UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Molecular docking as a tool to search vaccine candidates against *Salmonella* targeting poultry.

Trabajo de integración curricular presentado como requisito para la obtención del título de Ingeniero Biomédico.

Autor:

Cynthia Samantha Noble Miranda

Tutor:

PhD. Lenin Ramírez

Co-tutor:

M.Sc. Mariela Pérez

Urcuquí, mayo 2024

AUTORÍA

Yo, **Cynthia Samantha Noble Miranda**, con cédula de identidad 172379297-2, declaro que lasideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absolutaresponsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, mayo de 2024.

Cynthia Samantha Noble Miranda
CI:172379297-2

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Cynthia Samantha Noble Miranda,** con cédula de identidad 172379297-2, cedo a laUniversidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajode titulación no podrá ser cedido a ninguna empresa editorial para su publicaciónu otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación deeste trabajo de integración curricular en el repositorio virtual, de conformidad alo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, mayo de 2024.

Cynthia Samantha Noble Miranda

CI:172379297-2

Agradecimientos

Este trabajo representa no solo el punto culminante de esta etapa de mi vida, sino también el resultado de los desafíos y experiencias que he atravesado para llegar hasta aquí. Reconozco que sin todo lo vivido durante este tiempo, mi persona y trayectoria serían distintas. Por ello, deseo expresar mi profundo agradecimiento a todas las personas que han sido parte de este viaje junto a mí. En primer lugar, agradezco a Dios, pues su presencia y guía me han dado la confianza necesaria para enfrentar cada día con determinación. Sin su influencia, no estaría rodeado por las maravillosas personas que hoy en día forman parte de mi vida.

A mis padres, Teresa y César, quienes han hecho de mí todo lo que soy, porque nunca he podido imaginar unos mejores padres que ellos para mí, por su amor y apoyo. Porque para ellos nunca nada fue demasiado y todo siempre fue poco, porque hicieron de sus hijos personas amorosas, amables, respetuosas y agradables en la vida de los demás.

A mis hermanos, José, Mercedes y Jahir, quienes han hecho de mi vida la más bonita de todas, porque su sola existencia y compañía han calmado todos mis miedos y dudas, porque han estado y sé que siempre estarán conmigo sin importar la situación y la distancia. Siento que las palabras limitan todo lo que siento por ustedes, así que solo quiero decirles que los amo y agradezco su vida.

También quiero agradecer a aquellas personas que llegaron a sentirse como mi familia, a Eduardo por estar, por los consejos y las palabras de aliento. A Eder, por acompañarme estos 5 años ya que tu compañía hizo que la ausencia de mi familia sea llevadera, porque siempre viste en mí cosas que yo no veía y, porque en muchos de los momentos más bonitos de la universidad tú estabas conmigo.

A mis amigos, Karo y Javi, con quienes compartí desde el primer día esta aventura, porque aun cuando no nos conocíamos me sonrieron, acogieron y se quedaron conmigo. Siempre he pensado que las cosas pasan cuando deben pasar y las personas llegan cuando deben llegar, y agradezco por los amigos que más adelante llegaron a Danny, Mafer, Leslie, Alicia, Emilio, Ronny y Kevin. Quiero decirles que fue muy bonito compartir mi felicidad con ustedes y que ustedes también hayan compartido la suya conmigo, anhelo que tengan una vida llena de bendiciones y mucha felicidad, los llevaré siempre en mi corazón.

Finalmente, pero no menos importante, quiero agradecer a mis profesores, Lenin y Mariel. Gracias por siempre estar, por tenerme paciencia y confiar en mí más de lo que alguna vez yo llegué a hacerlo. Siento que sin su apoyo y guía hacer este trabajo habría sido mucho más complicado, quiero agradecerles una vez más por haberme aceptado y entendido durante todo este tiempo.

Una vez más, gracias porque la persona que soy ahora tiene pedacitos de cada uno de ustedes, de su cariño, palabras, risas y locuras.

Samantha Noble.

Dedicatoria

A Dios y mi familia.

Resumen

La *Salmonella* es una bacteria gramnegativa conocida por ser la causante de enfermedades gastrointestinales en animales y humanos. Una parte determinante de la patogenicidad de *Salmonella* es la unión bacteriana a las células o estructuras del huésped, que está mediada por la fimbria de esta, más específicamente por FimH. Determinar dónde se produce la interacción entre la FimH de *Salmonella* y el huésped, en este caso el TLR4 del pollo podría ser decisivo para su control y mitigación de esta bacteria. El objetivo de este estudio es encontrar interacciones que sirvan como posibles epítopos para el desarrollo de una vacuna contra la *Salmonella*. Para ello se realizó un estudio in silico de la estructura. El acoplamiento molecular se realizó en Autodock Vina, el cual generó una serie de tablas con los diez mejores modelos de cada simulación. Como resultado, se determinó que siete de estos modelos podrían usarse como candidatos para el desarrollo de vacunas, antes de pasar a las pruebas experimentales es necesario realizar un análisis más detallado de estos modelos.

Palabras claves: Receptor tipo Toll 4, Fim H, acoplamiento molecular, energía de unión, análisis 2D, interacciones favorables e interacciones no favorables.

Abstract

Salmonella is a gram-negative bacteria known to cause gastrointestinal diseases in animals and humans. A determining part of the pathogenicity of *Salmonella* is bacterial attachment to host cells or structures, which is mediated by *Salmonella* fimbriae, more specifically by FimH. Determining where the interaction between the FimH of *Salmonella* and the host occurs, in this case, the TLR4 of the chicken could be decisive for its control and mitigation. The objective of this study is to find interactions that serve as possible epitopes for the development of a vaccine against *Salmonella*. For this, an in silico study of the structure was carried out. The molecular docking was carried out in Autodock Vina, which as a result produced a series of tables with the ten best models of each simulation. As a result, it was determined that seven of these models could be used as candidates for vaccine development. It is necessary that a more detailed analysis of these models be carried out before moving on to experimental tests.

Keywords: Toll-like receptor 4, Fim H,Molecular docking, Binding energy,2D analysis, favorable interactions and Non favorable interactions.

Contents

List of Figures **xii**

List of Tables **xv**

1 Introduction **1**

2 Theoretical Background **3**

2.1 Computer-aided drug design 6

2.2 Molecular docking bases 7

2.3 Problem Statement 7

2.4 General and Specific Objectives 9

3 Methodology **11**

3.1 Molecular docking 11

3.2 Structure validation 12

3.2.1 Preparation of coordinated files 12

3.2.2 Molecular Docking Procedure and Interaction Analysis 13

3.3 Evaluation parameters 17

3.3.1 Binding energy 17

3.3.2 RMSD (Root Mean Square Deviation) 19

3.3.3 Visual Inspection 19

4 Results & Discussion **21**

4.1 3D Structure Validation 21

4.2 Molecular docking results 23
4.3 2D Model Generation 34
4.3.1 Models selection 34
4.3.2 Models description 37

5 Conclusions **51**

Bibliography **53**

List of Figures

2.1 *Salmonella* type 1 fimbriae (T1F) structure. Adapted from Kolenda et al.[1]. 5

2.2 Workflow for computer-aided drug design.Tacken from Macalino et al.[2] 8

3.1 The 3D structures utilized in the molecular docking simulation include: a) TLR4, featuring a sequence length of 843 amino acids, serving as the receptor; and b) *Salmonella* FimH, with a sequence length of 335 amino acids, employed as the ligand for the simulation. 14

3.2 The workflow diagram outlines the essential steps for conducting the molecular docking simulation. It specifies the three required files and their respective specifications. 16

3.3 The graphical representation illustrates the transitions of the grid during molecular docking on TLR4, as detailed in the accompanying Table3.2. 18

4.1 Chicken TLR4 and *Salmonella* FimH structures evaluation. 22

4.2 Visualization of the best receptor-ligand models based on the binding energy resulted of the molecular docking simulation. Where on the x axis it can be found the number of ligand and on the y axis the binding energy. 24

4.3 Schematic representation of the interaction model between the receptor and the ligands 11(a) and 14(b). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain. 35

4.4 Schematic representation of the interaction model between the receptor and the ligands 16(c) and 17(d). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain. 36

4.5 Schematic representation of the interaction model between the receptor and the ligands 18(e) and 19(f).The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 38
4.6 Schematic representation of the interaction model between the receptor and ligands 21(g) and 23(h). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 39
4.7 Schematic representation of the interaction model between the receptor and the ligands 26(i) and 27(j). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 40
4.8 Schematic representation of the interaction model between the receptor and the ligands 28(k) and 28.1(l). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 41
4.9 Schematic representation of the interaction model between the receptor and the ligands 32(m) and 40(n). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 42
4.10 Schematic representation of the interaction model between the receptor and the ligands 41(o) and 43(p). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 43
4.11 Schematic representation of the interaction model between the receptor and the ligands 44(q) and 46(r). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 45
4.12 Schematic representation of the interaction model between the receptor and the ligands 48(s) and 51(t). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 46
4.13 Schematic representation of the interaction model between the receptor and the ligands 57(u) and 64(v). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 47
4.14 Schematic representation of the interaction model between the receptor and the ligands 65(w) and 65.1(x). The receptor is depicted in green and is referred to as the ‘A’ chain, while the ligand is identified as the ‘B’ chain. 48

4.15 Favorable ligand segments: Ligands are differentiated by different colors. For example, ligand 16 is blue, 19 is magenta, 27 is red, 40 is salmon, 44 is orange, 48 is marine, and L54 is white. 49

List of Tables

3.1 Amino acid subdivision of the FimH ligand. 15

3.2 Coordinates each grid box of size $126 \times 126 \times 126$ used in the molecular docking, to a better understanding it can be seen a graphic representation in Figure 3.3 17

4.1 Shows information on the 24 models considered viable. Important information is also detailed here, such as the ligands involved in the simulation, the affinity energy, the amino acid residues, the body of the grid where the model simulation was carried out, as well as the favorable and not favorable interactions of the model. 26

Chapter 1

Introduction

Food-borne illnesses are considered one of the major public health problems. According to the World Health Organization (WHO), more than 200 diseases have been involved with contaminated food, leading not only to health problems but also the socioeconomic ones. In May 2022 it was reported that at least 1 in 10 people in the world fall ill after eating contaminated food, resulting in the death of 420,000 people and the economic loss of 110 billion dollars a year. Due to its adaptability, profitability, and efficient breeding and production, worldwide, poultry has been seen as the main source of animal protein[3]. A study about antimicrobial resistance in the globalized food chain mentions that in 2022 there was an increase of 2.6% in poultry meat production in contrast to 2019[4]. However, it is this rapid expansion that could have increased the risk of poultry disease transmission. A study that evaluated the increase in foodborne diseases mentions that large-scale production and wide distribution of food is one of the main factors in the appearance of diseases[5]. In Ecuador, poultry production is one of the most important industries. According to the National Corporation of Poultry Farmers of Ecuador (CONAVE), poultry farming contributes 3% to the national gross domestic product, in 2021 the gross value of production was 3.7 billion dollars. They also reported that 15 tons more chicken meat was produced in 2022 than in 2021. With this increase in production, the rate of spread of bacteria and its cultivation also increased, making the use of antibiotics essential that could contribute to antibiotic resistance. Studies in Ecuador have revealed that at least half of producers expose broiler chickens, which are intended for local consumption, to high levels of antibiotics, which has led to resistance to antibiotics, making it difficult to contain the spread of these[6]. Studies carried out in 20 provinces of Ecuador where 383 samples were obtained determined that they were resistant to at least one of the antibiotics[7].

To overcome this wave of antibiotic resistance and contain the spread of bacteria, there has been a need to

highlight the crucial role that vaccines play in preventing infections caused by viruses and bacteria, which directly reduces the use of antibiotics and its resistance. The success of this will depend directly on the type of vaccine to be used, the method of administration, and where they are directed[8]. To contribute to the mitigation of this problem, this work will focus on the in silico study of structures related to the colonization of bacteria in chickens, in this case, *Salmonella Typhimurium*, to find a prospect for the design of a vaccine against *S. Typhimurium* for chickens.

Chapter 2

Theoretical Background

The emergence of new pathogens in humans, domestic animals, wildlife, and plants is inherently unpredictable, as well as their biological invasion dynamics, which remains unknown. An emerging pathogen refers to an infectious disease agent or strain characterized by increasing incidence after first appearance in a new host and, in turn, increasing incidence in an existing host population as a result of infection strain adaptation to survive environmental stress. Among the emerging pathogens, three can be distinguished: *new, evolving* and *emerging foodborne* pathogens[9], these are mainly classified by their appearance and prevalence in the host. *New foodborne* pathogens are those that have not yet been described and therefore represent a serious danger to human health. *Evolving foodborne*are those that become more potent or are associated with other pathogens that have caused disease and have been wrongly attributed to other foodborne pathogens. *Emerging foodborne* pathogens are related to pathogens that have been recently identified, have always been recognized as pathogens for a long time, but have only recently been associated with foodborne transmission. One of the recognized reasons for favoring the potential appearance of pathogens is the change of host species. The host change has had several consequences, including some of the most devastating epidemics, such as HIV/AIDS, rinderpest in African cuds, and COVID-19, among others[10].Some prevalent and serious emerging pathogens of meat, poultry and meat products including *Campylobacter jejuni, Salmonella Typhimurium, Escherichia coli, Mycobacterium avium* subsp.*paratuberculosis*, and prions[9].

Salmonella, considered as one of the major zoonotic pathogens is also considered as the main cause of foodborne illnesses[11].WHO has been reported that *Salmonella* is considered as 1 of the 4 cause of diarrhea diseases globally; as a result, 550 million people become ill each year, which is why it is important to implement, prevent and control

strategies of animal spread disease, which also include vaccine use[3]. *Salmonella* is a gram-negative bacillus belonging to the Enterobacteriaceae family, characterized mainly by being anaerobic and facultative intracellular pathogens, favoring its ability to survive, develop, and multiply inside and outside the host, in the presence or absence of oxygen. To date, two species of *Salmonella* are recognized: *Salmonella enterica* and *Salmonella bongori* each having six and one subspecies, respectively, and enclosing more than 2500 serotypes[12]. *Salmonella enterica* subs. *enterica* serovar *Typhimurium* is the sub-specie with the most incidence in animals and humans-feed illness, becoming one of the more common in the species. It is also known to possess fimbriae, a filament structure present on its surface. In *Salmonella* genus the fimbriae will be determined by the putative fimbrial operons. Studies have shown that *Salmonella enterica* serovar *Entenritis* could contain fourteen of this operons, however only a few of them will be expressed. In *Salmonella* serovar *Enteritis* there are present SEF21, SEF17, SEF14, LPF and PEF fimbriae enconded by *fim, agf, sef, lpf and pef* operons, respectively[13]. Their mechanism is similar to the enzymatic mechanism, of the horseshoe-key type, which is important to determine the host and the specific organotropism of the bacterium. The most common fimbriae expression is SEF21, also known as type 1 fimbriae, which, as mentioned before, is the form of a six-gene operon (*fimAICDHF)* encoding structural and assembly components and three regulatory genes, *fimZ, fimY and fimW*[14]. FimA protein subunits are the main component of type 1 fimbriae, followed by FimC and FimD subunits, which are used to assemble fimbriae on the cell surface. However, the adhesive capability of the fimbriae is given by FimH[15]. FimH is a lectin-like adhesin located at the tip of the fimbriae, which is mainly involved in the adhesion to the host cell[13] and then immune system activation (Figure 2.2).

Some studies have shown that *Salmonella* fimbria can produce an immune response in the host due to interaction with immune cells such as macrophages and dendritic cells[16]. It has also been mentioned that it could act as an antigen, studies have shown that fimbria can activate the Toll-like receptor on dendritic cells and macrophages, which could lead to the production of inflammatory cytokines and also to the activation of T cells[17].

Pathogen-associated molecular patterns (PAMPs) are components of bacteria and viruses, induce the expression of mediators by several types of cells, including macrophages, and influence the host immune system[18]. PAMPs include cell wall components derived from Gram-positive bacteria, lipopolysaccharides (LPS) from Gram-negative bacteria, lipoteichoic acid, flagella, and fimbriae, all of which are recognized by pattern recognition receptors such as Toll-like receptors (TLRs).

Toll-like receptors (TLRs) family plays a crucial role in the innate immune response by recognizing PAMPs. Toll-like receptor 4 (TLR4) is commonly found in the cell membrane of the intestine. It is primarily characterized

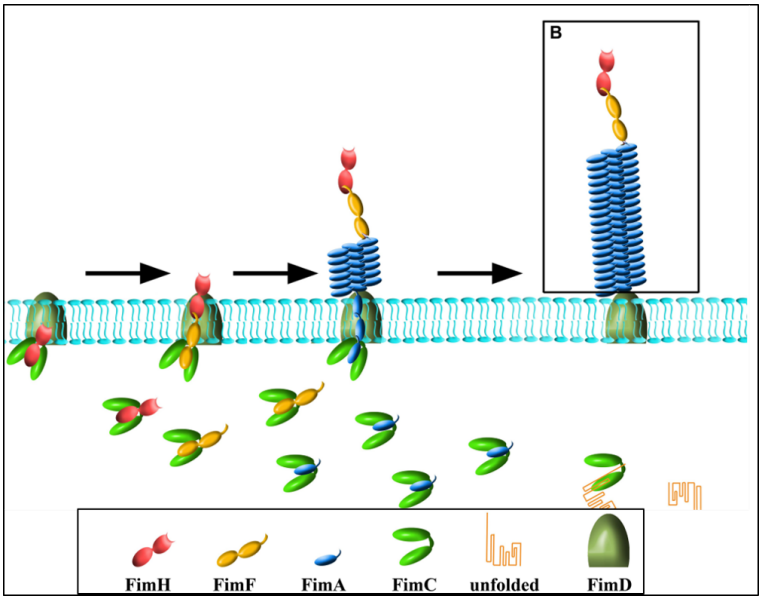


Figure 2.1: *Salmonella* type 1 fimbriae (T1F) structure. Adapted from Kolenda et al.[1].

for recognizing lipopolysaccharides (LPS) from Gram-negative bacteria and, nowadays, is also involved in the interaction with other molecules, including certain host-derived proteins.

Currently, multiple tactics have been implemented to combat diseases in animals destined for production. These range from cleansing measures to the use of additives designed to optimize gastrointestinal health and strengthen immune defenses[19]. In terms of hygiene, sanitation procedures are applied in breeding sites and food processing plants to reduce microorganisms and block the transmission of infectious agents. The use of probiotics and prebiotics has proven to be a valuable tactic, promoting healthier digestion[20]. On the other hand, it has been recognized that beta-glucan plays an essential role in directly activating the immune defenses of animals, thus offering an extra level of protection against diseases[21]. Although these methodologies have been beneficial, they still face obstacles such as antibiotic resistance. For this reason, vaccines are also essential to preserve animal health. Historically, vaccines have been provided to prevent common pathologies in livestock, such as foot-and-mouth disease or avian flu, ensuring direct defense against these diseases[22]. The integration of vaccines into a comprehensive disease control strategy is vital to minimize the possibility of contagion and ensure the well-being of animals within production systems.

Over the years, vaccines have been seen as a key strategy to prevent and control the spread of animal diseases. There are studies on protein-protein coupling with TLRs that reveal their effectiveness against *Salmonella*[23]. How-

ever, the development of new *Salmonella* vaccines still faces limitations and challenges during and after production. The insufficient knowledge of the resistant mechanism and long-term protective immunity are the main challenges to overcome, along with the emergence of new *Salmonella* serovars.Thus, computational methods have been considered a promising option for designing vaccines against the evolving threats that Salmonella poses to the chicken population. Among these methods, molecular docking and molecular dynamics simulations have become indispensable tools to accurately predict antigen-antibody interactions and the dynamic behavior of vaccine candidates.

2.1 Computer-aided drug design

In today's pharmaceutical industry, drug design stands out as one of the most expensive, demanding, and slow processes. The emergence of new pathogens poses a significant challenge to society, as the likelihood of acquiring an illness or condition has increased, and effective tools to prevent or combat these infections are often lacking. In response to these challenges, Computer-Aided Drug Design (CADD) has emerged as a promising tool for the discovery, design, and development of drugs, effectively reducing drug discovery and development costs by up to 50%.

CADD, through virtual screening, primarily focuses on predicting the potential of a compound or agent after interacting with a specific biological target. The goal is to identify a reduced number of compounds that can then undergo experimental evaluation. CADD evaluation can be approached in two ways: 1) Structure-Based Drug Design (SBDD) and 2) Ligand-Based Drug Design (LBDD). SBDD relies on the availability of the target's structure, and approaches such as virtual screening, structure-based pharmacophore modeling, and molecular docking are commonly employed. On the other hand, LBDD approaches center around similarity-based screening, including ligand-based pharmacophore modeling, scaffold hopping, and quantitative structure-activity relationship (QSAR) studies[2].

As previously mentioned, SBDD requires the 3D structure of the target, also known as the receptor. It is crucial to assess whether the biological behavior of the receptor is influenced by the binding of a molecule. Additionally, consideration must be given to the binding pocket of the target, which should not be excessively deep, large, or possess a highly charged binding pocket, as this could hinder interactions with molecules. Obtaining 3D structures involves using resources such as the Protein Data Bank (PDB), (http://www.rcsb.org/pdb), particularly when X-ray crystallography or nuclear magnetic resonance (NMR) methods are employed. In cases where computational approaches are used to determine protein structures, homology modeling using software like MODELLER, Swiss-Model, Phyre2,

I-TASSER, etc., is common[24].

2.2 Molecular docking bases

Considered a novel technology in drug design, molecular docking is a structure-based drug design method that, by simulating optimal conformations, can predict and determine the binding affinity and interactive mode of a receptor-ligand complex. In this method, a protein or nucleic acid serves as the receptor, while a small molecule or even another protein may function as the ligand. The location and orientation of the binding site can be influenced by protein flexibility[25].

The modeling of molecular interactions is inherently complex due to the various forces involved, such as hydrophobic interactions, van der Waals forces, hydrogen bonding, electrostatic forces, and stacking interactions between aromatic amino acids. The goal of receptor-ligand interaction is to predict the orientation of the molecule that forms a stable complex when it binds to another. Different methods can be employed to evaluate the effectiveness of docking performance, as mentioned earlier. Binding affinity is commonly assessed due to its importance in establishing the binding site. Other methods involve comparing values such as real space R-factor (RSR) and root-mean-square distance (RMSD)[26].

2.3 Problem Statement

Salmonella infection has been a challenge in the Ecuadorian poultry industry. The complications at both a health and economic level have been significant. The increase in demand for poultry products has forced industries to increase the poultry population, which has contributed to the expansion, contamination, and antimicrobial resistance. Faced with these challenges, the need to find a way to help control and mitigate how this bacteria has increased. The in silico study of the structures involved in the infection and colonization of *Salmonella* in chickens can contribute to the development of vaccines, which can be used as a strategy for the control and mitigation of *Salmonella* in the poultry population, in turn, it would mitigate economic problems in the industry as well as human and animal health.

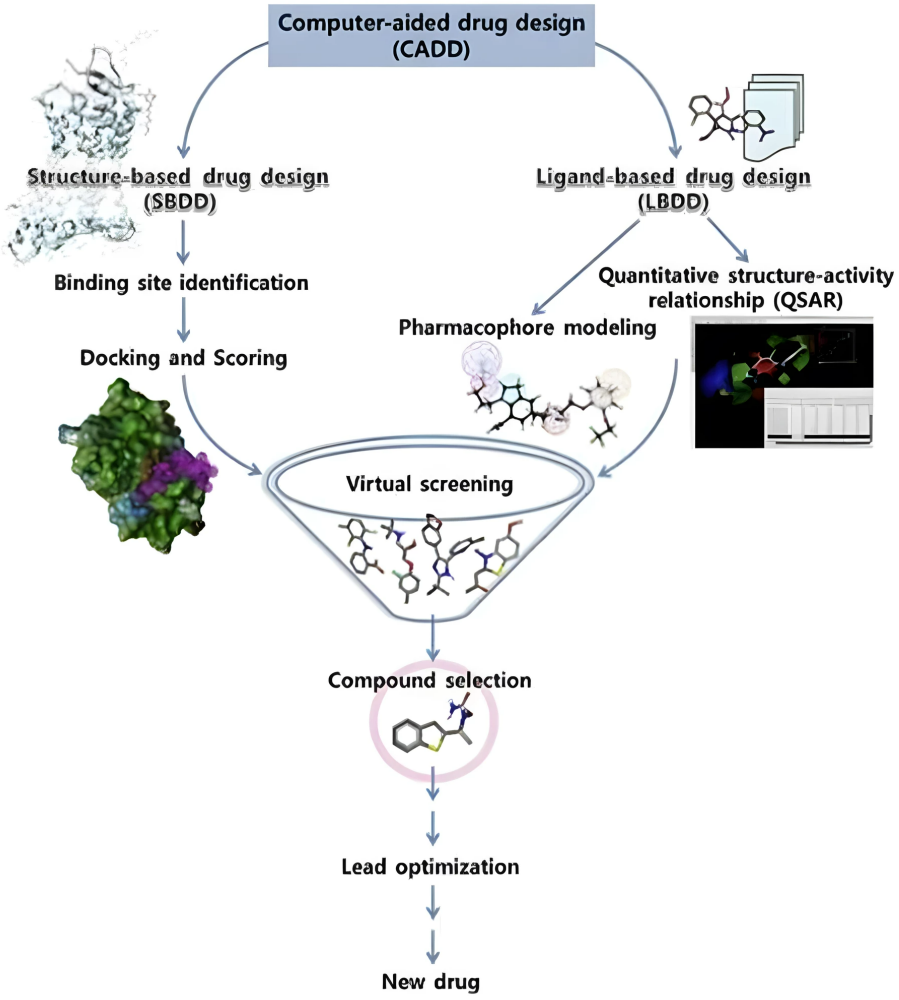


Figure 2.2: Workflow for computer-aided drug design.Tacken from Macalino et al.[2]

2.4 General and Specific Objectives

The main objective of this work is to find possible epitopes that can be used as a candidate for a vaccine against *Salmonella enterica* subs. *enterica* serovar *Typhimurium* for veterinary use by carrying out an in silico study of the structures between chicken TLR4 and *Salmonella enterica* subs. *enterica* serovar *Typhimurium* FimH, using molecular docking.

Specific objectives of this work include:

- Predict the possible binding sites between the Fim H of *Salmonella enterica* subs. *enterica* serovar *Typhimurium* and the Toll-like receptor 4 of the chicken, through simulations carried out in Autodock Vina.
- Determine if the models resulting from the simulation between *Salmonella enterica* subs. *enterica* serovar *Typhimurium* Fim H and the Toll-like receptor 4 are viable.
- Examine the interactions that occur in each model and thereby determine which models are viable to be used as a vaccine candidate against *Salmonella enterica* subs. *enterica* serovar *Typhimurium*.

Chapter 3

Methodology

This chapter provides a detailed account of the procedures undertaken during the development of this study. It outlines the essential files and parameters required for the molecular docking of the structures.

3.1 Molecular docking

Two main software tools were utilized to conduct the molecular docking: Autodock Tools for preparing the PDBQT files, and Autodock Vina for performing the molecular docking simulation. The initial step in the simulation process involved identifying the structures that would serve as the receptor and ligand. After identifying these structures, prerequisite files were generated to initiate the simulation, specifically, extended PDB files (PDBQT) for both the receptor and the ligand, along with the configuration file. Figure 3.2 shows a general workflow of the process steps used in this work. To obtain the 3D structures of the receptor and the ligand, homology modeling of the structures was carried out, this mainly focuses on building 3D protein models from the amino acid sequence, which is aligned with a similar protein structure of which its 3D structure is already known[27]. To carry out the homology modeling of the structures, the SWISS MODEL workspace (https://swissmodel.expasy.org/interactive) was used; For this, it was important to consider the percentage of sequence similarity between the structure of the template and those to be modeled. According to the literature it is considered that if the percentage of similarity between the sequences falls below 30% the estimated 3D structure could be unviable since they become more dispersed. On the other hand if this percentage is greater than 50% the obtained 3D structure is reliable enough to be used in the study[27]. To achieve this, the amino acid sequence was initially uploaded in FASTA format. Once uploaded, the structures are

built in 3D. The obtained structures are the ones that will be analyzed and compared against the acceptance criteria.

3.2 Structure validation

For structures modeled in the SWISS MODEL workspace, it is important to consider the following parameters[28][29]:

MolProbity Score: This measure informs us about the quality of the protein structure. A value less than 1.5 is considered very good, while values between 1.5 and 2 are deemed acceptable.

Clash score: This parameter assesses the aesthetic clashes between the atoms of the structure. A score less than 5 is considered good, with lower scores indicating better quality.

Rachamandran Favoured: This parameter reflects information about favorable regions, mainly considering their steric interactions and conformations.

Rachamandran Outliers: This parameter speaks to the regions where the dihedral angles phi and psi of the structure are not compatible with the steric constraints.

Rotamer outliers: This value primarily describes the preferred conformations of the dihedral ring of the side chains. Ideally, the values should be below 1-2 %.

C-beta deviations: This value indicates deviations in the positions of the C-beta atoms in the side chains. For this parameter, lower values are preferable.

Bad bonds and bad angles: These parameters refer to the geometry of the structure. The ideal value for both parameters is zero; hence, the lower the value, the better, considering zero as the best value.

Cis Prolines: This measurement indicates the percentage of proline residues in the cis conformation. Generally, a low percentage is considered ideal.

Twisted Non-Proline: It evaluates the twisting of non-proline peptide bonds. Similar to the previous parameter, a low percentage is better, and a value of 1 to 2 % is considered acceptable.

3.2.1 Preparation of coordinated files

Protein Structure Preparation

The 3D structure of the protein was modeled using SWISS-MODEL workspace, utilizing the FASTA file of Toll-like receptor 4 (TLR4) from the organism *Gallus gallus* (accession number AFK08535.1. Sequence length: 843 a.a.). First, the program will search for templates for modeling, then align the target sequence to finally build the model. Homology modeling was performed with *Perdix perdix*(Gray partridge) as a template, which exhibited 94.42%

sequence identity(Figure 3.1a).

Once the 3D structure of the protein was obtained, the PDBQT file of the protein was generated. Only the extracellular part was taken, which is found in amino acids 23 to 641. This file was instrumental for the molecular docking, involving the addition of necessary molecules and the removal of unnecessary ones. Autodock Tools facilitated this step by incorporating functions that streamlined the protein preparation process. Specifically, all hydrogens were added, with subsequent removal of only the non-polar hydrogens. Additionally, Kollman and Computer Gasteiger charges were included. It is imperative to verify the absence of water molecules in the structure; however, in this case, their elimination was unnecessary as the 3D structure inherently lacked water molecules due to its nature.

Ligand Structure Preparation

The 3D structure of both the ligand and the protein was acquired through 3D modeling in SWISS-MODEL workspace. The FASTA file of Type 1 fimbrin D-mannose specific adhesin (precursor), also known as FimH, from *Salmonella enterica* subs. *enterica* serovar *Typhimurium*. str. LT2 (primary accession number, P37925; Sequence length: 335 aa) was employed (Figure 3.1b).

In the PDBQT file of the ligand, the optimal root atoms were primarily established, and torsion numbers were assigned. These steps enabled the determination of the final total number of active torsions.

Configure file

This file encompasses determinant information for the molecular docking simulation including the input receptor file, grid box specification parameters, and also the numbers of runs and energy range. The grid box determines the specific site where molecular docking will take place. To mitigate computational costs, six docking boxes were employed. Although all of them shared the same size, they had different coordinates, as illustrated in Table3.2.

3.2.2 Molecular Docking Procedure and Interaction Analysis

Molecular docking was conducted using Autodock Vina, a software that allows ligand flexibility, enabling exploration of various ligand orientations and conformations. Six simulations were performed per ligand with a spacing of 1

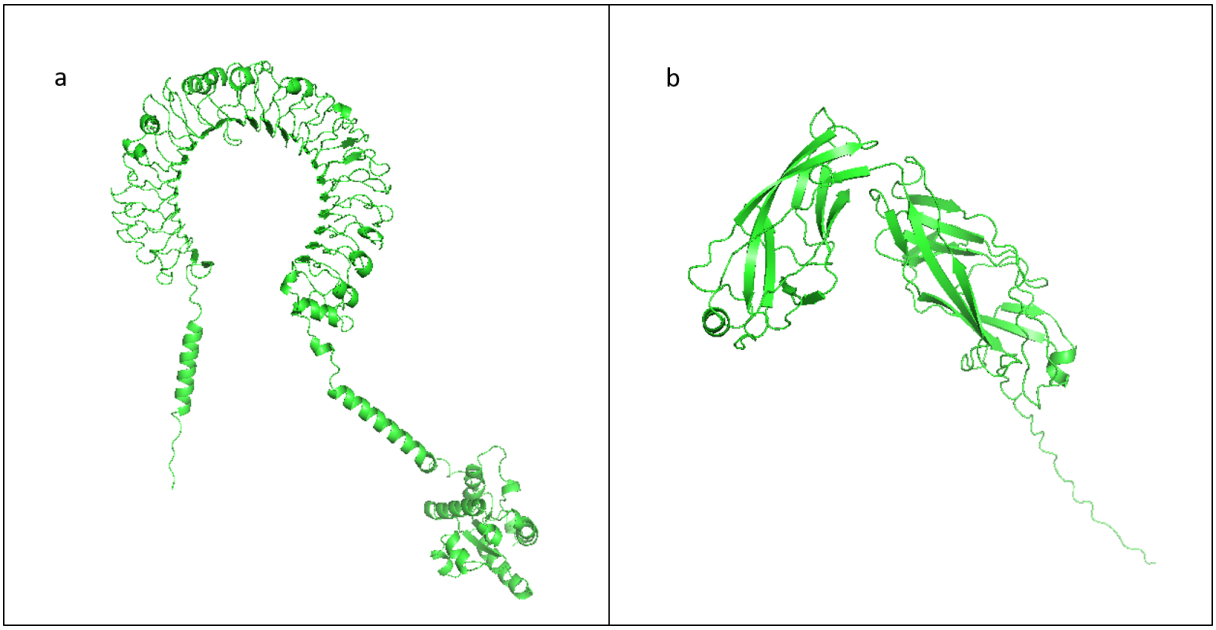


Figure 3.1: The 3D structures utilized in the molecular docking simulation include: a) TLR4, featuring a sequence length of 843 amino acids, serving as the receptor; and b) *Salmonella* FimH, with a sequence length of 335 amino acids, employed as the ligand for the simulation.

Table 3.1: Amino acid subdivision of the FimH ligand.

Ligand	**Ligand sequence**	**Ligand**	**Ligand sequence**	**Ligand**	**Ligand sequence**
Ligand 1	MKIYSALLLA	Ligand 23	ITDSVAGVFYP	Ligand 45	FSQAGAGNRPQ
Ligand 2	SALLLAGTALF	Ligand 24	AGVFYPPRNYI	Ligand 46	AGNRPQGVTPQ
Ligand 3	AGTALFFTHPA	Ligand 25	PPRNYILMGVD	Ligand 47	QGVTPQTKTIA
Ligand 4	FFTHPALATVC	Ligand 26	ILMGVDYNVSQ	Ligand 48	QTKTIAIKCTN
Ligand 5	ALATVCRNSNG	Ligand 27	DYNVSQQKPFG	Ligand 49	AIKCTNVAAQA
Ligand 6	CRNSNGTATDI	Ligand 28	QQKPFGVQDSK	Ligand 50	NVAAQAYLSMR
Ligand 7	GTATDIFYDLS	Ligand 29	GVQDSKLVFKL	Ligand 51	AYLSMRLEAEK
Ligand 8	IFYDLSDVFTS	Ligand 30	KLVFKLKVIRP	Ligand 52	RLEAEKASGQA
Ligand 9	SDVFTSGNNQP	Ligand 31	LKVIRPFINMV	Ligand 53	KASGQAMVSDN
Ligand 10	SGNNQPGQVV	Ligand 32	RPFINMVTIPR	Ligand 54	AMVSDNPDLGF
Ligand 11	PGQVVTLPEKS	Ligand 33	MVTIPRQTMFT	Ligand 55	NPDLGFVVANS
Ligand 12	TLPEKSGWVGV	Ligand 34	RQTMFTVYVTT	Ligand 56	FVVANSNGTPL
Ligand 13	SGWVGVNATCP	Ligand 35	TVYVTTSTGDA	Ligand 57	SNGTPLTPNNL
Ligand 14	VNATCPAGTTV	Ligand 36	TSTGDALSTPV	Ligand 58	LTPNNLSSKIP
Ligand 15	PAGTTVNYTYR	Ligand 37	ALSTPVYTISY	Ligand 59	LSSKIPFHLDD
Ligand 16	VNYTYRSYVSE	Ligand 38	VYTISYSGKVE	Ligand 60	PFHLDDNAAAR
Ligand 17	RSYVSELPVQS	Ligand 39	YSGKVEVPQNC	Ligand 61	DNAAARVGIRA
Ligand 18	ELPVQSTEGNF	Ligand 40	EVPQNCEVNAG	Ligand 62	RVGIRAWPISV
Ligand 19	STEGNFKYLKL	Ligand 41	CEVNAGQVVEF	Ligand 63	AWPISVTGIKP
Ligand 20	FKYLKLNDYLL	Ligand 42	GQVVEFDGGDI	Ligand 64	VTGIKPAEGPF
Ligand 21	LNDYLLGAMS	Ligand 43	FDFGDIGASLF	Ligand 65	PAEGPFTARGY
Ligand 22	LGAMSITDSVA	Ligand 44	IGASLFSQAGA	Ligand 66	TARGYLRVDYD

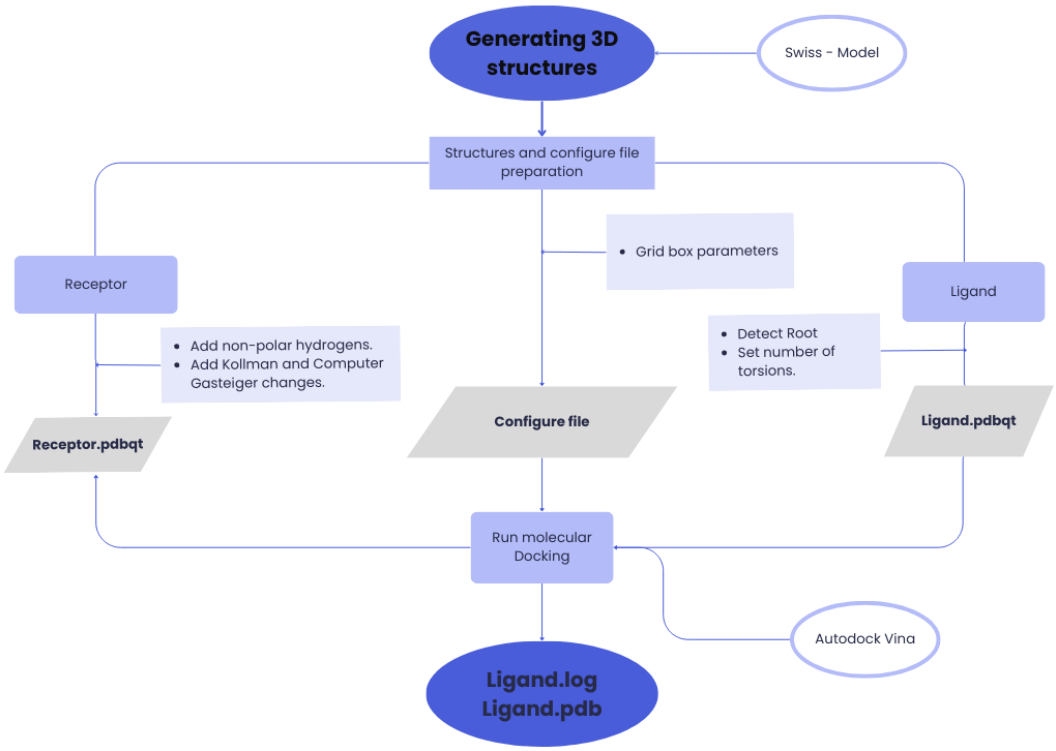


Figure 3.2: The workflow diagram outlines the essential steps for conducting the molecular docking simulation. It specifies the three required files and their respective specifications.

Table 3.2: Coordinates each grid box of size $126 \times 126 \times 126$ used in the molecular docking, to a better understanding it can be seen a graphic representation in Figure 3.3

Grid Box	**1**	**2**	**3**	**4**	**5**	**6**
x	-32.097	-32.097	-19.637	19.637	19.637	19.637
y	-3.065	-3.065	3.060	3.060	3.060	-0.493
z	-20.309	7.781	24.880	24.880	-2.364	-33.442

Å,a completeness of 15, and 10 outputs (models) per simulation. Subsequently, the best results in terms of binding energy were selected for a detailed analysis of the interactions.

The highest-quality models were initially visualized in PyMol and later in Discovery Studio Visualizer, generating detailed 2D images of the observed interactions.

As mentioned above, given the predictions made by molecular docking, the computational work required is high. Therefore, the computer hardware must be suitable. For a reduction in computational time and better performance, it is recommended that the computer used have a minimum RAM capacity of 32 GB, and a CPU with the greatest number of cores possible, for some cases it is essential to have a Nvidia graphics card. In this work, a computer was used with an 8-core processor of the 11th generation of Intel Core i7, with a speed of 2.30GHz and the capacity to handle up to 16 processing threads, and with a RAM of 16 GB, with a system 64-bit operating system.

3.3 Evaluation parameters

To assess the quality of the positions of predicted binding poses and the estimated binding affinities between a ligand and its target protein, it was necessary to evaluate certain parameters.

3.3.1 Binding energy

Considered one of the crucial parameters in the assessment of molecular docking, binding energy provides information on the strength and stability of the receptor-ligand interaction. A more negative binding energy usually indicates a strong binding affinity. In this study, the affinity energy values determined by the FDA were considered to be in the range of -5.63 kcal/mol to -6.85 kcal/mol[30].

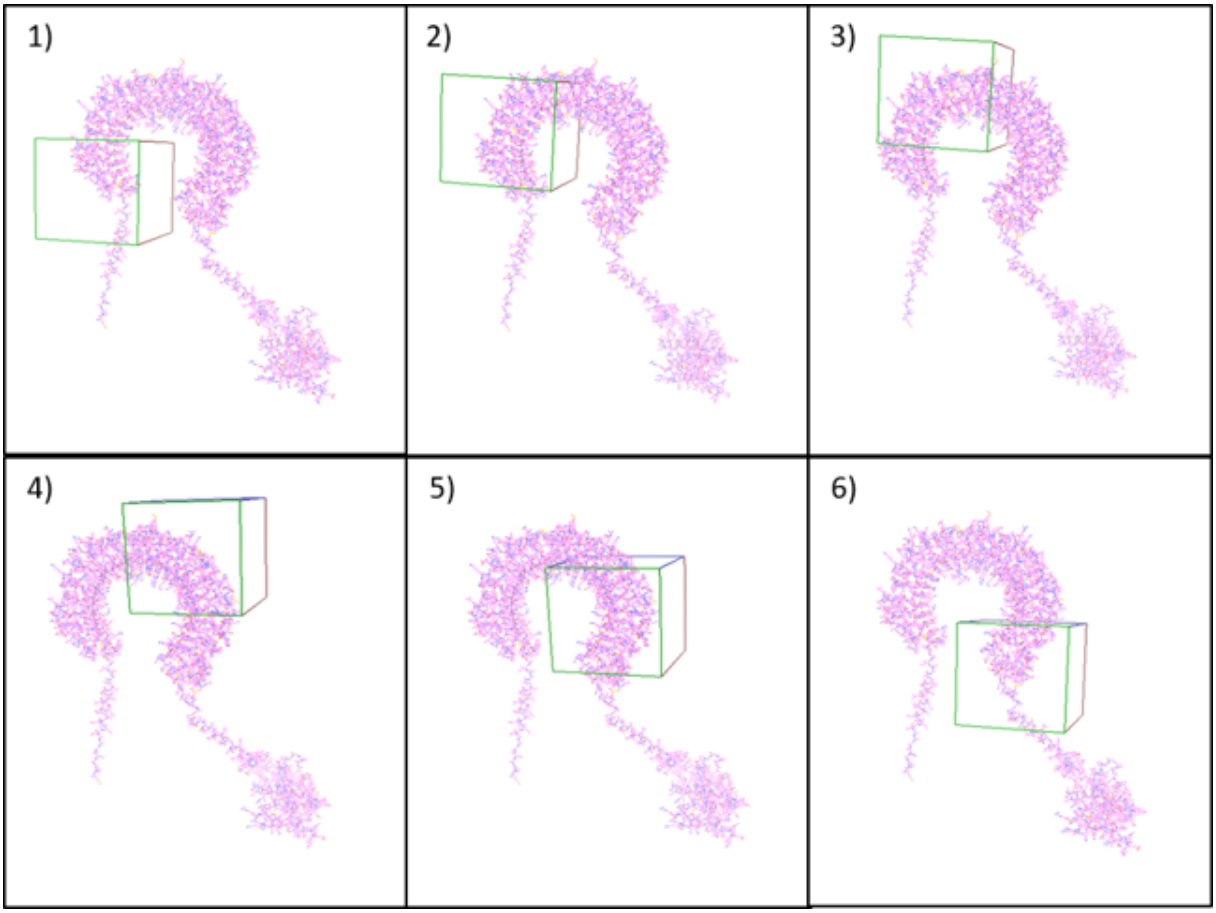


Figure 3.3: The graphical representation illustrates the transitions of the grid during molecular docking on TLR4, as detailed in the accompanying Table3.2.

3.3.2 RMSD (Root Mean Square Deviation)

RMSD is a quantitative measure that allows us to determine how well the simulated structure aligns with the experimental or native structure, enabling the evaluation of prediction coherence. Lower RMSD values suggest more consistent predictions, while larger RMSD values could indicate significant structural changes[31].

3.3.3 Visual Inspection

This parameter will enable the evaluation of the quality of the predicted binding poses. Important considerations include the binding site interactions, looking for the formation of hydrogen bonds, electrostatic interactions, van der Waals contacts, and other relevant interactions. Additionally, the assessment involves examining the binding posture, checking the orientations and positioning of the ligand at the binding site, and evaluating the shape and size of the binding pocket. Consideration of conformation changes is crucial in determining the reasonableness of any alterations in the protein or ligand.

Chapter 4

Results & Discussion

4.1 3D Structure Validation

To initiate molecular docking and determine the optimal model to serve as a vaccine candidate, the 3D structures of the ligand and receptor used for this study were first determined. The 3D modeling was carried out on the SWISS-MODEL workspace (https://swissmodel.expasy.org), these were made from the FASTA files of chicken TLR4 and *Salmonella enterica* subs. *enterica* serovar *Typhimurium* FimH. Once the 3D structures of chicken TLR4 and Salmonella FimH were obtained, they were validated. This evaluation focused on the parameters detailed in the methodology, which provide information on the molecular geometry, steric collisions, the conformation of the dihedral angles phi and shi, the orientation that the groups can take, the chemical bonds, and the atypical angles[29]. The validation of the structures is of utmost importance as the outcome of molecular docking is directly related to the quality and reliability of the model[32]. Figure 4.1 details the parameters evaluated in each structure, thus it can be seen that in the two structures, the parameters MolProbity Score, Clash Score, Ramachandran Favoured, Ramachandran Outliers, Bad Bonds, Twisted Non-Prolines and Cis Non-Proline are within the ranges considered as ideal; on the other hand, the values of C-Beta Deviations, Bad Angles, Cis Prolines, and Twisted Prolines, although not ideal, are within the acceptable values for the validation of the 3D structures. Once the structures were determined, the required files for docking were prepared, which included the receptor and ligand in PDBQT format and the configuration file, to finally start the simulation. In order to do a thorough study, blind molecular docking was carried out to explore possible interactions in more detail. For this purpose,*Salmonella enterica* subs. *enterica* serovar *Typhimurium* FimH was divided into 66 ligand segments, which are made up of 11 amino acids, except for the first segment (1), which

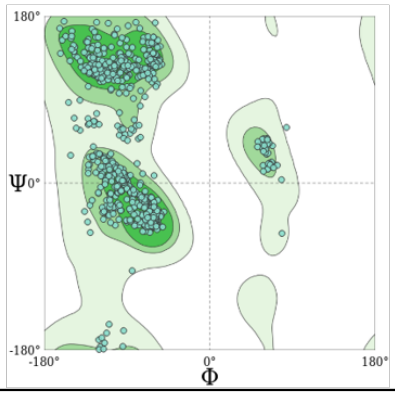
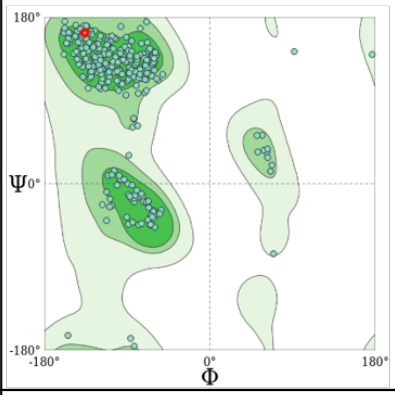
Ideal values	Chicken TLR4	Salmonella FimH
		
MolProbity Score: less than 1.5	0.96	0.73
Clash Score: less than 5.	0.22	0.20
Ramachandran Favoured: >98%.	94.53%	97.30%
Ramachandran Outliers: as low as possible.	0.36%	0.60%
Rotamer Outliers: bellow 1-2%.	0.90%	1.05%
C-Beta Deviations: as low as possible.	3	2
Bad Bonds: zero.	0/6884	0/2603
Bad Angles: zero.	46/9334	8/3555
Cis Non-Proline: as low as possible.	1/815	1/313
Cis Prolines: as low as possible.	2/27	1/21
Twisted Non-Prolines: zero.	1/815	1/313
Twisted Prolines: zero.	2/27	

Figure 4.1: Chicken TLR4 and *Salmonella* FimH structures evaluation.

contains only 10 amino acids, as detailed in Table3.1. The smaller conformational size of the ligands allows rapid and efficient scanning in contrast to the use of the entire ligand structure and allows the binding sites on the receptor to be determined clearly and specifically[33].

4.2 Molecular docking results

Molecular docking was executed using Autodock Vina from the Command Prompt window, where the grid box played a key role. This was specified so that the entire surface of the receptor was scanned. The configuration of this grid was defined in the configuration file according to the section that was desired to be evaluated. As a result, the program generated a table with the 10 best models per simulation, which resulted in a total of 66 tables, corresponding to the number of ligand segments for the 6 grids mentioned above.

In this part, the analysis focused on affinity energies and RMSD values detailed in the tables, crucial parameters when evaluating receptor-ligand complexes. Affinity energy, as provided by the FDA, is vital for assessing the strength and specificity of binding. While a more negative affinity energy suggests a stronger interaction, it is crucial to avoid excessively negative values, which may indicate non-specific binding and the formation of biologically irrelevant complexes[34]. Compliance with recommended energy ranges (-5.63 to -6.8 Kcal/mol) is essential, ensuring both stability and immunogenicity without deactivating the binding site[30], conversely, RMSD values serve to compare structural conformations between predicted and reference structures. A low RMSD value indicates close structural similarity, facilitating structural superposition[31].This comparison aids in assessing the accuracy and reliability of the predicted complex structures.

The tables show the models in descending sequence, the model with the most favorable (negative energy) affinity energy is at the top, in this sense the first model is better than the second model, the second model is better than the third, and so on. To determine in which part of the structures the interaction occurs, we begin by comparing the best models of each simulation, in this case, we select the first model of each table, the one with the highest affinity energy and an RMSD with a value equal to zero. First, the performance of each of the ligand segments in all the grid boxes was compared and the best of these was selected and graphed. This selection was made with the 66 ligand segments.

Figure 4.2 shows the affinity energy of the best models (receptor-ligand complex), one per ligand, where the

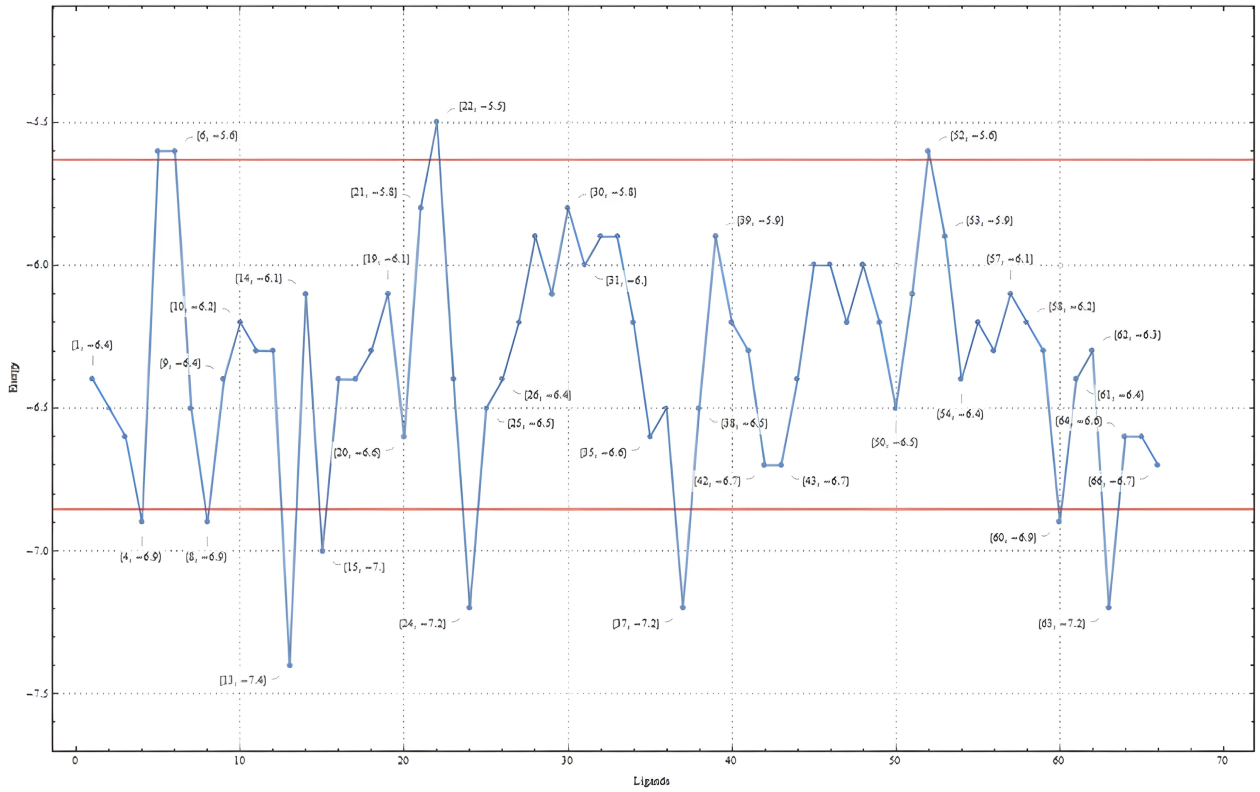


Figure 4.2: Visualization of the best receptor-ligand models based on the binding energy resulted of the molecular docking simulation. Where on the x axis it can be found the number of ligand and on the y axis the binding energy.

binding energy varies between -5.5 to -7.4 kcal/mol, but only 54 of the 66 models have an affinity energy inside the range established as feasible. Thus, models involving ligands 4, 6, 8, 13, 15, 22, 24, 37, 53, 60 and 63 were excluded as possible candidates for vaccine development.

Parallel to the affinity energy selection, a visual analysis of the same models was performed in PyMol, mainly to determine if the interactions were visually reasonable, focusing on the conformation and orientation of the structures, verifying if the model is biophysically plausible, as well as looking for steric hindrances between the structures. Identify possible crashes between structures, the presence of these steric hindrances could interfere with the stability of the model, which is why it is important to identify the existence of these and if it is possible to minimize them, change the orientation of ligand[35 36]. As a result of this analysis, only 24 of the 66 models were considered to be visually reasonable and had affinity energies within the FDA range. Table 4.1 details the ligands and the Grid Box to which the model belongs, as well as their affinity energy. In this, it can be seen that the ligand segments of the 24 favorable models are consecutive or close to each other, which would tell us about the spatial proximity between them, which could favor the interaction of the ligand with the receptor protein[37].The total binding affinity to the receptor could be favored since the interactions can be cumulative and occur synergistically, enhancing the stability of the interaction or in turn favoring the signaling cascade. In turn, by increasing the interaction of the ligands with the receptor, possible interactions with other organisms can be blocked, which allows greater selectivity and efficiency when developing a drug[38 39].

Table 4.1: Shows information on the 24 models considered viable. Important information is also detailed here, such as the ligands involved in the simulation, the affinity energy, the amino acid residues, the body of the grid where the model simulation was carried out, as well as the favorable and not favorable interactions of the model.

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid center**	**Number of favorable interactions**	**Number of Unfavorable interactions**
11(a)	-6.3	A:LYS59:NZ - B:GLU291:OE2, A:LYS59:NZ - B:GLU291:O, A:PRO50:CD - B:ASN206:OD1, B:ARG261 - A:LYS59.	3	4	36
14(b)	-6.1	A:VAL65:N - B:LEU423:O, A:VAL65:N - B:TYR449:OH, A:THR68:N - B:THR398:O, A:THR68:OG1 - B:THR398:OG1, A:VAL75:N - B:LEU454:O,B:THR424:OG1 - A:ASN66:OD1, B:LEU430:N - A:GLY72:O, B:LEU458:N - A:THR74:O, A:CYS69:SG - B:PHE404, A:ALA67 - B:PRO402, A:CYS69 - B:CYS400, A:CYS69 - B:PRO402, A:PRO70 - B:PRO402, A:VAL75 - B:LEU458, B:PHE427 - A:ALA71.	3	16	308
16(c)	-6.4	A:TYR77:OH - B:LEU271:O	2	1	36

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
17(d)	-6.4	B:ASN500:N - A:TYR82:O, B:ARG505:NH1 - A:SER90:O,B:ASP521:OD2 - A:TYR82, A:VAL83:O - B:PHE525, A:VAL83 - B:LEU517, A:VAL83 - B:VAL520, A:LEU86 - B:LEU510, A:LEU86 - B:LEU534, A:PRO87 - B:LEU510,A:PRO87 - B:LEU531, B:ALA527 - A:VAL88	5	11	307
18(e)	-6.3	B:SER334:CB - A:LEU86:O, A:LEU86:CD1 - B:PHE308, A:PRO87 -B:MET332, A:VAL88 - B:MET332, B:VAL330 - A:LEU86, B:PHE333 - A:PRO87,B:PHE357 - A:LEU86.	4	7	147
19(f)	-6.1	B:LYS170 - A:LEU100	2	1	0
21(g)	-5.8	B:LEU153:N - A:SER109:O, B:THR155:OG1 - A:GLY106:O, B:SER180:N - A:ALA107:O, A:LEU104 - B:LEU153, A:LEU105 - B:LEU129, A:LEU105 - B:LEU153, A:LEU105 - B:LEU156, A:ALA107 - B:LEU156, A:MET108 - B:LEU181, A:ILE110 - B:ILE150, B:PRO149 - A:ILE110, B:HIS127 - A:LEU104.	4	12	247

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
23(h)	-6.4	A:ILE110:N - B:GLU301:OE2, B:ASN302:ND2 - A:TYR119:O, A:SER113:OG - B:HIS248, B:GLU301:C,O;ASN302:N - A:PHE118, A:VAL114 - B:LEU244,A:ALA115 - B:ALA275, B:PHE276 - A:VAL114, B:PHE276 - A:ALA115.	4	8	103
26(i)	-6.4	B:ILE125:N - A:ASP99:OD2, A:THR130:OG1 - B:ASP130:O, B:GLN135:C - A:THR110:OG1, A:SER131:CB - B:ASP130:OD1, A:THR130:OG1 - B:TYR131, B:MET127:CE - A:HIS127, A:THR130:CG2 - B:TYR131, B:GLY128:O - A:TYR106, B:VAL129 - A:LEU108, A:VAL103 - B:LEU126, A:ALA124 - B:ILE125.	1	11	149

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
27(j)	-6.2	A:ASP130:N - B:SER628:OG, A:GLN135:NE2 - B:LEU558:O,B:HIS614:N - A:ASN132:O, B:SER630:N - A:ASP130:OD2, B:ASP559:CA - A:GLN135:OE1, B:HIS614:CE1 - A:TYR131:O, B:LYS613:O - A:TYR131, A:TYR131 - B:LEU610.	6	8	108
28(k)	-5.9	A:LYS145:NZ - B:ASP147:OD2, B:LYS173:NZ - A:ASP143:OD2, B:SER146:OG - A:ASP143:O, A:PRO138:CD - A:GLN136:O, B:LYS173:CE - A:ASP143:OD2, A:PRO138 - B:LEU123, A:VAL141 - B:PRO172, A:LYS145 - B:PRO172, B:ALA124 - A:PRO138, A:PHE139 - B:LEU121, A:PHE139 - B:LEU123, A:PHE139 - B:ALA125.	1	12	145
28.1(l)	-5.9	A:LYS145:NZ - B:GLU533:OE2, A:LYS145 - B:ILE509, B:ARG505 - A:LYS137	6	3	72

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
32(m)	-5.9	A:ARG164:NE - B:ARG207:O, A:ARG164:NH2 - B:GLN258:O, A:GLN165:NE2 - B:ASN209:OD1, A:PRO163:CD - B:THR179:O, A:PHE156 - B:TYR174, A:PRO155 - B:LEU171, A:PRO155 - B:ILE230, B:LYS173 - A:ILE157, B:ALA176 - A:ILE157, B:ALA176 - A:MET159, B:ARG207 - A:ILE162, A:PHE156 - B:PRO172, A:PHE156 - B:LYS173, A:PHE156 - B:ALA176.	2	14	215
40(n)	-6.	B:ASN269:ND2 - A:PRO196:O, B:SER188:CB - A:VAL201:O, A:PRO196 - B:ILE270, B:ARG238 - A:VAL195, B:HIS163 - A:CYS199, B:PHE267 - A:PRO196.	4	6	68

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
41(o)	-6.3	B: LYS71:NZ - A:ASP210:OD2, B:HIS95:N - A:PHE209:O, B:THR96:N - A:PHE209:O, B:THR96:OG1 - A:GLU208:OE2, A:GLU208:CA - B:GLU98:OE1, B:HIS95:CD2 - A:ASP210:O, B:HIS127:CE1 - A:GLU200:OE1, B:PRO149:CD - A:GLN205:OE1, B:HIS152:CD2 - A:ASN202:OD1, B:HIS152:CE1 - A:ALA203:O, B:LEU73:CD2 - A:PHE209,A:ALA203 - B:ALA125, B:ALA125 - A:VAL206, A:PHE209 - B:LEU70, A:PHE209 - B:LEU72, B:PHE126 - A:ALA203.	2	16	303
43(p)	-6.7	A:ASP210:N - B:ASN206:OD1, A:PHE211:N - B:ASN206:OD1, A:ASP213:N - B:LYS229:O, B:ARG207:NE - A:PHE211, A:ASP213:OD2 - B:HIS231, A:ASP213:O - B:HIS231, A:ILE214 - B:ILE230, A:ALA216 - B:LYS229, B:ALA228 - A:ILE214, B:LYS229 - A:ILE214, A:PHE211 - B:ARG207.	4	11	184

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
44(q)	-6.4	A:GLY225:N - B:GLU329:OE2, B:SER328:CB - A:GLY225:O.	2	2	20
46(r)	-6	A:ARG227:CD - B:SER282:O.	4	1	28
48(s)	-6	A:THR235:OG1 - B:SER636:O, B:SER571:N - A:ASN244:O, B:SER571:OG - A:ASN244:O, B:SER569:CB - A:THR243:OG1, B:THR595:CA - A:ALA239:O, B:CYS637:CA - A:THR235:O, B:CYS637:CA - A:THR235:OG1, A:CYS242:SG - B:PHE567.	6	8	9
51(t)	-6.1	A:ALA257:N - B:ASN190:O, B:SER195:OG - A:TYR250:O, A:ALA257:CA - B:ASN165:O, B:SER144:CA - A:GLU258:O, A:MET253:SD - B:THR192:O, A:MET253:SD - B:TYR221, A:ARG254 - B:LYS220, A:ALA257 - B:ILE166, A:ALA257 - B:ALA167, A:ALA257 - B:ILE191, B:LYS220 - A:MET253, B:TYR193 - A:LEU255.	1	12	163

Continued on the next page

Table 4.1 – Continued from previous page

Ligand	**Binding affinity (kcal/mol)**	**TLR 4 amino acid residues**	**Grid Box**	**Number of favorable interactions**	**Number of Unfavorable interactions**
57(u)	-6.1	A:ASN280:N - B:GLU375:OE2, A:THR285:OG1 - B:GLU426:O, B:SER401:CB - A:THR285:OG1, A:LEU289 - B:LEU423, A:LEU289 - B:ILE444.	3	5	213
64(v)	-6.6	B:ASN243:N - A:GLU321:OE2, B:ASN243:ND2 - A:ALA320:O, A:ALA320 - B:LYS220, B:ALA167 - A:ILE317	2	4	140
65(w)	-6.6	B:SER188:OG - A:ARG327:O, B:ARG238:NH2 - A:THR325:OG1, A:THR325:CA - B:ASN269:OD1, :SER187:OG - A:PHE324, A:ALA320 - B:LEU271, A:ARG327 - B:LEU216,B:PHE267 - A:ALA326.	2	7	133
65.1(x)	-6.6	A:ALA320:N - B:ASP147:OD2, A:ALA320:N - B:SER143:OG, A:GLY328:N - B:LEU161:O, A:ARG327:CD - B:VAL213:O, B:ASN217:OD1 - A:PHE324, A:PRO323 - B:ILE166, A:PHE324 - B:ILE166, B:HIS163 - A:LEU330, B:PHE186 - A:ARG327.	3	9	339

4.3 2D Model Generation

4.3.1 Models selection

Having determined the favorable models in the previous section, the next step was to generate a 2D plot of each model. 2D graphics are useful in the analysis of the results since they provide a flat representation of the structures allowing a better view of the structures and the interactions that occur between them, facilitating the identification of the type of interactions and amino acid residues involved[37].

The Figures 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9,4.10, 4.11, 4.12, 4.13 and 4.14,show the 24 selected models and the given interactions between TLR4 and the ligand segments, detailed in Table 4.1. To determine which models are most viable, it is necessary to take into consideration the favorable and unfavorable interactions present in the models.

In the 2D model representation, the interactions between the structures are depicted in various colors, indicating both favorable and unfavorable contacts. Notably, the Unfavorable Bump interactions suggest a steric clash between the atoms of the structures, likely due to misaligned ligand groups or their improper positioning within the receptor's active site. This misalignment diminishes the likelihood of stable and efficient binding for model[40]. Additionally, the Unfavorable Donor-Donor interaction denotes a detrimental overlap between hydrogen bond donors, hindering the establishment of hydrogen bonds. The presence of these unfavorable interactions could detrimentally impact the model's accuracy, complicate the identification of interactions, and potentially result in inaccuracies in the assessment of binding energy[41 38 42].

Additionally, the study highlights favorable interactions, such as the Attractive Charge interaction, which involves a beneficial electrostatic interaction between oppositely charged entities[40], suggesting a region where these charges are drawn together, thereby enhancing the stability of the interaction. The Conventional Hydrogen Bond type is also observed, signifying a hydrogen bond where a hydrogen atom engages with another hydrogen donor atom[43]. This interaction is crucial for the specificity and strength of the bond within the system. The images further reveal Carbon Hydrogen Bond interactions, where a hydrogen atom connected to a carbon atom interacts with another atom in the receptor molecule, contributing to the model's stability, akin to the Conventional Hydrogen Bond. Moreover, the Alkyl interaction type, associated with hydrophobic interactions between the ligand's alkyl groups and the receptor, provides an attractive force in the hydrophobic regions, thus offering hydrophobic stability to the model[40]. The model also exhibits Pi-Sulfur, Pi-Alkyl, Pi-Anion, and Pi-Lone Pair interactions. These involve the specific engagement of the ligand's aromatic ring with a sulfur atom, an alkyl group, a negative ion, and a lone pair of electrons on the

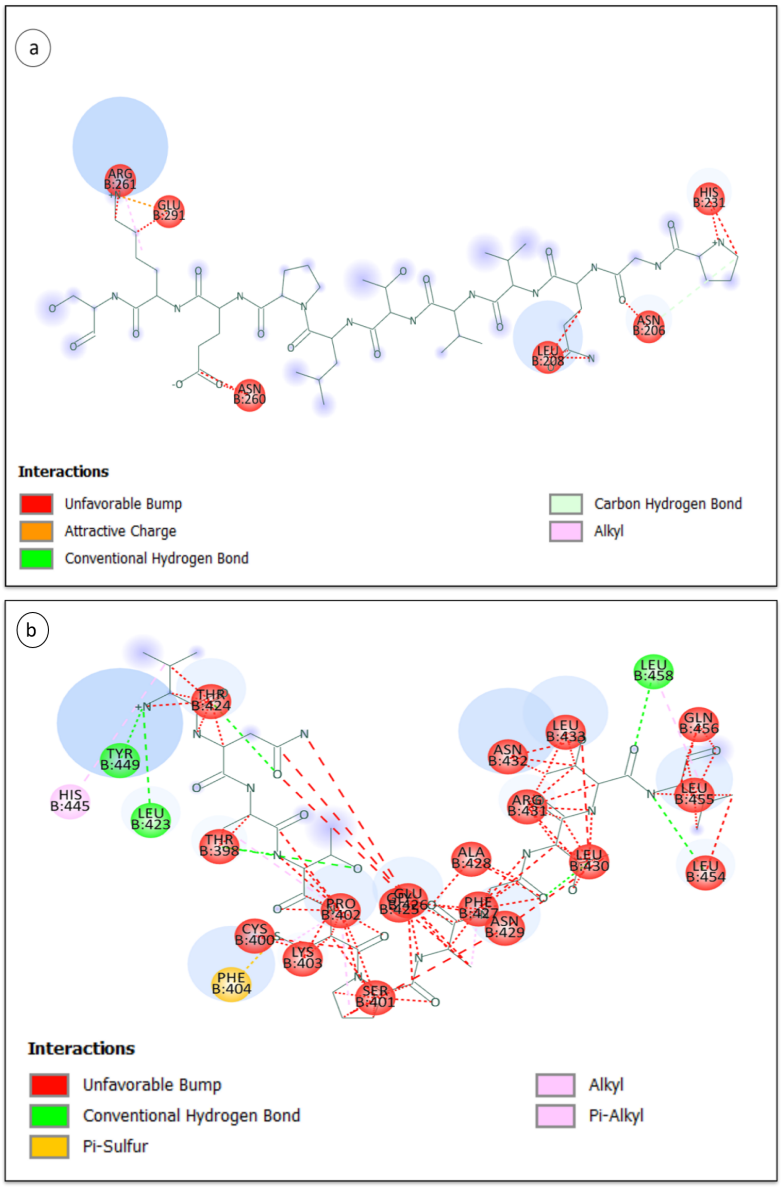


Figure 4.3: Schematic representation of the interaction model between the receptor and the ligands 11(a) and 14(b). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

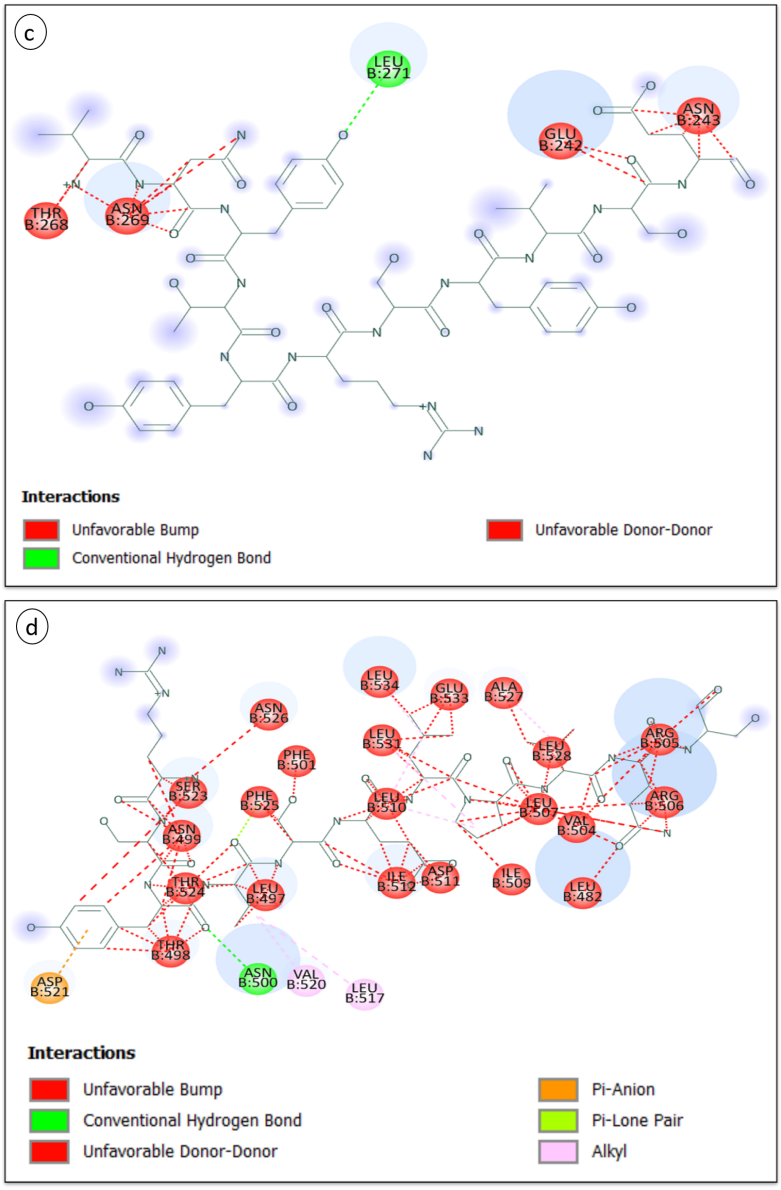


Figure 4.4: Schematic representation of the interaction model between the receptor and the ligands 16(c) and 17(d). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

receptor, respectively. Such interactions underscore the molecular complementarity and the precise recognition of the binding site by the orientation of the molecules[40].

4.3.2 Models description

Upon identifying the types of interactions depicted in the 2D images, the subsequent phase involves selecting the most appropriate model for developing a vaccine against *Salmonella enterica* subs. *enterica* serovar *Typhimurium* in chickens. Although the numerical data may show a predominance of unfavorable over favorable interactions, suggesting potential non-viability, it is imperative to consider that a model's feasibility also relies on the quality of other interactions between the structures. These play a pivotal role in precisely predicting the model's behavior and ensuring its stability[37].

Figure 4.3 and 4.4 shows a predominance of unfavourable interactions in the four models; at the same time, it can be seen that in the 2D image (b), (c) and (d) of Figure 4.3 and 4.4 there are favorable interactions of the hydrogen bond type and in two of these (b) and (d) models a hydrophobic type, the occurrence of these interactions could certainly represent a positive appearance in the model. Therefore, despite the fact that the unfavorable interactions are greater in number, this does not mean that there is no part of the molecules in which an interaction occurs, which is the case in models (b), and (d) in Figure 4.3 and 4.4. Since model (a) only shows unfavorable interactions, it is discarded as a possible candidate for the indicated purposes.

Figures 4.5 and 4.6show the 2D structures of the models involving the ligands 18(e), 19(f), 21(g) and 23(h). Here, the number of unfavorable interactions is smaller than those of the models in Figure 4.3 and 4.4. One of the most remarkable models is model (f) Figure 4.5. This model, contrary to the other 23 models, is the only one that has no unfavorable interactions and, in addition to this, presents a hydrophobic interaction. On the contrary, models (e), (g), and (h) present favorable and unfavorable interactions; the presence of favorable interactions of the hydrophobic type and one of the hydrogen bond types in model (g), points to the possibility that there is affinity in some part of the structures so that all the models present in this figure will be considered as viable.

Following the analysis of the previous figures, it can be seen that of the models in Figures 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, 4.13 and 4.14, only models (p) and (s) are considered unfavorable, given that all the interactions present in them are of the unfavorable type, meaning that they do not provide stability to the model. Regarding the remaining models, these ones have at least one type of interaction, either hydrophobic or hydrogen bonding type, which does

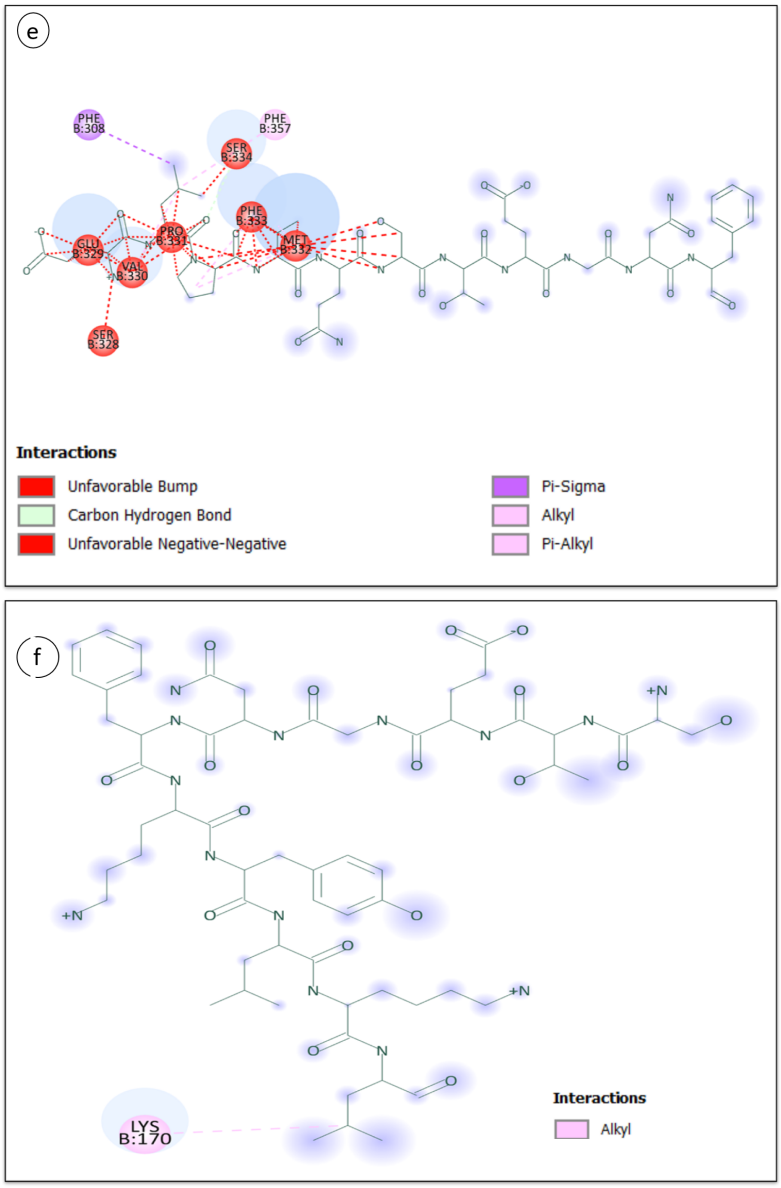


Figure 4.5: Schematic representation of the interaction model between the receptor and the ligands 18(e) and 19(f).The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

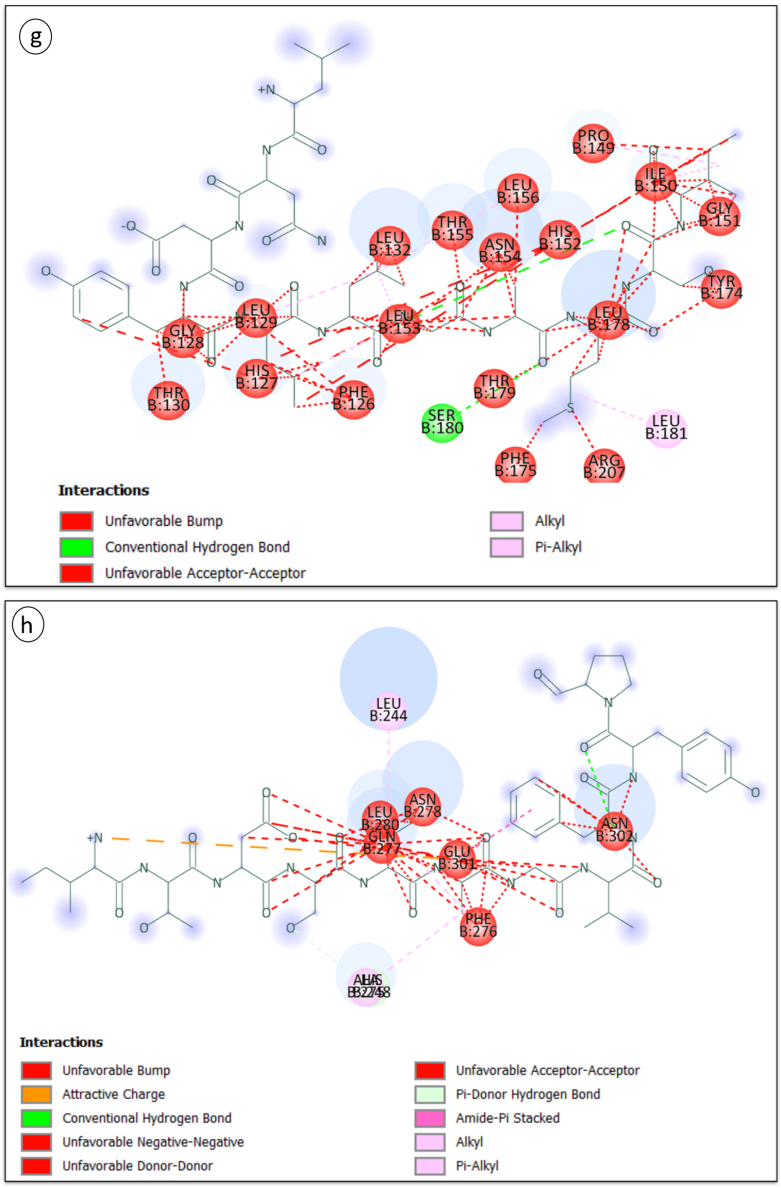


Figure 4.6: Schematic representation of the interaction model between the receptor and ligands 21(g) and 23(h). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

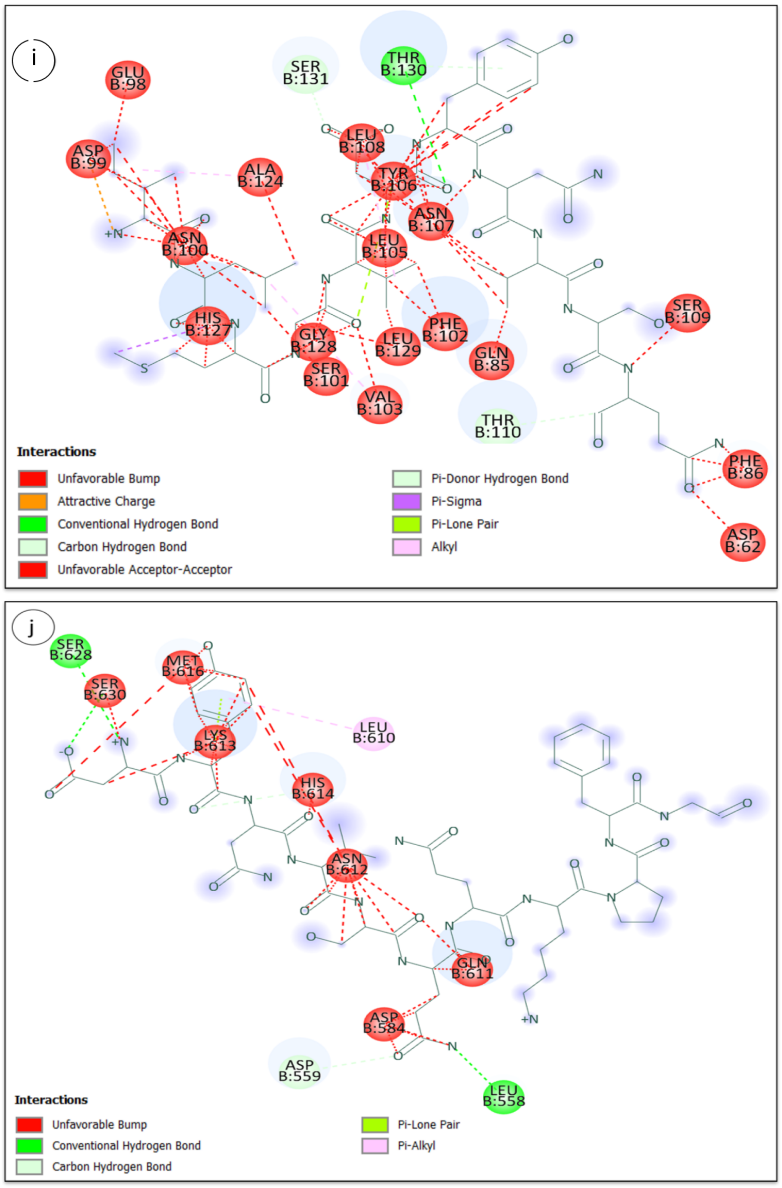


Figure 4.7: Schematic representation of the interaction model between the receptor and the ligands 26(i) and 27(j). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

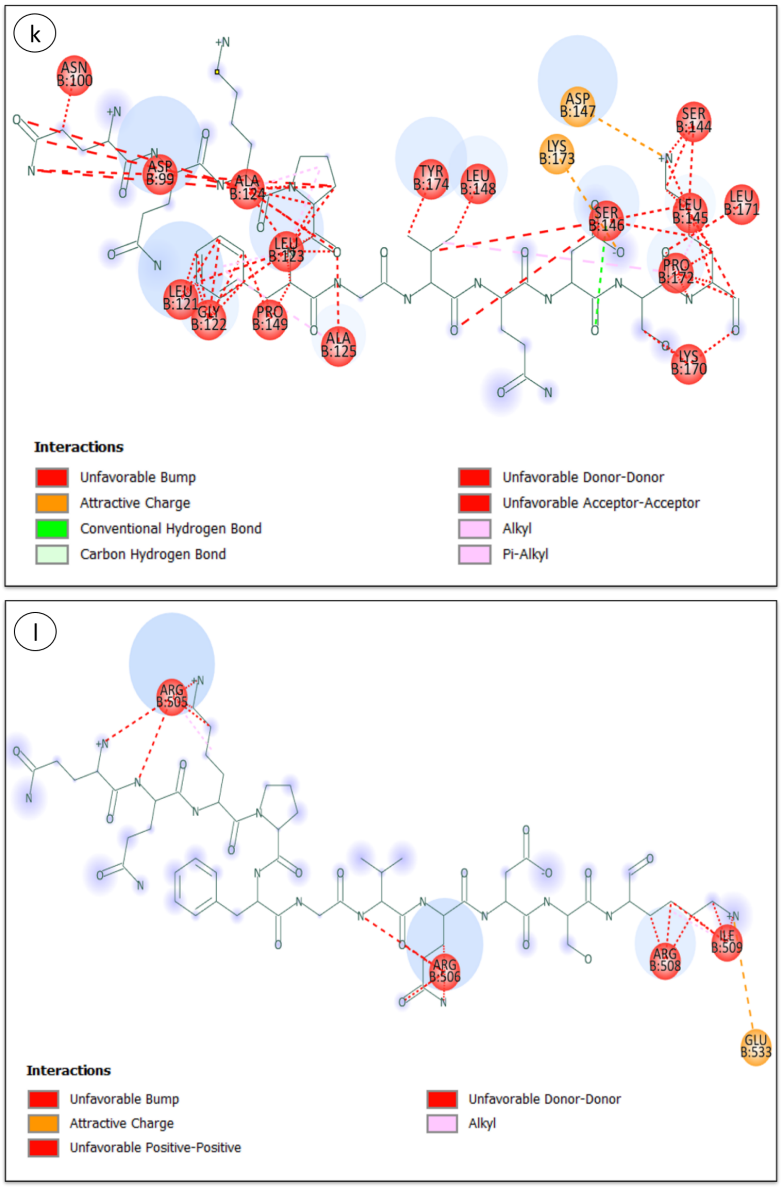


Figure 4.8: Schematic representation of the interaction model between the receptor and the ligands 28(k) and 28.1(l). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

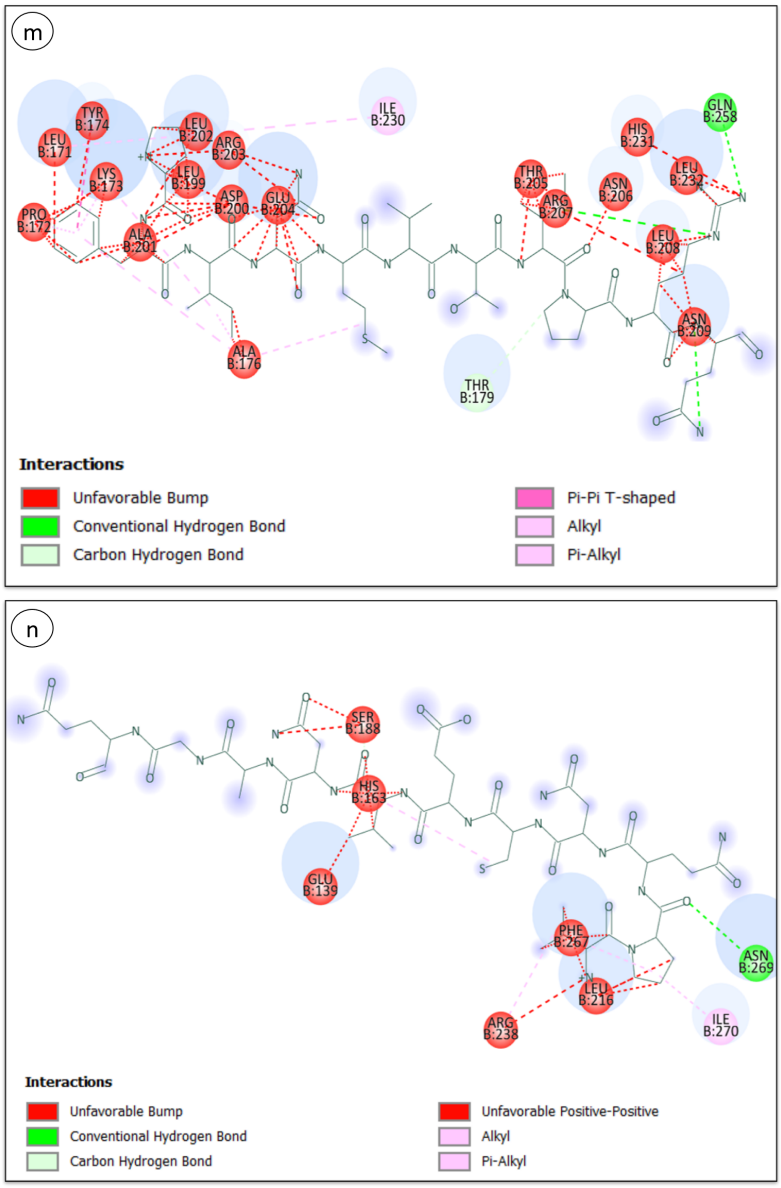


Figure 4.9: Schematic representation of the interaction model between the receptor and the ligands 32(m) and 40(n). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

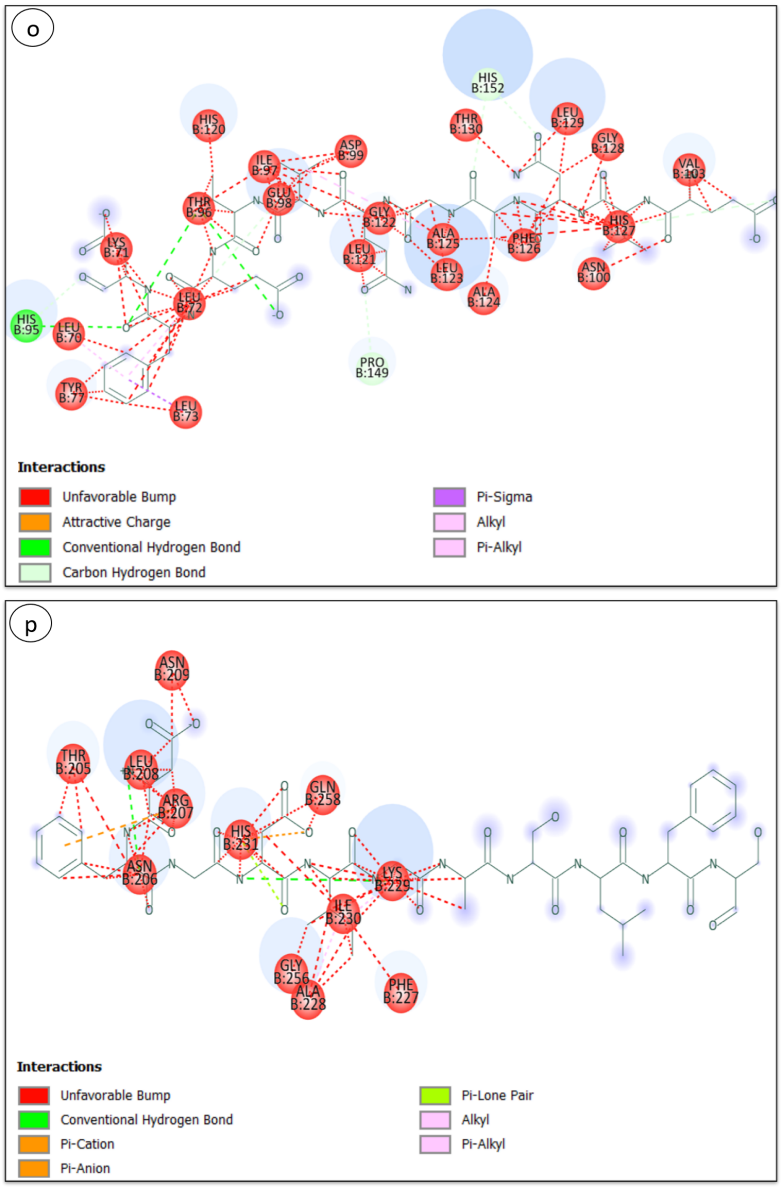


Figure 4.10: Schematic representation of the interaction model between the receptor and the ligands 41(o) and 43(p). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

not allow them to be discarded completely.

As mentioned above, the presence of hydrogen bonds and hydrophobic interactions will be a determining factor in the selection of models for the study. The greater the presence of hydrogen bonds with the amino acid residues, the stronger the bond. Based on the 2D structures of the models, the hydrogen bond type interactions were presented in the models (b), (c), (d), (g), (i), (j), (m), (n), (o), (q), (s), (t), (u), (w) and (x). On the other hand, the presence of hydrophobic interactions is important since these are non-polar amino acid residues that, in an attempt to avoid contact with water, tend to form groups within the structures. These interactions generally include alkyl, Pi-sulfur, Pi-alkyl, and Pi-anion. Based on the 2D structures of the models, these types of interactions were found in models (b), (d), (e), (f), (g), (h), (j,) (m), (n), (t), and (v). Furthermore, the presence of unfavorable bump or donor-donor type interactions can reduce the stability of the other types of bonds, thus affecting the models that will be used as candidates for the development of the vaccine, as mentioned previously. The presence of these interactions does not rule out the model directly, but it is important to find a balance between favorable and unfavorable interactions. The analysis of the simulation results revealed that the models where the amino acid residues of the receptor and the ligand interact mainly through hydrogen bonds are hydrophobic and that in addition, the number of unfavorable interactions is not very large, models (c), (f), (j), (n), (q), (s) and (t). The (c) model presents a single hydrogen bond between A: TYR77:OH - B: LEU271:O and four unfavorable interactions of the bump and donor-donor type, these interactions are far from the favorable interaction so they would not be directly interfering with the stability of this link, which allows the model to retain stability and be considered a possible vaccine candidate.

The model (f) could be considered the most reliable candidate since it does not present unfavorable interactions and also presents a hydrophobic type bond between B: LYS170 - A: LEU100. Model (j) presents three hydrogen bonds between A: GLN135:NE2 - B: LEU558:O, B: ASP559:CA - A: GLN135:OE1y A: ASP130:N - B: SER628:OG, the presence of two consecutive hydrogen bonds could speak of a strong and stable interaction, and a hydrophobic interaction A: TYR131 - B: LEU610, seven unfavorable interactions were also observed. In the model (n), two favorable interactions were produced, one of the hydrogen bond type between B: ASN269:ND2 - A: PRO196:O and the other hydrophobic between A: PRO196 - B: ILE270, the presence of these two types of interaction provides stability and reliability of the models. In this model, only six non-favorable interactions were presented. In the (q) model, only three interactions are presented, two of the non-favorable type and one of the favorable type where a hydrogen bond is formed between B: SER328:CB - A: GLY225:O.

In model (s), there are four favorable and two unfavorable interactions. These favorable interactions are three hydrogen bonds between A:THR235:OG1 - B:SER636:O, B:SER571:OG - A:ASN244:O, B:SER569:CB

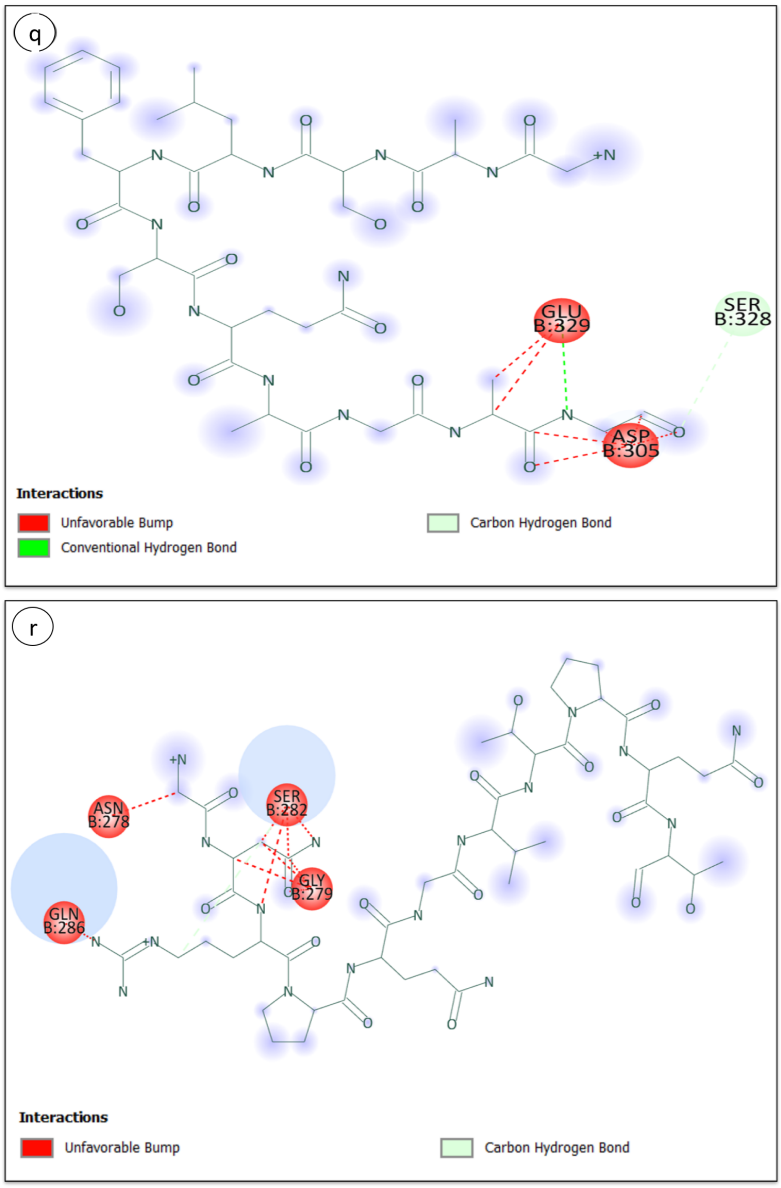


Figure 4.11: Schematic representation of the interaction model between the receptor and the ligands 44(q) and 46(r). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

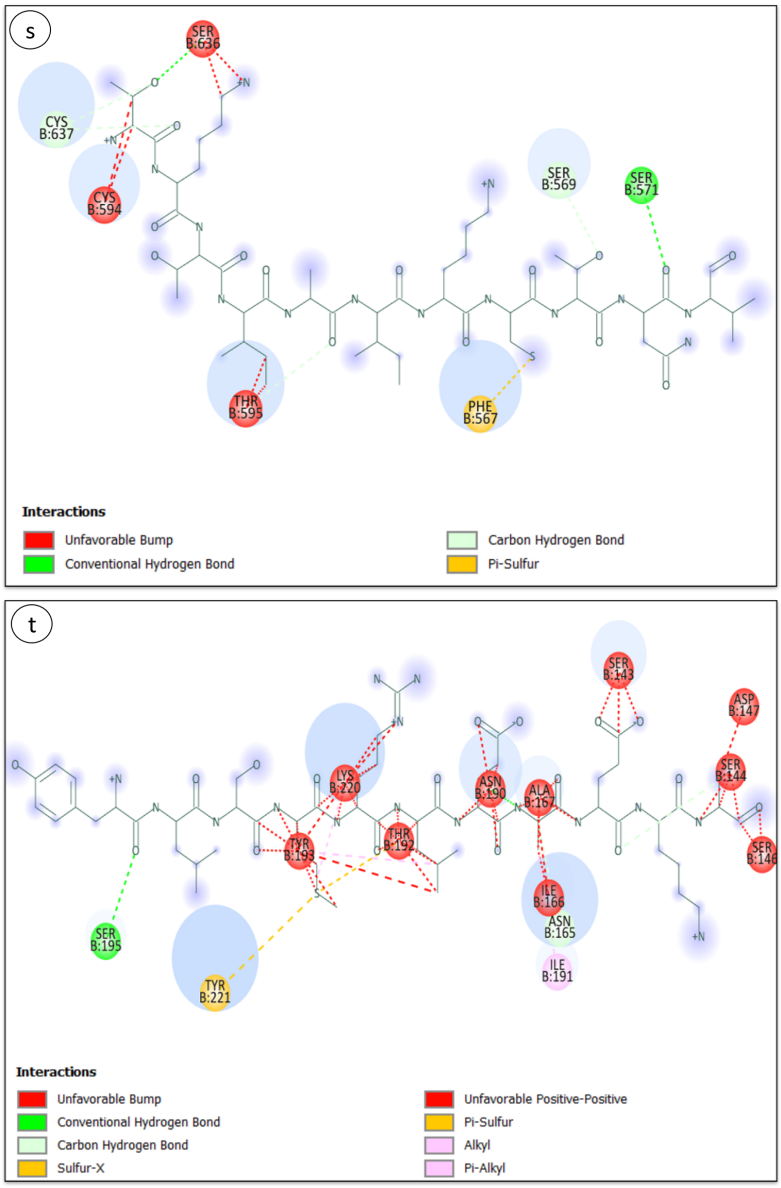


Figure 4.12: Schematic representation of the interaction model between the receptor and the ligands 48(s) and 51(t). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

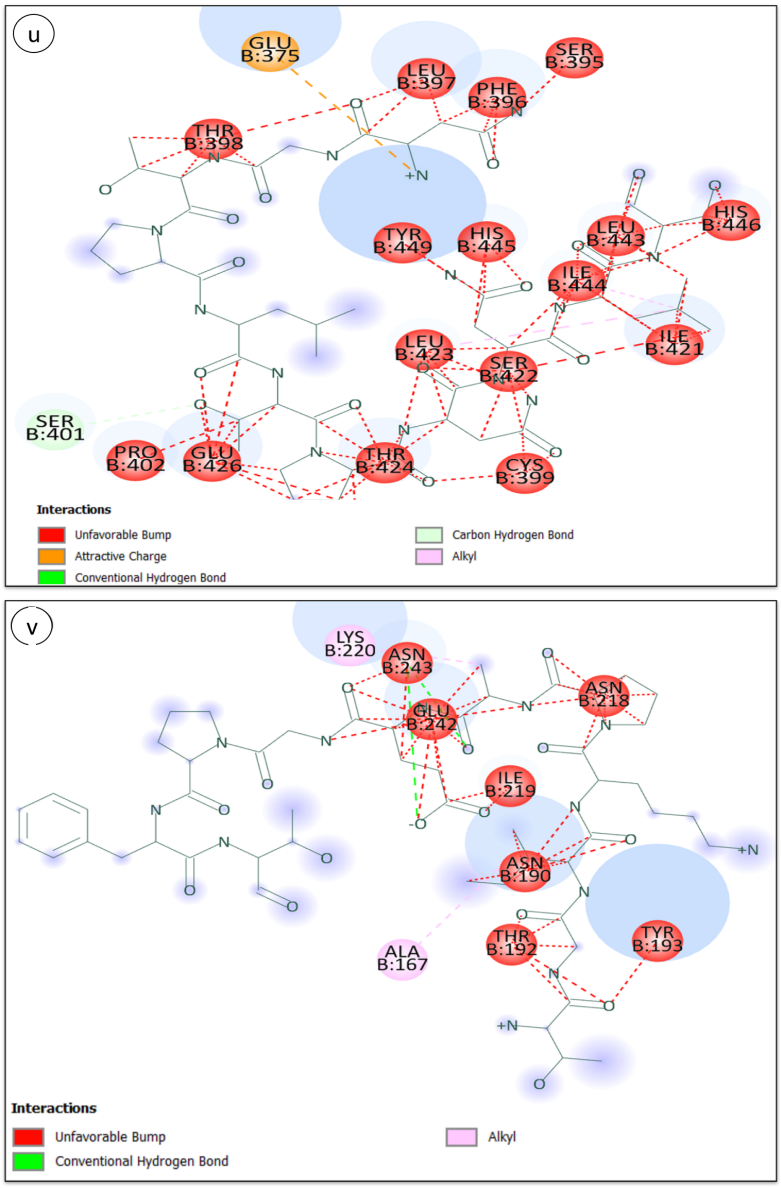


Figure 4.13: Schematic representation of the interaction model between the receptor and the ligands 57(u) and 64(v). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

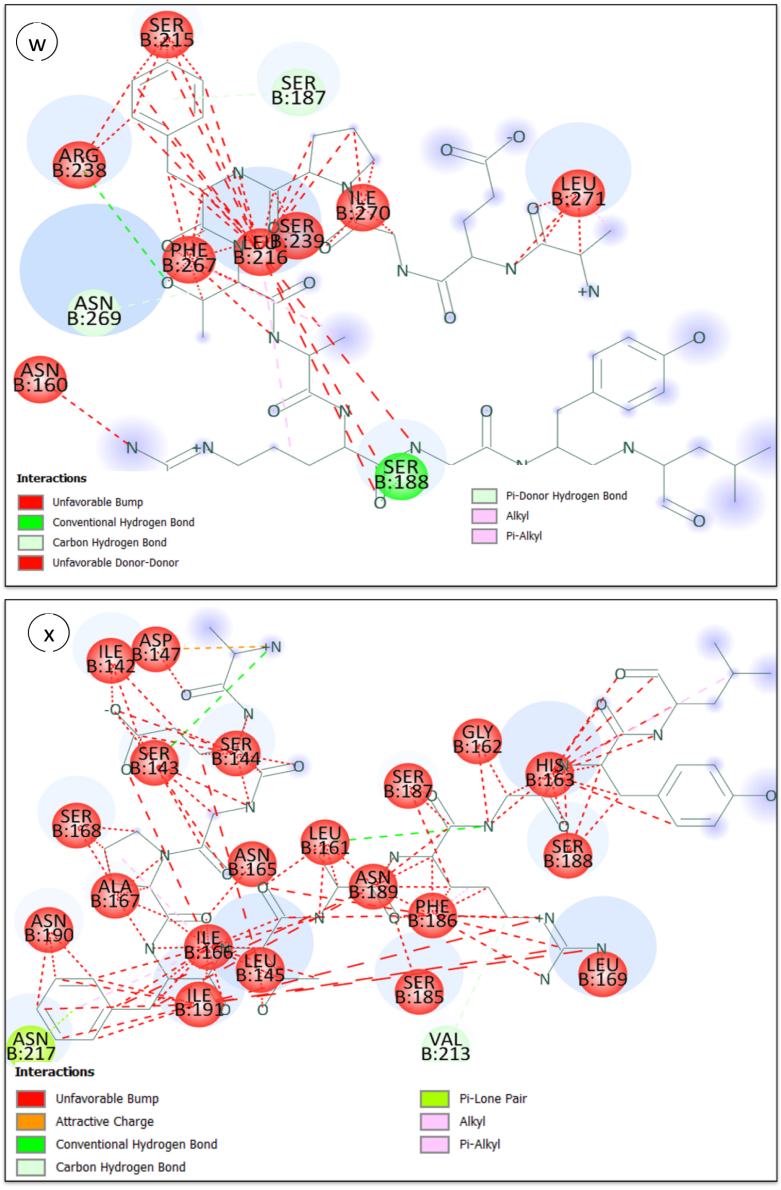


Figure 4.14: Schematic representation of the interaction model between the receptor and the ligands 65(w) and 65.1(x). The receptor is depicted in green and is referred to as the 'A' chain, while the ligand is identified as the 'B' chain.

- A:THR243:OG1 and a Pi bond -sulfur between A:CYS242:SG - B:PHE567.

The (t) model has a total of four favorable interactions, two interactions of the hydrogen bond type, one of the hydrophobic type, and one of the Pi-sulfur type, these interactions occur between B: SER195:OG - A: TYR250:O, A: ALA257:CA - B: ASN165:O, A: ALA257 - B: ILE191 and A: MET253:SD - B: TYR221, respectively, and ten unfavorable interactions.

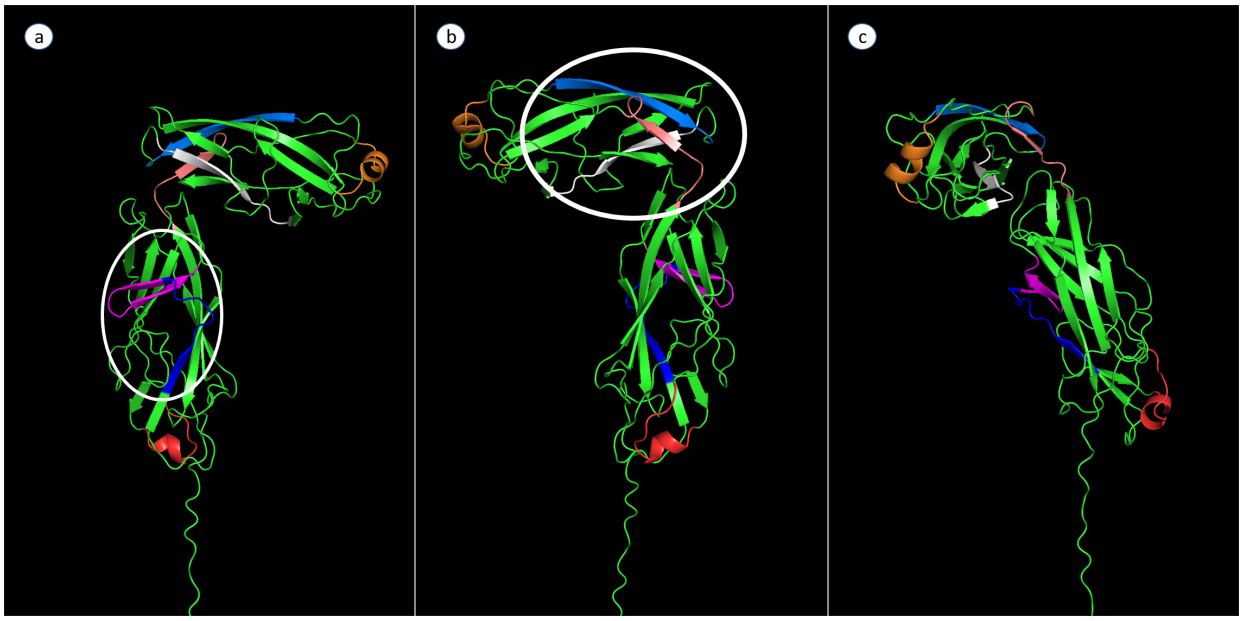


Figure 4.15: Favorable ligand segments: Ligands are differentiated by different colors. For example, ligand 16 is blue, 19 is magenta, 27 is red, 40 is salmon, 44 is orange, 48 is marine, and L54 is white.

To gain a better understanding of the favorable models, the figure 4.15 can display the ligand segments utilized in these models. Within this figure, it becomes evident that certain segments are spatially proximate and consecutive to one another. This is observed in the case of ligands 40 and 48 (see 4.15, image b), as well as 16 and 19 (see see 4.15, image a). he proximity between these ligands may facilitate the determination of the most favorable ligand segment and binding site. Upon examination of the details provided in Table 4.1 , it is apparent that the favorable binding site for 16 and 19 resides within grid box 2. These models exemplify the commutative nature of interactions. In contrast, the ligand segments 40 and 48 exhibit favorable binding sites located in distinct grid boxes. Consequently, it is advisable to prioritize the examination of models (c) and (f) in subsequent studies. Given their binding energy, the interactions depicted, the spatial proximity of the ligand segments, and the grid box within which these models

were simulated, they emerge as strong candidates for the development of a *Salmonella* vaccine for chickens.

Chapter 5

Conclusions

To conclude, the present study explored the interactions between TLR4 and Salmonella FimH. To do this, we first determined that the ligand and receptor structures were reliable and stable. Once the structures were obtained, the molecular docking of 66 ligands was run in 6 different grid boxes, resulting in a total of 396 result tables, each of which included the ten best models from each simulation, yielding 3960 models as molecular docking results, of which only 24 were viable in the first run. These models were analyzed with the help of Discovery Studio Visualizer, which was used to generate a 2D image of the interactions in each model. What was sought in the models was that they were stable and that the interactions present in them were strong. In this sense, it was found that 17 of these models had no interactions relevant to the model and, in addition, the unfavorable interactions greatly outnumbered the favorable ones. It was the evaluation of interactions that finally determined which models could be considered viable for vaccine development, and these were models (c), (f), (j), (n), (q), (s), and (t).Given the spatial proximity of the ligand segments used and that the grid box where the interaction occurred was the same, it is important to pay greater attention to models (c) and (f).

On the other hand, given the number of unfavorable interactions present in almost all the models, it is essential to verify whether the initial structures of the ligand and the receptor are adequate. Given the complexity of the interactions and the conditions under which the study has been carried out, it is important that the results obtained are subjected to other types of analysis, such as molecular dynamics, the calculation of surface areas of interaction, as well as the validation of these results experimentally, such as in vitro tests or more detailed structural studies.

Given the significant impact of Salmonella infections on both public health and the economy, it is imperative to underscore the necessity of developing new vaccines. While existing vaccines are available, they often fall short in

effectively combating all strains of Salmonella, especially those that have developed resistance to antibiotics.

Bibliography

[1] Kolenda, R.; Ugorski, M.; Grzymajlo, K. Everything you always wanted to know about Salmonella type 1 fimbriae, but were afraid to ask. *Frontiers in microbiology* **2019**, *10*, 1017.

[2] Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research* **2015**, *38*, 1686–1701.

[3] Marangon, S.; Busani, L. The use of vaccination in poultry production. *Revue Scientifique et Technique-Office International des Epizooties* **2007**, *26*, 265.

[4] de Mesquita Souza Saraiva, M.; Lim, K.; do Monte, D. F. M.; Givisiez, P. E. N.; Alves, L. B. R.; de Freitas Neto, O. C.; Kariuki, S.; Júnior, A. B.; de Oliveira, C. J. B.; Gebreyes, W. A. Antimicrobial resistance in the globalized food chain: A One Health perspective applied to the poultry industry. *Brazilian Journal of Microbiology* **2022**, 1–22.

[5] Nyachuba, D. G. Foodborne illness: is it on the rise? *Nutrition reviews* **2010**, *68*, 257–269.

[6] Hedman, H. D.; Eisenberg, J. N.; Vasco, K. A.; Blair, C. N.; Trueba, G.; Berrocal, V. J.; Zhang, L. High prevalence of extended-spectrum beta-lactamase CTX-M–producing Escherichia coli in small-scale poultry farming in rural Ecuador. *The American journal of tropical medicine and hygiene* **2019**, *100*, 374.

[7] Amancha, G.; Celis, Y.; Irazabal, J.; Falconi, M.; Villacis, K.; Thekkur, P.; Nair, D.; Perez, F.; Verdonck, K. High levels of antimicrobial resistance in Escherichia coli and Salmonella from poultry in Ecuador. *Revista Panamericana de Salud Pública* **2023**, *47*, e15.

[8] Rabie, N. S.; Amin Girh, Z. M. Bacterial vaccines in poultry. *Bulletin of the National Research Centre* **2020**, *44*, 1–7.

[9] Mor-Mur, M.; Yuste, J. Emerging bacterial pathogens in meat and poultry: an overview. *Food and Bioprocess Technology* **2010**, *3*, 24–35.

[10] Woolhouse, M. E.; Haydon, D. T.; Antia, R. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in ecology & evolution* **2005**, *20*, 238–244.

[11] others,, *et al.* Is it time to move on to gene-based Salmonella typing: Evidence and implications. *Indian Journal of Medical Microbiology* **2023**, *44*, 100359.

[12] Agbaje, M.; Begum, R.; Oyekunle, M.; Ojo, O.; Adenubi, O. Evolution of Salmonella nomenclature: a critical note. *Folia microbiologica* **2011**, *56*, 497–503.

[13] Kisiela, D.; Laskowska, A.; Sapeta, A.; Kuczkowski, M.; Wieliczko, A.; Ugorski, M. Functional characterization of the FimH adhesin from Salmonella enterica serovar Enteritidis. *Microbiology* **2006**, *152*, 1337–1346.

[14] Zeiner, S. A.; Dwyer, B. E.; Clegg, S. FimA, FimF, and FimH are necessary for assembly of type 1 fimbriae on Salmonella enterica serovar Typhimurium. *Infection and immunity* **2012**, *80*, 3289–3296.

[15] Rehman, T.; Yin, L.; Latif, M. B.; Chen, J.; Wang, K.; Geng, Y.; Huang, X.; Abaidullah, M.; Guo, H.; Ouyang, P. Adhesive mechanism of different Salmonella fimbrial adhesins. *Microbial pathogenesis* **2019**, *137*, 103748.

[16] Zhang, X.-L.; Jeza, V. T.; Pan, Q. Salmonella typhi: from a human pathogen to a vaccine vector. *Cellular & molecular immunology* **2008**, *5*, 91–97.

[17] Humphries, A. D.; Townsend, S. M.; Kingsley, R. A.; Nicholson, T. L.; Tsolis, R. M.; Bäumler, A. J. Role of fimbriae as antigens and intestinal colonization factors of Salmonella serovars. *FEMS microbiology letters* **2001**, *201*, 121–125.

[18] Adib-Conquy, M.; Scott-Algara, D.; Cavaillon, J.-M.; Souza-Fonseca-Guimaraes, F. TLR-mediated activation of NK cells and their role in bacterial/viral immune responses in mammals. *Immunology and cell biology* **2014**, *92*, 256–262.

[19] Smits, C. H.; Li, D.; Patience, J. F.; den Hartog, L. *Animal nutrition strategies and options to reduce the use of antimicrobials in animal production*; FAO, 2021; Vol. 2021.

[20] Markowiak, P.; Śliżewska, K. The role of probiotics, prebiotics and synbiotics in animal nutrition. *Gut pathogens* **2018**, *10*, 1–20.

[21] Bearson, S. M.; Trachsel, J. M.; Bearson, B. L.; Loving, C. L.; Kerr, B. J.; Shippy, D. C.; Kiros, T. G. Effects of β-glucan on Salmonella enterica serovar Typhimurium swine colonization and microbiota alterations. *Porcine Health Management* **2023**, *9*, 7.

[22] Schat, K. Vaccines and vaccination practices: Key to sustainable animal production. *Encyclopedia of agriculture and food systems* **2014**, 315.

[23] Beikzadeh, B. Immunoinformatics design of multi-epitope vaccine using OmpA, OmpD and enterotoxin against non-typhoidal salmonellosis. *BMC bioinformatics* **2023**, *24*, 63.

[24] Coumar, M. S. *Molecular Docking for Computer-Aided Drug Design: Fundamentals, Techniques, Resources and Applications*; Academic Press, 2021.

[25] Jakhar, R.; Dangi, M.; Khichi, A.; Chhillar, A. K. Relevance of molecular docking studies in drug designing. *Current Bioinformatics* **2020**, *15*, 270–278.

[26] Fan, J.; Fu, A.; Zhang, L. Progress in molecular docking. *Quantitative Biology* **2019**, *7*, 83–89.

[27] Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols* **2009**, *4*, 1–13.

[28] Santhoshkumar, R.; Yusuf, A. In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of Curcuma longa. *Journal of Genetic Engineering and Biotechnology* **2020**, *18*, 1–9.

[29] Kleywegt, G. J.; Jones, T. A. Phi/psi-chology: Ramachandran revisited. *Structure* **1996**, *4*, 1395–1400.

[30] Terefe, E. M.; Ghosh, A. Molecular Docking, Validation, Dynamics Simulations, and Pharmacokinetic Prediction of Phytochemicals Isolated from Croton dichogamus Against the HIV-1 Reverse Transcriptase. *Bioinformatics and Biology Insights* **2022**, *16*, 11779322221125605.

[31] López-Camacho, E.; García-Godoy, M. J.; García-Nieto, J.; Nebro, A. J.; Aldana-Montes, J. F. A new multi-objective approach for molecular docking based on rmsd and binding energy. 2016.

[32] Hooft, R. W.; Sander, C.; Vriend, G. Objectively judging the quality of a protein structure from a Ramachandran plot. *Bioinformatics* **1997**, *13*, 425–430.

[33] Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins:"what you see" is not always "what you get". *Structure* **2009**, *17*, 489–498.

[34] Gilson, M. K.; Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

[35] Yamazaki, Y.; Naganuma, J.; Gotoh, H. A theoretical, dynamical evaluation method of the steric hindrance in nitroxide radicals using transition states of model reactions. *Scientific reports* **2019**, *9*, 20339.

[36] Patro, L. P. P.; Rathinavelan, T. STRIDER: steric hindrance and metal coordination identifier. *Computational Biology and Chemistry* **2022**, *98*, 107686.

[37] Kasmi, R.; Hadaji, E.; Chedadi, O.; El Aissouq, A.; Bouachrine, M.; Ouammou, A. 2D-QSAR and docking study of a series of coumarin derivatives as inhibitors of CDK (anticancer activity) with an application of the molecular docking method. *Heliyon* **2020**, *6*.

[38] Li, J.; Fu, A.; Zhang, L. An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences* **2019**, *11*, 320–328.

[39] Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS computational biology* **2015**, *11*, e1004586.

[40] others,, *et al.* Molecular docking study of fatty acids from Pliek U Oil in the inhibition of SARS-CoV-2 protein and enzymes. 2021.

[41] Maden, S. F.; Sezer, S.; Acuner, S. E. *Molecular Docking-Recent Advances*; IntechOpen, 2022.

[42] Mishra, N.; Forouzesh, N. *Algorithms and Methods in Structural Bioinformatics*; Springer, 2012; pp 1–16.

[43] Lins, L.; Brasseur, R. The hydrophobic effect in protein folding. *The FASEB journal* **1995**, *9*, 535–540.