms of Binding 45

2.2. Protein-Ligand Binding Models 49

3. MOLECULAR DOCKING 51

3.1. Definition 52

3.2. Search Algorithms 52

3.2.1. Rigid Body Search Algorithms 53

3.2.2. Flexible-Ligand Algorithms 53

3.3. Scoring Functions 57

3.3.1. Force-field or Physics-based Scoring Functions 58

3.3.2. Empirical Scoring Functions 58
3.3.3. Knowledge-based Scoring Functions 59
3.3.4. Consensus Scoring 59
3.4. Software Used 60
3.4.1. AutoDock Vina 60
3.4.2. GOLD 62
3.4.3. LeDock 67
4. VIRTUAL SCREENING 69
4.1. Introduction 69
4.2. Inverted Virtual Screening 70
C. RESULTS 72
1. LINE OF RESEARCH 73
2. THE STRUCTURAL DATABASE OF INSECTICIDE TARGETS 74
2.1. Scope 74
2.2. Methodology 75
2.2.1. Targeted physiology 77
2.2.2. Target 77
2.2.3. PDB code 77
2.2.4. Experimental Method 77
2.2.5. DOI 77
2.2.6. Release Date 77
2.2.7. Mutation(s) 77
2.2.9. Resolution 78
2.2.10. Source Organism 78
2.2.11. Macromolecular Name 78
2.2.12. Chain ID 78
2.2.13. EC Number 78
2.2.14. Ligand 78
2.2.15. Ligand Name 78
2.2.16. Ligand SMILES 78
2.3. Results – Current appearance of the Database 78
2.4. Conclusion and Future Works 81
3. IN SILICO IDENTIFICATION OF PROTEIN TARGETS ASSOCIATED TO THE INSECTICIDE ACTIVITY OF NATURAL PRODUCTS 81

3.1. CONTEXT 81
3.2. METHODOLOGY 82
3.3. RESULTS AND DISCUSSION 85

D. CONCLUSION AND FUTURE WORKS 90

E. REFERENCES 94

F. APPENDICES 107

Index of Figures

Figure 1. Schematic representation of the main targets and classes of insecticides that disrupt the insect nervous system. The diagram illustrates the common molecular targets, including (a) voltage-dependent sodium channels, (b) acetylcholinesterase, (c) GABA receptors, and (d) nicotinic acetylcholine receptors. The main classes of insecticides that act on these targets are also represented and include pyrethroids, DDT, organophosphates, carbamates, organochlorines, and neonicotinoids. These insecticides cause paralysis and the death of the insect by disrupting its nervous system.....30
Figure 2. Major known targets for insecticidal activity. 33
Figure 3. Traditional workflow of SBDD and LBDD [109].....44
Figure 4. Workflow for the CADD project performed in the present work. 45
Figure 5. Representation of the protein-ligand binding process. 46
Figure 6. Mechanism of protein-ligand binding reactions [113].....50
Figure 7. Systematic SA performs a systematic study of all available degrees of freedom. Each angle corresponds to an accepted or rejected value in the energy environment. 54
Figure 8. Scoring functions in docking and their respective equations.....60
Figure 9. A flowchart of a docking-based inverted virtual screening protocol. 71
Figure 10.Scheme of the metrics used during the development of the database.....75
Figure 11. Work in progress – protein targets grouped into broad categories based on the physiological functions that are affected. The chemical structure of the most used pesticides is also represented next to their main protein targets.....76
Figure 12. Work in progress – protein targets grouped into broad categories based on the physiological functions that are affected. The chemical structure of the most used pesticides is also represented next to their main protein targets.....80
Figure 13. Eugenol and derivatives used in this study. 84
Figure 14. Docking-predicted binding mode of EU3e to OBPs a) and docking-predicted binding modes of EU3e to Acetylcholinesterase b) with PLP scoring function. 88

Index of Equations

Equation 1. General equation for protein-ligand association reactions, where P represents the Protein, L represents the Ligand, and PL represents the complex formed between the two species. 46

Equation 2. Calculation of Gibbs Free Energy for the formation of the protein-ligand complex. 47

Equation 3. Mathematical relationship between the binding constant and the standard molar Gibbs binding energy for the protein-ligand complex. 47

Equation 4. Mathematical relationship between standard molar enthalpy (ΔH°), entropy and Gibbs energy of binding at constant temperature (T)..... 48

Equation 5. Number of total conformers resulting from the application of a fixed increment to all routable bonds, to generate all possible conformations..... 54

Equation 6. General Empirical based scoring function..... 58

Equation 7. General Knowledge based scoring. 59

Equation 8. AutoDock Vina scoring function. 61

Equation 9. Decomposition of the Autodock Vina scoring function. 61

Equation 10. Estimation of the standard molar Gibbs energy of association, for the lowest-scoring conformation obtained with the Autodock Vina search algorithm..... 62

Equation 11. Piecewise Linear Potential scoring function..... 63

Equation 12. ChemPLP scoring function..... 63

Equation 13. General decomposition of the CHEMSCORE scoring function..... 64

Equation 14. The ligand conformation rearrangement component of CHEMSCORE scoring function. 64

Equation 15. Decomposition of the protein-ligand component of CHEMSCORE scoring function..... 64

Equation 16. General equation for calculating the mixed entropic-energetic terms associated with the protein-ligand component of CHEMSCORE scoring function..... 65

Equation 17. Equation for calculating the rotamer term associated with the protein-ligand component of CHEMSCORE scoring function..... 65

Equation 18. Equation for calculating the clash energetic term associated with the protein-ligand component of CHEMSCORE scoring function. 66

Equation 19. Clash energy for an atomic pair, characterized by a distance $\mathbf{ri}$ and a clash distance $\mathbf{rclash}$. ..66

Equation 20. Equation for calculating the covalent term associated with the protein-ligand component of CHEMSCORE scoring function..... 67

Equation 21. General formulation for the GOLDSCORE scoring function. 67

Equation 22. LeDock scoring function..... 68

Equation 23. The Heaviside step function. 68

Equation 24.The Lennard-Jones ($\boldsymbol{EiLJ}$), hydrogen bond ($\boldsymbol{Eihb}$) and electrostatic ($\boldsymbol{Eiel}$) interactions of a ligand atoms i with the protein..... 69

Index of Tables

TABLE 1. STRUCTURAL AND CHEMICAL COMPOSITION, AS WELL AS CORRESPONDING ADVERSE EFFECTS, OF THE TEN CURRENTLY MOST USED INSECTICIDES IN WORLD. 25
TABLE 2. BIOCHEMICAL SITES OF ACTION OF NATURAL INSECTICIDES [105]. 39
TABLE 3. COMMON SBDD AND LBDD METHODS 43
TABLE 4. DOCKING PROGRAMS USING SYSTEMATIC AND STOCHASTIC METHODS. 56
TABLE 5. LIST OF TARGETS SELECTED FOR THE INVERTED VIRTUAL SCREENING STUDY. 82
TABLE 6. AVERAGE EUGENOL DERIVATIVE SCORES OBTAINED FOR ALL PDB STRUCTURES WITH THE SIX DIFFERENT SCORING FUNCTIONS. 85
TABLE 7. RANKING OF TARGETS OBTAINED WITH THE DIFFERENT DOCKING PROGRAMS/SCORING FUNCTIONS. 86

List of Abbreviations

ACh	Acetylcholine
AChE	Acetylcholinesterase
ACSII	American Standard Code for Information Interchange
ASP	Astex statistical potential
ATP	Adenosinetriphosphate
BFSG	Broyden-Fletcher-Goldfarb-Shanno
BHC	Benzene hexachloride
CADD	Computer-aided drug design
Chain ID	Chain identifier
CNS	Central nervous system
CPB	Colorado potato beetle
CUP	Central processing unit
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic acid
DOI	Digital Object Identifier.
EC number	Enzyme commission number
EFSA	European food safety authority
FSH	Follicle-stimulating hormone
GA	Genetic algorithms
GABA	ϒ-aminobutyric acid
GBSA	Generalized born surface area
GlmU	N-acetylglucosamine-1-phosphate uridyltransferase
GOLD	Genetic optimization for ligand docking
H	Enthalpy
HaSCP-2	Sterol carrier protein 2
Hb	Hemoglobin
IVS	Inverted virtual screening
L	Ligand

LBDD	Ligand-based drug design
LCT	λ-cyhalothrin
LH	Luteinizing hormone
MAL	Malathion
MD	Molecular Dynamics
MDA	Malondialdehyde
MGK-264	N-octylbicycloheptene dicarboximide
MlD	Molecular Docking
nAChR	Nicotinic acetylcholine receptor
NMR	Nuclear magnetic resonance
OBP	Odorant binding protein
P	Protein
PBO	Piperonyl butoxide
PBSA	Poisson Boltzmann surface area
PCV	Packed cell volume
PDB	Protein data bank
PL	Protein-ligand complex
PNS	Peripheral nervous system
PSMOs	P450-dependent polysubstrate monooxygenases
QSAR	Quantitative structure-activity relationship
RMSD	Root mean square deviation
S	Entropy
SA	Scoring Functions
SBDD	Structure-based drug design
SMILES	Simplified molecular-input line-entry system
T3	Triiodothyronine
TEC	Total erythrocyte count
TLC	Total leukocyte count
TMX	Thiamethoxam
TTX	Tetrodotoxin

VS	Virtual screening

A. INTRODUCTION

1. Insecticides: an overview

The definition of insecticide is any toxic substance that is used to eradicate and control insect populations (these include ovicides and larvicides for eggs and larvae, respectively). Such compounds are primarily used to control pests that infest cultivated plants, or to eliminate disease-carrying insects in specific areas. The earliest documented insecticide compounds were substances such as sulfur, heavy metals, salts, and even plant extracts (e.g., *Chrysanthemum cinerariifolium* formerly known as Dalmatian pyrethrum) [1–4]. The use of elemental and/or natural compounds for pest control started at the very dawn of agriculture and has continued, in some cases, to be used to the present day. The first record of insecticide usage dates ≈ 4500 years ago by Sumerian people, who used sulfur compounds in order to kill insects and mites. Additionally, ≈3200 years ago, the Chinese were using mercury and arsenical compounds to control body lice [5]. Botanical preparations are also amongst the first recorded pest controllers. For instance, the discovery of *C. cinerariifolium* insecticidal activity may have been accidental. A book about these common flowers tells us the story of a German woman of Dubrovnik who picked the flowers for their beauty, and after they withered, she noticed that dead insects had gathered around the plant's remnants, suggesting a possible connection between C. cinerariifolium and its ability to kill insects [4]. These flowers, formerly classified as pyrethrum flowers, contain up to 1.5% of a substance named pyrethrin, which is an active insecticidal compound [3]. This ingredient was used as an insecticide in ancient China and in the Middle Ages in Persia, and it was brought to Europe shortly after by Armenian traders, being sold as "Persian Dust" (around ≈200 years ago). This powder was produced from dried flowers of *Chrysanthemum roseum*, and the major constituents of these dried extracts were pyrethrin I and II, which compose some of today's household sprays [6].

In the 19th century, a vast variety of chemicals started to be used against crops' infestations. A farmer discovered that Paris green, a paint pigment (copper acetoarsenite), had supposedly insecticidal properties when discarding remaining paint

onto a potato plantation that was infested with the CPB (Colorado potato beetle) [7]. This substance was widely used in many countries of the world until the mid-20th century. In order to control the malaria vector, Paris green would be sprayed on the surface of breeding places, working as a larvicide [8]. Around the same time period, borax was also reported as an insecticide when used as a coating material for crop seeds such as corn [9].

During the late 1800s and early 1900s, scientists developed the first synthetic organic chemicals that served as insecticides. These modern synthetic insecticides were made in the form of organochloride compounds. Although benzene hexachloride (BHC) and dichlorodiphenyltrichloroethane (DDT) were synthesized in the 1800s, it was not until later that their insecticidal properties were fully discovered and utilized [10]. Michael Faraday, an English scientist, first produced BHC in 1825, while Othmar Ziedler, an Austrian chemist, synthesized DDT in the same year. However, it was not until Bender and Müller, respectively, in 1933 and 1939, that the insecticidal properties of BHC and DDT were first demonstrated [11]. This was probably the most significant development in the history of pest control and resulted in Müller being awarded the Nobel Prize in 1948 [12]. This chemical agent was designed to eliminate insects, weed, rodents, fungi, and other human annoyance trouble, but its adverse effects spread to every ecosystem it came into contact with. In fact, it still impacts the environment and human health to the present day, due to its long residual efficacy and accumulation throughout the food chain [13]. A milestone in environmental science is the publication of the book Silent Spring by Rachel Carson [14] that exposes the effects of the indiscriminate use of pesticides such as DDT; this book was considered one of the greatest science books of all time. Regarding human health, DDT is the cause of various ailments, including various types of cancer, acute and persistent injuries to the nervous system, lung damage, injury to reproductive organs, and dysfunction of the immune and endocrine systems, and it has also been linked to numerous birth defects [13]. DDT quickly lost its popularity as the USA, Japan, and Western Europe banned the production and application of the substance (in the Stockholm Convention, 2001), classifying it as a priority pollutant. However, it is continuing to be illegally used in third-world countries [15].

Shortly thereafter, at the beginning of the 20th century, researchers began exploring modifications to natural pyrethrins' structure. In 1949, Schechter and LaForge discovered allethrin, the first pyrethroid compound, which improved the effectiveness of insecticides over time [16]. These compounds were divided into two types, Type I and Type II, based on their chemical structure. The discovery of allethrin, which belongs to the Type I pyrethroid compound group, renewed interest in pyrethrins as insecticides [17]. It also inspired chemists worldwide to investigate modifications to the pyrethroid alcohol and acid moieties, and eventually to the essential ester function [16]. These derivatives proved to be significantly more effective, cost-effective, and stable than their natural pyrethrin counterparts [18]. Despite the fact these synthetic compounds lose their activity rather quickly when exposed to ultraviolet light, this photodegradation property of pyrethroids [19] helped to prevent their accumulation in the environment, and, therefore, this class of insecticides still finds wide application in plant protection.

In more recent years, we have seen the appearance of new insecticides, such as neonicotinoids, a class of neuro-active insecticides that are chemically similar to nicotine; they act by systematically moving in the plant tissues and protecting all parts of the plant. Reportedly, their discovery is connected to the Shell and Bayer companies, which started their development in the 1980s and 1990s, respectively [20]. However, imidacloprid was the very first neonicotinoid that appeared on the insecticide market. It was registered as "Hachikusan" in Japan in 1993 [3]. Nowadays, it is possible to find a large number of represented neonicotinoids, such as acetamiprid, clothianidin, dinotefuran, imidacloprid, nitenpyram, nithiazine, thiacloprid, and thiamethoxam [21]. They quickly gained popularity, and neonicotinoids such as imidacloprid have been the most widely used insecticides in the world, from 1999 to at least 2018 [22]. In 2016, imidacloprid was banned alongside clothianidin, thiamethoxam, acetamiprid, and thiacloprid by the French government, and the EFSA (The European Food Safety Authority) concluded, in February 2018, that the most used neonicotinoid insecticides represent a risk to wild bees and honeybees [23].

Currently, besides neonicotinoids (especially imidacloprid), the other two most used insecticides are organophosphates (more specifically chlorpyrifos) and carbamates (more specifically carbaryl). Organophosphate insecticides correspond to roughly half of all insecticides used worldwide, with chlorpyrifos being the most widely used (approved to be used on more than 50 different crops) [24]. Regarding carbamate insecticides, there are ≈50 chemicals that belong to this family and are used as fungicides, herbicides, and nematicides, in addition to being used as insecticides [25,26]. Carbaryl, a white crystalline solid, was the first carbamate to be commercialized, and to this day, it is more widely used than all the other carbamates combined [24].

2. Insecticides: importance and increasing demand

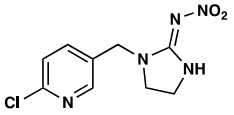
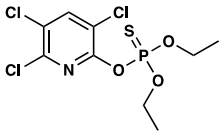
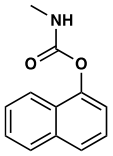
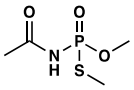
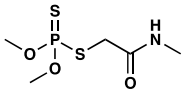
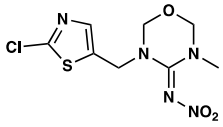
From ancient times to the present day, the use of pesticides such as insecticides has become an essential and strictly necessary agricultural component in order to assure crop yields and minimize post-harvest losses [27]. With a continuously increasing population, in addition to deteriorating environmental conditions (based on irrefutable and growing evidence of climate change coupled to increasing levels of pollution), the task of achieving long-term development without causing environmental harm has never been greater. Feeding a population of more than 9 billion people by 2050 will require food production to grow by 70 to 100%, making agriculture one of the most pressing challenges for sustainable development [28]. Due to its significant role in deforestation [29] and widespread land usage that includes 70% of grasslands, 50% of savannas, and 45% of temperate forests [28], it's becoming increasingly important to restrict crop cultivation areas while simultaneously boosting productivity to meet growing demand. Additionally, the changing of dietary habits of expanding middle classes has driven the need for higher quality products through the control of various insect pests [30]. Furthermore, insects are often hosts of devastating diseases. Vector-borne diseases are among the major causes of illness and death worldwide, particularly in tropical and subtropical regions; therefore, vector control, through the use of insecticides, is highly important for the prevention and control of infectious diseases

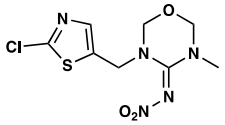
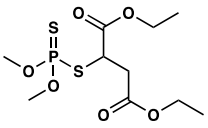
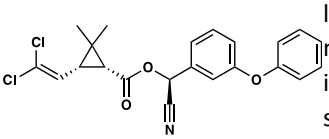
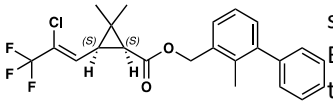
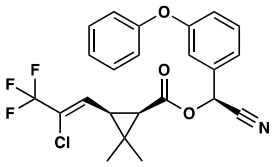
such as malaria, dengue, and filariasis [31]. Available approaches to control pest insects range from (natural or chemical) insecticide usage to cultural practices (e.g., crop rotation), genetically modified plants (e.g., increasing host plant resistance), biological control (e.g., the release of sterilized pests to disrupt reproduction), physical and mechanical control, and microbial control [32]. Sometimes, multiple approaches are needed in order to address certain infestation problems; however, for many pest control complications, insecticides have and continue to provide farmers and public health workers with the tools and means to predictably, quickly, and effectively address a specific pest problem [30]. Insecticides are often an easy and reliable solution, which results in an increasing demand for the compounds. Nonetheless, their toxicology should be thoroughly studied before being applied in order to prevent any more environmental residual and prolonged damage.

3. Prominent insecticides and their adverse effects

Environmental contamination is the main problem associated with these poisonous compounds, and they may be harmful to other organisms, including humans, rather than just exclusively killing insects. Many insecticides are short-lived and decompose quickly or are fully metabolized by the animals that ingest them, but some are persistent and, when administrated in higher quantities, could be devastating for ecosystems, as they travel across the food chain. When insecticides are applied to crops, much of it reaches the soil and consequently contaminates groundwater reserves from direct application or, in worst case scenarios, as runoff from treated areas. Furthermore, when poorly used, insecticides could create some levels of resistance amongst an insect population. They could also eliminate the natural predators that once held them back. The nonspecific nature of the currently used broad spectrum of chemicals makes them more likely to have such unintended effects on the abundance of both harmful and beneficial insects. In the following table (Table 1), the top ten most used insecticides in the world at the moment are listed. This table includes their structural and chemical description. Additionally, their adverse effects are also indicated.

Table 1. Structural and chemical composition, as well as corresponding adverse effects, of the ten currently most used insecticides in world.

[No]Insecticide	Chemical Formula	Chemical Structure	Adverse Effects
[1] **Imidacloprid (neonicotinoid)**	$C_9H_{10}ClN_5O_2$		The residues of this substance can make their way int the food chain and affect both the reproductiv capacity of lab rats and that of their offspring. It is chemical that disrupts endocrine and steroidogenes [33].
[2] **Chlorpyrifos (organophosphate)**	$C_9H_{11}Cl_3NO_3PS$		The laboratory rats that were exposed showed decrease in body weight and an increase in the relativ weights of their liver and kidney. The damage to the liver was significant, and there was a notable increas in total protein and uric acid levels. Additionally, ther was an increase in oxidative stress observed in th exposed rats [34,35].
[3] **Carbaryl (carbamate)**	$C_{12}H_{11}NO_2$		The toxicity observed is a result of cholinesteras inhibition. When pigs were exposed to this substanc for an extended period, it caused a progressiv neuromyopathy that resulted in structural damag which cannot be reversed acutely with atropin Similarly, in lab rats, there was a significant reductio in their overall weight. Moreover, there was a notab decrease in the number of germ cells, spermatocyte spermatids, and Leydig cells. Additionally, th testosterone levels significantly declined, while th levels of LH and FSH increased significantly [36,37].
[4] **Acephate (Organophosphate)**	$C_4H_{10}NO_3PS$		The highest doses administered to lab rats inhibite the activity of acetylcholinesterase in the brain an skeletal muscles. In the same group, there was decrease in the number of implantations and liv fetuses, along with an increase in the number of ear resorptions observed. Furthermore, there was decrease in sperm motility and count in the expose rats. Dose-dependent histologic changes, includir the degeneration of muscle fibers, were also observe [38].
[5] **Dimethoate (Organophosphate)**	$C_5H_{12}NO_3PS_2$		This substance, like other organophosphates, is know to inhibit acetylcholinesterase (AChE) activity, leadir to severe nerve damage. In plants, its effects ar reflected in reduced photosynthesis and growth, whi in birds, the activity of brain enzymes is inhibited resulting in sublethal effects. Aquatic organisms ar expected to be highly affected by direct exposur leading to changes in their swimming behavior [39].
[6] **Thiamethoxam (Neonicotinoid)**	$C_8H_{10}ClN_5O_3S$		In cockerels, exposure to thiamethoxam (TMX) at sul lethal levels resulted in a dose-dependent reduction i key hematological parameters, including tot

		(E isomer)  (Z isomer)	erythrocyte count, hemoglobin, packed cell volum and total leukocyte count. The biochemistry of th birds was also impacted, with significant alterations i total proteins, albumin, and globulin. The stud indicated that TMX caused substantial changes in th hematological profile and liver and kidney function o the birds. In addition, TMX increased oxidativ damage to lipids and DNA in these organs, whil reducing the antioxidant activities in liver and kidne cells, leading to oxidative stress [40,41].
[7] **Malathion (Organophosphate)**	$C_{10}H_{19}O_6PS_2$		Malathion (MAL) was found to have adverse effects o frog oocyte maturation, resulting in reduced levels o Emi2, a critical factor for oocyte maturation. I addition, embryos fertilized under the influence o MAL showed a higher rate of abnormal divisio leading to embryo death during early embryogenesi The toxicity mechanisms of MAL include inhibition o acetylcholinesterase, oxidative stress, DNA damag and apoptotic cell damage. Its toxic effects on th central nervous system are well documented, but also affects the liver, kidney, testis, ovaries, lun pancreas, and blood. MAL is considered a genotox and carcinogenic chemical compound and evidenc shows adverse effects associated with prenatal, an postnatal exposure in both animals and human These findings are supported by various studie [42,43].
[8] **Zeta-cypermethrin (Pyrethroid)**	$C_{22}H_{19}Cl_2NO_3$		In common guppies (*Leporinus reticulatus*), exposu to various doses of zeta-cypermethrin resulted in th lifting of the epithelial layer from gill lamellae an necrosis. Other observed histopathological effec included exudation, hyperplasia, and shortening o secondary lamellae. Additionally, in vitro experimen showed that zeta-cypermethrin caused DNA damag in human peripheral lymphocytes, indicating i genotoxic properties [44,45].
[9] **Bifenthrin (Pyrethroid)**	$C_{23}H_{22}ClF_3O_2$		The hepatic function of tadpoles is negatively affecte by cis-bifenthrin. Aquatic species are high susceptible to the acute lethal toxicity of bifenthri Bifenthrin also has sublethal toxic effects on no target organisms, such as developmental toxicit neurobehavioral toxicity, oxidative damage, immun toxicity, and endocrine-disrupting effects [46,47].
[10] **λ-cyhalothrin (Pyrethroid)**	$C_{23}H_{19}ClF_3NO_3$		Previously conducted research has indicated tha synthetic pyrethroids, such as λ-cyhalothrin (LCT have high levels of aquatic toxicity. Exposure o zebrafish to synthetic pyrethroids, including LC resulted in a dose-dependent increase in mortalit higher malformation rates, and lower hatching rate This exposure to LCT led to a significant decrease i thyroid hormone triiodothyronine (T3) level

	indicating potential developmental toxicity b disrupting endocrine signaling at concentration present in the environment. In other studie administration of LCT to laboratory rats led t decreased functional sperm parameters, enzymat and non-enzymatic antioxidant levels, and th presence of irregular seminiferous tubules containir only Sertoli cells [48,49].

4. Major known targets for insecticidal activity

To fully comprehend how an insecticide works, it is necessary to have knowledge about its particular target(s) within an organism. Typically, this is a crucial protein or enzyme. Consequently, insecticides are usually classified based on their structure and mode of action. Most insecticides act on (1) the insect's nervous systems, (2) metabolic targets, and (3) growth regulators and others.

4.1. Molecules disrupting insect's nervous system

The primary target observed for most insecticides is the peripheral nervous system (PNS) and central nervous system (CNS).

Organochlorines are a type of insecticide that is made up of organic compounds that contain one or more covalently bonded chlorine atoms. The most well-known type of organochlorine is chlorinated hydrocarbons, which include DDT, chlordane, lindane, and endosulfan [50]. Through an imbalance of sodium and potassium ions, these insecticides disrupt nerve impulse transmission. Furthermore, some organochlorines act on GABA receptors, preventing ions from entering neurons and resulting in a hyperexcitable state characterized by tremors and convulsions [51]. Other insecticides that target GABA receptors include antibiotic insecticides and pyrethroids.

Antibiotic insecticides, also known as microbial insecticides, are derived from bacteria or fungi and are effective against tough greenhouse pests such as spider mites and leaf miners. These insecticides block neurotransmitters at the neuromuscular junction,

impairing the insects' ability to feed and lay eggs and ultimately leading to their death [52–54]. Spinosyns are a very special type of microbial insecticides with a complex structure that includes a large macrocyclic lactone ring, a tetra-hydrogen ring, and a dihydropyranone group, as well as oxygen and nitrogen atoms and a sugar moiety [55]. These compounds are extremely specific and can effectively target a wide variety of pests, including caterpillars, lepidopteran larvae, leaf miners, thrips, and termites [56].

On the other hand, pyrethroids (Table 1, molecules 8–10) are made up of a cyclopropane or cyclohexane ring with a carboxylic acid group attached to it and two aryl or heteroaryl groups that may contain halogen or other substituents [57]. This structure enables the molecule to bind to the sodium channel in the nerve cell membrane of insects, causing a sodium/potassium imbalance that causes the insect to become hyperexcitable. Tremors, incoordination, hyperactivity, and, finally, paralysis are symptoms [58]. Pyrethroids are typically classified into two types: type I and type II. Type I pyrethroids have an alfa-cyano group attached to the cyclopropane or cyclohexane ring, whereas type II pyrethroids have a beta-cyano group attached to the same ring, making them more potent and persisting in the environment for longer [57]. They, like others, are extremely toxic to fish, in addition to being effective against most agricultural pests [58].

Both neonicotinoids and formamidines are new classes of insecticides that are applied at low dosages and are extremely effective.

Neonicotinoids (Table 1, molecules 1 and 6), named for their chemical similarity to nicotine, consist of a heterocyclic ring structure to which a nitro group and a cyan group are attached. Neonicotinoids act as an activator of nicotinic acetylcholine receptors (nAChRs) in the nervous system of insects, leading to overstimulation and paralysis. Their high affinity for insect nAChRs and longer half-life compared to nicotine make them highly effective insecticides. One of the most significant advantages of neonicotinoids is their high selectivity in toxicity, meaning that they have a minimal impact on non-target organisms such as birds and mammals. However, their use has been the subject of controversy due to their potential impact on pollinators such as bees [59].

Formamidines are a type of pesticide that works by blocking the monoamine oxidase enzyme, which is responsible for the breakdown of neurotransmitters in insects, and formamidines create a buildup of these molecules by blocking it. This causes infected insects to become dormant, eventually leading to death. Some of the most often used formamidine compounds in pesticides are chlordimeform, amitraz, and thiacloprid. Formamidines have various advantages over conventional insecticides, including low mammalian toxicity and the fact that they do not last long in the environment. These chemicals are commonly used to manage pests that have evolved resistance to other pesticide classes, such as organophosphates and carbamates [60].

Organophosphates (Table 1, molecules 2, 4, 5, and 7) and carbamates (Table 1, molecule 3) both block AChE and produce acetylcholine buildup at NMJs, resulting in the fast twitching of voluntary muscles and paralysis [61]. They share a similar chemical structure, but with some key differences—both contain a central phosphorus or carbamate functional group, respectively, that is important for their insecticidal properties.

Organophosphates contain a phosphorus atom that is typically bonded to two oxygen atoms (forming a carbonyl group). This carbon-oxygen double bond is essential for organophosphates' ability to inhibit AChE activity. Malathion, chlorpyrifos, and parathion are examples of organophosphate insecticides [61].

Carbamates, on the other hand, have a carbamate functional group consisting of a nitrogen atom bonded to both a carbonyl group and an oxygen atom. Like organophosphates, carbamates inhibit the activity of AChE but in a slightly different way—instead of forming a covalent bond with the enzyme like organophosphates do, carbamates form a reversible bond with AChE. Examples of carbamates include carbaryl, methyl and aldicarb [62].

Botanical preparations remain a popular choice as insecticides, with some even serving as the basis for synthetic insecticides such as pyrethroids (derived from pyrethrum) and

neonicotinoids (inspired by the use of tobacco for crop protection). One such compound is limonene, a terpene found in the essential oils of citrus fruits, rosemary, and peppermint. Limonene targets the sensory nerves of the peripheral nervous system and can effectively control fleas, lice, mites, and ticks. Furthermore, it has low toxicity to warm-blooded animals and only minor toxicity to fish. However, botanical preparations may require more frequent application than synthetic insecticides. Despite this, they can be a viable option for integrated pest management programs that prioritize non-chemical control methods. Figure 1 illustrates the main targets and classes of insecticides that disrupt the insect nervous system.

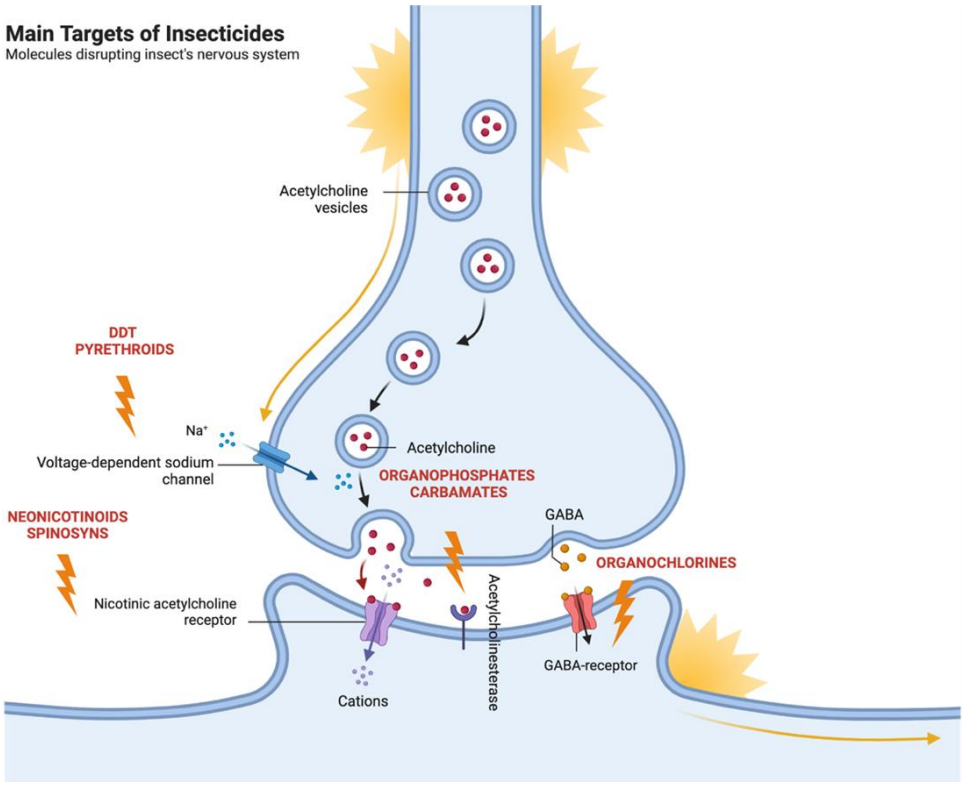


Figure 1. Schematic representation of the main targets and classes of insecticides that disrupt the insect nervous system. The diagram illustrates the common molecular targets, including (a) voltage-dependent sodium channels, (b) acetylcholinesterase, (c) GABA receptors, and (d) nicotinic acetylcholine receptors. The main classes of insecticides that act on these targets are also represented and include pyrethroids, DDT, organophosphates, carbamates, organochlorines, and neonicotinoids. These insecticides cause paralysis and the death of the insect by disrupting its nervous system.

4.2. Metabolic targets

Other substances used as insecticides work as endotoxins and as highly toxic molecules; they interfere with the normal function of insect's metabolisms. For instance, organosulfur compounds act as ovicides, eliminating the pest in the egg stage. They usually carry low toxicity to other organisms [63].

Dinitrophenols act by uncoupling or inhibiting oxidative phosphorylation, preventing the creation of the essential adenosine triphosphate (ATP) [64].

Organotins work similarly to dinitrophenol, attacking and inhibiting the same binding sites, preventing ATP formation. They are extensively used against mites on fruit trees, and they were formerly used as an antifouling agent and molluscicide, being highly toxic to aquatic life [65].

Pyrazoles act by inhibiting the NADH-CoQ reductase site of mitochondrial electron transport, which disrupts ATP formation [66].

Pyridazinones interrupt mitochondrial electron transport at site one and are mainly used as a miticide. However, like most others, they showcase toxicity to aquatic arthropods and fish [67,68].

Botanical preparations can also work as endotoxins and, depending on the type, can have various effects. Rotenone is a naturally occurring compound found in the roots, stems, and leaves of certain plants, including the Derris and Lonchocarpus species; acts as a respiratory enzyme inhibitor, and is used as a piscicide that kills fish at doses that are non-toxic to fish food organisms [69]. Neem is a tree that is native to India. The leaves, bark and seeds of the neem tree contain compounds with medicinal properties that are used as insecticides since they reduce feeding and disrupt moulting by inhibiting biosynthesis or metabolism of ecdysone, the moulting hormone [70]. It is commonly used against moth and butterfly larvae.

Fumigants act by releasing gas into the air, which penetrates the treated space and targets the metabolism of pests. These agrochemicals' effect on metabolism can vary depending on the specific chemical and the target organism. However, in general, fumigants disrupt key metabolic pathways in target organisms, such as energy production, DNA synthesis, and protein synthesis, leading to cell death. Additionally, some fumigants can react with cellular components, such as proteins and DNA, causing structural damage that impairs metabolic function. Methyl bromide, phosphine, sulfuryl fluoride, and ethylene oxide constitute some commonly used fumigants [71].

Inorganic compounds can also be used as insecticides. Their mode of action is dependent on the type of inorganic compound. Typical examples include uncoupling oxidative phosphorylation (arsenicals), inhibition of enzymes involved in energy production, and acting as desiccants. For each pest group, there is a different compound to be applied according to its efficacy; for example, for mites, sulfur should be used, and for cockroaches, boric acid [72].

4.3. Growth regulators and others

Biochemicals, which are classified as biorational compounds, have low toxicity to non-targeted species [73] and consist of various substances such as hormones; enzymes; pheromones; growth regulators; and microbials such as viruses, bacteria, fungi, protozoa, and nematodes. They act as either attractants, growth regulators, or endotoxins and can also function as attractants to specific species [74,75]. As an example, benzoylureas act as insect growth regulators by interfering with chitin synthesis [76].

Quinazolines have been shown to affect the larval stages of many insects by hindering the production of chitin in the exoskeleton, which can lead to the breaking of the cuticle or death due to lack of food in the affected larvae [77,78].

There is also another set of compounds classified as synergists/activators, which inhibit cytochrome P450-dependent polysubstrate monooxygenases (PSMOs), preventing the degradation of toxicants and enhancing the activity of insecticides when used alongside them; synergists and activators are not themselves considered toxic or insecticidal. Examples include piperonyl butoxide (PBO) and N-octyl bicy-cloheptene dicarboximide (MGK-264) [79–81].

Figure 2 shows the major known targets for insecticidal activity.

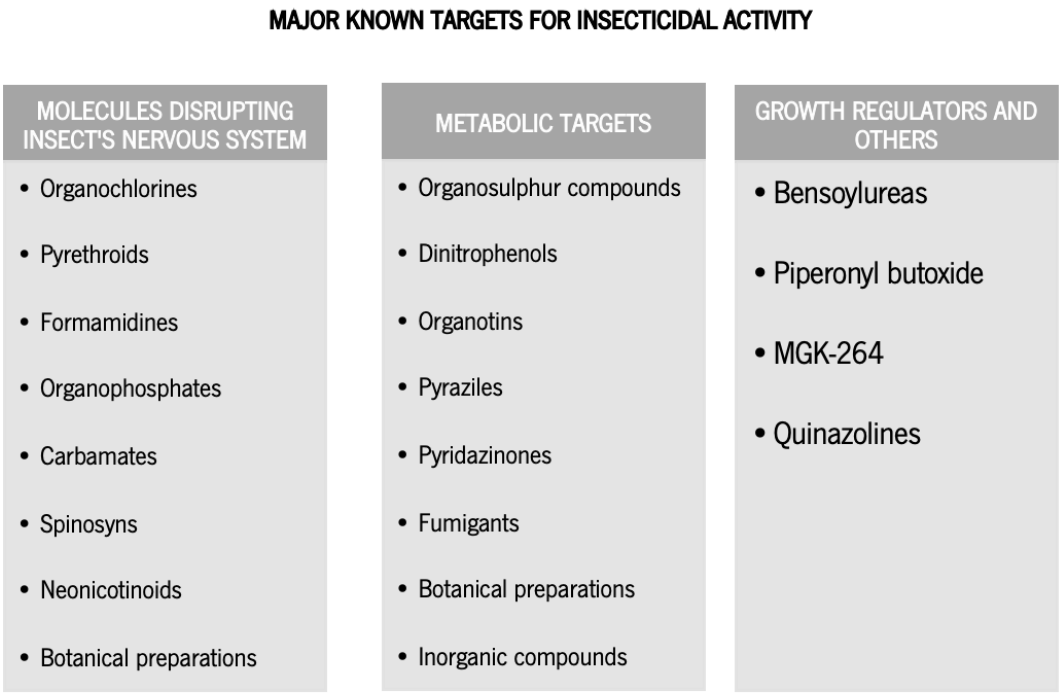


Figure 2. Major known targets for insecticidal activity.

5. The problem of resistance

As a serious threat to human health and agriculture, insect pests can be controlled using insecticides; however, during this ongoing war, insects evolved and found ways to retaliate through the development of multiple resistances. Insects have short life cycles and produce numerous offspring, enabling them to adapt rapidly to stressful situations, such as exposure to insecticides. Their adaptability is due to their high potential for

genetic variation, allowing for the evolution of traits that promote resistance. When an insect pest that is infesting a cultivation becomes resistant, farmers tend to increase the insecticide's usage in quantity and on a larger scale. However, they must be replaced by other types of insect control as soon as possible, right when pest control diminishes [81]. The development of insect resistance can be attributed to three primary factors: genetics, biological and ecological factors, and operational practices. Genetics encompasses various elements, such as the frequency, number, and dominance of resistance alleles; penetrance, expressivity, and interactions of resistance alleles; past genetic selection through exposure to other chemicals; and the extent of integration of a resistant genome with fitness factors. Biological and ecological factors include biotic and generation turnover, offspring per generation, monogamy or polygamy, parthenogenesis, behavioral, isolation, mobility, migration, monophagy or polyphagy, fortuitous survival, and refugia. Lastly, operational practices encompass factors such as the chemical nature of the pesticide, interaction with previously used chemicals, persistence of residues, formulations, application threshold, selection threshold, life stages targeted, mode of application, space-limited selection, and alternating selection [82-86].

Typically, when an insect is exposed to an insecticide, the compound can rapidly penetrate the insect's integument through various routes, such as contact, inhalation, or ingestion [87], ultimately reaching the intended target area. This could be a vital enzyme, nerve tissue, or receptor protein. After binding successfully, and when it reaches certain threshold concentrations, they cause a wide variety of symptoms, resulting in the insect's death [88]. Resistance can be acquired at any step of the insecticide's pathway. Thus, the rate of absorption could be lowered through acquiring higher levels of impermeability, while modifications could appear on target sites where the insecticide's molecules no longer bind. In addition, the appearance of new or more enzymes to help break down the toxic compound is another way organisms can adapt to insecticides. When an organism is exposed to an insecticide, it may initially have low levels of the enzymes needed to break down the compound. However, over time, the organism may adapt and produce more of these enzymes, which can help to reduce the toxicity of the insecticide. This adaptation can be seen as an evolutionary response to the selective pressure of the

insecticide, and it can lead to the emergence of resistant populations [89]. There is also a phenomenon called cross-resistance, where, when an arthropod develops a certain degree of resistance in relation to a compound, it is most likely that the same individual would be resistant to similarly acting insecticides [90]. For instance, there is the possibility that previous selection with insecticides can confer resistance to relatively new insecticides through cross-resistance [91]. One example of this is the diamondback moth, a common pest of cabbage and other cruciferous vegetables. Diamondback moths have become resistant to many insecticides over time, due in part to the extensive use of insecticides in agriculture. In one study, diamondback moths that had been exposed to pyrethroid insecticides were found to be cross-resistant to the newer neonicotinoid insecticides, even though the neonicotinoids had not been used extensively in the area. This suggests that the use of pyrethroids had selected for moths that were already resistant to neonicotinoids or that the two types of insecticides share similar mechanisms of action that contribute to cross-resistance [82].

Target-site resistance occurs when a specific insecticide's site of action is modified within resistant insects, resulting in the molecules no longer binding effectively to those same sites [92]. This is often seen in important pest species with mutations at the nicotinic acetylcholine receptor (nAChR), which can lead to insensitivity to neonicotinoids [93]. Another form of resistance is altered target-site resistance, which happens when the site where the toxin usually binds becomes modified to reduce the insecticide's effects. Insects achieve resistance through the modification of the target protein, resulting in a reduction in the binding of the insecticide and decreased effectiveness of the compound [94,95].

Metabolic resistance is the most common type of resistance and is characterized by a large set of enzymes that are used to breakdown the insecticide, normally used as a way for the insect to detoxify foreign materials. Resistant strains may possess higher levels or more effective forms of these enzymes [92].

The three main enzyme systems are esterases, mono-oxygenases, and glutathione S-transferases, and while metabolic resistance is important for all four insecticide classes (organophosphates, carbamates, pyrethroids, and neonicotinoids), different enzymes affect each class differently [96].

For example, multiple cytochrome P450s, which can structurally metabolize diverse substrates, are also known to play an important role in several biosynthetic pathways. These enzymes are involved in xenobiotic (insecticides and plant toxins) detoxification, hence promoting the development of insecticide resistance and insects' adaptation to their hostplants [97].

Another type of resistance is behavioral [98]. As simple as it may seem, there are various reports of insects who stop feeding when they detect certain insecticides, even leaving the area where the toxic compound is abundant [99,100]. The capability of some insects to recognize danger and act accordingly is notable. Some may move to the underside of a sprayed leaf, move deeper into the crop's canopy, or fly away from the contaminated area [100–102].

Cuticular resistance is traduced by the permeability level of the insect's integument in relation to the toxic compound. The cuticle is the first and major barrier that protects the insect from penetration of external compounds. A reduced penetration of toxic compounds culminates in a reduced uptake of noxious chemicals; hence, modifications in the insect's cuticle (such as in thickness) prevent or slow the absorption/penetration of insecticides [103].

Lastly, there is another physical resistance mechanism whereby the rate of excretion is increased. The excretion process can occur via reflex vomiting of the insecticide and/or defecation of the insecticide, with or without entry into the hemocoel. If the toxic compound is not able to enter hemocoel, the insecticide passes directly through the gut being excreted during defecation; if it does enter, however, the insecticide must be

moved back into the lumen of the gut via filtration and efflux by the Malpighian tubules [104].

For circumventing the described resistance problems, it is highly advised to employ a synergistic approach combining alternative options for conventional frequently used insecticides.

6. Alternatives for conventional Insecticides

Prolonged usage of synthetic insecticides has caused environmental damage, health problems, and biodiversity problems (such as loss of species diversity). Synthetic pesticides have also harmed farmers in the export trade, especially in the horticultural sector. Both farmers and exporters in developing countries have lost market and profits if banned insecticides were detected above the established tolerable level by law [83].

Some of the most rustic and/or outdated solutions for this expanding problem include cultural practices such as the implementation of physical or mechanical barriers (e.g., aluminum foil mulches, see-through nets, temperature/relative humidity manipulation, physical shock and electric discharges, etc.), crop rotations and intercropping (increasing crop diversity), alternative seeding patterns, and companion planting (companion plants which improve crop's performance or work as a lure/repellent for insect pests) [84–86].

Genetically modified plants have long been considered a potential solution to the challenge of feeding a growing population while minimizing the negative impacts on the environment. One promising approach has been to develop plants with intrinsic insect resistance, which can help to reduce the use of chemical pesticides and ultimately decrease carbon dioxide emissions. To achieve this goal, researchers in plant resistance need to utilize advanced technologies such as genotyping by sequencing and high-throughput phenotyping. These tools allow for the identification, mapping, and tracking

of important resistance genes in plants, which are essential for the development of new, improved crop varieties [87].

Additionally, besides engineered pest-resistant crops, genetic pest management includes genetic control of the pest itself. This is focused on sterility resulting from hybrid crosses between different species or different genetic strains that result from the action of mutagenic ionizing radiation. Presumably, this might be used to induce dominant lethal mutations in insects, which, when released into the wild, could sterilize female insects [88,89].

Microbial control is another promising method that involves using insect pathogens to manage pest populations. This approach identifies and utilizes host-virus associations, including various microbial agents, such as fungi, protozoa, bacteria, and nematodes. However, only a few of these entomopathogens have been developed into effective biocontrol agents, and developing new microbial agents requires rigorous testing to determine their efficacy and safety [90,91].

Nowadays, most studies focus on the exploitation of natural products since they offer several advantages over synthetic insecticides. One of the advantages of botanical insecticides is their efficacy against a broad range of insect pests. This is due to their diverse modes of action, which can include disrupting the insect's nervous system, interfering with its metabolism, or causing physical damage to its outer shell. Another advantage is their biodegradability, which means they break down naturally and do not persist in the environment.

Botanical insecticides are also known for their low toxicity to non-target organisms, including humans, pets, and beneficial insects such as bees and butterflies. This makes them a safer option for pest control in areas where non-target organisms may be present [68].

One example of a botanical insecticide is eugenol, which is derived from clove oil. Eugenol has shown efficacy against a variety of insect pests, including aphids, mites, and whiteflies. It works by disrupting the insect's nervous system, causing paralysis and death. Other examples of botanical insecticides include pyrethrins, which are derived from chrysanthemum flowers and are effective against a range of insects such as mosquitoes, flies, and moths. Another example is rotenone, which is extracted from the roots of several tropical plants and has been used for insect control for centuries. Table 2 compiles the biochemical sites of action of the most prominent natural insecticides [105].

Table 2. **Biochemical sites of action of natural insecticides [105].**

Common Name	Class of Insecticide	Targeted System	Mode of Action
Abamectin	Avermectin	Nervous system	Chloride channel activator
Azadirachtin	Botanical from neem oil	Growth and development/ metabolic processes	Prothoracicotropic hormone (PTTH) inhibitor; phagostimulant disruptor
Bacillus thuringiensis	Microbial	Metabolic processes	Insect midgut membrane disruptor
Cinnamaldehyde	Botanical	Energy production	Exact mode of action not well understood; possibly interference with glucose uptake or utilization
Decalesides I and II	Botanical (natural trisaccharides)	Nervous system	Inhibition of sodium pump
Emamectin benzoate	Avermectin	GABA-gated chloride channels	Chloride channel activator
Pyrethrins I and II	Botanical (pyrethrum)	Nervous system	Sodium channel modulato
Rotenone	Botanical	Mitochondrial electron transport system	Electron transport inhibitor—site 1
Ryanodine	Botanical	Calcium channels (ryanodine receptor)	Activation

Spinosad	Spinosyn	Nervous system	Nicotinic acetylcholine receptor agonist (mimic)

In addition to their efficacy and safety, botanical insecticides offer a high accessibility of source materials, as many of the plants from which they are derived can be easily grown and harvested. This makes them a sustainable option for pest control, particularly in developing countries, where access to synthetic insecticides may be limited [105].

7. Conclusions

The development of insecticides has played a critical role in modern agriculture by providing effective control of pests, thereby ensuring food security and improving crop yields. As the world's population continues to grow, it is imperative that agriculture remains productive and sustainable. However, the increasing resistance of pests to existing insecticides, as well as concerns over their environmental impact, highlights the need for continued research and innovation in this field. To meet this challenge, future insecticide development is likely to focus on several key areas, including eco-friendly alternatives such as biopesticides and insect growth regulators, effective resistance management strategies, precision agriculture technologies that minimize the use of insecticides, combination products that target multiple modes of action, and the discovery of novel modes of action that will lead to the development of more effective and safer insecticides.

By addressing these areas, we can ensure that the future of insecticide development will not only maintain effective pest control but also promote sustainable agriculture and minimize negative impacts on the environment. Ultimately, this will help secure the long-term productivity of agriculture and the well-being of our planet and its inhabitants.

B. COMPUTATIONAL METHODS

1. Computer-aided drug design

Computer-aided drug design (CADD) combines a number of chemical-molecular and quantum approaches to aid in the design, discovery, and optimization of new drugs. Some common techniques used in CADD include molecular dynamics simulations, which simulate the movement of atoms in a molecule, and docking simulations, which predict how a drug molecule will bind to a target protein. CADD can also be used in conjunction with experimental methods such as high-throughput screening to identify new drug candidates. Overall, CADD can significantly speed up and optimize the drug discovery process (reducing the time and cost of the process by up to 30%), making it an important tool in the pharmaceutical industry [106].

These techniques have been broadly divided in two groups: Structure-Based Drug Design (SBDD) and Ligand-Based Drug Design (LBDD). Structure-based drug design relies on the three-dimensional structure of the target protein, obtained through X-ray crystallography or NMR spectroscopy. These methods are very useful to characterize the binding site, to elucidate the molecular mechanism of action of active molecules and to evaluate the kinetics and thermodynamics of ligand-target recognition. SBDD's most notable examples include docking and molecular dynamics (MD) [107]. Ligand-based drug design is based on the information of the chemical structures of a set of ligands with known biological activity. One of the main goals of these methods is identifying bioactive compounds or improving the activity of active molecules. Typical examples of ligand-based methods are similarity searching and QSAR modeling [108].

Table 3. Common SBDD and LBDD methods.

Structure-based Drug Design

Molecular Docking	De-novo design	Pharmacophore modelling	Molecular Dynamics
Predicts the interaction between a ligand and a macromolecule based on the energy associated with their conformation.	Generates new molecules with desired biological activity using computational algorithms.	Identifies critical structural components of a molecule required for its biological activity.	Simulates the dynamic behavior of molecular systems, allowing for the analysis of their stability and interaction with other compounds.

Ligand-based Drug Design

QSARs	Pharmacophore modelling	LB-Virtual screening
Quantitative Structure-Activity Relationships Predicts the biological activity of molecules based on their structural characteristics.	Identifies critical structural components of a molecule required for its biological activity.	Uses the structure of a reference ligand to identify other compounds with similar activity.

Both SBDD and LBDD can be used to identify new drug candidates and optimize their properties. However, SBDD is more focused on finding new compounds that bind to a specific target protein, while LBDD is more focused on finding new compounds that have similar activity to known drugs [106].

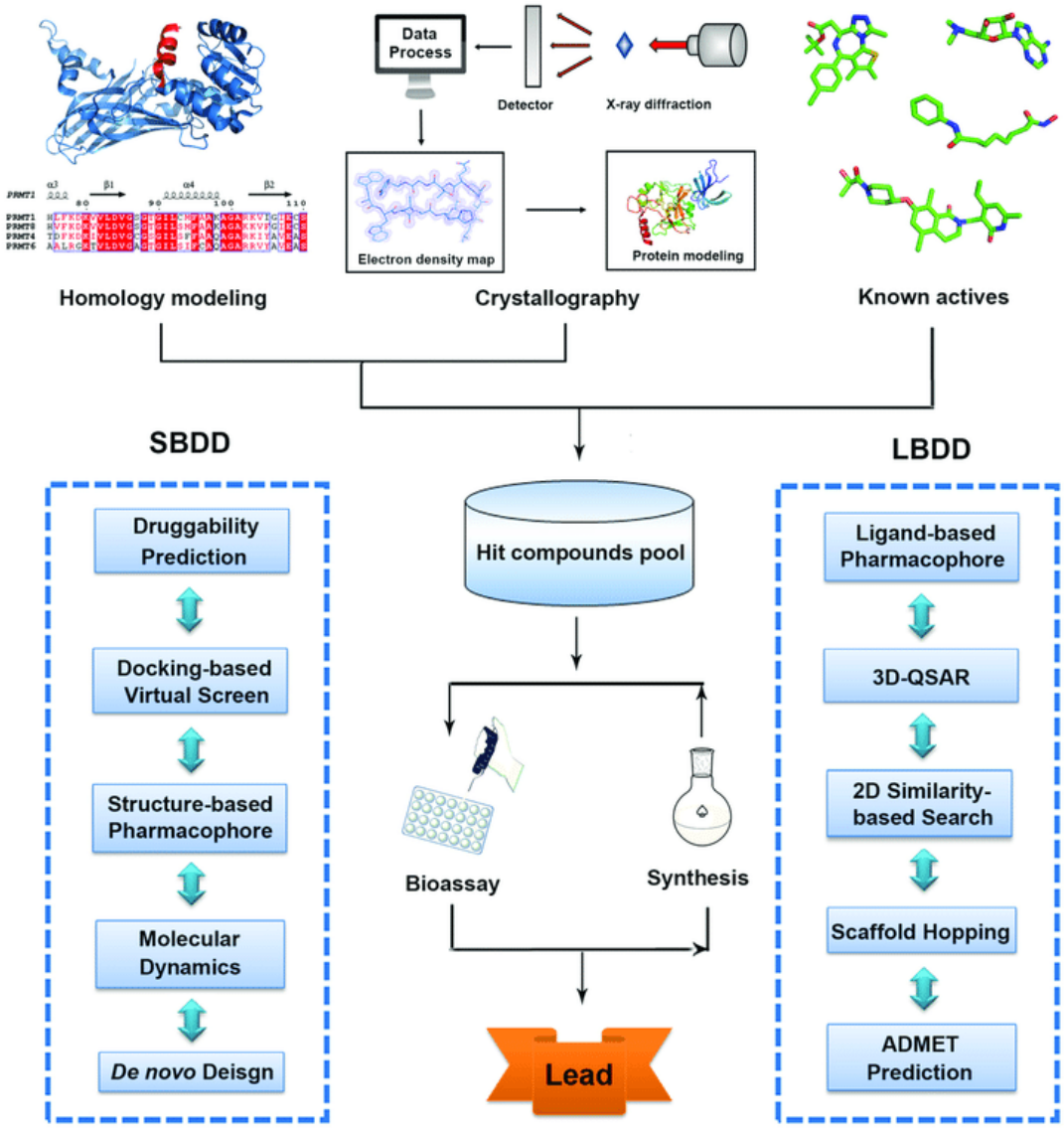


Figure 3. Traditional workflow of SBDD and LBDD [109].

The CADD project performed in the present work integrates a Structure-Based approach. The workflow used is described in figure 4.

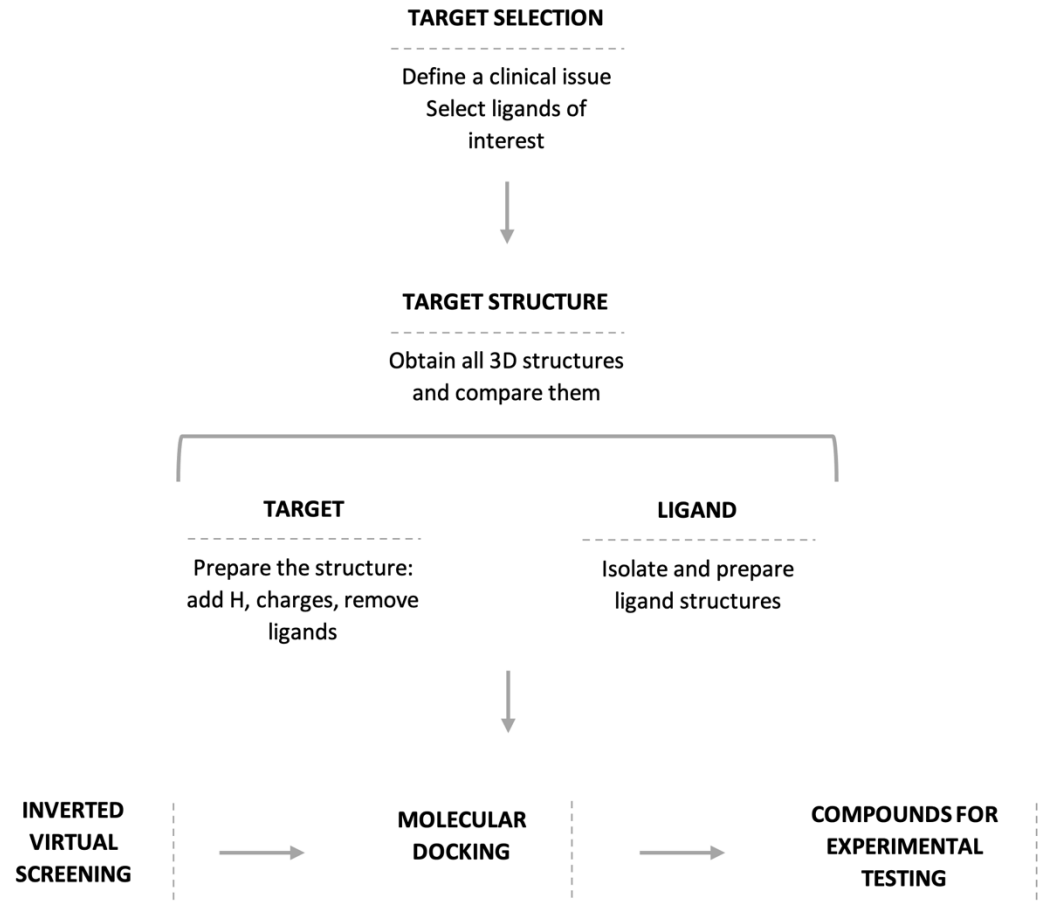


Figure 4. Workflow for the CADD project performed in the present work.

2. Protein-ligand interactions

2.1. Physicochemical Mechanisms of Binding

The success of *in silico* methods in drug development depends, in part, on our understanding of the physical-chemical fundamentals involved during the binding process between molecules.

Proteins are dynamic macromolecules that perform a wide variety of essential processes dependent on their interaction with other molecules or ions, known as ligands.

Protein-ligand interactions have a high degree of affinity and specificity, so even though a particular protein is surrounded by several potential ligands, often it will only establish

a bond with one, in a specific area of the protein known as the binding site. The protein binding site is determined by the arrangement of its amino acids, which configures its shape and chemical reactivity.

A non-covalent interaction (dipole-dipole interactions, hydrogen bonds, electrostatic interactions, among others) is often established between a ligand and a protein, which is naturally weak and reversible. This often means that several interactions occur simultaneously so that the complex formed between the protein and the ligand is stable.

To describe a protein-ligand interaction, we use binding kinetics, which essentially considers the speed at which the binding reaction occurs, and is represented by the following equation:

$$P + L \underset{}{\overset{\Delta G_{Binding}}{\leftrightharpoons}} PL$$

Equation 1. General equation for protein-ligand association reactions, where P represents the Protein, L represents the Ligand, and PL represents the complex formed between the two species.

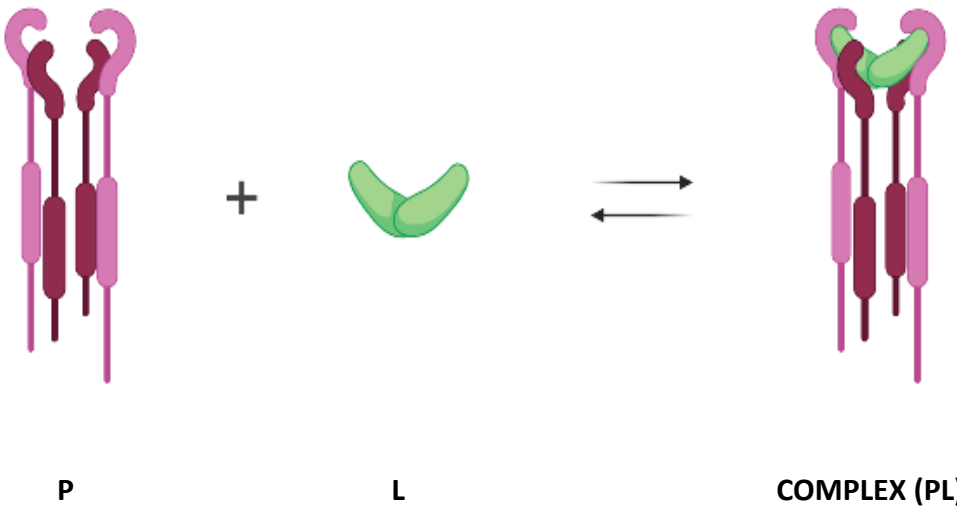


Figure 5. Representation of the protein-ligand binding process.

However, the protein and the ligand are not the only components of the binding system, since the solvent (usually water in the liquid or buffer solution) also plays a crucial role. So, in the target-ligand binding, attention must be paid to the thermodynamics of the Protein-Ligand-Solvent system, and it must be taken into account that the association between the two molecules results from interactions and energy exchanges between P, L and also the solvent molecules.

The Protein-Ligand binding is a thermodynamic process, occurring only when the change in Gibbs free energy (ΔG) of the system is negative. The absolute value of ΔG also determines the strength of the binding, that is, the affinity of a particular ligand with the target, which translates into the stability of the formed complex. Despite the solvent having a great influence on binding, its contribution to the process is extremely difficult to calculate, so its effect is largely ignored, with an approximation of the binding energy being calculated using equation 2:

$$\Delta \boldsymbol{G_{binding}} = \Delta \boldsymbol{G_{complex}} - \Delta \boldsymbol{G_{protein}} - \Delta \boldsymbol{G_{ligand}}$$

Equation 2. Calculation of Gibbs Free Energy for the formation of the protein-ligand complex.

The standard molar Gibbs binding energy can be determined from the experimentally calculated binding constant ($\boldsymbol{K_{binding}}$) through the following equation:

$$\Delta \boldsymbol{G^0_{Binding}} = -\boldsymbol{RTln}\Delta \boldsymbol{K_{binding}}$$

Equation 3. Mathematical relationship between the binding constant and the standard molar Gibbs binding energy for the protein-ligand complex.

This thermodynamic property can be decomposed into its enthalpy ($\Delta \boldsymbol{H^0_{binding}}$) and entropy ($\Delta \boldsymbol{S^0_{binding}}$) components. In this way:

$$\Delta \boldsymbol{G^0_{binding}} = \Delta \boldsymbol{H^0_{binding}} - \boldsymbol{T}\Delta \boldsymbol{S^0_{binding}}$$

Equation 4. Mathematical relationship between standard molar enthalpy (ΔH°), entropy and Gibbs energy of binding at constant temperature (T).

The enthalpic and entropic contributions of protein binding are difficult to calculate, therefore, to reduce the computational time of calculation, these contributions are often simplified in computational algorithms. Enthalpy (H) is a thermodynamic property of a system, defined as the sum of its internal energy plus the product of its volume and the external pressure.

Entropy (S) is the quantification of heat change entering the system at constant temperature in a reversible process, divided by temperature.

In a protein-ligand binding process, enthalpy has a large energetic contribution. Almost always, a non-covalent interaction (London dispersion interactions, hydrogen bonds, electrostatic interactions) is established between a ligand and a protein, which can be defined as a chemical interaction capable of stabilizing the conformation of interacting molecules. The binding enthalpy ($\Delta H^0_{binding}$) is, in a simplified manner, the sum of all energy changes resulting from the formation of non-covalent interactions, and its contribution may or may not be favorable for the binding process. The protein-ligand association enthalpy implies the replacement of the initial interaction between the protein and water molecules at the binding site and the interaction between the free ligand and the surrounding water molecules with the interaction of the ligand with the protein and the complex with the new environment formed around it. From an enthalpic point of view, favorable protein-ligand binding occurs when protein-ligand interactions are dominant. Otherwise, spontaneous binding of this type may be driven by entropy (hydrophobic binding processes).

The entropy of protein-ligand binding ($\Delta S^0_{binding}$) is the sum of two components: the internal component ($\Delta S^{int}_{binding}$) and the solvent reorganization component ($\Delta S^{sol}_{binding}$). The first component is almost always negative due to the reduction in particle number during the binding process and is also responsible for changes in the conformational freedom of both the ligand and protein. The entropy of the solvent changes after binding due to the reorganization of its molecules, which can cause it to be released.

The enthalpy, entropy, and change in Gibbs energy resulting from both contributions drive the thermodynamic process of ligand binding to protein, and the concepts explained are used in CADD for the prediction of binding strength and affinity.

2.2. Protein-Ligand Binding Models

The mechanics of protein-ligand binding are described through several models, including the Lock-and-Key, Induced Fit, and Conformational Selection models. The selection of a specific model has an impact on the choice of docking protocol and the accuracy of the resulting predictions. In this context, it is imperative to provide a brief overview of these models.

The Lock-and-Key model is a classic explanation of enzyme-substrate interaction, first introduced by Emil Fischer in 1894. It suggests that enzymes act like a lock and the substrate acts as a key, where the shape of the active site on the enzyme fits precisely with the substrate, allowing it to bind and undergo a chemical reaction. This model helps explain the specificity of enzymes, as only the correctly shaped substrate can fit into the active site, allowing it to participate in a reaction [110,111].

While the Lock-and-Key model is still widely taught, it has been modified over time to the “Induced Fit” model. In this model (introduced by Daniel Koshland in 1958), the active site of an enzyme is flexible and can change shape slightly upon substrate binding, resulting in a tighter fit and a more efficient reaction. The Induced Fit model takes into account the dynamic nature of enzyme-substrate interactions and provides a more accurate description of the process compared to the Lock-and-Key model [110,111].

The Conformational Selection model is another explanation of enzyme-substrate interaction, which suggests that enzymes exist in a range of conformations and that the substrate selects the appropriate conformation for interaction. In this model, the

enzyme does not change its shape upon substrate binding, but rather the substrate selects the conformation of the enzyme that is most favorable for interaction. This model emphasizes the role of the substrate in determining the interaction with the enzyme, rather than the active site of the enzyme dictating the interaction as described in the Lock-and-Key and Induced Fit models.

The Conformational Selection model suggests that the enzyme exists in an ensemble of conformations and that the substrate selectively binds to a particular conformation that is best suited for the reaction to occur. This model helps explain the observation that enzymes can bind to multiple substrates with different structures, as the correct conformation of the enzyme is selected for each substrate.

It is important to note that the Conformational Selection model is not mutually exclusive with the Induced Fit model and both models can contribute to the overall understanding of enzyme-substrate interactions [111,112].

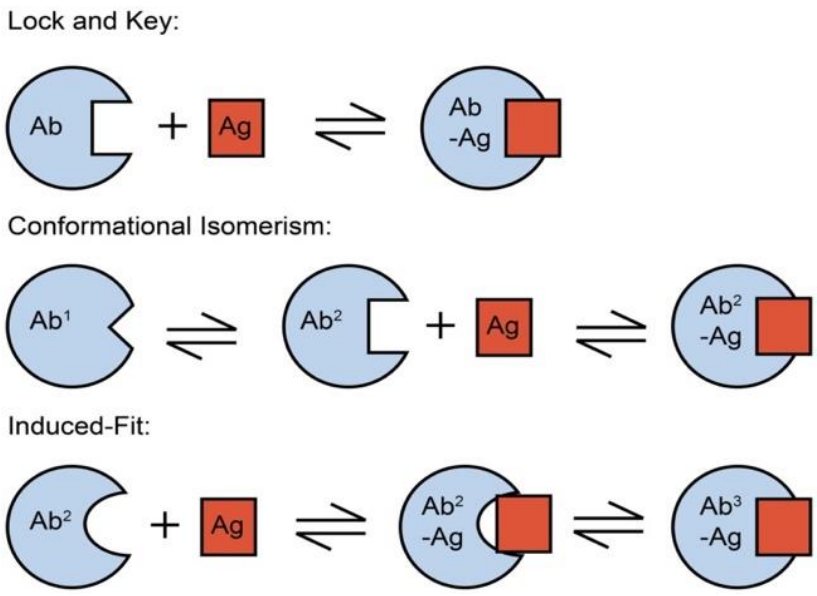


Figure 6. Mechanism of protein-ligand binding reactions [113].

3. Molecular docking

During the past three decades, Molecular Docking (MlD) has been indispensable in structure-based drug development due to its accuracy in predicting the best binding pose between two molecules. Depending on the nature of these molecules, MlD can be categorized as Protein-Protein or Protein-Ligand Docking. We will be focusing on Protein-Ligand Docking which goal is to predict and rank the preferred position, orientation, and conformation of a given ligand in relation to a known 3D structural target protein, when the two bind to form a stable complex [114].

The area on the protein where the ligand may bind, also referred as the active center or the binding area of the complex, may or may not be known. When the binding site's location is known, the program allows the user to easily limit the binding region to a specific section of the protein. However, in the absence of this information, as a last resort, a blind docking can be performed. Blind docking scans the entire target surface to explore putative binding pockets [115]

Molecular docking programs must run quickly as they are designed to be applied to vast databases. This involves a simplification of the search algorithms (SA) and the scoring functions (SF). Nevertheless, the trade-off between accuracy and speed must be conserved. Therefore, several combinations of known SF and SA should be evaluated to discover the best potential procedure for the situation at hands. A binding center and search space for the algorithm must also be defined. It is important to remember that the best-scored pose must coincide with a real binding conformation, which means that if experimental data exists, this pose must match to what is observed [116]. Hence, a docking protocol is a combination of a search algorithm, a scoring function, and a defined binding center [117]. Following that, these terms will be thoroughly discussed.

3.1. Definition

During the past three decades, Molecular Docking (MlD) has been indispensable in structure-based drug development due to its accuracy in predicting the best binding pose between two molecules. Depending on the nature of these molecules, MlD can be categorized as Protein-Protein or Protein-Ligand Docking. We will be focusing on Protein-Ligand Docking which goal is to predict and rank the preferred position, orientation, and conformation of a given ligand in relation to a known 3D structural target protein, when the two bind to form a stable complex [114].

The area on the protein where the ligand may bind, also referred as the active center or the binding area of the complex, may or may not be known. When the binding site's location is known, the program allows the user to easily limit the binding region to a specific section of the protein. However, in the absence of this information, as a last resort, a blind docking can be performed. Blind docking scans the entire target surface to explore putative binding pockets [115]

Molecular docking programs must run quickly as they are designed to be applied to vast databases. This involves a simplification of the search algorithms (SA) and the scoring functions (SF). Nevertheless, the trade-off between accuracy and speed must be conserved. Therefore, several combinations of known SF and SA should be evaluated to discover the best potential procedure for the situation at hands. A binding center and search space for the algorithm must also be defined. It is important to remember that the best-scored pose must coincide with a real binding conformation, which means that if experimental data exists, this pose must match to what is observed [116]. Hence, a docking protocol is a combination of a search algorithm, a scoring function, and a defined binding center [117]. Following that, these terms will be thoroughly discussed.

3.2. Search Algorithms

Search Algorithms are used to generate several poses for the protein-ligand complex within a conformational space that results from both systems' degrees of freedom and from the spatial arrangement of the interaction between them two. This search results in a space too large to be accurate in an acceptable time span – as told before, the several existing SA employ certain approximations and simplifications to reduce computational time without compromising on accuracy [115,118].

These algorithms are classified as Rigid-Body, Flexible-Ligand, and Flexible-Protein, depending on the number of degrees of freedom considered for each molecule in the complex, i.e., how much flexibility is considered during the Docking experiment [115,118].

3.2.1. Rigid Body Search Algorithms

These were widely used in the early stages of Protein-Ligand docking investigations and are still used in certain Protein-Protein docking research today. They represent the simplest and quickest algorithms – these approaches only analyze geometrical complementarities between the two molecules of the complex and only explore the six degrees of freedom (rotational and translational space), neglecting flexibility. Both the receptor and the ligand are taken into account, which means they're fairly restrictive [115,117].

3.2.2. Flexible-Ligand Algorithms

Flexible-Ligand Algorithms are currently the most extensively used ones. These perceive the protein to have a stiff body and the ligand to be completely flexible. They investigate the complex's six translational and rotational degrees of freedom, as well as the ligand's conformational degrees of freedom. Because these methods are more computationally intensive, they rely on a number of approximations to ensure that they can be used efficiently. Systematic methods, random or stochastic methods, and molecular simulation methods, are the three categories [115].

a) Systematic methods are split into three categories: conformational search methods, fragmentation methods, and database approaches. They attempt to investigate all conformational degrees of freedom of the ligand [115].

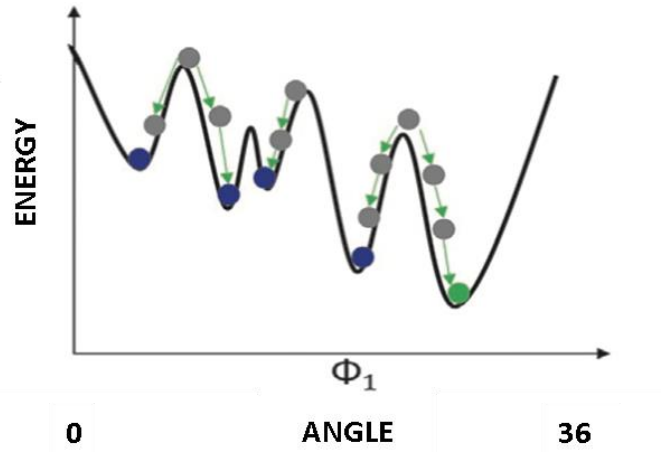


Figure 7. Systematic SA performs a systematic study of all available degrees of freedom. Each angle corresponds to an accepted or rejected value in the energy environment.

- Conformational search methods systematically explore all rotatable bonds by 360º in modest, predetermined steps to create all conceivable conformations. With a greater number of rotatable bonds, a greater number of conformations are formed, with the ultimate outcome being unachievable with present computational capacity. As a result, multiple constraints on the ligand bonds are used to minimize the number of conformers formed [115,119]. This number, N, can be calculated using the following equation:

$$N_{TC} = \prod_{i}^{N} \frac{360}{\theta_i}$$

Equation 5. Number of total conformers resulting from the application of a fixed increment to all routable bonds, to generate all possible conformations.

- Fragmentation search methods divide the ligand into numerous pieces that are subsequently docked into the binding site and covalently bonded to reconstruct the original ligand. Instead, the ligand can be separated into a core fragment that is docked initially, followed by the other fragments in a process known as "incremental construction" or "anchor and grow technique".

- Database search methods rely on pre-generated conformational ensemble databases to add flexibility to the docking process while taking intra and intermolecular distances into account. The algorithm uses the databases to generate a given number of variant compounds of the desired ligand and perform rigid docking on these molecules [115,119]

b) Random or stochastic algorithms explore the ligand conformational space by making random changes to its conformation, which are then accepted or rejected by a predetermined probability function. Yet, by doing so, the chances of the final solution being locked in a local minimum are significantly reduced, and the chances of finding a desirable global minimum are greatly increased [120]. Random algorithms are used in six types of docking methods: Monte Carlo, Genetic Algorithms, Tabu Search, Particle Swarm Optimization, Differential Evolutionary Algorithms, and Evolutionary Gaussian Algorithms [115,119].

- Monte Carlo methods dock the ligand into the binding site utilizing a large number of random translations and rotations, reducing the likelihood of being caught in a local minimum. These approaches employ basic energy minimization functions that do not require any derivative information and are highly effective at stepping energy barriers, allowing for a good sample of the conformational space. A Boltzmann probability function is used to assess the produced conformations [115,119].

- Genetic algorithms (GA) are based on genetics and the theory of biological evolution. It starts with an initial population of several ligand poses (chromosomes) generated randomly. Each pose is represented by an individual. Each individual is defined by a set of genes, which describe the ligand conformation and its translation and orientation in relation to the protein. The full set of these variables is called the genotype and the atomic coordinates of the ligand are the phenotype. Through various generations, or cycles, genetic operators such as mutations, crossovers and migrations are applied to random individuals of the population to explore the conformational space. At the end of each generation, at random, individuals are evaluated with conformations with negative evolution being excluded. The process continues until the population satisfies a predefined fitness function. Various programs use these

algorithms including GOLD and AutoDock which were used in this work. Differential Evolutionary algorithms are derived from GA methods [115,119].

- Tabu Search algorithms move from one pose to another, imposing several restrictions to make sure that previous poses are not revisited. The Root Mean Square Deviation (RMSD) of a new conformation is calculated in relation to a "tabu list" featuring the visited poses and used to accept or reject the new conformation [115,119].

- Particle Swarm Optimization is a simpler and faster process than GA. Many molecular docking studies showed that the conventional algorithms can give satisfactory results, even when the protein is considered as a rigid entity, (lock and key model of molecular recognition). However, many proteins undergo a range of structural changes upon ligand binding. These range from a local rearrangement of side chains near the binding site to less common backbone movements (induced fit model). In order to address this issue, specialized search algorithms were developed to account for the partial flexibility of the protein [115,119].

A few examples of docking programs using systematic and stochastic methods are listed in the table below.

Table 4. Docking programs using systematic and stochastic methods.

SYSTEMATIC METHODS	STOCHASTIC METHODS
Glide	GOLD
FlexX	DockVision
Dock	PLANTS
Superflex-Dock	AutoDock
FRED	PSI-DOCK

3.2.3. Flexible-protein algorithms

Although traditional algorithms produced good results even when the protein was regarded a rigid entity, specific search algorithms were created to account for a protein's partial flexibility. These can overcome the modifications that proteins go through following ligand binding, which can range from local side chain rearrangement near the binding site to less typical backbone movements (induced fit model) [115].

3.3. Scoring Functions

Scoring functions are quick approximation mathematical approaches for predicting the strength of interaction (or binding affinity) between two or more molecules, outlining the correct poses from the incorrect ones [119]. Because many physical phenomena involved in molecular recognition are not taken into account, accuracy might be compromised. Therefore, an accurate scoring function would ideally perform equally well on all four tasks:

- scoring power - the capacity to provide scores that are linearly related to experimental data on binding affinity.
- ranking power - the capacity to appropriately rank a given set of ligands that bind to a common target protein based on their binding affinities when their binding postures are known.
- docking power - the capacity to identify a ligand's natural binding pose as the one with the highest score while screening a vast number of produced decoy poses.
- screening power - the capacity to choose genuine binders to a certain target protein from a pool of random molecules.

The present number of scoring functions is huge and constantly growing. They can be grouped in four main classes: force-field scoring functions, empirical scoring functions, knowledge-based potentials, and consensus scoring [115].

3.3.1. Force-field or Physics-based Scoring Functions

Force-field or physics-based scoring functions are based on molecular mechanics force fields such as AMBER ([115].They compute binding energy by adding the contributions of bonded interactions (bond stretching, angle bending, and torsion angles) and non-bonded interactions (van der Waals and electrostatic interactions) within the protein-ligand complex, which accounts for the contribution of enthalpy to energy [119].

The force field was created to model the enthalpy gas-phase contributions to structure and energetics. As a result, essential factors for the ligand-receptor interaction, such as solvation and entropic parameters, were omitted. This is remedied by the insertion of solvation approaches like GBSA and PBSA that account for desolvation energies, and a torsional entropy term assesses the conformational entropy lost during the binding process.

The disadvantage of this scoring function is that it needs the usage of cut-off distances to address non-bonded interactions. These are selected randomly, which compromises the binding process's appropriate treatment of long-term impacts [119].

3.3.2. Empirical Scoring Functions

Empirical scoring functions quantify a protein-ligand complex's binding affinity by adding up several energy components involved in protein-ligand binding, such as hydrogen bonds, hydrophobic effects, protein-ligand clashes, and so on. Each element is multiplied by a coefficient derived from a series of linear regression calculations performed on a training set of protein-ligand complexes with known binding affinities.

$$Score = \sum_{i} w_i \, \Delta G_i$$

Equation 6. General Empirical based scoring function.

Because of their simplified treatment of the energy components, empirical scoring functions are substantially faster in binding score computations than force-field or physics-based scoring systems. The accuracy of empirical scoring functions, on the other hand, is directly connected to the quality and coverage of the protein-ligand training set used to create the model. ChemScore, GlideScore, and ChemPLP are examples of empirical scoring functions [119].

3.3.3. Knowledge-based Scoring Functions

Instead of replicating binding affinities, knowledge-based scoring functions use statistical approaches to recreate experimental data. These functions employ statistic potentials to forecast the likelihood of occurrence of common interactions, such as various atom-atom pair connections, based on empirically observed structures. This technique believes that if an interatomic distance is greater than the average, it signifies a favorable interaction [119].

The benefit of knowledge-based scoring functions is their computational simplicity, needing just knowledge of a limited range of protein-ligand complex structures, which has grown in number over time. It is also as quick as empirical scoring functions. The key issue is that their parametrization is constrained by the known sets of complicated structures utilized to create the method [119].

$$Score = \sum_{i,j}^{N} u_{ij}(r)$$

Equation 7. General Knowledge based scoring.

3.3.4. Consensus Scoring

Consensus scoring functions combine information from many scoring functions to increase the likelihood of finding the correct solution. Each scoring function may predict the pose but not the binding affinity since the terminology utilized to describe this interaction are insufficient. Several studies have shown that employing consensus scoring functions can increase performance by correcting for the shortcomings of each scoring function, allowing for a reduction in the number of false positives [119].

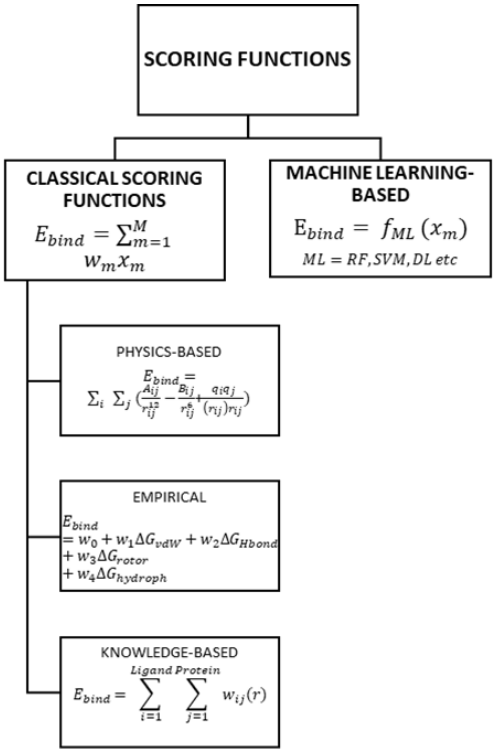


Figure 8. Scoring functions in docking and their respective equations.

3.4. Software Used

3.4.1. AutoDock Vina

Trott and Olson created AutoDock Vina, in 2009, at the Scripps Research Facility in California. This program maintains some of the original ideas from AutoDock 4 (another

software that was not employed in this project), but its conceptually different, showing to be faster and more accurate [118,121]

Vina employs an Iterated Local Search Global Optimizer, which repeats two processes - mutation and local optimization - and each step is either approved or rejected. The Broyden-Fletcher-Goldfarb-Shanno (BFSG) approach is used for local optimization, which applies both the value of the scoring function and its derivative to its arguments (position, orientation, and torsion values for rotatable bonds). The number of steps in the run is determined by the difficulty of the search issue, and the runs may be done on multithreading and multicore Central Processing Unit (CPUs) [121].

It uses a mixed empirical and KBSF scoring function, which can be expressed as:

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij})$$

Equation 8. AutoDock Vina scoring function.

In equation 8, t_i and t_j are the types assigned to atoms *i* and *j* respectively. The score value (c) represents the free energy of binding between a ligand and a protein, and can be decomposed in an intramolecular (c_{intra}) and an intermolecular (c_{inter}) [118,121]:

$$c = c_{intra} + c_{inter}$$

Equation 9. Decomposition of the Autodock Vina scoring function.

The score is determined by considering the interactions between atoms of the protein and the ligand, with a specific type assigned to each atom. Pairwise potentials ($f_{t_i t_j}$) between the different atom types are defined and used in the calculation of the score, considering all pairs of moving atoms, excluding those separated by three consecutive covalent bonds. The goal of the search algorithm used by Vina is to find the global

minimum of the score, which corresponds to the most energetically favorable conformation of the ligand-protein complex. The standard molar Gibbs energy of association ($\Delta_a G_k^0$) is then estimated from the intermolecular scoring function (c_{inter_k}) [118,121]:

$$\Delta_a G_k^0 \approx g(c_{inter_k}) \Leftrightarrow \Delta_a G_k^0 \approx g(c_k - c_{inter_k})$$

Equation 10. Estimation of the standard molar Gibbs energy of association, for the lowest-scoring conformation obtained with the Autodock Vina search algorithm.

3.4.2. GOLD

In 1997, a collaborative project between the University of Sheffield, the GlaxoSmithKline and the Cambridge Crystallographic Data Center resulted in one of the most cited docking programs in the literature today, GOLD (or Genetic Optimization for Ligand Docking). This software generates ligand poses using a genetic algorithm and offers four score systems to choose from, namely, The Astex Statistical Potential (ASP), ChemPLP, CHEMSCORE, and GOLDSCORE. The docking method is configured using the Hermes graphical user interface [118,122].

a) The Astex Statistical Potential is an atom-atom potential derived from a protein-ligand complex database. It creates statistical potentials based on information from existing ligand-protein structures on the frequency of interaction between ligand and protein atoms. ASP differs from other statistical potentials in that it employs a reference state that runs how the raw distribution of data is turned into potentials. If there are no interactions between the atoms, the reference state is the predicted number of contacts [123].
b) ChemPLP is an empirical scoring function that uses the Piecewise Linear Potential to represent the steric complementarity between the protein and the ligand (PLP). The PLP scoring function defines two piecewise functions: plp for repulsive/attractive interactions and rep for completely repulsive interactions.

$$f_{plp} = \sum_{p \in P_{prot-lig-plp}} plp(p_r, p_A, p_B, p_C, p_D, p_E, p_F,) + \sum_{p \in P_{prot-lig-rep}} rep(p_r, p_A, p_B, p_C, p_D, p_E, p_F,)$$

Equation 11. Piecewise Linear Potential scoring function.

This equation is the current version of GOLD's default scoring function, where $P_{prot-lig-plp}$ and $P_{prot-lig-rep}$ represents sets of protein-ligand atom pairs that are utilized to evaluate each function and p_r denotes the distance between a ligand and a protein atom. The parameters p_A to p_F depend on the interaction potential chosen [124].

ChemPLP (present below in equation 2.4.4 - 5) makes use of several CHEMSCORE scoring function terms; amongst them, the distance and angle-dependent factors associated with hydrogen and metal binding and the Tripos force field and heavy-atom collision used to calculate intraligand interactions.

$$f_{CHEMPLP} = f_{plp} + f_{hb} + f_{hb-ch} + f_{hb-CHO} + f_{met} + f_{met-coord} + f_{met-ch} + f_{met-coord-ch} + f_{clash} + f_{tors} + c_{site}$$

Equation 12. ChemPLP scoring function.

In this equation, f_{plp} represents the piece linear potential, f_{hb} for the hydrogen bonds, f_{met} represents the metal interactions, f_{clash} represents the ligand clash potential, f_{tors} represents the ligand torsional potential and c_{site} represents a quadratic potential that guides the calculations to the binding site [124].

c) CHEMSCORE is an empirical scoring function comprised of 82 protein complexes with empirically established binding affinities. Through equation 13, it calculates the free standard molar Gibbs energy of association.

$$ChemScore \approx \Delta_a G^0 \Leftrightarrow ChemScore = \Delta_a G^0_{ref} + \Delta_a G^0(L) + \Delta_a G^0(PL)$$

Equation 13. General decomposition of the CHEMSCORE scoring function.

In equation 13 $\Delta_a G^0_{ref}$ is a reference value, $\Delta_a G^0(L)$ is the component is the component connected with the conformational rearrangement of the ligand upon binding and $\Delta_a G^0(PL)$ is the component related with the protein-ligand interactions.

The ligand conformation rearrangement component of CHEMSCORE, $\Delta_a G^0(L)$, represented in equation 14, is thought to be entirely energetic. It indicates the energy cost associated with the ligand's use of a non-optimized shape in the PL complex in order to maximize its interactions with the protein target [125].

$$\Delta_a S^0(L) \approx 0 \Rightarrow \Delta_a G^0(L) \approx \Delta_a E(L) \Leftrightarrow \Delta_a G^0(L) = E_{PL}(L) - E(L)$$

Equation 14. The ligand conformation rearrangement component of CHEMSCORE scoring function.

In the preceding equation, $\Delta_a S^0(L)$ represents the ligand conformation rearrangement entropy, $\Delta_a E(L)$ the ligand conformation rearrangement energy, $E_{PL}(L)$ the energy of the ligand in the geometry adopted upon binding with the protein, and $E(L)$ the energy of the ligand in its optimized geometry [125].

In contrast, the protein-ligand component $\Delta_a G^0(PL)$ is determined using the following equation:

$$\begin{aligned}\Delta_a G^0(PL) \\ &= \Delta_A G^0_{hbond}(PL) + \Delta_a G^0_{Ma}(PL) + \Delta_a G^0_{lipo}(PL) + \Delta_a S^0_{rot}(PL) \\ &+ E_{clash}(PL) + E_{cov}(PL)\end{aligned}$$

Equation 15. Decomposition of the protein-ligand component of CHEMSCORE scoring function.

There are entropic-energetic, pure entropic, and pure energetic variables in this equation. The hydrogen-bond $\Delta_A G^0_{hbond}(PL)$ the metal-acceptor $\Delta_a G^0_{Ma}(PL)$ and the

lipophilic $\Delta_a G^0_{lipo}(PL)$ are the entropic-energetic terms, and they can be calculated using equation 16 (see below). The rotameter term that is pure entropic is $\Delta_a S^0_{rot}(PL)$. The clash $E_{clash}(PL)$ and the covalent $E_{cov}(PL)$ energies are the pure energetic expressions [125].

$$\Delta_a G^0_X(PL) = \Delta_a G^0_{X,opt}(PL) \sum_{i=1}^{n_X} f_i \,; X = hbond, Ma\ or\ lipo\ and\ 0 \leq f_i \leq 1$$

Equation 16. General equation for calculating the mixed entropic-energetic terms associated with the protein-ligand component of CHEMSCORE scoring function.

In equation 16, n_X is the number of atomic pairs (one atom belonging to the ligand and other to the protein) associated with an interaction of the type X, f_i is the effectiveness factor of the *i*-th of these interactions and $\Delta_a G^0_{X,opt}(PL)$ is the standard molar Gibbs energy for an optimal interaction (f = 1) of this type.

The rotamer term $\Delta_a S^0_{rot}(PL)$ represents the entropy penalty, associated with the rotamers of ligand that are constrained due to interactions with the protein. This term is calculated in a similar way to those used in equation 16.

$$\Delta_a S^0_{rot}(PL) = \Delta_a S^0_{rot,max}(PL) \sum_{i=1}^{n_{rot}} f_i \,; 0 \leq f_i \leq 1$$

Equation 17. Equation for calculating the rotamer term associated with the protein-ligand component of CHEMSCORE scoring function.

In equation 17, n_{rot} is the number of ligand rotamers that are constrained upon binding, f_i is the effectiveness factor of the i-th of these rotamers and $\Delta_a S^0_{rot,max}(PL)$ is the maximum entropy penalty (correspondent to f = 1) associated with a rotamer of this type [125].

The clash energy term ($E_{clash}(PL)$) is associated with the repulsive interactions involving atomic pairs dominant at short distances ($r \leq r_{clash}$). This term is calculated by equation 18 [125].

$$E_{clash} = \sum_{i=1}^{n_{clash}} \varepsilon_i\left(r_i, r_{clash_i}\right)$$

Equation 18. Equation for calculating the clash energetic term associated with the protein-ligand component of CHEMSCORE scoring function.

In equation 18, n_{clash} is the number of heavy atomic pairs (one atom belonging to ligand and the other to protein), that are close in contact and $\varepsilon_i\left(r_i, r_{clash_i}\right)$ is the clash energy of the *i*-th of these pairs that is characterized by a distance r_i and a clash distance r_{clash_i}. This energetic quantity can be calculated by equation 19.

$$\varepsilon_i(r_i > r_{clash_i}) = \begin{cases} 0;\ r_i > r_{clash_i} \\ \dfrac{20}{\Delta_a G^0_{hbond,opt}} \dfrac{(r_{clash_i} - r_i}{r_{clash_i}};\ r_i \leq r_{clash_i}\ and\ the\ atomic\ pair\ i\ is\ involved\ in\ a\ hydrogen-bond \\ \dfrac{20}{\Delta_a G^0_{hbond,opt}} \dfrac{(r_{clash_i} - r_i}{r_{clash_i}};\ r_i \leq r_{clash_i}\ and\ the\ atomic\ pair\ i\ is\ involved\ in\ a\ metal-acceptor\ interaction \\ 1 + \dfrac{4\left(r_{clash_i} - r_i\right)}{r_{clash_i}};\ r_i \leq r_{clash_i}\ and\ the\ atomic\ pair\ i\ is\ not\ involved\ in\ any\ of\ the\ previous\ interactions. \end{cases}$$

Equation 19. Clash energy for an atomic pair, characterized by a distance $\mathbf{r_i}$ and a clash distance $\mathbf{r_{clash}}$.

The covalent energy term ($E_{cov}(PL)$) is associated with the covalent bonds eventually established between the ligand and the protein. This term can be calculated using equation 2.4.4 – 14 [125].

$$E_{cov} = \sum_{i=1}^{n_{tc}} \varepsilon tors\left(\omega_i\right) + c_{cov} \sum_{j=1}^{n_{ac}} K_j\left(\theta_j - \theta_{o,j}\right)^2$$

Equation 20. Equation for calculating the covalent term associated with the protein-ligand component of CHEMSCORE scoring function.

In equation 20, the first summation is over all n_{tc} dihedral angles involved in the covalent linkage and the second one is extended to all n_{ac} covalent bond angles around the same linkage. In this equation, $\varepsilon tors(\omega_i)$ is the torsional energy associated with the dihedral angle (ω_i), K_j is the force constant of the bond angle number, j of magnitude, θ_j, $\theta_{o,j}$ the ideal magnitude for this angle and c_{cov} a constant used to balance the covalent bond term against the rest of the CHEMSCORE scoring function.

d) The GOLDSCORE force field scoring function, estimates the association energy $(\Delta_a E)$. This scoring function is the original used by GOLD and can be calculated using the following equation:

$$GoldScore \approx \Delta_a E \Leftrightarrow GoldScore \approx \Delta_a E_1 + \Delta_a E_2 \ (21a)$$
$$\Delta_a E_1 = \Delta E_{rearr}(L) + \Delta E_{rearr}(P) \Leftrightarrow \Delta_a E_1$$
$$\approx E(L_{bound}) - E(L) + E(P_{bound}) - E(P) (21b)$$
$$\Delta_a E_2 = \Delta_a E_{2,L-J} + \Delta_a E_{2,hbond} \ (21c)$$

Equation 21. **General formulation for the GOLDSCORE scoring function.**

If a rigid-protein search algorithm is adopted, E(P_{bound}) = E(P) and second term of equation 21b is neglected.

3.4.3. LeDock

LeDock is flexible small molecule docking software developed by Hongtao Zhao and co-workers (Zhang & Zhao, 2016). LeDock combines a genetic algorithm with simulated annealing search to generate the first generation of docking poses. The conformation of the ligand is randomly changed at the start of each simulated annealing search, so that each search starts with a different pose. This software uses an empirical scoring function which can be calculated by the following equation [126].

$$LeDockScore \approx \Delta_a G^0 \Leftrightarrow LeDockScore$$
$$= \alpha \sum_{i=1}^{n_L} \left(E_i^{LJ} + E_i^{hb}\right) \times H(\left|E^{LJ}{}_i + E_i^{hb}\right| - E_{cut})$$
$$+ \beta(r) \sum_{i=1}^{n_L} E_i^{el} + \gamma E_L^{str}$$

Equation 22. **LeDock scoring function.**

In this equation, the summations are extended to all the nL atoms of the ligand. Each of their terms represents a specific interaction (Lennard-Jones + hydrogen bond in the first summation and electrostatic in the second summation) between an atom i of the ligand with all atoms of the protein. In the first summation, H is the Heaviside step function (see equation 23) and E_{cut} is the cut-off energy for Lennard-Jones + hydrogen bond interactions.

$$H(x) = \begin{cases} 0; x < 0 \\ 1; x \geq 0 \end{cases}$$

Equation 23. **The Heaviside step function.**

As E_{cut} is a positive value, H ($|ELJ + Ehb| - E$) prevents the docking algorithm of calculating negligible interactions of this type. The coefficients α, β(r) and γ are fitted, using a least squares procedure, for reproducing experimental values of $\Delta_a G^0$ obtained for a large number of protein-ligand complexes. In particular, β(r) is a distance dependent function, which accounts for both electrostatic screening and desolvation. The Lennard-Jones (E_i^{LJ}), hydrogen bond (E_i^{hb}) and electrostatic E_i^{el} interactions of the ligand atom iii i with the protein are calculated respectively as:

$$E_i^{LJ} = \sum_{j=1}^{np} 4\varepsilon_{ij}((\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6 \ (24a)$$

$$E_i^{hb} = \sum_{j=1}^{np} w_{ij}(r_{ij} - r_{cut})\, H(r_{cut} - r_{ij}) \quad (24b)$$

$$E_i^{el} = \sum_{j=1}^{np} \frac{q_i q_j}{r_{ij}} \quad (24c)$$

Equation 24.The Lennard-Jones (E_i^{LJ}), hydrogen bond (E_i^{hb}) and electrostatic (E_i^{el}) interactions of a ligand atoms i with the protein.

In general, for these equations, r_{ij} is the distance between the ligand's atom i and the protein's atom j. In equation $24a$, σ_{ij} is the distance for which the Lennard-Jones interaction energy between atoms i and j is null and εij is the symmetrical of the minimum value for this interaction energy. In equation $24b$ $w_{ij}(r_{ij} - r_{cut})$ is the energy of the hydrogen-bond that depends on the nature of the atoms involved and, on the distance, while r_{cut} is the cut-off distance (minimum distance for a non-null hydrogen bond interaction). $H(r_{ij} - r_{cut})$ is the Heaviside step function that imposes this constrain. In equation $24c$ q_i and q_j are the charges of atoms $_i$ and $_j$ respectively.

4. Virtual screening

4.1. Introduction

Virtual screening (VS) is a computational method used in drug discovery to identify potential drug candidates by screening a virtual library of compounds against a target protein of interest. The goal of virtual screening is to identify compounds that have a high likelihood of binding to the target protein, which can then be further tested in experimental assays. It is very useful since it allows to evaluate thousands of compounds in a matter of hours and reduce the number of compounds that have to be synthesized or purchased [127,128].

There are two main types of virtual screening: structure-based and ligand-based. Structure-based virtual screening uses the 3D structure of the target protein to predict

how compounds will bind to it. This can be done using molecular docking, which involves predicting the binding pose and binding energy of a compound to a target protein. The scores obtained are then used to distinguish the ligands that bind strongly to the target from those that do not [116].

Ligand-based virtual screening, on the other hand, uses the properties of known ligands that bind to the target protein to predict the properties of new potential ligands. The main disadvantage of these methods is the need of a significant amount of activity data for the compounds that are studied in order to get reasonable results [116].

4.2. Inverted Virtual Screening

Inverted virtual screening (IVS), also known as reverse virtual screening or target fishing, is a variation of virtual screening that is used in various stages of drug discovery to identify potential targets for a given compound or set of compounds. In this approach, a query ligand is screened against a virtual library of proteins to identify those that have a high likelihood of binding to the compound [129]

The process of inverted virtual screening typically requires a protein database, that is a collection of structures of proteins or active sites and a molecular docking program and involves three steps [129]:

- Docking: the compound of interest is docked into the active site of each protein in the database using molecular docking software.
- Scoring: the docked poses of the compound are then scored to determine the binding affinity between the compound and the protein.
- Filtering: based on the binding affinity scores, a threshold is set, and the proteins that have a binding affinity score above the threshold are considered as potential targets.

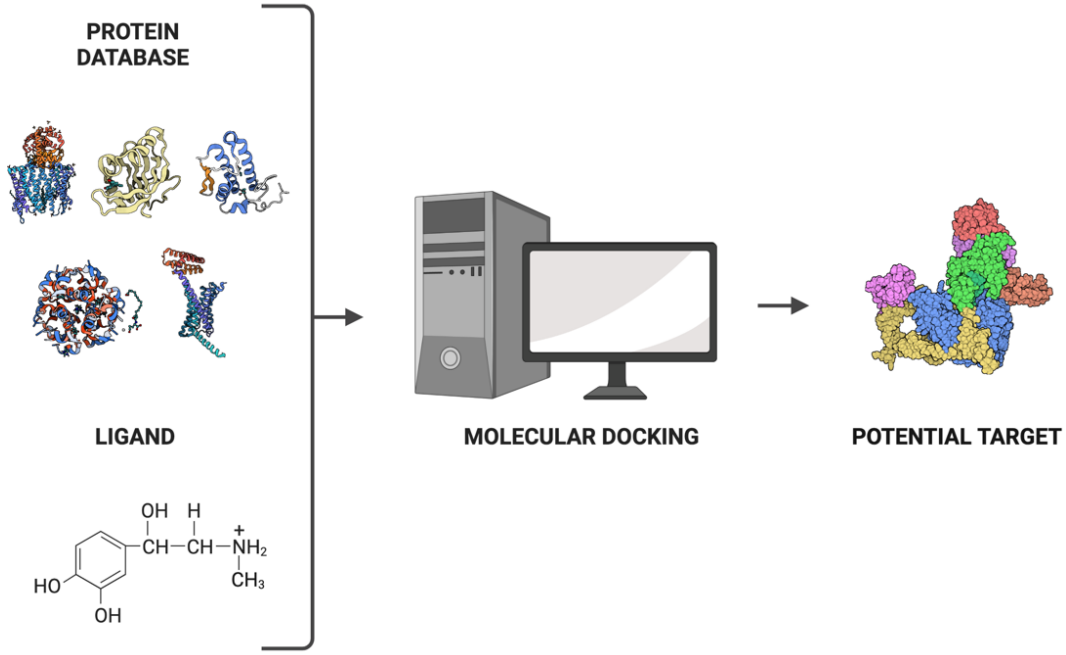


Figure 9. A flowchart of a docking-based inverted virtual screening protocol.

C. RESULTS

The results will be presented and discussed in this section. The chapter begins with a discussion of the research methodology used in this study, as well as the goal and rationale for it. It then presents background, methodology, results, and conclusions for each phase of the study process. Firstly, the results for The Structural Database of Insecticide Targets are presented, followed by the application of an integrated molecular modelling – inverted virtual screening protocol on a collection of eugenol derivatives with confirmed insecticide activity against a molecular library of chosen protein targets typically associated with the insecticide activity of natural compounds.

1. Line of Research

To meet the needs of an exponentially growing population in terms of sustainable food production, an increase in the use of pesticides is inevitable to ensure a greater production and a safe food supply [130,131]. However, despite their beneficial role, some of these agrochemicals have been associated to dangerous characteristics, including carcinogenicity, teratogenicity, high and acute residual toxicity, interference with the hormonal and reproductive systems of mammals and long environmental persistence. Thus, the development of alternative pesticides that are eco-friendly, safe to humans and non-target organisms and that can circumvent the evolution of resistance has been an important topic of research in recent years.

This was the rationale behind this work – to develop new insecticides that could avoid all the prejudicial effects of conventional insecticides. In particular, to provide a strategy that could be used to understand and identify the most likely protein targets of molecules with insecticidal activity.

Two main goals were defined for this project - the development of a structural database of insecticide targets and its application in an integrated inverted virtual screening methodology to a collection of molecules with confirmed insecticide activity in order to identify their most likely protein targets responsible for the insecticide activity.

To accomplish these goals, the project began with a detailed review of the literature available on the most used insecticides. From then on, we built the database, through careful research of structural information on the most relevant protein targets with insecticidal activity. This database will be useful to all researchers in the field.

At the same time, we explored the literature for other virtual screening studies performed on known proteins targets associated to insecticide activity to minimize the candidate pool. Of the 18 studies found in the literature, 14 different protein targets were selected to continue the study. After careful optimization of the VS protocol, the eugenol derivatives were docked into each of these targets with six different scoring functions. The consistency of the scores was evaluated and a ranked list of most likely targets was created to be validated experimentally in the future.

2. The structural database of insecticide targets

2.1. Scope

The Structural Database of Insecticide Targets aims to compile all known protein target structures associated with insecticide activity in order to create a highly useful tool that integrates molecular, biological, and atomic level information that may be used by both experimentalists and theoreticals entirely for free. Currently, the database contains 307 entries, including 31 different protein targets, and 125 molecular ligands from 31 different organisms.

This is an *ongoing* project that will continue to be improved and curated as the number of resolved protein targets directly associated to insecticide action increases.

Figure 10.Scheme of the metrics used during the development of the database.

2.2. Methodology

To lead us through the initial stages of developing the Structural Database of Insecticide Targets, the IRAC Mode of Action Classification Scheme was crucial – it clarified the primary known protein targets of conventional insecticides and was also used as a template for the database architecture.

After intensive manual review of the available literature on the topic, data acquisition began with the direct download of a broad sample of potential pesticide targets structures from PDB. The acquired information was then carefully curated and verified, considering certain criteria, such as the relevance of the target, the source organism and the experimental method used. Subsequently, the significant structures were extracted and organized to be added to the database.

To get further useful information for each item, external links were made to other databases including Brenda, ChEMBL, and UniProt.

As previously mentioned, the database design was sculpted after the IRAC Mode of Action Classification Scheme, which divides protein targets into broad categories based on the physiological functions that they influence. These categories include: Growth and Development, Nerve and Muscle, Respiration, Inhibition of Blood Coagulation and Other.

The figure below represents an outline of the work in progress. The chemical structure of the most used pesticides in the present days is also represented next to their main protein targets.

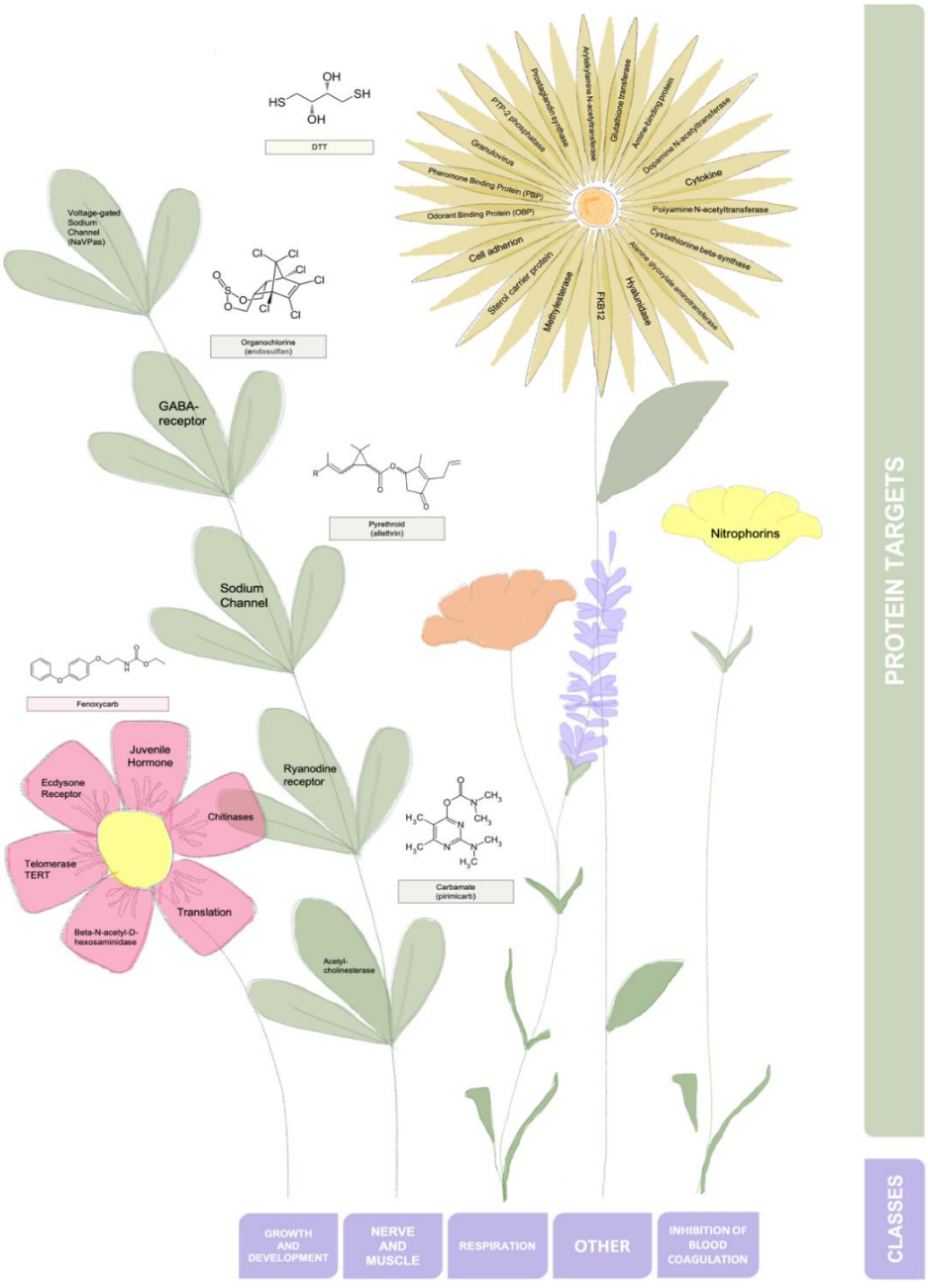


Figure 11. Work in progress – protein targets grouped into broad categories based on the physiological functions that are affected. The chemical structure of the most used pesticides is also represented next to their main protein targets.

The Structural Database of Insecticide Targets is composed by the following sections:

2.2.1. Targeted physiology

This section groups the protein targets based on the physiological functions they affect.

2.2.2. Target

Name of the protein target responsible for the insecticidal activity.

2.2.3. PDB code

4-letter code assigned to each protein's crystallographic structure, as in the Protein Data Bank (PDB).

2.2.4. Experimental Method

Methodology applied to solve the crystallographic structure, including X-Ray diffraction, Nuclear Magnetic Resonance (NMR), and Electron Microscopy.

2.2.5. DOI

DOI of the article describing the associated experimental structure.

2.2.6. Release Date

The year the structure entry was released in the PDB archive.

2.2.7. Mutation(s)

For some structures, only the mutated form of some three-dimensional structures is available. This occurs because enzymes are frequently insoluble in their wild-type form, requiring the application of *in-site* mutagenesis in crucial residue positions. Additionally, some variants are engineered variants that are claimed to affect thermostability and enzymatic activity. This database specifies whether or not we are in the presence of a structure in its mutated form.

2.2.8. Molecular Weight

Mass of the macromolecule.

2.2.9. Resolution

Resolution an important measure of the resolvability in the electron density map of a molecule. The resolution range can influence the geometrical and structural studies meaning that low resolution shows a structure with almost no errors.

2.2.10. Source Organism

This name includes naturally occurring, artificially created, and synthetic source organisms, and it corresponds to the taxonomy identifier given by the PDB depositor.

2.2.11. Macromolecular Name

Scientific name of the macromolecule.

2.2.12. Chain ID

Several information relating the chains of each enzyme (number of chains, length, sequence and missing positions) available on PDB databank are included.

2.2.13. EC Number

Enzyme can be divided according to its Enzyme Commission number (EC number) or by its gene name. Both methodologies are commonly used and were added to the database. The data were obtained in PDB database and Brenda.

2.2.14. Ligand

This section determines whether or not the protein has crystallographic ligands.

2.2.15. Ligand Name

Name of the crystallographic ligand present in the protein.

2.2.16. Ligand SMILES

The simplified molecular-input line-entry system (SMILES) is a specification in the form of a line notation for describing the structure of chemical species using short ASCII strings.

2.3. Results – Current appearance of the Database

Currently, the database created contains X-ray crystallographic data for 307 molecular systems of 31 different protein targets directly associated to insecticide action. Each line corresponds to one entry – a structure characterized by its PDB code. The 20 columns that make up the database are: Targeted Physiology, Target, PDB Code, Experimental Method, DOI, Release Date, Mutations, Resolution, Source Organism, Macromolecule Name, Chain ID, EC Number, Ligand, Ligand Name and Ligand SMILES. It is important noting that some targets have more than one crystallographic ligand. In these cases, each ligand has two columns, one with its name and the other with its SMILES, and they are arranged in decreasing order of molecular weight. In the sample below of the current state of the database, only one ligand per target is shown, although it could contain more. To easily offer more pertinent information for a researcher seeking for a certain protein structure, the columns were carefully chosen and curated.

<table>
<tr><th>TARGETED PHYSIOLOGY</th><th>TARGET</th><th>PDB CODE</th><th>EXPERIMENTAL METHOD</th><th>DOI</th><th>RELEASE DATE</th><th>MUTATIONS</th><th>MOLECULAR WEGHT</th><th>RESOLUTION (Å)</th><th>SOURCE ORGANISM</th><th>MACROMOLECULE NAME</th><th>CHAIN ID</th><th>EC NUMBER</th><th>LIGANDS</th><th>LIGAND NAME</th><th>LIGAND SMILES</th></tr>
<tr><td>Growth and Development</td><td>Ecdysone receptor</td><td>1R20</td><td>X-RAY DIFFRACTION</td><td>10.1038/nature02112</td><td>18/11/03</td><td>no</td><td>61.26</td><td>3.0</td><td>Heliothis virescens</td><td>Ecdysone receptor, Ultraspiracle protein</td><td>D, A</td><td></td><td>yes</td><td>L-ALPHA-PHOSPHATIDYL-BETA-OLEOYL-GAMMA-PALMITOYL-PHOSPHATIDYLETHANOLAMINE, A</td><td>CCCC=CCC=CCCCCCCCC(=O)O[CH](COC(=O)CCCCCCC=CCCC=CCCC=CC)CO[P](O)(=O)OCCN</td></tr>
<tr><td>Growth and Development</td><td>Group I chitinase</td><td>3WL1</td><td>X-RAY DIFFRACTION</td><td>10.1107/S139900471303 3841</td><td>09/04/14</td><td>no</td><td>46.49</td><td>1.8</td><td>Ostrinia furnacalis</td><td>Chitinase</td><td>A</td><td>3.2.1.14</td><td>yes</td><td>2-acetamido-2-deoxy-beta-D-glucopyranose</td><td>CC(=O)N[CH]1[CH](O)O[CH](CO)[CH](O)[CH]1O</td></tr>
<tr><td>Growth and Development</td><td>Beta-N-acetyl-D-hexosaminidase OfHex6</td><td>3NSN</td><td>X-RAY DIFFRACTION</td><td>10.1074/jbc.M110.184796</td><td>24/11/10</td><td>no</td><td>66.38</td><td>2.1</td><td>Ostrinia furnacalis</td><td>N-acetylglucosaminidase</td><td>A</td><td>3.2.1.52</td><td>no</td><td></td><td></td></tr>
<tr><td>Nerve and Muscle</td><td>Acetylcholinesterase</td><td>1QON</td><td>X-RAY DIFFRACTION</td><td>10.1110/ps.9.6.1063</td><td>20/07/00</td><td>no</td><td>66.38</td><td>2.7</td><td>Drosophila melanogaster</td><td>Acetylcholinesterase</td><td>A</td><td>3.1.1.7</td><td>yes</td><td>9-(3-IODOBENZYLAMINO)-1,2,3,4-TETRAHYDROACRIDINE</td><td>Ic1cccc(CNc2c3CCCCc3nc4ccccc24)c1</td></tr>
<tr><td>Nerve and Muscle</td><td>Alpha-esterase-7</td><td>5TYP</td><td>X-RAY DIFFRACTION</td><td>10.1073/pnas.1909130116</td><td>06/12/17</td><td>yes</td><td>66.39</td><td>1.9</td><td>Lucilia cuprina</td><td>Carboxylic ester hydrolase</td><td>A</td><td>3.1.1</td><td>yes</td><td>(3-bromo-4-methylphenyl)boronic acid</td><td>B(c1ccc(c(c1)Br)C)(O)O</td></tr>
<tr><td>Nerve and Muscle</td><td>Voltage-gated sodium channel NavPaS</td><td>6A95</td><td>ELECTRON MICROSCOPY</td><td>10.1126/science.aau2596</td><td>08/08/18</td><td>no</td><td>192.23</td><td>2.6</td><td>Periplaneta americana | Diguetia canities</td><td>Sodium channel protein PaFPC1 | Mu-diguetoxin-Dc1a</td><td>A | B</td><td></td><td>yes</td><td>Sodium ion</td><td>[Na+]</td></tr>
<tr><td>Inhibition of Blood Coagulation</td><td>Nitrophorin</td><td>3TGA</td><td>X-RAY DIFFRACTION</td><td>10.1002/anie.201108691</td><td>30/05/12</td><td>yes</td><td>20.337</td><td>1.3</td><td>Rhodnius prolixus</td><td>Nitrophorin-4</td><td>A</td><td>1.7.6.1</td><td>yes</td><td>PROTOPORPHYRIN IX CONTAINING FE</td><td>CC1=C(CCC(O)=O)C2=Cc3n4[Fe]5|6|N2=C1C=c7n5c(=CC8=N|6C(=Cc4c(C)c3CCC(O)=O)C(=C8C=C)C)c(C)c7C=C</td></tr>
<tr><td>Respiration</td><td>Kinase domain</td><td>5YJ9</td><td>X-RAY DIFFRACTION</td><td>10.1038/s41598-018-28656-8</td><td>20/08/25</td><td>yes</td><td>48.199</td><td>2,5</td><td>Tribolium castaneum</td><td>Serine/threonine-protein kinase PINK1, mitochondrial-like Protein</td><td>D</td><td></td><td>yes</td><td>PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER</td><td>Nc1ncnc2n(cnc12)[CH]3O[CH](CO[P](O)(=O)O[P](O)(=O)N[P](O)(O)=O)[CH](O)[CH]3O</td></tr>
<tr><td>Other</td><td>Sterol Carrier Protein</td><td>4UEI</td><td>SOLUTION NMR</td><td>10.1038/srep18186</td><td>30/12/15</td><td>no</td><td>14.05</td><td></td><td>Helicoverpa armigera</td><td>Sterol carrier protein 2/3-oxoacyl-CoA thiolase</td><td>A</td><td>2.3.1.176</td><td>no</td><td></td><td></td></tr>
<tr><td>Other</td><td>Odorant binding protein</td><td>3K1E</td><td>X-RAY DIFFRACTION</td><td>10.1371/journal.pone.0008006</td><td>08/12/09</td><td>no</td><td>30.59</td><td>1.9</td><td>Aedes aegypti</td><td>Odorant binding protein</td><td>A, B</td><td></td><td>yes</td><td>2,5,8,11,14,17,20,23,26,29,32,35,38,41,44,47,50,53,56,59,62,65,68,71,74,77,80-HEPTACOSAOXADOOCTACONTAN-82-OL</td><td>COCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCO</td></tr>
<tr><td>Other</td><td>Polyphenol oxidase</td><td>1BUG</td><td>X-RAY DIFFRACTION</td><td>10.1038/4193</td><td>02/09/99</td><td>no</td><td>38.815</td><td>2.7</td><td>Ipomoea batatas</td><td>PROTEIN (CATECHOL OXIDASE)</td><td>A, B</td><td>1.10.3.1</td><td>yes</td><td>COPPER (II) ION</td><td>[Cu+2]</td></tr>
<tr><td>Other</td><td>Oxidoreductase</td><td>1YVE</td><td>X-RAY DIFFRACTION</td><td>10.1093/emboj/16.12.3405</td><td>04/09/97</td><td>no</td><td>57.046</td><td>1.65</td><td>Spinacia oleracea</td><td>ACETOHYDROXY ACID ISOMEROREDUCTASE</td><td>A,B,C, D</td><td>1.1.1.86</td><td>yes</td><td>NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE</td><td>c1nc(c2c(n1)n(cn2)C3C(C(C(O3)COP(=O)(O)OP(=O)(O)OCC4C(C(C(O4)N5C=CCC(=C5)C(=O)N)O)O)O)OP(=O)(O)O)N</td></tr>
<tr><td>Other</td><td>N-Acetylglucosamine-1-phosphate uridyltransferase</td><td>2V0K</td><td>X-RAY DIFFRACTION</td><td>10.1110/ps.073135107</td><td>15/01/08</td><td>no</td><td>49.346</td><td>2.3</td><td>Haemophilus influenzae</td><td>BIFUNCTIONAL PROTEIN GLMU</td><td>A</td><td>2, 2.7.7.23, 2.3.1.157</td><td>yes</td><td>URIDINE-5'-DIPHOSPHATE</td><td>C1=CN(C(=O)NC1=O)C2C(C(C(O2)COP(=O)(O)OP(=O)(O)O)O)O</td></tr>
<tr><td>Other</td><td>Octopamine receptor</td><td>4N7C</td><td>X-RAY DIFFRACTION</td><td>10.1016/j.molimm.2014.03.016</td><td>21/05/14</td><td>no</td><td>20.376</td><td>1.75</td><td>Blattella germanica</td><td>Bla g 4 allergen variant 1</td><td>A</td><td>3.5.1.88</td><td>yes</td><td>CITRIC ACID</td><td>C(C(=O)O)C(CC(=O)O)(C(=O)O)O</td></tr>
<tr><td>Other</td><td>Peptide deformylase</td><td>5CY8</td><td>X-RAY DIFFRACTION</td><td></td><td>03/08/16</td><td>no</td><td>19.098</td><td>2.38</td><td>Xanthomonas oryzae pv. oryzae KACC 10331</td><td>Peptide deformylase</td><td>A</td><td></td><td>yes</td><td>(3R)-2,3-dihydro[1,3]thiazolo[3,2-a]benzimidazol-3-ol</td><td>c1ccc2c(c1)nc3n2C(CS3)O</td></tr>
</table>

Figure 12. Work in progress – protein targets grouped into broad categories based on the physiological functions that are affected. The chemical structure of the most used pesticides is also represented next to their main protein targets.

2.4. Conclusion and Future Works

The database is currently being developed and we hope to publish it soon. It will follow the Fair Principles, ensuring accessibility, interoperability, and reusability of the data. Our aim is to provide a valuable tool for visualizing, exploring, and understanding these targets, which may lead to the design and development of more effective insecticides.

New structures are constantly being deposited on the PDB, which makes it crucial to continuously update this tool. Also, in the future, plans exist to add modelled structured based on the AlphaFold, and curated 3D representations of these forms after modelling and molecular dynamics simulations.

3. In silico identification of protein targets associated to the insecticide activity of natural products

3.1. Context

One of the difficulties associated with the rational development of new eugenol derivatives with enhanced insecticidal activity lies in the lack of knowledge of the specific protein target responsible for their activity and to the binding conformation of these molecules. Here, we report the application of an integrated molecular modelling – inverted virtual screening protocol of a collection of eugenol derivatives with confirmed insecticide activity against a molecular library of protein targets typically associated with the insecticide activity of natural compounds. The protocol included six different scoring functions from popular docking software alternatives. The results consistently show a marked preference for interaction of the eugenol derivatives with the odorant binding proteins (OBPs) in insect species.

3.2. Methodology

A search on Scopus was performed for papers describing VS studies, involving targets and molecules with insecticidal and herbicide activity. The selection criteria placed relevance of the target and year of publication. In the eighteen studies found, fourteen targets were identified and listed below:

Table 5. List of targets selected for the inverted virtual screening study.

TARGET	ORGANISM	PDB	RESOLUTION	DESCRIPTION	REFERENCE
Ecdysone receptor	*Heliothis virescens*	1R20	3.00 Å	A receptor-based pharmacophore model was developed using VS based on 1R20 bound to an agonist.	[27]
		1R1K	2.90 Å	2 million different compounds were tested against the 1R1K ecdysone receptor, which was bound to its known ligand, Ponasterone A.	[28]
Chitinase	*Ostrinia furnacalis*	3WL1	1.77 Å	Two crystal structures of chitinases, 3WL1 bound to its reaction product and 3WQV bound to an inhibitor, were used for pharmacophore-based screening.	[29]
		3WQV	2.04 Å		
Beta-N-acetyl-D-hexosaminidase OFHex1		3NSN	2.10 Å	Considering the 3NSN crystal structure coupled to a recognized inhibitor, VS of the ZINC database is used to find OfHex1 inhibitors.	[30]
		3OZP	2.00 Å	3OZP, a crystal structure of OfHex1 coupled to an inhibitor, was targeted by VS from the ZINC database.	[31]
N-Acetylglucosamine-1-phosphate uridyltransferase (GlmU)	*Xanthomonas oryzae*	2V0K	2.30 Å	Docking homology model developed with 2V0K and 2VD4 templates. The crystal structure of 2V0K is attached to a recognized ligand, whereas 2VD4 is bound to a potential inhibitor.	[32]
		2VD4	1.90 Å		
Acetylcholinesterase	*Aedes aegypti*	1QON	2.72 Å	With the crystal structures of acetylcholinesterase 1QON and 4EY6 bound to potential inhibitors, search for novel compounds with insecticidal efficacy against *A. aegypti*.	[33]
		4EY6	2.40 Å		
	Drosophila melanogaster	1DX4	2.70 Å	Instead of utilizing 1DX4 as a template, the homology 3D model was created. A powerful inhibitor is attached to the 1DX4 crystal structure.	[34]
Polyphenol oxidase	*Ipomoea batatas*	1BUG	2.70 Å	Docking simulations utilizing the homologous sweet potato polyphenol oxidase crystal structure in combination with the pesticide phenylthiourea.	[35]

p-hydroxyphenylpyruvate dioxygenase	*Arabidopsis thaliana*	6ISD	2.40 Å	A receptor-ligand pharmacophore model based on the crystal structure of 6ISD bound to a routinely used herbicide was created. The best model developed was then used to VS experiments.	[36]
Oxidoreductase	*Spinacia oleracea*	1YVE	1.65 Å	The crystal structure of a plant oxireductase, 1YVE, coupled to its cofactor, NADPH, was exploited to uncover novel inhibitors in VS tests.	[37]
Voltage-gated sodium channel	*Periplaneta americana*	6A95	2.60 Å	The crystal structure of NavPaS, a voltage-gated sodium channel, linked to tetrodotoxin, a pore blocker (TTX).	
Octopamine receptor	*Blatella germanica*	4N7C	1.75 Å	Octopamine receptor Bla g 4's crystal structure when it is coupled to tyramine.	[38]
Sterol carrier protein-2 (HaSCP-2)	*Helicoverpa armígera*	4UEI	Solution NMR	Structure-based VS. of a database of commercially available compounds to find potential inhibitors of HaSCP-2. For the binding cavity, the residues Phe53, Thr128, and Gln131 were chosen.	[39]
Peptide deformylase	*Xanthomonas oryzae*	5CY8	2.38 Å	A potential inhibitor is linked to the crystal structure of 5CY8 after docking and VS of a library of 318 phytochemicals.	[40]
Alpha-esterase-7 (αE7)	*Lucilia cuprina*	5TYJ	1.75 Å	5TYJ and 5TYP crystal structures are attached to the inhibitors (3-bromo-5-phenoxylphenyl) boronic acid and (3-bromo-4-methylphenyl) boronic acid, respectively. Computational design of effective and selective covalent inhibitors of E7.	[41]
		5TYP	1.88 Å		
Odorant binding protein	*Aedes aegypti*	5V13	1.84 Å	Using the crystal structure of the hormone-binding protein 5V13 of a juvenile mosquito with insecticidal action against *A. aegypti*, search for novel compounds.	[42]
	Drosophila melanogaster	2GTE	1.40 Å	2GTE crystal structure is bound to its natural ligand.	[43]
	Anopheles gambiae	3N7H	1.60 Å	QSAR and docking experiments on the crystal structure for the rational design of insect repellents 3K1E crystal structure attached to a polyethylene glycol molecule; 3N7H crystal structure bound to a common repellent.	[43]
	Aedes aegypti	3K1E	1.85 Å		

Eugenol and eleven derivatives (Figure 13 EU1-EU3e) were selected as new potential insecticides. These molecules have been previously synthesized and validated experimentally with good insecticidal activity.

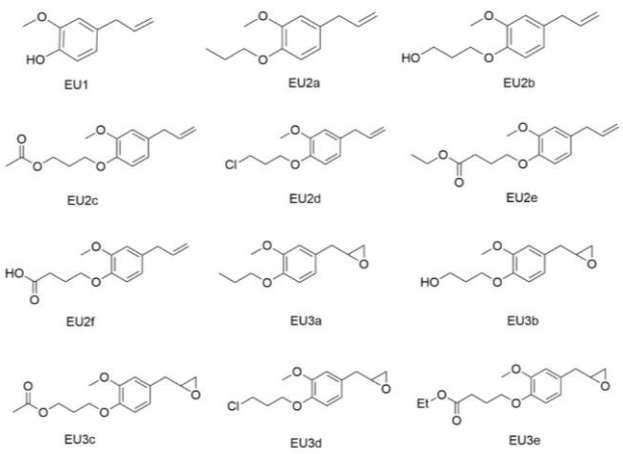


Figure 13. **Eugenol and derivatives used in this study.**

Each Protein Databank (PDB) structure was prepared for docking using the AutoDock Vina plugin for PyMOL. Crystallographic waters and cofactors were removed. The ligands were extracted and saved in separate files to be used for the re-docking and as a reference site for the docking coordinates. When there were no crystallographic ligands present, a selection based on the most important active site residues was made. Re-docking was used to evaluate the ability of the docking software to reproduce the geometry and orientation of the crystallographic pose, as well as the quality of the docking protocol, and to optimize the docking protocol.

The docking programs/scoring functions used were GOLD (PLP, ASP, ChemScore, and GoldScore scoring functions), AutoDock Vina, and LeDock. With each docking program/scoring function, the protocol was optimized for each protein target, to minimize the RMSD in the docking predictions of the reference ligand in redocking, by comparison with the crystallographic structure of the corresponding complex.

The optimized parameters for each program/scoring function were: Vina — docking box position, docking box dimension, exhaustiveness; LeDock — docking box position, docking box dimension; GOLD (PLP, ASP, ChemScore, GoldScore) — binding pocket center, docking region radius, search efficiency, number of runs. The final optimized conditions were used for the subsequent stages. Eugenol and derivatives were prepared

for docking using Datawarrior and OpenBabel and were docked into each structure with the optimized protocol with all the six scoring functions. A ranked list was prepared based on the average scores of each target.

3.3. Results and Discussion

Table 6 presents the average scores obtained for of all the eugenol derivatives for each potential target with each scoring function. The score for all of the GOLD scoring functions is dimensionless, and the higher the score, the better the binding affinity. Vina and LeDock scoring functions, on the other hand, use a metric that is a more precise approximation of binding free energy, so a more negative value means better affinity.

Table 6. Average eugenol derivative scores obtained for all PDB structures with the six different scoring functions.

TARGET	PDB	PLP	ASP	ChemScore	GoldScore	Vina	LeDock
Ecdysone receptor	1R20	57.3	27.5	28.1	52.5	-6.4	-4.7
	1R1K	59.3	26.4	28.3	54.5	-7.1	-5.2
Chitinase	3WL1	63.0	40.8	30.1	60.0	-6.9	-4.8
	3WQV	63.4	40.7	30.6	55.7	-6.5	-4.3
Beta-n-acetyl-d-hexosaminidase ofhex1	3NSN	66.7	46.7	29.1	62.8	-6.1	-4.4
	3OZP	63.3	43.7	28.3	58.7	-7.1	-4.3
N-acetylglucosamine-1-phosphate uridyltransferase (glmu)	2V0K	55.0	24.1	23.3	54.3	-5.9	-4.6
	2VD4	46.9	22.2	21.6	43.8	-5.2	-3.7
Acetylcholinesterase	1QON	73.3	48.2	35.3	62.2	-7.6	-5.0
	4EY6	72.6	41.2	32.4	55.2	-7.1	-5.0
	1DX4	70.0	43.2	32.2	55.3	-7.2	-4.9
Polyphenol oxidase	1BUG	56.7	27.2	25.9	56.2	-5.2	-4.1
P-hydroxyphenylpyruvate dioxygenase	6ISD	57.9	31.6	24.8	47.8	-6.3	-4.3
Oxidoreductase	1YVE	66.0	25.6	32.1	59.5	-6.3	-5.2
Voltage-gated sodium channel	6A95	53.1	23.6	22.3	56.5	-5.8	-4.5
Octopamine receptor	4N7C	68.1	37.9	35.1	65.2	-7.1	-4.5
Sterol carrier protein-2 (hascp-2)	4UEI	54.1	28.2	29.4	45.8	-6.4	-4.9
Peptide deformylase	5CY8	64.0	26.4	24.3	62.5	-6.8	-5.6
Alpha-esterase-7 (αe7)	5TYJ	62.9	34.6	29.3	52.1	-6.4	-4.3
	5TYP	59.9	35.2	29.4	53.1	-6.4	-4.8
Odorant binding protein	5V13	72.1	43.2	35.9	59.4	-7.6	-5.1
	2GTE	63.1	33.8	34.3	56.9	-6.5	-3.1

3N7H	64.8	34.5	28.9	56.6	-6.3	-4.6
3K1E	73.4	39.6	35.8	62.4	-6.0	-5.5

Overall, the results show good consistency, with odorant binding proteins, acetylcholinesterases, octopamine receptors, and chitinases yielding better scores. On the other hand, targets, such as voltage-gated sodium channels, sterol carrier protein-2 (HaSCP-2), and GlmU, are consistently presenting lower scores for all scoring functions.

The structure with the best score was selected for each potential target and they were ranked from the best target to worst, according to the predictions of the different docking programs/scoring functions. The results are listed in Table 7. Globally, considering the results obtained with the several scoring functions, odorant binding proteins are the target with the highest affinity towards eugenol derivatives, followed closely by acetylcholinesterase, chitinases, and octopamine receptors. Enan in 2001 suggested that the insecticidal activity of eugenol was mediated by octopamine receptors. Our study implies that there might be other targets involved as well, as the binding affinity of eugenol derivatives was higher for odorant binding proteins (OBPs) and acetylcholinesterase.

Some variations between the predictions of different scoring functions exist. For example, for the PLP and ChemScore scoring function, odorant binding proteins, and acetylcholinesterase come in first and second as preferable targets for eugenol derivatives. However, for ASP and Vina, the preferable target is the acetylcholinesterase, and for both Vina and LeDock, odorant binding proteins are the second preferable targets. The discrepancy is even higher for GoldScore, with odorant binding proteins coming in third place and octopamine receptors presenting the highest binding affinity for eugenol derivatives. This may be explained by the own nature of each scoring function, as they consider different aspects of protein-ligand binding.

Table 7. Ranking of targets obtained with the different docking programs/scoring functions.

Ranking	PLP	ASP	ChemScore	GoldScore	Vina	LeDock	Overall Ranking

Odorant binding protein	1	4	1	3	2	2	1
Acetylcholinesterase	2	1	2	5	1	5	2
Chitinase	4	2	5	2	6	7	3
Octopamine receptor	3	5	3	1	5	10	4
Peptide deformylase	6	11	12	4	7	1	5
Oxidoreductase	5	12	4	6	11	4	6
β-N-acetyl-D-hexosaminidase OfHex1	7	3	9	7	3	13	7
Ecdysone receptor	9	9	8	10	4	3	8
α-esterease-7	8	6	7	12	9	8	9
Sterol carrier protein-2 (hascp-2)	13	8	6	14	8	6	10
p-hydroxyphenylpyruvate dioxygenase	10	7	11	13	10	12	11
Polyphenol oxidase (PPO)	11	10	10	9	14	14	12
N-acetylglucosamine-1-phosphate uridyltransferase (GlmU)	12	13	13	11	12	9	13
Voltage-gated sodium channel	14	14	14	8	13	11	14

The consistency of the results was visually confirmed by the analysis of the corresponding poses. The hypothesis formed is that eugenol and eugenol derivatives can be used as repellents because they can bind to odorant binding proteins or be used as pesticides, inhibiting insect acetylcholinesterase. As observed in figure 14, they are very different targets, both in size and in function.

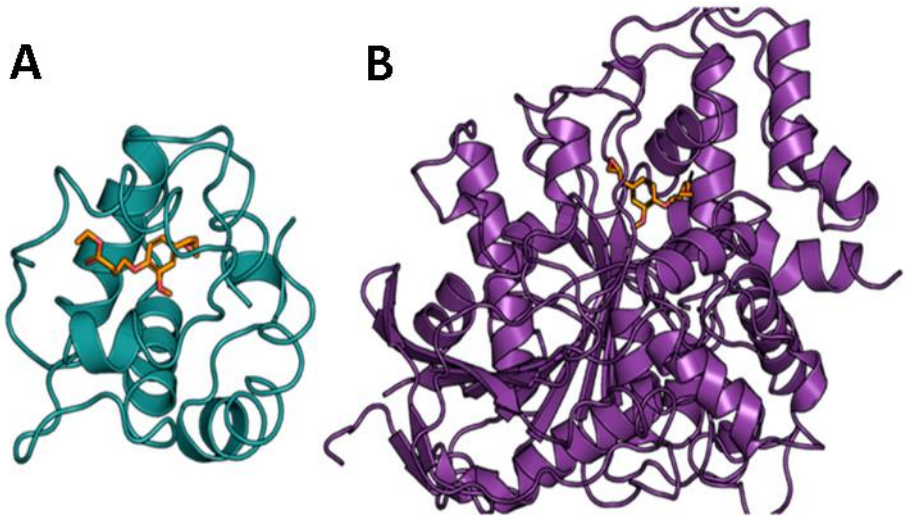


Figure 14. Docking-predicted binding mode of EU3e to OBPs a) and docking-predicted binding modes of EU3e to Acetylcholinesterase b) with PLP scoring function.

Odorant binding proteins (OBPs) are a large family of insect proteins that are crucial for species survival and reproduction, as they use pheromones, plant volatiles, and other odorant molecules to mate, find food, and avoid predators. OBPs are present in a variety of organisms, are highly expressed and highly divergent in sequence. They do however, present a few common features, such as their small size and the presence of six conserved cysteines. These features also make them good targets for rapid screenings. There is not enough consensus regarding the specificity of these proteins and further studies must be performed to better understand the sensitivity of OBPs [131].

Acetylcholinesterase (AChE) is one of the most common targets of synthetic pesticides, such as organophosphates and carbamate and has been a target of reference for over 50 years. This enzyme is a serine hydrolase and is responsible for regulating the levels of acetylcholine in a variety of organisms, from mammals to insects Due to its extensive “attack”, some pests have become resistant to organophosphates, and the search for new and effective alternatives is currently being promoted [132].

Interestingly, during a search in the Protein Data Bank for eugenol, a structure of an odorant binding protein complexed with eugenol was found. It is an OBP of *Apis mellifera* (PDB: 3S0E) that exhibits high affinity for eugenol. This reinforces the proposed

theory that eugenol and derivatives can, in fact, bind to OBPs and could potentially work as repellents. Still, additional computational and experimental studies need to be performed to further optimize and develop this hypothesis.

D. CONCLUSION AND FUTURE WORKS

Insecticides play a crucial role in controlling pest populations, but their increased use has led to the development of resistance in pests and the emergence of adverse effects on the environment and non-target species. As a result, there is a rising demand for the development of alternative pesticides based on natural substances that are potentially less toxic to other organisms and/or biodegradable. In this project, an inverted virtual screening protocol was applied to eugenol derivatives (chemical compounds derived from eugenol, which is a naturally occurring substance found in various plants, including cloves and cinnamon) in order to identify potential protein targets for insecticidal activity.

The study revealed these compounds have increased binding affinity for odorant binding proteins and acetylcholinesterases, suggesting that eugenol derivatives may have potential as repellents, which are an alternative to conventional insecticides that aim to deter pests without causing harm.

The inverted virtual screening protocol used in this study proved to be a simple and effective approach for the identification of new insecticidal targets. By carefully optimizing the protocol and using multiple scoring functions, the results of this study were able to provide a ranked list of the most likely targets for insecticidal activity. These findings provide a valuable starting point for further experimentation and optimization in the development of new insecticides.

In addition, in this project the creation of a large database of insect 3D-structures of protein targets associated with insecticide activity was developed. While in its first steps, this database will constitute an important starting point for future large scale inverted virtual screening projects for the development and identification of insecticides.

Going forward, there are several directions in which this work could be further developed:

1. Experimental validation: The results of this study provide a valuable starting point for further experimentation and optimization in the development of new insecticides or repellents. It would be valuable to validate the findings of this study through experimental means, such as bioassays, to determine the efficacy of eugenol derivatives as insecticides or repellents.

2. Optimization of eugenol derivatives: Based on the results of this study, further optimization of eugenol derivatives may be necessary to improve their efficacy as insecticides or repellents. This could involve modifying the molecular structure of the eugenol derivatives or exploring new derivatives with improved binding affinity for the identified protein targets.

3. Expansion to other insect species: The current study focused on a limited set of insect species. Further work could expand the scope of the study to include other insect species to determine the generalizability of the results and the potential of eugenol derivatives as insecticides or repellents.

4. Application of other virtual screening methods: The current study used an inverted virtual screening protocol, but there are other virtual screening methods that could be applied to eugenol derivatives to identify additional protein targets for insecticidal activity. Further work could explore the application of other virtual screening methods to eugenol derivatives to identify new insecticidal targets.

5. Development of the Insecticide Targets database: The current study includes a limited number of structures and insecticide targets. The recent release of Alphafold and of the Alphafold Database offers high quality 3D structures predicted from artificial intelligence with almost X-ray quality. These structures can be included in the database for species and protein targets not previously represented.

In conclusion, this master's thesis highlights the importance of virtual screening protocols in the development of new and alternative insecticides. The results of this study provide a simple and effective approach for the identification of new insecticidal targets and suggest that eugenol derivatives may have potential as new insecticides or repellents. Overall, this work contributes to the ongoing efforts to find alternative insecticides that are more environmentally friendly and less likely to lead to resistance.

E. REFERENCES

1. Corso, I.C.; Gazzoni, D.L. Sodium Chloride: An Insecticide Enhancer for Controlling Pentatomids on Soybeans. *Pesqui. Agropecu. Bras.* **1998**, *33*, 1563-1571.
2. Wang, C.; Yang, Y.; Wu, N.; Gao, M.; Tan, Y. Combined Toxicity of Pyrethroid Insecticides and Heavy Metals: A Review. *Environ. Chem. Lett.* **2019**, *17*, 1693–1706.
3. Oberemok, V.V.; Laikova, K.V.; Gninenko, Y.I.; Zaitsev, A.S.; Nyadar, P.M.; Adeyemi, T.A. A Short History of Insecticides. *J. Plant. Prot. Res.* **2015**, *55*, 221–226.
4. McLaughlin, G.A. History of Pyrethrum. In *Pyrethrum: The Natural Insecticide*, Casida, J.E., Ed.; Academic Press: New York, NY, USA, 1973, pp. 3-15.
5. Unsworth, J. History of Pesticide Use. Available online: agrochemicals.iupac.org/index.php?option=com_sobi2&sobi2Task=sobi2Details&sobi2Id=31&ItemId=19 / (accessed on 4 October 2022).
6. Davies, T.G.E.; Field, L.M.; Usherwood, P.N.R.; Williamson, M.S. DDT, Pyrethrins, Pyrethroids and Insect Sodium Channels. *IUBMB Life* **2007**, *59*, 151–162.
7. Casagrande, R.A. The Colorado Potato Beetle: 125 Years of Mismanagement. *Bull. Entomol. Soc. Am.* **1987**, *33*, 142–150.
8. Majori, G. Short History of Malaria and Its Eradication in Italy. *Mediterr. J. Hematol. Infect. Dis.* **2012**, *4*, e2012016.
9. Schwardt, H.H. Borax as an Insecticide for Protecting Seed. *J. Econ. Entomol.* **1930**, *23*, 401–404.
10. Tanaka, K. γ-BHC: Its History and Mystery - Why Is Only γ-BHC Insecticidal? *Pestic. Biochem. Physiol.* **2015**, *120*, 91–100.
11. Metcalf, R.L. A Century of DDT. *J. Agric. Food Chem*, **1973**, *21*, 511-519.
12. Nobel Prize Outreach The Nobel Prize in Physiology or Medicine 1948. Available online: https://www.nobelprize.org/prizes/medicine/1948/muller/biographical/ (accessed on 5 October 2022).
13. Thuy, T.T. Effects of DDT on Environment and Human Health. *J. Educ. Hum. Soc. Sci.* **2015**, *2*.
14. Carson, R. *Silent Spring*, Fawcett Publications: Greenwich, CT, USA, 1964.

15. Ullah, S.; Faiz, P.; Aamir, M.; Sabir, M.A.; Mahmood, Q. Occurrence and Spatio-Vertical Distribution of DDT in Soils of Abandoned DDT Factory Area, Amangarh, Pakistan. *SN App.l Sc.i* **2019**, *1*.

16. Matsuo, N. Discovery and Development of Pyrethroid Insecticides. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 2019, *95*, 378–400.

17. Castillo, G.; Barrios-Arpi, L.; Ramos-Gonzalez, M.; Vidal, P.; Gonzales-Irribarren, A.; Ramos-Cevallos, N.; Rodríguez, J.L. Neurotoxicity Associated with Oxidative Stress and Inflammasome Gene Expression Induced by Allethrin in SH-SY5Y Cells. *Toxicol. Ind. Health.* **2022**, *38*, 777-788.

18. Elliott, M. The Pyrethroids: Early Discovery, Recent Advances and the Future. *Pestic. Sci*. **1989**, *27*, 337-351.

19. Fernández-Álvarez, M.; Lores, M.; Llompart, M.; García-Jares, C.; Cela, R. The Photochemical Behaviour of Five Household Pyrethroid Insecticides and a Synergist as Studied by Photo-Solid-Phase Microextraction. *Anal. Bioanal. Chem.* **2007**, *388*, 1235–1247.

20. Kollmeyer, W.D.; Flattum, R.F.; Foster, J.P.; Powell, J.E.; Schroeder, M.E.; Soloway, S.B. Discovery of the Nitromethylene Heterocycle Insecticides. In *Nicotinoid Insecticides and the Nicotinic Acetylcholine Receptor*, Yamamoto, I., Casida, J.E. Eds.; Springer: Tokyo, Japan, 1999, pp. 71–89.

21. Casida, J.E. Neonicotinoids and Other Insect Nicotinic Receptor Competitive Modulators: Progress and Prospects. *Annu. Rev. Entomol.* **2018**, *63*, 125–144.

22. Ihara, M.; Matsuda, K. Neonicotinoids: Molecular Mechanisms of Action, Insights into Resistance and Impact on Pollinators. *Curr. Opin. Insect. Sci.* **2018**, *30*, 86–92.

23. Jactel, H.; Verheggen, F.; Thiéry, D.; Escobar-Gutiérrez, A.J.; Gachet, E.; Desneux, N. Alternatives to Neonicotinoids. *Environ. Int.* **2019**, *129*, 423–429.

24. Mora-Gutiérrez, A.; Rubio, C.; Romero-López, Á.A.; Rubio-Osornio, M. Neurotoxic Effects of Insecticides Chlorpyrifos, Carbaryl, Imidacloprid, in Different Animal Species. In *Neurotoxicity-New Advances*, Sabuncuoglu, S., Ed.; IntechOpen: Rijeka, Croatia, 2021.

25. Knight, J.L.; Weaver, D.F. A Computational Quantitative Structure-Activity Relationship Study of Carbamate Anticonvulsants Using Quantum Pharmacological Methods. *Seizure* **1998**, 7, 347-354.

26. Mustapha, M.U.; Halimoon, N.B.; Lutfi, W.; Johari, W.; Yunus, M.; Shukor, A.; Umar Mustapha, M.; Halimoon, N.; Johar, W.; Yunus, M. An Overview on Biodegradation of Carbamate Pesticides by Soil Bacteria. *Pertanika J. Sci. Technol.* **2019**, *27*, 547–563.

27. Rano, S.; Singh, M. Strategy for the Inspection of Pesticide Residues in Food and Agriculture. In *Sustainable Agriculture Reviews 47*, Inamuddin, Ahamed, M.I., Lichtfouse, E., Eds.; Springer, 2020, pp. 309–333.

28. Macke, J.; Bozhikin, I.; Sarate, J.A.R. Feeding a Growing Population without Deforestation: Agroforestry System Partnerships and Mechanisms. *Agrofor. Syst.* **2021**, *95*, 687–706.

29. Doggart, N.; Morgan-Brown, T.; Lyimo, E.; Mbilinyi, B.; Meshack, C.K.; Sallu, S.M.; Spracklen, D.V. Agriculture Is the Main Driver of Deforestation in Tanzania. *Environ. Res. Lett.* **2020**, *15*.

30. Sparks, T.C. Insecticide Discovery: An Evaluation and Analysis. *Pestic. Biochem. Physiol.* **2013**, *107*, 8–17.

31. Rivero, A.; Vézilier, J.; Weill, M.; Read, A.F.; Gandon, S. Insecticide Control of Vector-Borne Diseases: When Is Insecticide Resistance a Problem? *PLoS Pathog.* **2010**, *6*, 5–6.

32. Göldel, B.; Lemic, D.; Bažok, R. Alternatives to Synthetic Insecticides in the Control of the Colorado Potato Beetle (*Leptinotarsa Decemlineata* Say) and Their Environmental Benefits. *Agriculture* **2020**, *10*, 1–27.

33. Mikolić, A.; Karačonji, I.B. Imidacloprid as Reproductive Toxicant and Endocrine Disruptor: Investigations in Laboratory Animals. *Arh. Hig. Rada. Toksikol* **2018**, *69*, 103–108.

34. Mansour, S.A.; Mossa, A.T.H. Adverse Effects of Exposure to Low Doses of Chlorpyrifos in Lactating Rats. *Toxicol. Ind. Health* **2011**, *27*, 213–224.

35. Uchendu, C.; Ambali, S.F.; Ayo, A.J. The Organophosphate, Chlorpyrifos, Oxidative Stress and the Role of Some Antioxidants: A Review. *Afr. J. Agric. Res.* **2012**, *7*, 2720-2728.

36. Branch, R.A.; Jacqz, E. Is Carbaryl as Safe as Its Reputation? Does It Have a Potential for Causing Chronic Neurotoxicity in Humans? *Am. J. Med.* **1986**, *80*, 659-64.

37. Fattahi, E.; Gholam Ali Jorsaraei, S.; Gardaneh, M. The Effect of Carbaryl on the Pituitary-Gonad Axis in Male Rats. *Iran. J. Reprod. Med.* **2012**, *10*, 419-424.

38. Farag, A.T.; Eweidah, M.H.; El-Okazy, A.M. Reproductive Toxicology of Acephate in Male Mice. Reprod. Toxicol. **2000**, *14*, 457-462.

39. Van Scoy, A.; Pennell, A.; Zhang, X. Environmental Fate and Toxicology of Dimethoate. *Rev. Environ. Contam. Toxicol.* **2016**, *237*, 53-70.

40. Elhamalawy, O.H.; Al-Anany, F.S.; el Makawy, A.I. Thiamethoxam-Induced Hematological, Biochemical, and Genetic Alterations and the Ameliorated Effect of Moringa Oleifera in Male Mice. *Toxicol. Rep.* **2022**, *9*, 94–101.

41. Khan, A.; Khatoon, A. Effect of Sub Lethal Doses of Thiamethoxam (A Pesticide) on Hemato-Biochemical Values in Cockerels. *Pak. Vet. J.* **2017**, *37*, 135–138.

42. Badr, A.M. Organophosphate Toxicity: Updates of Malathion Potential Toxic Effects in Mammals and Potential Treatments. *Environ. Sci. Pollut. Res.* **2020**, *27*, 26036–2657.

43. Ji, Q.; Lee, J.; Lin, Y.H.; Jing, G.; Tsai, L.J.; Chen, A.; Hetrick, L.; Jocoy, D.; Liu, J. Atrazine and Malathion Shorten the Maturation Process of Xenopus Laevis Oocytes and Have an Adverse Effect on Early Embryo Development. *Toxicol. In Vitro***2016**, *32*, 63–69.

44. Çaliskan, M.; Erkmen, B.; Yerli, S. v. The Effects of Zeta Cypermethrin on the Gills of Common Guppy Lebistes Reticulatus. *Environ. Toxicol. Pharmacol.* **2003**, *14*, 117–120.

45. Calderón-Segura, M.E.; Gómez-Arroyo, S.; Cortés-Eslava, J.; Martínez-Valenzuela, C.; Mojica-Vázquez, L.H.; Sosa-López, M.; Flores-Ramírez, D.; Romero-Velázquez, Z.E. In Vitro Cytotoxicity and Genotoxicity of Furia®180 SC (Zeta-Cypermethrin) and Bulldock 125®SC (β-Cyfluthrin) Pyrethroid Insecticides in Human Peripheral Blood Lymphocytes. *Toxicol. Mech. Methods.* **2018**, *28*, 268–278.

46. Yang, Y.; Wu, N.; Wang, C. Toxicity of the Pyrethroid Bifenthrin Insecticide. *Environ. Chem. Lett.* **2018**, *16*, 1377–1391.

47. Li, M.; Zhu, J.; Wu, Q.; Wang, Q. The Combined Adverse Effects of Cis-Bifenthrin and Graphene Oxide on Lipid Homeostasis in Xenopus Laevis. *J. Hazard. Mater.* **2021**, *407*.

48. Zhang, Q.; Zhang, Y.; Du, J.; Zhao, M. Environmentally Relevant Levels of Λ-Cyhalothrin, Fenvalerate, and Permethrin Cause Developmental Toxicity and Disrupt Endocrine System in Zebrafish (Danio Rerio) Embryo. *Chemosphere* **2017**, *185*, 1173–1180.

49. Ben Abdallah, F.; Fetoui, H.; Zribi, N.; Fakhfakh, F.; Keskes, L. Quercetin Attenuates Lambda Cyhalothrin-Induced Reproductive Toxicity in Male Rats. *Environ. Toxicol.* **2013**, *28*, 673–680.

50. Jayaraj, R.; Megha, P.; Sreedev, P. Review Article. Organochlorine Pesticides, Their Toxic Effects on Living Organisms and Their Fate in the Environment. *Interdiscip. Toxicol.* **2016**, *9*, 90–100.

51. Kaushik, P.; Kaushik, G. An Assessment of Structure and Toxicity Correlation in Organochlorine Pesticides. *J. Hazard. Mater.* **2007**, *143*, 102–111.

52. Wolfe, Z.M.; Scharf, M.E. Differential Microbial Responses to Antibiotic Treatments by Insecticide-Resistant and Susceptible Cockroach Strains (*Blattella Germanica* L.). *Sci. Rep.* **2021**, *11*.

53. Zhang, Y.; Meng, X.; Yang, Y.; Li, H.; Wang, X.; Yang, B.; Zhang, J.; Li, C.; Millar, N.S.; Liu, Z. Synergistic and Compensatory Effects of Two Point Mutations Conferring Target-Site Resistance to Fipronil in the Insect GABA Receptor RDL. *Sci. Rep.* **2016**, *6*.

54. Gaio, A.D.O.; Gusmão, D.S.; Santos, A. v.; Berbert-Molina, M.A.; Pimenta, P.F.P.; Lemos, F.J.A. Contribution of Midgut Bacteria to Blood Digestion and Egg Production in Aedes Aegypti (Diptera: Culicidae) (L.). *Parasit. Vectors.* **2011**, *4*.

55. Kirst, H.A. The Spinosyn Family of Insecticides: Realizing the Potential of Natural Products Research. *J. Antibiot.* **2010**, *63*, 101–111.

56. Salgado, V.L.; Sparks, T.C. The Spinosyns: Chemistry, Biochemistry, Mode of Action, and Resistance. In *Comprehensive Molecular Insect Science*, Gilbert. L.I., Latrou, K. and Gill, S.S., Eds.; Elsevier: Oxford, UK, 2005, 137-173.

57. Ujihara, K. The History of Extensive Structural Modifications of Pyrethroids. *J. Pestic. Sci.* **2019**, *44*, 215–224.

58. Clark, J.M.; Symington, S.B. Advances in the Mode of Action of Pyrethroids. *Top. Curr. Chem.* **2012**, *314*, 49–72.

59. Taillebois, E.; Cartereau, A.; Jones, A.K.; Thany, S.H. Neonicotinoid Insecticides Mode of Action on Insect Nicotinic Acetylcholine Receptors Using Binding Studies. *Pestic. Biochem. Physiol.* **2018**, *151*, 59–66.

60. Hollingworth, R.M.; Leister, J.; Ghali, G. Mode of Action of Formamidine Pesticides: an Evaluation of Monoamine Oxidase as the Target*. Chem. Biol. Interact.* **1979**, *24*,35-49.

61. Rathnayake, L.K.; Northrup, S.H. Structure and Mode of Action of Organophosphate Pesticides: A Computational Study. *Comput. Theor. Chem.* **2016**.

62. Kuhr, R.J.; Dorough, H.W. *Carbamate Insecticides: Chemistry, Biochemistry, and Toxicology.*; CRC Press, Inc.: Cleveland, OH, USA, 1976.

63. Auger, P.; Guichou, S.; Kreiter, S. Variations in Acaricidal Effect of Wettable Sulfur on Tetranychus Urticae (Acari: Tetranychidae): Effect of Temperature, Humidity and Life Stage. *Pest. Manag. Sci.* **2003**, *59*, 559–565.

64. Ware, G.W.; Whitacre, D.M. *The Pesticide Book: An Introduction to Insecticides*, Meister Publications: Willoughby, OH, USA, 2004.

65. Pieper, G.R.; Casida, J.E. House Fly Adenosine Triphosphatases and Their Inhibition by Insecticidal Organotin Compounds. *J. Econ. Entomol.* **1965**, *58*, 392–400.

66. Lummen, P. Complex I Inhibitors as Insecticides and Acaricides. *Biochim. Biophys. Acta.* **1998**, *1364*, 287-296*.*

67. Das, S.K. Mode of Action of Pesticides and the Novel Trends – A critical Review. *Int. Res. J. Agric. Soil Sci.* **2013**, *3*, 393-401.

68. Ahmed, N.E.; Asif, M. Various Chemical and Biological Activities of Pyridazinone Derivatives. Eur. J. Exp. Biol. **2017**, *5*, 1–19.

69. Chou, A.P.; Li, S.; Fitzmaurice, A.G.; Bronstein, J.M. Mechanisms of Rotenone-Induced Proteasome Inhibition. *Neurotoxicology* **2010**, *31*, 367–372.

70. Jennifer Mordue, A.; Nisbet, A.J. Azadirachtin from the Neem Tree Azadirachta Indica: Its Action Against Insect*s*. An. Soc. Entomol. Bras. **2000**, 29, 615-632.

71. Price, N.R. The Mode of Action of Fumigants. *J. Stored Prod. Res*. **1985**, *21*, 157-164.

72. Sarwar, M. Inorganic Insecticides Used in Landscape Settings and Insect Pests. *Chem. Res. J.* **2016**, *1*, 50–57.

73. Sullivan, J.J.; Goh, K.S. Environmental Fate and Properties of Pyriproxyfen. *J. Pestic. Sci.* **2008**, *33*, 339–350.

74. Rosell, G.; Quero, C.; Coll, J.; Guerrero, A. Biorational Insecticides in Pest Management. *J. Pestic. Sci.* **2008**, *33*, 103–121.

75. Khater, H.F. Ecosmart Biorational Insecticides: Alternative Insect Control Strategies. In *Insecticides-Advances in Integrated Pest Management*, Perveen F., Ed.; IntechOpen: Rijeka, Croatia, 2012; pp. 17–60.

76. Sun, R.; Liu, C.; Zhang, H.; Wang, Q. Benzoylurea Chitin Synthesis Inhibitors. *J. Agric. Food Chem.* **2015**, *63*, 6847–6865.

77. Joo, S.H.; Keum, Y.S. Oxidative Metabolism of Quinazoline Insecticide Fenazaquin by Aspergillus Niger. *Appl. Biol. Chem.* **2018**, *61*, 681–687.

78. Guo, X.C.; Zhang, Y.H.; Gao, W. bin; Pan, L.; Zhu, H.J.; Cao, F. Absolute Configurations and Chitinase Inhibitions of Quinazoline-Containing Diketopiperazines from the Marine-Derived Fungus *Penicillium Polonicum*. *Mar. Drugs* **2020**, *18*.

79. Tozzi, A. A Brief History of the Development of Piperonyl Butoxide as an Insecticide Synergist. In *Piperonyl Butoxide - The Insecticide Synergist*, Jones, D.G., Ed.; Academic Press, 1998, pp. 1–5.

80. Tak, J.H.; Isman, M.B. Enhanced Cuticular Penetration as the Mechanism for Synergy of Insecticidal Constituents of Rosemary Essential Oil in Trichoplusia Ni. *Sci. Rep.* **2015**, *5*.

81. B-Bernard, C.; Philogfene, B.J.R. Insecticide Synergists: Role, Importance, and Perspectives. *J. Toxicol. Environ. Health* **1993**, *38*, 199–233.

82. Park, N.J.; Oh, S.C.; Choi, Y.H.; Choi, K.R.; Cho, K.Y. Inheritance and Cross Resistance of Phenthoate-Selected Diamondback Moth, *Plutella xylostella* (Lepidoptera: Yponomeutidae). *J. Asia Pac. Entomol.* **2004**, *7*, 233–237.

83. Ahmed, N.; Alam, M.; Saeed, M.; Ullah, H.; Iqbal, T.; Al-Mutairi, K.A.; Salman, M. Botanical Insecticides Are a Non-Toxic Alternative to Conventional Pesticides in the

Control of Insects and Pests. In *Global Decline of Insects*, El-Shafie, H.A.F., Ed.; IntechOpen: Rijeka, Croatia, 2021.

84. Marasigan, K.; Toews, M.; Kemerait, R.; Abney, M.R.; Culbreath, A.; Srinivasan, R. Evaluation of Alternatives to an Organophosphate Insecticide with Selected Cultural Practices: Effects on Thrips, Frankliniella Fusca, and Incidence of Spotted Wilt in Peanut Farmscapes. *J. Econ. Entomol.* **2018**, *111*, 1030-1041.

85. Ben-Issa, R.; Gomez, L.; Gautier, H. Companion Plants for Aphid Pest Management. *Insects* **2017**, *8*.

86. Endersby, N.M.; Morgan, W.C. Alternatives to Synthetic Chemical Insecticides for Use in Crucifer Crops. *Biol. Agric. Hortic.* **1991**, *8*, 33–52.

87. Smith, C.M. Conventional Breeding of Insect-Resistant Crop Plants: Still the Best Way to Feed the World Population. *Curr. Opin. Insect. Sci.* **2021**, *45*, 7–13.

88. Mustafa, M.; Rehman Bilal, A.; Hamza, M.A.; Hamid, A.; Ijaz, M.; Ghilman, S.; Sherazi, M.; Rashid, H.; Bilal, A.R. Transgenic Crops and Sterile Insect Releases Are Eco-Friendly Approaches for the Management of Major Insect Pests of Crops. *Life Sci. J.* **2022**, *19*, 45-51.

89. Robinson, A.S. Mutations and Their Use in Insect Control. *Mutat. Res.* **2002,** *511*, 113-132.

90. Kunimi, Y. Current Status and Prospects on Microbial Control in Japan. *J. Invertebr. Pathol.* **2007**, *95*, 181–186.

91. Guo, J.; Wu, S.; Zhang, F.; Huang, C.; He, K.; Babendreier, D.; Wang, Z. Prospects for Microbial Control of the Fall Armyworm Spodoptera Frugiperda: A Review. *BioControl* **2020**, *65*, 647–662.

92. Bass, C.; Field, L.M. Gene Amplification and Insecticide Resistance. *Pest. Manag. Sci*. **2011**, *67*, 886–890.

93. Crossthwaite, A.J.; Rendine, S.; Stenta, M.; Slater, R. Target-Site Resistance to Neonicotinoids. *J. Chem. Biol.* **2014**, *7*, 125–128.

94. Wilson, T.G.; Ashok, M. Insecticide Resistance Resulting from an Absence of Target-Site Gene Product. *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 14040-14044.

95. Ffrench-Constant, RH. The molecular genetics of insecticide resistance. *Genetics* **2013**,*194*, 807-815.

96. Jensen, S.E. Insecticide Resistance in the Western Flower Thrips, Frankliniella Occidentalis. *Integr. pest manag. rev*. **2000**, *5*, 131–146.

97. Vontas, J.; Katsavou, E.; Mavridis, K. Cytochrome P450-Based Metabolic Insecticide Resistance in Anopheles and Aedes Mosquito Vectors: Muddying the Waters. *Pestic. Biochem. Physiol*. **2020**, 170.

98. Carrasco, D.; Lefèvre, T.; Moiroux, N.; Pennetier, C.; Chandre, F.; Cohuet, A. Behavioural Adaptations of Mosquito Vectors to Insecticide Control. *Curr. Opin. Insect. Sci*. **2019**, *34*, 48–54.

99. Takken, W. Do Insecticide-Treated Bednets Have an Effect on Malaria Vectors? *Trop. Med. Int. Health* **2002**, *7*, 1022–1030.

100. Zalucki, M.P.; Furlong, M.J. Behavior as a Mechanism of Insecticide Resistance: Evaluation of the Evidence. *Curr. Opin. Insect. Sci.* **2017**, *21*, 19–25.

101. Corrêa, A.S.; Pereira, E.J.G.; Cordeiro, E.M.G.; Braga, L.S.; Guedes, R.N.C. Insecticide Resistance, Mixture Potentiation and Fitness in Populations of the Maize Weevil (*Sitophilus Zeamais*). *Crop Protect*. **2011**, *30*, 1655–1666.

102. Foster, S.P.; Tomiczek, M.; Thompson, R.; Denholm, I.; Poppy, G.; Kraaijeveld, A.R.; Powell, W. Behavioural Side-Effects of Insecticide Resistance in Aphids Increase Their Vulnerability to Parasitoid Attack. *Anim. Behav***. 2007**, *74*, 621–632.

103. Balabanidou, V.; Grigoraki, L.; Vontas, J. Insect Cuticle: A Critical Determinant of Insecticide Resistance. *Curr. Opin. Insect. Sci*. **2018**, *27*, 68–74.

104. Gott, R.C.; Kunkel, G.R.; Zobel, E.S.; Lovett, B.R.; Hawthorne, D.J. Implicating ABC Transporters in Insecticide Resistance: Research Strategies and a Decision Framework. *J. Econ. Entomol.* **2017**, *110*, 667–677.

105. Shivanandappa, T.; Rajashekar, Y. Mode of Action of Plant-Derived Natural Insecticides. In *Advances in Plant Biopesticides*, Singh, D., Ed.; Springer: New Delhi, India, 2014, pp. 323–345.

106. Prieto-Martínez, F.D.; López-López, E.; Eurídice Juárez-Mercado, K.; Medina-Franco, J.L. Computational Drug Design Methods—Current and Future Perspectives. In *In Silico Drug Design: Repurposing Techniques and Methodologies*; Elsevier, 2019; pp. 19–44

107. Klebe, G. Recent Developments in Structure-Based Drug Design. *J Mol Med* **2000**, *78*, 269–281.

108. Kenneth M. Merz, Jr.; Dagmar Ringe; Charles H. Reynolds *Drug Design: Structure- and Ligand-Based Approaches*; Cambridge University Press 2010: Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Dubai, Tokyo, 2010;

109. Lu, W.; Zhang, R.; Jiang, H.; Zhang, H.; Luo, C. Computer-Aided Drug Design in Epigenetics. *Front Chem* **2018**, *6*.

110. Tripathi, A.; Bankaitis, V.A. *Molecular Docking: From Lock and Key to Combination Lock*;

111. Du, X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **2016**, *17*.

112. Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G. *Theory of Protein Folding: The Energy Landscape Perspective*; **1997**; Vol. 48;.

113. Bax, H.J.; Keeble, A.H.; Gould, H.J. Cytokinergic IgE Action in Mast Cell Activation. *Front Immunol* **2012**, *3*, doi:10.3389/fimmu.2012.00229.

114. Morris, G.M.; Lim-Wilby, M. *Molecular Docking*;

115. Brás, N.F.; Cerqueira, N.M.F.S.A.; Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Protein Ligand Docking in Drug Discovery. In *Protein Modelling*; Springer International Publishing, **2014**; Vol. 9783319099767, pp. 249–286.

116. Cerqueira, N.M.F.S.A.; Gesto, D.; Oliveira, E.F.; Santos-Martins, D.; Brás, N.F.; Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Receptor-Based Virtual Screening Protocol for Drug Discovery. *Arch Biochem Biophys* **2015**, *582*, 56–67.

117. Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins: Structure, Function and Genetics* **2006**, *65*, 15–26.

118. Sousa, S.F.; Ribeiro, A.J.M.; Coimbra, J.T.S.; Neves, R.P.P.; Martins, S.A.; Moorthy, N.S.H.N.; Fernandes, P.A.; Ramos, M.J. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Curr Med Chem* **2013**, *20*, 2296–2314, doi:10.2174/0929867311320180002.

119. Stanzione, F.; Giangreco, I.; Cole, J.C. Use of Molecular Docking Computational Tools in Drug Discovery. In *Progress in Medicinal Chemistry*; Elsevier B.V., **2021**; Vol. 60, pp. 273–343.

120. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384–13421.

121. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **2009**, NA-NA, doi:10.1002/jcc.21334.

122. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *Development and Validation of a Genetic Algorithm for Flexible Docking*;

123. Mooij, W.T.M.; Verdonk, M.L. General and Targeted Statistical Potentials for Protein-Ligand Interactions. *Proteins: Structure, Function and Genetics* **2005**, *61*, 272–287, doi:10.1002/prot.20588.

124. Korb, O.; Stützle, T.; Exner, T.E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J Chem Inf Model* **2009**, *49*, 84–96, doi:10.1021/ci800298z.

125. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. *Improved Protein-Ligand Docking Using GOLD*; **2003**;

126. Zhang, Y.; Meng, X.; Yang, Y.; Li, H.; Wang, X.; Yang, B.; Zhang, J.; Li, C.; Millar, N.S.; Liu, Z. Synergistic and Compensatory Effects of Two Point Mutations Conferring Target-Site Resistance to Fipronil in the Insect GABA Receptor RDL. *Sci Rep* **2016**, *6*, doi:10.1038/srep32335.

127. Vieira, T.F.; Sousa, S.F. Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. *Applied Sciences (Switzerland)* **2019**, *9*, doi:10.3390/app9214538.

128. Gimeno, A.; Ojeda-Montes, M.J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int J Mol Sci* **2019**, *20*.

129. Xu, X.; Huang, M.; Zou, X. Docking-Based Inverse Virtual Screening: Methods, Applications, and Challenges. *Biophys Rep* **2018**, *4*, 1–16, doi:10.1007/s41048-017-0045-8.

130. Liu, X.; Cao, A.; Yan, D.; Ouyang, C.; Wang, Q.; Li, Y. Overview of Mechanisms and Uses of Biopesticides. *Int J Pest Manag* **2021**, *67*, 65–72.

131. Venthur, H.; Zhou, J.J. Odorant Receptors and Odorant-Binding Proteins as Insect Pest Control Targets: A Comparative Analysis. *Front Physiol* **2018**, *9*.

132. Hu, X.; Yin, B.; Cappelle, K.; Swevers, L.; Smagghe, G.; Yang, X.; Zhang, L. Identification of Novel Agonists and Antagonists of the Ecdysone Receptor by Virtual Screening. *J Mol Graph Model* **2018**, *81*, 77–85, doi:10.1016/j.jmgm.2018.02.016.

133. Liu, J.; Liu, M.; Yao, Y.; Wang, J.; Li, Y.; Li, G.; Wang, Y. Identification of Novel Potential β-N-Acetyl-D-Hexosaminidase Inhibitors by Virtual Screening, Molecular Dynamics Simulation and MM-PBSA Calculations. *Int J Mol Sci* **2012**, *13*, 4545–4563, doi:10.3390/ijms13044545.

F. APPENDICES

1. List of articles

Proceeding paper:

Tatiana F. Vieira, Maria F. Araújo, Maria José G. Fernandes, David M. Pereira, A. Gil Fortes, Elisabete M. S. Castanheira, M. Sameiro T. Gonçalves, Sérgio F. Sousa, "In Silico Identification of Protein Targets Associated to the

Insecticide Activity of Eugenol Derivatives". The 24th International Electronic Conference on Synthetic Organic Chemistry, 2020 doi:10.3390/ecsoc-24-08333

2. Conference communication

Poster presentation:

Maria F. Araújo, Tatiana F. Vieira, Elisabete M.S. Castanheira Coutinho and Sérgio F. Sousa, "Development of a Structural Database of Insecticide Targets". X Bioinformatics Open Days, Braga/Portugal, 2021 May 6th

Maria F. Araújo, Tatiana F. Vieira, Elisabete M.S. Castanheira Coutinho and Sérgio F. Sousa, "Development of a Structural Database of Insecticide Targets". IJUP'2021 - 14º Encontro de jovens investigadores da Universidade do Porto (IJUP), Portugal/Porto, 2021 May 6th.

Development of a Structural Database of Insecticide Targets

Maria F. Araújo[1,2]*, Tatiana F. Vieira[1], Elisabete M.S. Castanheira Coutinho[2] and Sérgio F. Sousa[1]

[1]UCIBIO/REQUIMTE, BioSIM Departamento de Biomedicina, Faculdade de Medicina, Universidade do Porto
[2]Departamento de Física, Escola de Ciências, Universidade do Minho

*mffaraujo@live.com.pt

Biomolecular SIMulations Research Group
info@biosim.pt www.biosim.pt

INTRODUCTION

To meet the needs of an exponentially growing population in terms of sustainable food production, an increase in the use of pesticides is inevitable to ensure a greater production and a safe food supply.[1,2]

However, despite their beneficial role, some of these agrochemicals have been associated to dangerous characteristics, including carcinogenicity, teratogenicity, high and acute residual toxicity, interference with the hormonal and reproductive systems of mammals and long environmental persistence.[3]

Thus, the development of alternative pesticides that are eco-friendly, safe to humans and non-target organisms and that can circumvent the evolution of resistance has been an important topic of research in recent years. This implies understanding the mode of action of conventional pesticides and knowing their targets.[4]

INSECTICIDE TARGETS DATABASE

Development of a database gathering atomic level information on the protein targets directly associated to insecticide action.

METHODOLOGY

- Protein Data Bank
- UniProt
- Brenda
- ChEMBL

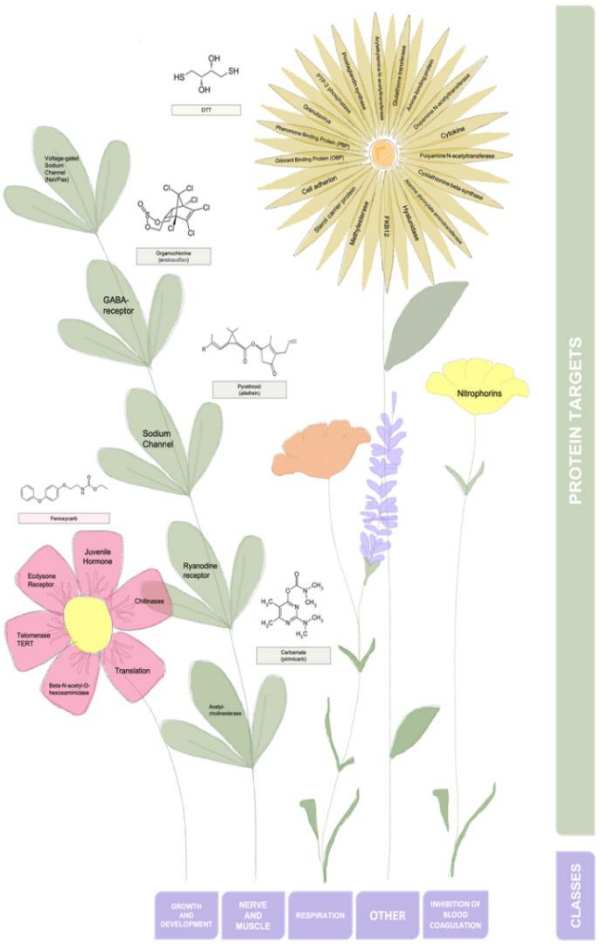
FEATURES

Structure	Resolution	Known substrate
Binding pocket	Key amino acid residues	Known binders

METRICS

307	31	125	31
ENTRIES	TARGETS	LIGANDS	ORGANISMS

WORK IN PROGRESS



This database is organized in a scheme that associates the mode of action in broad categories based on the affected physiological functions.

FUTURE PERSPECTIVES

This is a free access catalogue containing the known structures of all protein targets directly associated to insecticide action. We aim to be a valuable tool to help visualise, explore and understand these targets and possibly lead to the design and development of new and more effective insecticides.

REFERENCES

(1) Mnif, W.; Hassine, A. I. H.; Bouaziz, A.; Bartegi, A.; Thomas, O.; Roig, B. Effect of Endocrine Disruptor Pesticides: A Review. Int. J. Environ. Res. Public Health **2011**, 8 (6), 2265–2303. https://doi.org/10.3390/ijerph8062265.

(2) Umetsu, N.; Shirai, Y. Development of Novel Pesticides in the 21st Century. J. Pestic. Sci. **2020**, 45 (2), 54–74. https://doi.org/10.1584/JPESTICS.D20-201.

(3) Liu, X.; Cao, A.; Yan, D.; Ouyang, C.; Wang, Q.; Li. Y. Overview of Mechanisms and Uses of Biopesticides. Int. J. Pest Manag. **2019**, 0 (0), 1–8. https://doi.org/10.1080/09670874.2019.1664789.

(4) Swale, D. R. Perspectives on New Strategies for the Identification and Development of Insecticide Targets; Elsevier Inc, **2019**; Vol. 161. https://doi.org/10.1016/j.pestbp.2019.07.001.

This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDP/04378/2020, PTDC/ASP-AGR/30154/2017 and UIDB/04378/2020) and was produced with the support of INCD funded by FCT and FEDER under the project 01/SAICT/2016 nº 022153, CPCA/A00/7140/2020 and CPCA/A00/7145/2020.

Supplementary Figure 1. Poster presentation at X Bioinformatics Open Days, Braga/Portugal, 2021 May 6th.

Development of a Structural Database of Insecticide Targets

Maria F. Araújo[1,2*], Tatiana F. Vieira[1], Elisabete M.S. Castanheira Coutinho[2] and Sérgio F. Sousa[1]

[1]UCIBIO/REQUIMTE, BioSIM Departamento de Biomedicina, Faculdade de Medicina, Universidade do Porto
[2]Departamento de Física, Escola de Ciências, Universidade do Minho

*mffaraujo@live.com.pt

Biomolecular SIMulations Research Group
info@biosim.pt www.biosim.pt

INTRODUCTION

To meet the needs of an exponentially growing population in terms of sustainable food production, an increase in the use of pesticides is inevitable to ensure a greater production and a safe food supply.[1,2]

However, despite their beneficial role, some of these agrochemicals have been associated to dangerous characteristics, including carcinogenicity, teratogenicity, high and acute residual toxicity, interference with the hormonal and reproductive systems of mammals and long environmental persistence.[3]

Thus, the development of alternative pesticides that are eco-friendly, safe to humans and non-target organisms and that can circumvent the evolution of resistance has been an important topic of research in recent years. This implies understanding the mode of action of conventional pesticides and knowing their targets.[4]

METRICS

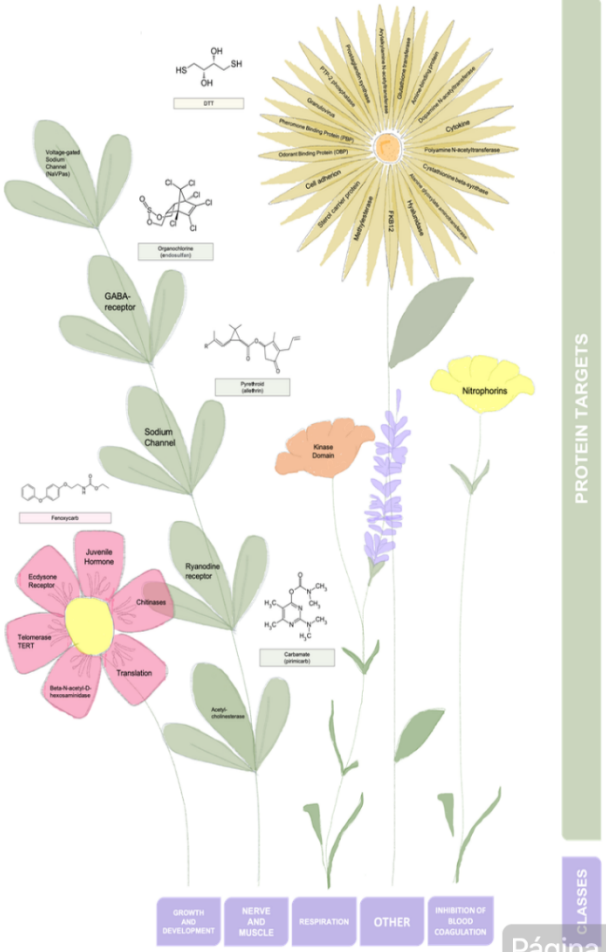
Development of a database gathering atomic level information on the protein targets directly associated to insecticide action.

METHODOLOGY

FEATURES

- Structure
- Resolution
- Binding Pocket
- Known Substrate
- Key Amino Acid Residues
- Known Binders

WORK IN PROGRESS



FUTURE PERSPECTIVES

This is a free access catalogue containing the known structures of all protein targets directly associated to insecticide action. We aim to be a valuable tool to help visualise, explore and understand these targets and possibly lead to the design and development of new and more effective insecticides.

REFERENCES

(1) Mnif, W.; Hassine, A. I. H.; Bouaziz, A.; Bartegi, A.; Thomas, O.; Roig, B. Effect of Endocrine Disruptor Pesticides: A Review. Int. J. Environ. Res. Public Health **2011**, 8 (6), 2265–2303. https://doi.org/10.3390/ijerph8062265.

(2) Umetsu, N.; Shirai, Y. Development of Novel Pesticides in the 21st Century. J. Pestic. Sci. **2020**, 45 (2), 54–74. https://doi.org/10.1584/JPESTICS.D20-201.

(3) Liu, X.; Cao, A.; Yan, D.; Ouyang, C.; Wang, Q.; Li, Y. Overview of Mechanisms and Uses of Biopesticides. Int. J. Pest Manag. **2019**, 0 (0), 1–8. https://doi.org/10.1080/09670874.2019.1664789.

(4) Swale, D. R. Perspectives on New Strategies for the Identification and Development of Insecticide Targets; Elsevier Inc, **2019**; Vol. 131. https://doi.org/10.1016/j.pestbp.2019.07.001.

This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDP/04378/2020, PTDC/ASP-AGR/30154/2017 and UIDB/04378/2020) and was produced with the support of INCD funded by FCT and FEDER under the project 01/SAICT/2016 nº 022153, CPCA/A00/7140/2020 and CPCA/A00/7145/2020.

Supplementary Figure 2. Poster presentation at 14º IJUP, Porto/Portugal, 2021 May 6th.

"Trago dentro do meu coração,

Como num cofre que se não pode fechar de cheio,

Todos os lugares onde estive,

Todos os portos a que cheguei,

Todas as paisagens que vi através de janelas ou vigias,

Ou de tombadilhos, sonhando,

E tudo isso, que é tanto, é pouco para o que eu quero."

- Álvaro de Campos
Passagem das horas