JOANA ROCHA DA SILVEIRA BARRETO DE AGUIAR

Designing antivirals against rabies using molecular docking and protein modeling

São Paulo – SP

2021

JOANA ROCHA DA SILVEIRA BARRETO DE AGUIAR

Designing antivirals against rabies using molecular docking and protein modeling

Dissertação apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para a obtenção do título de Mestre em Ciências

Departamento:

Medicina Veterinária Preventiva e Saúde Animal

Área de Concentração:

Epidemiologia Experimental Aplicada às Zoonoses

Orientador:

Professor Dr. Paulo Eduardo Brandão

São Paulo – SP

2021

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO

(Biblioteca Virginie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo)

T. 4104
FMVZ

Aguiar, Joana Rocha da Silveira Barreto de
Designing antivirals against rabies using molecular docking and protein modeling / Joana Rocha da Silveira Barreto de Aguiar. – 2021.
138 f. : il.

Título traduzido: Design de antivirais contra raiva utilizando docagem molecular e modelagem de proteínas.

Dissertação (Mestrado) – Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2021.

Programa de Pós-Graduação: Epidemiologia Experimental Aplicada às Zoonoses.
Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.
Orientador: Prof. Dr. Paulo Eduardo Brandão.

1. Raiva. 2. *Docking* molecular. 3. Homologia. 4. Antivirais. 5. Proteína. I. Título.

Ficha catalográfica elaborada pela bibliotecária Maria Aparecida Laet, CRB 5673-8, da FMVZ/USP.

Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia
Universidade de São Paulo

São Paulo, 12th August 2021

CERTIFIED

We certify that the Research "Designing antivirals against rabies using molecular docking and protein modeling", protocol number CEUAx 3927180219 (ID 001077), under the responsibility Paulo Eduardo Brandão, agree with Ethical Principles in Animal Research adopted by Ethic Committee in the Use of Animals of School of Veterinary Medicine and Animal Science (University of São Paulo), and was approved in the meeting of day March 07, 2019.

Certificamos que o protocolo do Projeto de Pesquisa intitulado "Designing antivirals against rabies using molecular docking and protein modeling", protocolado sob o CEUAx nº 3927180219, sob a responsabilidade de Paulo Eduardo Brandão, está de acordo com os princípios éticos de experimentação animal da Comissão de Ética no Uso de Animais da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, e foi aprovado na reunião de 07 de março de 2019.

Prof. Dr. Marcelo Bahia Labruna
Coordenador da Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo

Camilla Mota Mendes
Vice-Coordenadora da Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo

FOLHA DE AVALIAÇÃO

Autor: AGUIAR, Joana Rocha da Silveira Barreto de

Título: Designing antivirals against rabies using molecular docking and protein modeling

Dissertação apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para a obtenção do título de Mestre em Ciências

Data: ____/____/____

Banca examinadora

Prof. Dr. _____

Instituição:_____ Julgamento:_____

Prof. Dr. _____

Instituição:_____ Julgamento:_____

Prof. Dr. _____

Instituição:_____ Julgamento:_____

DEDICATÓRIA

À minha família, orientador e amigos.

AGRADECIMENTOS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

RESUMO

Aguiar, J. R. S. B. **Design de antivirais contra raiva utilizando docagem molecular e modelagem de proteínas.** [Designing antivirals against rabies using molecular docking and protein modeling.]. 2021. 138 f. Dissertação (Mestre em Ciências) – Faculdade Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2021.

A raiva é uma doença zoonótica que afeta principalmente a população pobre de países em desenvolvimento. Embora exista uma vacina eficiente contra o vírus, muitas das pessoas nas áreas mais afetadas pela doença podem não ter conhecimento da vacina e de sua necessidade, não conseguem chegar a uma área onde as vacinas estão prontamente disponíveis, ou não podem pagar pelo alto custo da vacinação e da imunoglobulina. Este projeto teve como objetivo encontrar um tratamento potencial para a doença quando ela já atingiu seus estágios mais avançados, de modo que aqueles que não têm acesso à vacinação ainda tenham uma chance de sobreviver à esta doença letal. Este projeto teve como foco a capacidade da bioinformática de encontrar potenciais ligantes que pudessem inativar ou bloquear todas as cinco proteínas ou raiva, de modo que fossem incapazes de se ligar aos receptores do hospedeiro, ou de danificar o sistema imunológico dos pacientes. Todas as cinco proteínas da raiva de dezenove cepas diferentes (noventa e cinco proteínas) foram modeladas por meio de modelagem de homologia, e vinte e seis ligantes que passaram pela regra de cinco de Lipinski foram escolhidos. A primeira etapa de docagem foi docar todas as 95 proteínas por meio de uma docagem cega com cada um dos 26 ligantes para reduzir o número de ligantes e analisar potenciais sítios ativos. Após a conclusão de todos as docagens cegas, foram escolhidos dezessete ligantes para a docagem ativa, que considerou resíduos específicos (encontrados na literatura, ferramentas de bioinformática e análise da docagem cega) como potenciais sítios ativos de cada proteína. Os resultados da docagem cega também foram visualizados e analisados quanto às energias de ligação, bem como a localização das conexões. O número de vezes que cada um dos resíduos dos potenciais sítios ativos foi acessado também foi analisado. Os ligantes que tiveram os melhores resultados gerais foram reduzidos a quatro e tiveram suas estruturas e farmacologia analisadas no contexto da infecção por raiva. Os sítios

ativos potenciais também foram analisados e limitados aos sítios ativos mais prováveis para cada proteína.

Palavras-chave: Raiva. Docking molecular. Homologia. Antivirais. Proteína

ABSTRACT

Aguiar, J. R. S. B. **Designing antivirals against rabies using molecular docking and protein modeling** [Design de antivirais contra raiva utilizando docagem molecular e modelagem de proteínas.]. 2021. 138 f. Dissertação (Mestre em Ciências) – Faculdade Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2021.

Rabies is a zoonotic disease that mainly affects poor population in developing countries. While there is an efficient vaccine for it, many of those in the areas most affected by the disease can be unaware of the vaccine and their need for it, may be unable to reach an area where the vaccines are readily available, or may be unable to pay for the high cost of vaccination and immunoglobulin. This project aimed to find a potential treatment for the disease when it has already reached its later stages, so those who can't access vaccination still have a chance of surviving this lethal disease. This project focused on the bioinformatics capability of finding potential ligands that could inactivate or block all five proteins or rabies, so that they would be unable to bind to host receptors, or to damage the immune system of patients. All five proteins of rabies from nineteen different strains (ninety-five proteins) were modeled through homology modeling, and twenty-six drug-like ligands that passed Lipinski's rule of five were chosen. The first docking step was to put all ninety-five proteins through a blind docking with each of the twenty-six ligands to narrow down the number of ligands and analyze potential active sites. After all the blind dockings were concluded, seventeen ligands were chosen for the active site docking, which also considered specific residues (found from literature, bioinformatic tools, and analysis of blind docking) as potential active sites of each protein. The blind docking results were also visualized and analyzed for both the binding energies of each binding as well as the location of the connections. The number of times each of the residues from the potential active sites were accessed were also analyzed. The ligands that had the best results overall were narrowed down to four and analyzed both for their structure and pharmacology in the context of rabies infection. Potential active sites were also analyzed and narrowed down to the most likely active sites for each protein.

Keywords: Rabies. Molecular docking. Homology. Antivirals. Proteins

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

Rabies is a zoonotic disease caused by the rabies virus (RABV) that leads to an illness that is about 99% deadly. Specifically, it affects the brain and often causes fatal encephalomyelitis. It can also incur a host of neurological symptoms that are characteristic of rabies, such as hydrophobicity and sensitivity to light, as well as more common symptoms such as seizures, confusion, and fever. While there is a vaccine for rabies that could help both in pre- and post-exposure (if taken during the correct window of time), there is no treatment available that has proved to be effective. While the combination of five doses of vaccine and one of immunoglobulin saves the lives of many, RABV still causes approximately 59,000 human fatalities per year, especially in continents like Asia and Africa (WORLD HEALTH ORGANIZATION, 2021). However, this number fails to address that a large portion of rabies deaths go underreported, especially in areas with little resources or knowledge of what this disease is and the need to seek medical help (BAWASKAR, BAWASKAR e BAWASKAR, 2017). Globally, more than 29 million people receive post-exposure vaccination per year, preventing an even larger number of rabies fatalities (WORLD HEALTH ORGANIZATION, 2021).

Transmission of rabies generally occurs through the transfer of saliva from the bite of a mammal, usually a dog or bat. RABV then enters the central nervous system (CNS) by going from the muscle area of entrance to axons and using axonal transport to reach the brain (DU PONT, PLEMPER e SCHNELL, 2019). From there, it wreaks havoc in the brain and causes death quickly. Many of the specifics on how RABV proteins affect the brain, and where (or if) they bind is still unclear, as well as to how exactly it evades the immune system.

Molecular docking, protein modeling and protein homology are bioinformatic tools that can be used as a steppingstone for more in depth *in vitro* and *in vivo* tests. These tools can help to give a closer look into rabies' five proteins and find ligands that could inactivate their active sites and interrupt their activities in a patient's body. Molecular docking, specifically, works by predicting in which position a specific ligand is going to bind to a protein and how strong (or weak) this binding will be (PAGADALA, SYED e TUSZYNSKI, 2017). The lower the binding affinity, the better the ligand is at docking on the protein's active site. Medicine against certain viruses, such as HIV and

Influenza, have both been found through molecular docking (followed by *in vitro* and *in vivo* trials) or are the focus of new molecular docking studies (PHILIPS e STEWART, 2018) (ZHANG, ZHAO, *et al.*, 2021) (ADEBAMBO e GUNARATNAM, 2018).

1.1 Neglected Tropical Diseases

Neglected tropical diseases (NTDs) are a group of seventeen (CENTER FOR DISEASE CONTROL AND PREVENTION, 202) diseases that mainly affect underdeveloped, tropical areas (ENGLES e ZHOU, 2020). These diseases, because of the locations and population they damage, often do not get enough funding or research on new medications or vaccines that could advance either diagnosis, avoidance, or treatment of the disease. NTDs are thought to impact over 1 billion people (WHO), and are the cause of most infections in underdeveloped regions of Africa, Asia and Latin America, mainly because of tropical weather, where these infections thrive, and a lack of resources and information.

Thanks to where these diseases happen and cause most cases and deaths, less money and time is devoted to research about them. Developed countries, in their majority, have already managed rabies through mass vaccination of wildlife, as well as big household pet vaccination campaigns. There is also a wider knowledge of rabies and the need for vaccination in case a bite does happen.

Even with a small amount of cases per year, in the United States of America the annual cost for rabies prevention is US$300 million (CENTER FOR DISEASE CONTROL AND PREVENTION, 2020). Even with great wildlife and household pets' vaccination efforts, more than 50,000 Americans get post-exposure prophylaxis each year, with vaccines and rabies immune globulin pricing at about $3,800 (CENTER FOR DISEASE CONTROL AND PREVENTION, 2019). In some areas, the price of vaccines can be prohibitively expensive, with the full treatment being around $40 in places in Africa, where the daily income might be $1-2 (DU PONT, PLEMPER e SCHNELL, 2019).

Aside from the vaccines being expensive, they also need to be maintained in certain temperatures and under certain conditions, complicating its wide use in developing countries. Further, the vaccine is only effective in post exposure situations during a specific window of time (prior to the infection reaching the central nervous system) after

a bite. In areas where transportation and physical mobility are a problem, this window of time hinders people's abilities to reach a hospital or vaccination clinic on time.

1.2 Antivirals Against Rabies

While having a vaccine for such a lethal disease is important and necessary, it also comes with the problems mentioned in the previous section. Because of those issues, the scientific community has been looking for an antiviral for rabies, which would help thousands of individuals all over the world, but especially in less developed areas. Ideally, an antiviral would be able to fight against the infection after it already reached the central nervous system (CNS), which would add to the vaccine abilities, since the latter needs to be applied before the infection has reached its critical peak.

Several different approaches have been taken to find an antiviral for rabies that could be of help to the already existing vaccine and could also treat the disease when in a more advanced stage. Being able to affect the disease in its advanced stage means that any treatment against rabies would have to be effective in the CNS, which complicates matters as it would also have to cross the blood-brain barrier (BBB).

One way of circumventing the blood-brain barrier issue is transfecting antibodies straight to the brain of a patient. Studies have shown that transfecting anti-RABV antibodies into the brain can hinder or completely prevent dissipation of the virus in the brain, where the virus causes the most damage (AGOSTINHO e BRANDAO, 2020). This technique could be incredibly helpful since it would be of use after the vaccination window has already closed. However, transfecting antibodies has shown to be effective only after the first few days of symptoms (or of when the virus has reached the brain) and may be useless throughout the latter portion of the infection.

Repurposed drugs have been one of the main techniques used to find a potential antiviral. Ribavirin, an antiviral used in respiratory syncytial virus (RSV) infection that inhibits production of viral messenger RNA, has been researched and shown *in vitro* potential against rabies infection. *In vivo,* however, it seems ribavirin shifts immune response to an inflammatory Th1-type response, while rabies needs a balanced Th1/Th2 immune response (APPOLINARIO e JACKSON, 2014).This means that ribavirin could suppress antibody production, which helps the clearance of rabies from the patient to a certain degree.

Favipiravir is another type of repurposed drug that has been researched in the context of rabies. However, while it has interesting mechanisms of action and did improve symptoms and signs of rabies infection in mice, it doesn't seem to have any effect in the CNS, which is of the utmost importance considering that is where rabies infection really does its harm (JOCHMANS e NEYTS, 2019).

Amantadine is another known antiviral that has been repurposed for rabies treatment. It inhibits viral M2 protein, hindering viral uncoating and, *in vitro*, it showed inhibitory effects against RABV. Aside from its antiviral capabilities, amantadine may also have a neuroprotective role. No research has been done *in vivo*, although five patients who went through the Milwakee Protocol in 2006 and 2007 did receive amantadine and all five succumbed to the disease (NIGG e WALKER, 2009).

Following the path of the previous antivirals, favipiravir had similar results in research. Favipiravir is an RNA virus inhibitor that has shown effect against Ebola virus, which is in the same order as rabies (Mononegavirales) and, during *in vivo* research in mice, it did lower titers of the virus. However, the survival rate of the animals was not high, and they developed limb paralysis (DU PONT, PLEMPER e SCHNELL, 2019).

Antivirals are not the only compounds that can be repurposed. Interferon-alpha (IFN-α) works as a defense mechanism by the body, slowing down the spread of rabies and delaying mortality in mice. In cells, IFN-α has shown to prevent viral replication; however, in several patients, this effect was not seen and, whenever used, IFN-α has not slowed down disease or prevented death (APPOLINARIO e JACKSON, 2014).

While working with repurposed drugs is more straightforward and, therefore, more prevalent, there are other compounds and molecules that can be used as antivirals. An example of that are short-interfering RNAs (siRNAs), double-stranded RNAs that, operating through the RNA interference (RNAi) pathway, can trigger mRNA degradation. *In vitro*, siRNAs were able to inhibit rabies virus replication, a result that has also been seen for different viruses, such as Hepatitis C and HIV (BRANDÃO, CASTILHO, *et al.*, 2007) (APPOLINARIO e JACKSON, 2014).

Most of the antiviral research against rabies infection has shown less than promising results, especially in terms of repurposed antivirals. While some of these antivirals often work for a range of viruses, rabies seems to never fit in that equation. Other compounds, such as interferons and siRNAs, while showing potential, don't have

a lot of previous research or known mechanism of action, making them less interesting than potential ligands that would already follow certain rules and specifications. Transfecting antibodies is a helpful technique, but because it may only be effective during a specific part of the infection, it might be a great resource to pair with the use of antivirals, so that they could be used throughout the infection.

1.3 Virus Structure

RABV is a member of the order Mononegavirales, in the Rhabdoviridae family, in the *Lyssavirus* genus. Its genome is a single, negative-stranded RNA molecule of about 12kb, which encodes for five proteins: the matrix protein (M, with 202 amino acids), the glycoprotein (G, with 505 amino acids), the phosphoprotein (P, with 297 amino acids), the nucleoprotein (N, with 450 amino acids), and the large structural protein, or RNA-dependent RNA polymerase (L or RdRp, with 2130 amino acids) (ALBERTINI, RUIGROK e BLONDEL, 2011). Each one of them plays an important role in either virus entry in the host, virus replication once inside, or virus pathogenicity. RABV also has two structural components, a ribonucleoprotein core (RNP) and an envelope.

Rabies virions show a bullet-like shape, with the viral RNA surrounded by the nucleoprotein, forming a nucleocapsid. This nucleocapsid is also associated with an amount of phosphoprotein, with some carrying copies of the large protein. This large protein, together with the RNA, N and P proteins form the ribonucleoprotein, or the RNP, the component involved in replication and transcription. Figure 1 shows a schematic of the virus with its proteins and components.

Figure 1 - Schematic of the Rabies Virion

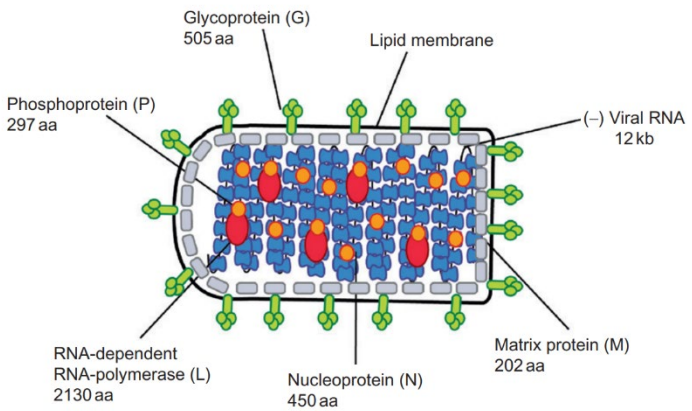


Figure 1 shows a schematic of the rabies virion and its components, with the length of each protein (in amino acids) shown. The glycoprotein surrounds the viral membrane, with the matrix protein right below the membrane. The nucleoprotein is associated with the viral RNA, and forms the ribonucleoprotein with P and L. Image from (ALBERTINI, RUIGROK e BLONDEL, 2011).

1.3.1 Proteins

The proteins encoded by RABV are of utmost importance for this project. Each one of them has a role to play in the pathogenicity of the virus, and each may be a target for antivirals. Understanding both the role the proteins play in the viral cycle and their structures is necessary since molecular docking will focus on structure to find a potential antiviral, while possible pharmacological capabilities may also be necessary for an antiviral to work as it should.

The glycoprotein is the only protein exhibited on the surface of RABV, and it binds the virus to a cellular receptor in the host. RVG is a trimer and the main contributor to pathogenicity, as it has a role in virus entry (interacting with receptors, such as the nicotinic acetylcholine receptor (YANG, LIN, *et al.*, 2020)), virus fusion and virus spread (mediating viral entry to the nervous system) in the CNS (TOMAR, SINGH, *et al.*, 2010). It is also involved in trans-synaptic spread when already in the CNS. G is also the only RABV protein that induces an antibody response, as it is the target for neutralizing antibodies.

The matrix protein is necessary for viral assembly and budding, as well as for covering the RNP and maintaining its condensed shape (MEBATSION, WEILAND e CONZELMANN, 1999). In terms of shape, the matrix protein also determines (or helps determine) the bullet-like shape of the virion (MEBATSION, WEILAND e CONZELMANN, 1999). It has been shown as well that the matrix protein targets mitochondria and, in the late stages of infection, induces apoptosis in neuronal cells, which allows for viral dissemination (ZAN, LIU, *et al.*, 2016). Finally, matrix proteins have an important role in regulating viral transcription and replication from RNPs. With M binding to RNPs and changing their shape, making RNPs tightly coiled, it’s possible the coil doesn’t work as a template for RNA synthesis, possibly halting both transcription and replication. M may also have a role as a regulatory protein, balancing replication and translation, inhibiting transcription and simultaneously stimulating replication (FINKE, MUELLER-WALDECK e CONZELMANN, 2003).

P is a regulatory protein that has an important pathogenicity role as it inhibits different host immune responses, being an interferon antagonist (VIDY, CHELBI-ALIX e BLONDEL, 2005), and interacts with mitochondria complex I, causing oxidation of

electrons and blocking certain mitochondrial processes vital for human life. Further, P has a role in viral transcription, being a transcription factor for the L protein, providing a connection between L and the RNP (FISHER, STREICKER e SCHNELL, 2018). In replication, the P protein forms a complex with N and makes it replication-competent, so that the nucleoprotein can encapsidate nascent RNA during replication (GUPTA, BLONDEL, *et al.*, 2000).

N is a nucleocapsid protein that binds to viral RNA and encapsidates it, resulting in the ribonucleoprotein complex (RNP), which is necessary for viral transcription and replication and protects the RNA from degradation (KOUZNETZOFF, BUCKLE e TORDO, 1998) (FISHER, STREICKER e SCHNELL, 2018). N is also connected to pathogenicity of the virus by downplaying the host's interferon activity and escaping activation of RIG-I, genes involved in virus recognition (MASATANI, , *et al.*, 2010).

Finally, L is an RNA-dependent RNA polymerase (RdRp) and has replicase activity, being involved in transcription, capping, and polyadenylation of viral mRNAs. In order to have activity as a RdRp and provide its functions, L must bind with its co-factor P, making a ribonucleoprotein (RNP) (NAKAGAWA, KOBAYASHI, *et al.*, 2017).

Rabies, therefore, necessitates all five proteins to accomplish its viral functions in a host body. With that it in mind, it is important to work towards a viral therapeutic design that takes all five proteins, its active sites, and how they might achieve pathogenicity into consideration.

1.4 Viral Cycle

The first step of the viral cycle of the rabies virus is to enter a host. The glycoprotein binds to a cell surface at a neuromuscular junction, using either an acetylcholine receptor, a phosphatidyl serine receptor, a neuronal cell adhesion molecule or P75 neurotropin receptor. After attachment, the virus enters the host cell through endocytosis, where the acidic environment of the endosome incites a conformational change in G that mediates fusion of the cellular membrane with the viral envelope (FISHER, STREICKER e SCHNELL, 2018). Post-fusion, the RNP is released (uncoating step) in the cytoplasm and can now be used as template for viral gene replication (YOUSAF, QASIM, *et al.*, 2012).

During transcription, a leader RNA and 5-capped and polyadenylated mRNAs are made. During replication, full-length nucleocapsids in antigenome-sense RNA are

made, and become templates for genome-sense RNA synthesis. Both of those are encapsidated by the nucleoprotein, and this new RNP can either serve as template for secondary transcription or go to the cell membrane to assemble with matrix and glycoprotein for the budding of virions (FISHER, STREICKER e SCHNELL, 2018). Figure 2 shows these steps.

Figure 2 - Schematic of the Viral Cycle of RABV

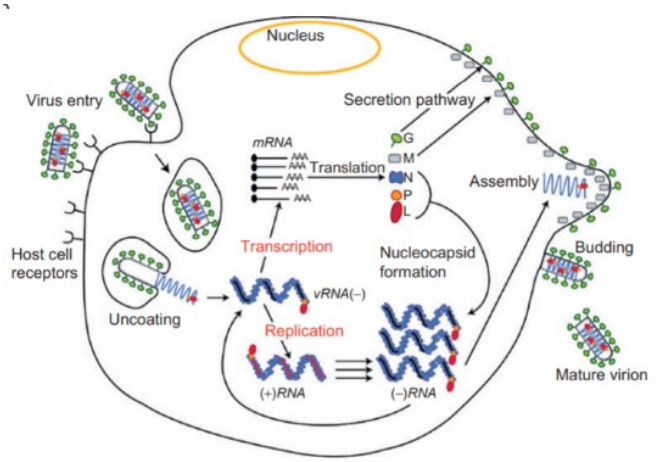


Figure 2 shows the viral cycle of the rabies virus, starting with virus entry, followed by endocytosis and uncoating of the RNP, and then transcription and/or replication. Figure from (FISHER, STREICKER e SCHNELL, 2018)

1.5 Bionformatics

Bioinformatics is a science interested in joining biology and informatics to analyze biological information, as well as discover more about protein or gene functions, three-dimensional shape of proteins and molecules, and the structural relationship between molecules. In drug design and discovery, bioinformatics can speed up the rate at which scientists find new drugs, learn how proteins will interact with these drugs, as well as repurpose already known drugs with known capabilities and toxicity (XIA, 2017).

1.5.1 Molecular Docking

Molecular docking is a tool for drug discovery. It is a bioinformatic resource that would come before any *in vitro* research, in order to potentially narrow down or select compounds to be used in further laboratorial analysis. Molecular docking can be used to find out more about the interaction between a protein and ligand, for example, which can make it easier to analyze both structures and the function of these structures.

AutoDock4.2 (MORRIS, HUEY, *et al.*, 2009) is one of a few software that can be used to predict the shape and binding of biomolecular complexes. It is useful in

research as it is a cheap and accessible method and can be the first step of drug discovery to both eliminate ligands that may not be useful and narrow down the scope of an *in vitro* and *in vivo* drug discovery research.

AutoDock uses the Lamarckian Genetic Algorithm (LGA) and a scoring function called empirical free energy to make its calculations and get a confident result for the ligand-receptor binding if the ligand has around 10 flexible bonds (MORRIS, GOODSELL, *et al.*, 2012). More than that and the software will not be able to run the docking and give accurate results.

Genetic algorithms are methods of solving optimization problems through the use of the genetic model of evolution. This means that these algorithms use concepts related to genetics. AutoDock's algorithm establishes state variables, which is set by values that describe conformation, translation, and orientation of the ligand-protein arrangement. Each state variable is considered a gene, with the ligand's state being considered a genotype. The phenotype is the atomic coordinates of each ligand. With these values, AutoDock sets a population, which is a number of ligand poses that are optimized and considered for the docking. Individuals in this population can go through crossover, when random pairs mate and their offspring can inherit new genes (new state variables), and through random mutation (one gene, or variable, changes). Selection of the best pose depends on fitness, or the total internal energy of the complex post-docking. In this way, at each generation there is a fight to choose the best option to move on to the next (natural selection) (MORRIS, GOODSELL, *et al.*, 2012). AutoDock also sets a "number of generations" and a value for "maximum evaluations," the lower of the two will determine when the docking will end. The Lamarckian aspect of the algorithm makes it so whenever a conformation finds a local minimum, this knowledge is passed on to the next conformation (MORRIS, HUEY, *et al.*, 2009). While AutoDock offers other types of algorithms, the LGA has shown to be the best in finding the lowest energy the system can acquire (AutoDock, 2006). In terms of the binding energy, AutoDock uses an energy force field based on a thermodynamic model which evaluates energy of the bound and unbound states of the protein and ligand (MORRIS, HUEY, *et al.*, 2009) to get a final value.

Plus, for the process to work as it should, AutoDock requires both the ligand and protein to be in PDBQT format, which includes partial charges (Q, a charge created due to the distribution of electron in bonds, which can often by asymmetrical, as in

polar covalent bonds) and atom types (T) that AutoDock recognizes. It also adds Gasteiger partial charges, merges non-polar hydrogens, and detects aromatic carbons. PyRx makes it easy for this to be done by simply clicking on the PDB ligand or protein (format from ZINC15 and from homology modeling) and converting it into a PDBQT ligand.

After preparation of both ligands and proteins (all went through energy minimization) and conversion of the files, AutoDock runs them through the first phase, an AutoGrid calculation. This step precalculated affinity potentials for all atom types in the ligand, which is done by inserting the protein into a grid and a probe atom goes around each grid point, checking for the energy of interaction of this atom with each area of the protein (MORRIS, GOODSELL, *et al.*, 2012). These grids are made for each type of atom that can be found in the molecule, and a sort of map of the energy is formed that the ligand can be measured against. In the actual docking phase, the binding energy of a specific ligand pose is calculated based on the values acquired from the AutoGrid phase. The ligand goes through different poses around the specified area of the protein (the whole protein in the binding docking, and specific areas in the active docking) and, following the algorithm and free binding energy methods already explained, the ligand finds configurations that have the lowest binding energies.

This project uses both blind docking and active docking in different stages of the process. Blind docking is used to narrow down the ligands that bind to each of the five proteins, and the active docking was used afterwards to not only narrow down the search even further, but also to understand the structure of active sites of each of the proteins.

Figure 3 shows an illustrative example of a blind docking, with the square search space using the whole protein (in this case, a nucleoprotein). Figure 4 shows an active docking, with the square search space focusing on specific residues and a specific area of the protein (in this case, a nucleoprotein). The square search space delimits where the ligand will search for potential binding residues.

Figure 3 - Blind Docking Example (Nucleoprotein)

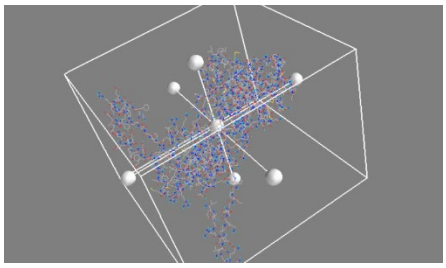


Figure 3 - Blind docking of the nucleoprotein, with the square box encompassing the whole search space. Image made on PyRx (2021).

Figure 4 - Active Docking Example (Nucleoprotein)

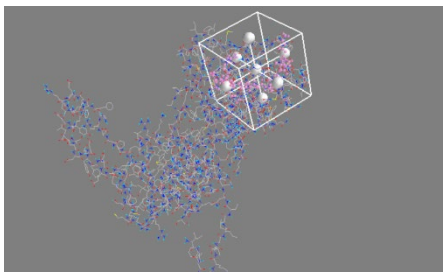


Figure 4 - Active docking of the nucleoprotein, with the square box encompassing only specific residues of the search space. Image made on PyRx (2021).

1.5.1.1 Ligands and Proteins

A ligand is a molecule that can bind to a protein molecule (a receptor). When a complementary ligand binds, it can change the shape of a protein, or induce a cellular response in the protein and organism. Ligands come in all shapes and sizes, but those most often used for drug design consider the Lipinski rule of five, a set of norms that state that an orally active drug can't violate more than one of the following criteria: less than 5 H-bond donors and 10 H-bond receptors, molecular weight smaller than 500, and a Log P of less than 5 (LIPINSKI, LOMBARDO, *et al.*, 1996).

Proteins can act as receptors of ligands. With the rabies virus having five proteins, and all five facilitating either virus entry, virus replication, or virus pathogenicity, all five were considered as potential receptors for a ligand to inactivate their functions. To get an accurate representation of all five proteins from different strains (not all have been crystallized yet), as well as accurate depiction of their potential active sites, different programs had to be used.

1.5.1.2 Homology Modeling

For the proteins to be used in molecular docking, they needed to have a 3D-elucidated molecule. While some of the strains used in this project had already been crystalized, that wasn't true for all of them, so a bioinformatic program had to be used to model each protein. Homology modeling can construct a protein structure model from its sequence of amino acids through matching of the structure with other 3D targets of similar proteins.

Homology modeling is only truly successful when there is a significant relationship between the sequence of interest with sequences of other, similar proteins or compounds. This made sense as there is a number of viruses related to RABV, such

as the vesicular stomatitis virus, and other lyssaviruses. With the programs chosen for this project, point mutations and possible deletions were conserved in the final model, regardless of which protein was chosen for the modeling, so that the 3D-RABV proteins were the most accurate as possible.

The program chosen to do homology modeling, Phyre^2 (Protein Homology/Analogy Recognition Engine V 2.0) (KELLEY LA, 2015), gathers the inputted sequence (in this case, amino acid sequences) from the user, aligns it with similar sequences which have their structures already elucidated, and then uses these alignments to find the most likely structure for the unknown protein based on the structure of a known 3D protein through homology modeling (KELLEY LA, 2015).

Another program, I-Tasser (ROY, KUCUKURAL e ZHANG, 2010), can also predict protein structure from its amino acid through homology modeling and was used to double-check the structures. I-Tasser can also extrapolate the function of the protein or of specific areas of the protein by comparing the protein with other known 3D protein structure, which can be important for deducing active sites, for example.

2. OBJECTIVES

The general objective of this project was to find ligands that could act as antivirals for the treatment of rabies using molecular docking and bioinformatics techniques, followed by the evaluation of which were the most efficient ligands for binding with each of the proteins, the potential active site of each protein, and an effort to understand what about these ligands worked so well, both structure and pharmacologically.

In order to do that, specific objectives also needed to be met:

1. To use homology modeling to get an accurate 3D-structure model of each of the five RABV proteins.

2. To identify, using protein databases, ligands that could have the ability of binding to the active sites of each of the five RABV proteins and to analyze which of the ligands had a better binding affinity to each of the five proteins.

3. MATERIALS AND METHODS

All five proteins of rabies (phosphoprotein, nucleoprotein, large protein, matrix protein, and glycoprotein) from nineteen different strains (Table 1) were docked through molecular docking with 26 different ligands in two rounds. The first round was a blind docking, a docking that does not focus on active sites, and the second round used potential active sites for each protein.

3.1 Proteins

The first step to prepare for the docking was to decide which strains of rabies would be used. It was important to use as many strains of rabies as possible, as they may have differences in its sequences and structures which could be relevant to understanding how one or more ligands interact with the proteins. These strains were chosen based on which had full genomes available, as well as based on how often they are found in both research and in the wild.

After deciding on nineteen relevant strains of rabies, it was necessary to acquire a 3D representation of all the five rabies proteins from those strains. This was done by employing homology modeling. The sequences from each of the nineteen proteins used were downloaded from UniProt, the Universal Protein Resource (CONSORTIUM, 2019). A list of all the strains used, as well as the countries where they are most found and their accession numbers from UniProt and NCBI are in Table 1. The sequences were downloaded in FASTA format, as a sequence of amino acids. The FASTA documents hold information about the amino acid sequence, and not structural information about the proteins, such as position of each amino acid, position of each atom, location of alpha helices and beta sheets, among others.

Table 1 - Chosen Strains of RABV

Strain Names	Country	Accession Number
Callithrix jacchus	Brazil	KM594025
Cerdocyon thous (Antigenic variant 2)	Brazil	KM594039
Mexico/DRV	Mexico	HQ450386
Desmodus rotundus (Antigenic variant 3)	Brazil	KM594042.1
Eptesicus furinalis	Brazil	KM594029

ERA	Fixed Strain	EF206707
HEP-Flury	Fixed Strain	AB085828
India	India	AY956319
China/MRV	China	DQ875050
Myotis nigricans	Brazil	KM594032
Nishigahara	Fixed Strain	AB044824
Nyctinomops laticaudatus	Brazil	KM594036
PM1503/AVO1	Laboratory Strain	DQ099525
Pasteur Vaccine	Fixed Strain	M13215
raccoon	United States of America	KY026482.1
Street Alabama Dufferin (SAD B19)	Fixed Strain	EF206715
Silver-haired bat (SHBRV)	United States of America	AY705373
Tadarida brasiliensis	Brazil	KM594038.1
Vulpes vulpes	Russia	KC595283

Table 1 shows the strains used for this project, as well as where they're commonly found and their accession numbers from UniProt.

3.1.1 Homology modeling and Energy Minimization

Using the FASTA documents from UniProt (CONSORTIUM, 2019), all sequences of amino acids from each protein from each strain were inputted on the Phyre^2 (KELLEY LA, 2015) website, which ran each one of them and came up with the best homology model for each. Rabies strains are generally well-conserved in their sequences, although most have at least one point-mutation when compared to one another.

For most strains of the same protein Phyre^2 returned the same result of homology modeling. This means that the model structure for each protein (regardless of its strain) was the same. However, point mutations and possible deletions were conserved in the final model, so while the structure in terms of alpha helices and beta sheets remained the same, the different residues that exist in different strains also remained the same, making it possible to visualize possible differences (or similarities) in terms of where and how ligands were bound to proteins of different strains.

With all the proteins chosen, the next step was to prepare them for docking. In molecular docking, energy minimization for compounds is important because the drawn molecules (especially if gotten through homology modeling, which is the case for all proteins in this study) may not be in the most stable shape. Energy minimization will make it so the protein will achieve a better and more accurate shape, with less torsion and more space for the bonds. All the proteins had their energy minimized with Swiss-PDBViewer (GUEX e PEITSCH, 1997), a software that allows for visualization and analysis of different proteins. It also has many tools that deal with energies and force fields, allowing the user to compute electrostatic potential, minimize energy, fix sidechains, compute energy, among others. Swiss-PDBViewer uses a force field that can check the energy of the protein and fix any parts of the geometry of the protein that might not be in its most beneficial shape. Figure 5 shows the difference between a compound with no energy minimization, and one that has gone through the process using Swiss-PDBViewer.

Figure 5 - Difference Between Molecules Before and After Energy Minimization

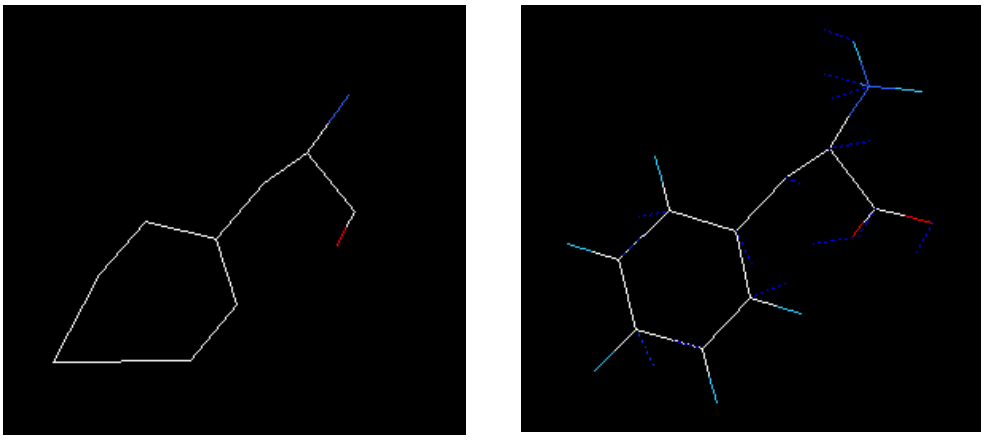


Figure 5 - figure shows compound without energy minimization (left) and after going through energy minimization (right). Image retrieved from http://www.pdg.cnb.uam.es/cursos/Leon_2003/pages/visualizacion/programas_manuales/spdbv_userguide/us.expasy.org/spdbv/text/energy.htm (2021).

In the case of all proteins modeled with the program they all had at least a 70% homology result with known rabies structures. Large Proteins had around 98% of coverage (98% of the molecule was modeled), with 100% confidence. Matrix proteins had around 80% coverage, with 100% confidence. Glycoproteins had around 70% coverage with 100% confidence. Nucleoprotein had around 90% coverage with 100% confidence. Finally, phosphoprotein had only 37% coverage with 100% confidence. There were homology options that had larger coverage for the residues of phosphoprotein, but the confidence was lower. Because of this low confidence, it was

decided to stick with the model that had very low coverage but 100% confidence. All these results were gathered from Phyre[2].

3.2 Ligands

This project was interested in finding possible ligands that, bound to one of the proteins of rabies, could inactivate the protein or stop it from fulfilling its function. It was important to have both a good variety and quantity of ligands but having in mind specific parameters that would make ligands more likely to function as drugs in the central nervous system, where rabies does its damage.

After securing 3D structures of all five proteins for the nineteen strains of RABV, relevant ligands were researched. The website ZINC15 (STERLING e IRWIN, 2015), a ligand library, was used for that purpose. To filter for potential ligands that could fulfill the needs of the project, considering relevant factors applicable to the disease, the 'Tranches' tool in ZINC15 was used. The first filter was to choose only 3D ligands, so they could be downloaded in PDB format, the format necessary for the subsequent docking. Only ligands considered to be "drug-like" were chosen, so they already had properties such as high potency, aqueous solubility, and hydrophobicity, among others (SCHNEIDER, 2000-2013).

To fulfill the basic requirement to cross the blood-brain barrier (BBB), filters such as molecular weight (375MW) and logP value (2.0), which measures how hydrophilic or hydrophobic a molecule is (higher than 0 means the molecule is more hydrophobic), were set (MIKITSH e CHACKO, 2014). Finally, the last filter was pH value, which was set as "Referenced", meaning the ligands had their dominant form at pH 7 (STERLING e IRWIN, 2015). After filtering for those, the number of ligands found by ZINC15 was 23,877,978 ligands.

A cutoff of 25 ligands (plus acetylcholine, explained below) was chosen for the fulfillment of this project. Out of the many ligands ZINC15 found, several of them were then analyzed one by one. This analysis now focused on even further requirements needed for the ligand to cross the BBB, such as H-bond donors, H-bond acceptors, parametrical polar surface area (tPSA), and number of rotatable bonds (MIKITSH e CHACKO, 2014). The values used for this filtering were: less than 5 H-bond donors, less than 10 H-bond acceptors, tPSA up until 100, and less than 8 rotatable bonds. These values are based on Lipinski's rule of five, which gives parameters for drugs that

can pass through the BBB via passive diffusion (LIPINSKI, LOMBARDO, *et al.*, 1996). Table 2 shows the ZINC codes for each ligand (except acetylcholine receptor, which has a RCSPDB code), the A-number given to each to facilitate the docking process, as well as names for any of the ligands that are known and/or are already used in medications and research.

The nicotinic acetylcholine receptor is a known receptor for the rabies virus (specifically for the glycoprotein). The specific residues of the receptor that may connect to the virus and help it establish is pathogenicity have been researched, and residues 179 to 192 of alpha 1-subunit of the receptor seems to be the potential site for where the virus binds (LENTZ, 1990). The full receptor cannot be docked to the virus, as it is too big and does not follow the same parameters of protein-ligand docking. Considering the specific residues that probably bind to the glycoprotein are known, the crystallization of acetylcholine receptor (PDB ID: 2BG9) (UNWIN, 2005) (BERMAN, WESTBROOK, *et al.*, 2000) was used, and then it was cut up into only the relevant residues using PyMOL (SCHRÖDINGER, LLC), a visualization software that has the capability of separating portions of a molecule, among other tools that will be explained further in this project.

Table 2 - Chosen Ligands

ZINC Number	Ligand Number in project	Name (if any)
ZINC538273	A1	Ofloxacin
ZINC430	A2	Lorapride
2BG9 (modified)	A3	Acetylcholine receptor (residues 179-192)
ZINC581	A4	
ZINC649	A5	
ZINC1622	A6	
ZINC1773	A7	
ZINC1894	A8	
ZINC2016	A9	
ZINC10698	A10	
ZINC538328	A11	Rufloxacin

ZINC538156	A12	Pelanserin
ZINC246272	A13	
ZINC309694	A14	
ZINC537983	A15	
ZINC584971	A16	
ZINC587083	A17	
ZINC587827	A18	
ZINC591669	A19	
ZINC595006	A20	
ZINC596282	A21	2,3 dehydro ofloxacin
ZINC601286	A22	
ZINC601298	A23	
ZINC605248	A24	
ZINC605274	A25	
ZINC13915654	A26	Favipiravir

Table 2 showing ligand codes from ZINC, as well as ligand A-numbers used for the project and names of any known ligands.

3.3 Docking

With the 26 ligands selected, the docking process was initiated. The first round of docking used all five proteins (glycoprotein, phosphoprotein, large protein, matrix protein, and nucleoprotein) from each of the nineteen strains and docked each with each of the 26 ligands, resulting in 2,470 dockings. This first type of docking is called blind docking, as it uses the whole protein instead of just focusing on a specific site. The second step of docking was the active docking, which focused on specific areas of each protein (potential active sites) as the search area for the ligands. Both phases of docking used the same or similar docking, visualization, and analysis software that will be laid down below.

3.3.1 PyRx

PyRx (DALLAYKAN e OLSON, 2015) is a software which packages many docking and visualization programs into one to make docking more efficient, accessible, and less time consuming. It includes a docking program called AutoDock4.2 (MORRIS, HUEY, *et al.*, 2009), a docking software widely used for predictions of energy between ligand and protein in bound conformations. Except for

ligand A3 (modified acetylcholine receptor), all dockings for this project were completed using AutoDock4.2 inside the software PyRx (DALLAYKAN e OLSON, 2015). For ligand A3, AutoDock VINA (TROTT e OLSON, 2010) was used, another docking program that is packaged in PyRx. The differences between these two and reasons why they were used will be outlined below.

3.3.1.1 AutoDock

For docking in AutoDock, all parameters (for the algorithm and the binding energies) were left in their standard configuration during the dockings in this work, with 10 runs of the docking, 150 individuals in the population, 250000 evaluations, a 27000-maximum number of generations, with only the top 1 surviving to the next generation. The rate of gene mutation is 0.02 and the crossover rate is 0.8. Generally, more runs are used per ligand-protein docking; however, since each ligand was docked to nineteen similar strains of the same protein, the number of each docking was kept at 10 since the comprehensiveness of this project minimized the possibility of getting incomplete results.

The results of each docking were saved in an excel file, and these results included ten values of binding energy, (the value that is relevant for this project, as it measures how well the ligand binds to the protein), as well as ten values of intermolecular energy (energy between atoms that are not bonded and separated either in the same molecule or between different molecules), internal energy (energy between atoms in the ligand or in the protein), torsional energy (energy needed to overcome torsional strain), and unbound energy (which are all the values needed to calculate the binding energy). Binding energy averages of each strain-ligand compound were found (for example average of Glycoprotein strain PV with ligand a1), as well as overall averages of a ligand (for example average of all Glycoproteins of all strains with ligand a1). The overall average was used to decide which ligands would be useful for the second round of docking (using active sites). The cutoff for the overall averages was that anything at or smaller (more negative) than -6.0 had a good enough result in the blind docking that it could yield useful results with the active site docking as well. A value that is -6.0 or more negative is a cutoff found to separate active from inactive drugs (SHITYAKOV e FORSTER, 2014), and in a pre-project docking testing, known antivirals (such as acyclovir) acquired averages of -6.0 or higher (more positive) with the virus they interact with.

When a docking is finished, aside from the results aforementioned, two documents are also outputted from AutoDock. One is a .dlg document, which can be read using a NotePad, and it has some important information to be retrieved. For example, it shows clusters, an important tool for understanding how likely a situation (or a pose, or a docking) is to happen. Events generally occur in clusters, not at random, so in molecular docking, a small number of clusters may mean that more ligands were in one specific area (forming a cluster), and since that is unlikely to happen, it may mean that the ligands are, in fact, in the binding area of the protein and in their best pose (KOZAKOV, CLODFELTER, *et al.*, 2005). It also gives out RMSD values, or root-mean-square deviation, which measures the difference between values. In AutoDock, specifically, it can measure the difference between one pose against a crystallographic observed binding (AutoDock, 2006). Since there is no definitive answer for where the active sites of RABV proteins are, or crystallographic complexes of such binding, these RMSD values are not necessarily relevant for this project.

3.3.1.2 AutoDock Vina

Ligand A3 used in this project was acetylcholine. It is well known that the rabies virus (specifically glycoprotein) somehow binds to acetylcholine in the host's body, although the full mechanism is not yet understood (LENTZ, BURRAGE, *et al.*, 1982). With that in mind, acetylcholine was used for the possibility of elucidating this mechanism and maybe finding a stable active site. However, acetylcholine is not a ligand and therefore follows different rules than ligands. It is bigger, for one, and has more rotatable bonds (a single bond, not in a ring, connected to an atom that is not hydrogen). The size problem was fixed by using PyMol, which allowed for a new molecule to be made by deleting the irrelevant parts of the receptor. It is believed the part of acetylcholine that binds to the Glycoprotein of rabies is somewhere between the amino acids 179 and 192 (LENTZ, 1990), therefore only that small part of the full receptor was used for binding. The issue with rotatable bonds cannot be fixed without significantly changing the molecule, which would defeat the purpose of using it in the first place. AutoDock has a restriction on the amount of rotatable bonds a ligand can have, and the small acetylcholine (amino acids 179 to 192) still had 21 rotatable bonds.

AutoDock VINA (TROTT e OLSON, 2010), a follower of AutoDock, is able to dock bigger molecules with more rotatable bonds, and it can also be found in the PyRx software. Therefore, the portion of the acetylcholine receptor was docked to all proteins

using AutoDock VINA, instead of AutoDock, keeping the same standard parameters. The main difference in using those methods is that the results output for AutoDock and AutoDock VINA are different (although both give the binding energy result, which is the most important result for this project); these differences will be discussed in the Results section.

3.3.2 Visualization

Post-docking, it was necessary to visualize the ligand and protein in order to see exactly where each ligand bound to in the protein. When a ligand finishes docking in AutoDock it is outputted as a .dlg document, which has all the important information of its binding energies, atoms that bound to the protein, clusters, etc. This file can be visualized as a 3D molecule using MglTools, but for visualization in any other software (such as PyMol or Discovery Studio), this file needs to be transformed back into either a .pdbqt or .pdb file. This was done by using the UNIX command (in a Cygwin terminal):

grep ‘^DOCKED’ my_docking.dlg | cut -c9- > my_docking.pdbqt (AutoDock, 2006)

This modifies the file so that the poses acquired by the ligand can still be visualized, but in different programs.

Most of the visualization used for this project was done in PyMol, which can show the areas where the ligand bound to the protein by checking for hydrogen bonds. In order to look for different types of bonds (such as hydrophobic bonds, but also pi-alkyl or pi-pi interactions), LigPlot (LASKOWSKI e SWINDELLS, 2011) and Discovery Studio (DASSAULT) were used. Figures of a number of these visualizations are shown in the Results section.

3.3.3 Active Docking

After blind docking, visualization and analysis of blind docking, certain ligands that had good results (energy binding equal or less than -6.0) passed on to the active docking phase. This was important to both have a result for possible active sites of all rabies proteins, and to achieve more complete results for the ligands and their potential in inhibiting RABV proteins.

This blind docking was completed in AutoDock in the PyRx software following the same specifications of the blind docking, with the difference that now the ligand

was also using a specific part of the protein to look for areas to bind instead of using the whole surface of the protein. In order to find possible active sites for this phase of the research, three methods were used: literature, active sites from blind docking, and I-Tasser.

Literature was researched to find possible already established binding sites of all the proteins of rabies. For each protein, the results of blind docking were analyzed to see where most ligands bound to in the protein. PyMol was used for this purpose, and polar hydrogen binding was analyzed. The positions where each ligand bound to in the protein were recorded and, when all ligands were analyzed, the values for each protein were put together. An application called Excel Counter was made using C# language, it works by going through the L column of each workbook in each excel file and grouping words (in this case, residue names and numbers). After grouping, it counts how many times each word (residue with number) appeared in each excel file.

Finally, I-Tasser (ROY, KUCUKURAL e ZHANG, 2010) was used as a bioinformatic tool to determine potential active sites for each of the proteins. I-Tasser compares structures of a protein with others that have had their active sites already established to make a comparison. It also considers known characteristics of binding sites to find something similar in the protein or molecule.

With all this information, the residues that showed up in all three of these methods were analyzed and put together to come up with a few different active sites to be researched for each protein. These active sites are shown in the Results section.

For the active docking, a specific window needed to be used for each active site in AutoDock. This window was determined based on the active site, so naturally the window would englobe the active site. However, because of protein folding, it’s impossible to focus only on the active site, so extra residues will unavoidably be a part of the area chosen as active site. It’s also important that the area for the ligand to bind be big enough, or the ligand will be strained and the docking won’t give accurate results. Table 3 shows the size and general area each box was created for each active site.

Table 3 - Position of Active Sites in Docking

Protein and Active Site	Size of box	Spacing	Box Position
Matrix 103-115	X=y=z=50	0.3750	X=16.3883, y=35.9671, z=88.1424
Matrix 172-188	X=y=z=50	0.3750	X:4.4166, y:35.1315, z:103.0660
Matrix 188-200	X=y=z=50	0.3750	X=6.1785, y=25.6764, z=114.6842
Matrix 84-100	X=50, y=z=70	0.3750	X=3.7188, y=22.9833, z=98.0499
Matrix 113-120	X=50, z=y=70	0.3750	X=16.3578, y=28.2437, z=93.6993
Nucleoprotein 349-403	X=y=100, z=70	0.3750	X=206.333, y=248.538, z=126.8282
Nucleoprotein 251-273	X=90, y=60, z=50	0.3750	X=230.176, y=244.582, z=115.1886
Nucleoprotein 300-328	X=75, y=65, z=50	0.3750	X=226.900, y=293.438, z=106.8526
Nucleoprotein 149-205	X=y=95, z=70	0.3750	X=216.2360, y=240.1538, z=75.7924
Nucleoprotein 223-240	X=90, y=55, z=70	0.3750	X=220.9692, y=247.8803, z=91.2683

Nucleoprotein 27-38	X=65, y=z=70	0.3750	X=233.2679, y=257.0023, z=94.6226
Phosphoprotein 265-290	X=y=z=50	0.3750	X=7.4988, y=33.7401, z=-6.2896
Phosphoprotein 240-260	X=y=50, z=80	0.3750	X=3.0772, y=34.9128, z=6.1101
Phosphoprotein 186-200	X=50, y=40, z=70	0.3750	X=-14.4415, y=38.3412, z=13.6135
Phosphoprotein 223-240	X=y=z=50	0.3750	X=-4.2422, y=37.9963, z=-8.8835
Glycoprotein 43-69	X=y=50, z=100	0.3750	X=2.0565, y:28.7116, z:-0.9101
Glycoprotein 91-104	X=y=50, z=80	0.3750	X=-4.9766, y=36.1392, z=-76.6080
Glycoprotein 120-143	X=60, y=50, z=110	0.3750	X=-3.0487, y=27.3474, z=-69.6848
Glycoprotein 166-197	X=40, y=50, z=85	0.3750	X=-6.9644, y=33.2406, z=-37.7948
Glycoprotein 217-240	X=60, y=50, z=75	0.3750	X=-2.8326, y=31.7418, z=2.0690
Glycoprotein 253-270	X=80, y=75, z=50	0.3750	X=-8.4228, y=28.6484, z=-8.7409

Glycoprotein 283-310	X=50, y=40, z=120	0.3750	X=16.7974, y=32.0454, z=9.3341
Large Protein 521-585	X=115, y=70, z=50	0.3750	x:71.2966, y:103.0919, z:64.3760
Large Protein 620-700	X=95, y=120, z=60	0.3750	x: 73.3576, y: 105.655, z: 54.5370
Large Protein 1112-1286	X=105, y=115, z=100	0.3750	x=66.8897, y=83.8271, z=95.5529
Large Protein 400-470	X=80, y=90, z=100	0.3750	x=48.6582, y=92.707, z=66.7172
Large Protein 808-908	X=80, y=90, z=100	0.3750	x: 101.5371, y: 90.0852, z: 84.4088

Table 3 showing box sizes and positions for each blind docking

The size of binding sites were dependent on the sizes of proteins. Large Protein, for example, had big active sites because it is such a huge protein, so it made sense for it to have bigger potential active sites. Phosphoprotein, on the other hand, had less residues modeled in the homology modeling phase, and therefore had smaller active sites.

The analysis for active docking was like the analysis for blind docking. The ligands were checked for binding energy and where they bound on the protein in PyMol, LigPlot, and Discovery Studio. These results and further explanation of them are in the Results and Discussion sections.

4. RESULTS

4.1 Blind Docking

After running the blind docking for all nineteen different strains of the five rabies proteins, the average binding energy of each ligand was taken, and the following results were obtained:

Table 4 - Glycoprotein Average Results

Ligand	Average Binding Energy	Strain with Highest Value
A1	-6.3	India -7.1
A2	-5.1	ERA -5.5
A3	-6.7	Eptesicus -7.6
A4	-4.5	Silver -5.1
A5	-6.7	India -7.4
A6	-5.4	India -6.0
A7	-6.3	DRV/Nishi -6.5
A8	-5.9	Nishi -6.4
A9	-6.8	Desmodus -7.3
A10	-5.6	ERA/PM -5.8
A11	-6.2	Nishi/Eptesicus -7.0
A12	-6.1	India/Eptesicus -6.6
A13	-6.0	Eptesicus -6.7
A14	-6.0	Cerdocyon -6.4
A15	-6.6	India/Tadarida -7.0
A16	-5.3	Callithrix/DRV/India -5.7
A17	-3.9	MRV -4.4
A18	-5.6	India -6.3
A19	-7.3	India -8.1
A20	-4.8	Eptesicus -5.1
A21	-6.5	India -7.22
A22	-5.4	India -6.2

A23	-5.0	India -5.5
A24	-5.4	Eptesicus -5.8
A25	-5.5	Silver -5.8
A26	-4.1	Nishi -4.5

Table 4 showing each ligand that was used for the blind docking of glycoprotein and their average binding energy. All nineteen strains were used to dock with each ligand.

Table 5 - Phosphoprotein Average Results

Ligand	Average Binding Energy	Strain with Highest Value
A1	-5.7	Desmodus -5.94
A2	-4.0	Racoon/Eptesicus -4.4
A3	-4.9	Vulpes -6.0
A4	-3.7	DRV -4.11
A5	-5.9	Eptesicus -6.5
A6	-4.7	Nycti -5.0
A7	-5.1	HEP -5.5
A8	-5.4	Tadarida -5.76
A9	-5.4	DRV -6.3
A10	-5.3	Racoon -5.5
A11	-5.8	Callithrix -6.0
A12	-5.3	Nycti -5.7
A13	-5.9	Eptesicus -6.4
A14	-5.9	Racoon -6.4
A15	-5.3	DRV -5.6
A16	-4.7	Callithrix -4.9
A17	-3.5	Racoon -4.1
A18	-4.5	Nycti -4.8
A19	-6.4	Tadarida -6.6
A20	-4.1	ERA -4.4
A21	-6.3	India -6.7
A22	-4.5	Nycti -4.9
A23	-4.2	Desmodus -4.7
A24	-4.7	Nishi -5.4

A25	-4.5	Desmodus -4.9
A26	-3.7	Nycti -3.9

Table 5 showing each ligand that was used for the blind docking of phosphoprotein and their average binding energy. All nineteen strains were used to dock with each ligand.

Table 6 - Nucleoprotein Average Results

Ligand	Average Binding Energy	Strain with Highest Value
A1	-5.9	Vulpes -6.3
A2	-4.4	Nycti -4.8
A3	-6.7	Vulpes -7.9
A4	-3.9	Eptesicus -4.4
A5	-6.4	India -6.6
A6	-5.1	Myotis -5.3
A7	-5.9	India -6.4
A8	-5.9	MRV -6.3
A9	-5.8	HEP -6.44
A10	-5.9	Silver -6.41
A11	-6.3	PV -6.6
A12	-5.5	ERA -5.9
A13	-5.8	Nycti -6.1
A14	-5.9	Vulpes -6.2
A15	-6.1	Callithrix -6.4
A16	-4.8	PV -5.2
A17	-3.8	PV -3.9
A18	-4.9	SAD -5.3
A19	-6.9	SAD -7.3
A20	-4.6	Callithrix -5.1
A21	-6.2	Cerdocyon -6.4
A22	-4.9	Tadarida -5.2
A23	-4.7	Myotis -5.1
A24	-4.9	MRV -5.1
A25	-4.9	SAD -5.3
A26	-3.7	Eptesicus -3.8

Table 6 showing each ligand that was used for the blind docking of nucleoprotein and their average binding energy. All nineteen strains were used to dock with each ligand.

Table 7 - Matrix Average Results

Ligand	Average Binding Energy	Strain with Highest Value
A1	-5.4	Nishi -5.8
A2	-4.2	ERA -4.5
A3	-5.1	Nishi -6.1
A4	-4.0	Tadarida -4.4
A5	-6.1	Nishi -6.5
A6	-4.9	Eptesicus -5.2
A7	-5.8	HEP -6.3
A8	-5.1	PV -5.5
A9	-5.3	Tadarida -5.8
A10	-4.9	ERA -5.3
A11	-5.4	SAD -5.6
A12	-5.4	DRV -6.0
A13	-5.5	ERA -5.9
A14	-5.9	Nishi -6.3
A15	-5.6	Nycti -5.9
A16	-4.9	PV -5.2
A17	-3.8	Nycti -4.3
A18	-4.9	Silver -5.3
A19	-6.9	Eptesicus -7.2
A20	-4.4	ERA -4.6
A21	-5.6	Cerdocyon -5.9
A22	-4.7	Callithrix -5.3
A23	-4.5	Cerdocyon -4.8
A24	-4.8	Tadarida -5.2
A25	-4.8	MRV -5.2
A26	-3.6	India -3.74

Table 7 showing each ligand that was used for the blind docking of matrix protein and their average binding energy. All nineteen strains were used to dock with each ligand.

Table 8 - Large Protein Average Results

Ligand	Average Binding Energy	Strain with Highest Value
A1	-6.6	Desmodus -6.9
A2	-5.1	DRV -5.6
A3	-7.8	Racoon -8.7
A4	-4.4	SAD -4.8
A5	-6.7	PM -7.1
A6	-6.1	Racoon/Silver -6.3
A7	-6.5	Desmodus -7.0
A8	-6.3	Cerdocyon -6.6
A9	-6.1	Tadarida -6.4
A10	-6.1	Desmodus -6.6
A11	-6.8	PM -7.2
A12	-6.9	Tadarida -7.5
A13	-6.6	ERA -7.0
A14	-6.7	MRV -7.0
A15	-6.8	Nyctinomops -7.3
A16	-6.1	ERA -6.6
A17	-5.0	PV -5.5
A18	-5.9	Callithrix -6.2
A19	-7.6	SAD -8.2
A20	-5.5	PM -5.9
A21	-6.9	Desmodus -7.2
A22	-6.2	Eptesicus -6.7
A23	-5.6	Eptesicus -6.1
A24	-5.5	Cerdocyon -6.2
A25	-5.8	HEP -6.5
A26	-4.1	Racoon -4.4

Table 8 showing each ligand that was used for the blind docking of large protein and their average binding energy. All nineteen strains were used to dock with each ligand.

The ligands which had averages equal to or below -6.0 (the threshold used for selecting ligands in this project) were taken to the second round of docking, the active (or targeted) docking phase. These ligands are shown in Table 9.

Table 9 - Ligands Selected for Active Docking

Protein	Ligands
Glycoprotein	A1, A3, A5, A7, A9, A11, A12, A13, A14, A15, A19, A21
Matrix Protein	A5, A19
Phosphoprotein	A19, A21
Nucleoprotein	A3, A5, A11, A15, A19, A21
Ligand Protein	A1, A3, A5, A6, A7, A8, A9, A10, A11, A12, A13, A14, A15, A16, A19, A21, A22

Table 9 shows which ligands achieved a value of -6.0 or less when connected to each protein. These were selected for the second round of docking, the active docking.

To run the active docking, potential active sites on each of the five proteins needed to be evaluated.

4.2 Potential Active Sites

A combination of literature, docking data, and data from a bioinformatic server were used to define potential active sites. In terms of docking data, each of the ligands in table 9 had the protein residues where they docked analyzed. Literature research was also conducted for each protein of RABV in case it indicated possible active sites. Lastly, a protein structure and function prediction server, I-Tasser (ROY, KUCUKURAL e ZHANG, 2010), was used to broaden the potential sites.

4.2.1 Literature

While RABV active sites for each of the proteins have not been successfully found, some potential active sites can be found in literature. For glycoprotein, the literature points out to residues 194-195-196-197 (TOMAR, SINGH, *et al.*, 2010) as well as residues 274-293 (BASTIDA-GONZÁLEZ, CELAYA-TREJO, *et al.*, 2016) as potential active sites of the protein.

For the matrix protein, residues 77, 100, 104, and 110 that may be important for matrix protein interaction with RelAp43 (BEN KHALIFA, LUCO, *et al.*, 2016), as well

as 67-79 (ZAN, LIU, *et al.*, 2016), which may have a role in mitochondrial targeting and apoptosis induction, were listed in the literature and used in this project.

For nucleoproteins, residues 273 and 394 (MASATANI, ITO, *et al.*, 2011) were potentially important for virus evasion, and residues 355-372 (RIEDEL, VASISHTAN, *et al.*, 2019) as potential active sites for N-N interaction.

For the phosphoprotein, literature showed that two N-binding sites may be located between residues 69 and 177 and 173 to 297 (carboxy-terminal region) of the phosphoprotein. Research also suggests that region between residues 268 and 297 need to function in order for the binding of the carboxy-terminal region to N work (CHENIK, CHEBLI, *et al.*, 1994). Region from 265 to 287 may be involved in evasion of host immunity, albeit weakly (SONTHONNAX, BESSON, *et al.*, 2019).

In terms of the large protein, less research has tackled the job of finding potential active sites for RABV's biggest protein. Residues G1112, T1170, W1201, H1241, R1242, F1285, and Q1286 were found to be important for RNA capping, and may constitute the active site of the PRNTase domain of RABV (OGINO, ITO, *et al.*, 2016). A molecular docking study found that residues M585, E620, K621, W622, N623, E696, L698, A726, and K778 were potential active sites based on where their ligands of study bound to in the protein (KIRIWAN e CHOOWONGKOMON, 2021).

4.2.2 Active Sites from Blind Docking

For each of the ligands that got an average result of equal to or less than -6.0 with one protein, the ligand and all the strains of the specific protein were analyzed in terms of their connections. For example, in Figure 6, the ligand (in light blue) is shown to have made two connections to glycoprotein strain Callithrix, specifically binding to the two residues in green: Asp266 and Ile262. This type of analysis was then made for that ligand to all nineteen strains of the specific protein, to evaluate where exactly the ligands were binding.

After analysis, Tables 10 to 14 were developed, showing the top residues that were employed in each specific protein. For example, Lys273 was the residue most used for docking in phosphoprotein after analysis of where ligands A19 and A21 (the two ligands that achieved -6.0 or less for phosphoprotein) most often bound to. These residues are in Table 10. Only residues that were used for connection to the ligands at least ten times were considered as potential residues for active site docking.

Figure 6 - Ligand A1 and Its Connections to Glycoprotein DRV

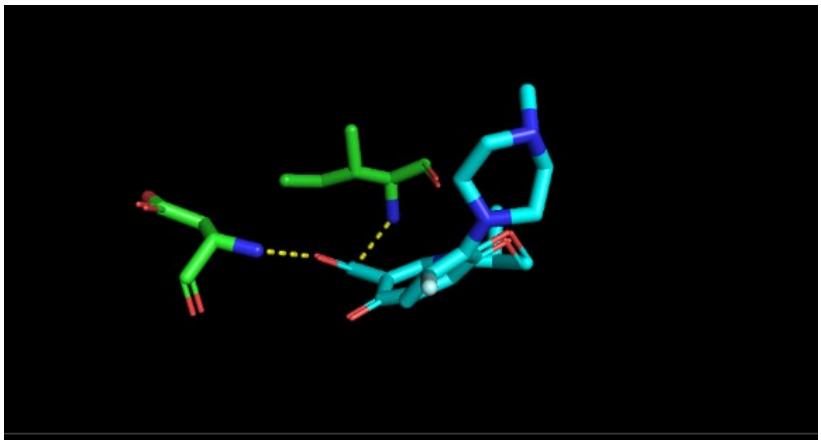


Figure 6 showing ligand a1 and its two polar connections to Callithrix strain of Glycoprotein. Image made in PyMol.

Table 10 - Residues Most Used for Docking in Phosphoprotein

Lys273	88
Lys231	82
Arg249	67
Asn243	29
Thr248	28
Lys272	28
Lys242	25
Leu259	23
Gly246	20
Arg293	19
Asn270	19
Tyr294	18
Leu224	15
Lys212	12
Lys214	11
Gln229	11
Arg260	10

Table 10 shows residues of phosphoprotein that were most used in docking all ligands which had an average binding value of -6.0 or lower.

Table 11 - Residues Most Used for Docking in Matrix Protein

Asn172	80

Thr188	68
Ser198	66
Arg118	56
Ser102	39
Asn174	28
Ile171	25
Lys115	24
Ser182	22
Asn52	20
Glu137	18
Cys178	18
Ser139	16
Arg51	16
Arg176	15
Asn86	15
Leu186	14
Asp30	14
Cys170	13
Ala104	12
Phe69	11
Ser84	11
Glu135	10

Table 11 shows residues of matrix protein that were most used in docking all ligands which had an average binding value of -6.0 or lower

Table 12 - Residues Most Used for Docking in Nucleoprotein

Arg400	125
Leu251	81
Phe260	71
Thr354	55
Glu127	46
Lys51	41
Leu258	40
Ser55	38

Lys29	35
Gly351	32
Lys38	31
Arg358	30
Glu356	27
Glu183	27
Tyr28	26
Ser58	26
Arg357	26
Arg290	25
Asn157	24
Ile257	23
Ile249	21
Phe359	20
Leu294	19
His262	19
Gln321	19
Arg149	18
Lys352	17
Glu266	17
Phe261	17
Phe349	17
Ile181	16
Arg168	15
Gln156	15
Gly125	15
Leu128	14
Leu257	14
Arg271	13
Tyr30	13
Asn61	12
Ala62	12
Lys263	12

Asn202	12
Pro275	12
Lys180	12
Glu337	12
Phe205	11
Gly276	10
Gly296	10
Lys64	10
Glu403	10

Table 12 shows residues of nucleoprotein that were most used in docking all ligands which had an average binding value of -6.0 or lower.

Table 13 - Residues Most Used for Docking in Glycoprotein

Lys236	242
Lys245	166
Ser409	137
Lys298	127
Leu341	109
Asp256	109
Trp270	101
Val49	99
Asp285	88
Glu294	81
Glu405	78
Ile36	73
Glu231	70
Asp274	68
Arg232	66
Gln275	65
Val411	64
Gln402	64
Tyr235	64
Leu47	57

Val315	56
Lys74	55
Glu50	55
Lys339	55
Lys166	52
Tyr24	50
Ser237	49
Arg253	48
Lys269	46
Asn278	44
Leu326	41
Asp165	39
Lys120	38
Ile154	36
Leu234	36
Thr258	35
Lys217	34
Arg352	34
Leu406	34
Asp209	33
Ile309	33
Gly240	32
Ser350	32
Tyr69	31
His416	31
Met255	30
Lys104	28
Tyr127	28
Met415	27
Glu128	26
Ser410	26
Lys148	25
Asp230	25

Thr212	24
Glu129	24
Lys313	23
Ser284	23
Leu276	22
Lys66	22
Thr225	22
Arg218	20
Pro328	20
Asn375	19
Lys325	18
Tyr164	18
Gly68	17
Val229	17
Arg166	17
Ala72	17
Asn76	17
Arg203	16
Glu288	16
Ser150	15
His289	15
Leu238	15
Pro272	15
His403	15
Ile73	14
Asp160	14
Arg126	14
Tyr62	14
Ile70	14
Glu201	14
Lys221	13
His280	13
Leu57	13

Ile152	13
Ala241	13
Thr182	13
Thr340	12
Asp304	12
Leu204	12
Glu149	12
Lys239	11
Val152	11
Ser71	11
Phe60	11
Arg215	11
Lys218	11
Ile382	11
His373	11
Cys226	11
Val327	10
Ser314	10
Ile336	10
Lys361	10
Glu267	10
Lys349	10
Lys275	10
Gly274	10
Leu179	10
Pro44	10

Table 13 shows residues of glycoprotein that were most used in docking all ligands which had an average binding value of -6.0 or lower.

Table 14 - Residues Most Used for Docking in Large Protein

Glu1473	53
Asp1269	34
Thr1268	32
Asn1280	28

Asp1762	25
Pro1474	23
Lys418	22
Arg1235	21
Arg1437	21
Arg312	21
Ile1404	20
Arg1628	20
Arg808	17
Arg811	17
Ser420	17
Val901	17
Trp444	17
Asn833	17
Thr1270	17
Arg1477	16
His1178	16
Trp1765	16
Thr832	15
Asp417	15
Lys506	15
Arg721	14
Arg545	14
Ser1272	14
Ile969	13
Arg1143	13
Ile845	13
Lys1687	13
Ser1194	13
Arg1712	12
Asn1408	12
Cys1632	12
Ser1475	12

Lys1244	12
Arg1242	11
Tyr1399	11
Met1271	11
Ile1409	11
Asn817	11
Thr836	11
Ser1158	11
Ile1528	11
Asn560	11
Lys1127	10
Lys1182	10
Pro896	10
Ser1737	10
Ser1745	10
Arg1305	10
Ser1530	10
Pro1406	10
Arg1051	10
Asp1277	10
Glu1799	10

Table 14 shows residues of large protein that were most used in docking all ligands which had an average binding value of -6.0 or lower.

4.2.3 Active Sites from I-Tasser

For phosphoprotein, the strain used to make a prediction of active sites was Tadarida (which had the lowest binding value in blind docking, as can be seen in Table 5). I-Tasser potential active site results are in Table 15.

Table 15 - Potential Phosphoprotein Active Sites According to I-Tasser

Cluster	Residue Numbers
Cluster Size 2	94, 98
Cluster Size 2	202,205

Cluster Size 2	97,100,101
Cluster Size 2	191,192
Cluster Size 1	48,54,80,132,139,148

Table 15 shows cluster sizes and residue numbers of potential active sites of phosphoprotein according to I-Tasser.

For nucleoprotein, the strain used to make a prediction of active sites was Vulpes (which had the lowest binding value in blind docking, as can be seen in Table 6). I-Tasser potential active site results are in Table 16.

Table 16 - Potential Nucleoprotein Active Sites According to I-Tasser

Cluster	Residue Numbers
Cluster Size 72	24,149,152,155,160,164,199,223,225,226,229,230,235,236,237,297,298,301,302,303,323,326,328,434
Cluster Size 4	149,153,154,155,160,230,232,233,234,239,240,242,243,244,252,253,254,273,274,275,276,277,279,280,323
Cluster Size 1	234,235,237,302,303
Cluster Size 1	275,276,278
Cluster Size 1	344,347,396,397,405,409,416,417

Table 16 shows cluster sizes and residue numbers of potential active sites of phosphoprotein according to I-Tasser.

For glycoprotein, the strain used to make a prediction of active sites was India (which had the lowest binding value in blind docking, as can be seen in Table 4). I-Tasser potential active site results are in Table 17.

Table 17 - Potential Glycoproteins Active Sites According to I-Tasser

Cluster	Residue Numbers
Cluster Size 3	303, 306, 307, 310
Cluster Size 2	41, 43, 305, 308, 319, 322
Cluster Size 1	16,64,65,232,236,256,258,259,278,279,280,284,285,288

Cluster Size 1	91,95,96,97,139,140,143
Cluster Size 1	137,138,140,143,144,145,146

Table 17 shows cluster sizes and residue numbers of potential active sites of glycoprotein according to I-Tasser.

For matrix protein, the strain used to make a prediction of active sites was Eptesicus (which had the lowest binding value in blind docking, as can be seen in Table 7). I-Tasser potential active site results are shown in Table 18.

Table 18 - Potential Matrix Protein Active Sites According to I-Tasser

Cluster	Residue Numbers
Cluster Size 2	117,173
Cluster Size 2	71,101,142,153
Cluster Size 2	93,96,118,119,120,121
Cluster Size 2	91,92
Cluster Size 1	94,97

Table 18 shows cluster sizes and residue numbers of potential active sites of matrix protein according to I-Tasser.

For Large Proteins, the strain used to make a prediction of active sites was SAD (which had the lowest binding value in blind docking, as can be seen in Table 8). Thanks to the size of the L Protein (2127 residues), I-Tasser couldn't analyze the full protein at once, and so the protein was divided between L_SAD_1 (amino acid 1M through E780) and L_SAD_2 (amino acid E781 through Y2121). I-Tasser potential active site results are shown in Tables 19 and 20.

Table 19 - Potential Large Protein Active Sites According to I-Tasser – L_SAD_1

Cluster	Residue Numbers
Cluster Size 4	381, 542, 552, 618, 619, 621, 622, 623, 696, 699, 703, 729
Cluster Size 3	552, 618, 729, 730
Cluster Size 1	368, 369, 481, 543, 544, 546, 548, 552, 555, 557, 566, 569, 618, 620, 621, 696, 697, 700, 727, 729, 730
Cluster Size 1	270, 273
Cluster Size 1	521, 524, 525, 526, 527, 528, 529

Table 19 shows cluster sizes and residue numbers of potential active sites of one half of the large protein according to I-Tasser.

Table 20 - Potential Large Protein Active Sites According to I-Tasser – L_SAD_2

Cluster	Residue Numbers
Cluster Size 2	385, 421
Cluster Size 3	46, 47, 48, 49, 50, 51, 52
Cluster Size 1	24, 28, 48, 52, 53, 56, 116, 117, 124, 132, 134, 137, 182, 183, 185, 194, 196, 246, 253, 268, 393, 467, 468
Cluster Size 2	76, 83
Cluster Size 2	11, 13

Table 20 shows cluster sizes and residue numbers of potential active sites of the second half of the large protein according to I-Tasser.

These three tools (literature, blind docking, and I-Tasser) were the basis for a decision on which binding sites would be used for the active site (non-blind) molecular docking. The residues from each tool were compared and combined, as to group and minimize the amount of potential active sites needed to be used. Table 21 shows the active sites that were used for each protein.

Table 21 - Potential Active Sites for Each Protein

Matrix Protein Active Sites	Glycoprotein Active Sites	Phosphoprotein Active Sites	Nucleoprotein Active Sites	Large Protein Active Sites
84-100	43-69	186-200	27-38	521-585
103-115	91-104	223-240	149-205	620-700
113-120	120-143	240-260	223-240	1112-1286
172-188	166-197	265-290	251-273	400-470
188-200	217-240		300-328	
	253-270		349-403	
	283-310			

Table 21 showing active sites that were used for active site docking for each protein. These active sites were combined from different sources: literature, the previous blind docking, and bioinformatic software.

4.3 Active Site Dockings

After the active site dockings, tables 22, 23, 24, 25, and 26 show the results acquired. The tables only show results that were lower or equal to -6.0.

Table 22 - Glycoprotein Results from Active Site Docking

Ligand	Strain	Active site	Binding energy	Clusters	RMSD	Hydrogen bonding	Strength of H-bond	Hydrophobic Interactions
A5	DRV	43-69	-6.5	3	0.0-1.77	2	Moderate	11 (3 aromatic)
A9	Nishi	43-69	-6.6	4	0.0-0.80	4	3 Moderate, 1 Weak	8 (1 aromatic)
A11	Nishi	43-69	-6.4	1	0.0-0.80	1	Weak	8 (2 aromatic)
A12	DRV	43-69	-6.1	8	0.0-1.72	0	-	15 (4 aromatic)
A13	DRV	43-69	-6.3	7	0.0-1.83	0	-	12 (3 aromatic)
A14	Nishi	43-69	-6.4	6	0.0-1.74	3	2 Moderate, 1 Weak	7 (3 aromatic)
A15	Nishi	43-69	-6.0	7	0.0-1.51	2	Moderate	12 (3 aromatic)
A19	DRV	43-69	-6.5	5	0.0-1.40	1	Strong	6 (1 aromatic)
A21	Nishi	43-69	-6.5	3	0.0-1.16	1	Moderate	9 (3 aromatic)
A9	MRV	120-143	-6.9	8	0.0-1.02	2	Moderate	9 (2 aromatic)
A11	Myotis	120-143	-6.7	4	0.0-1.10	2	Moderate	11 (2 aromatic)
A12	Nycti	120-143	-6.2	8	0.0-1.60	0	-	11 (2 aromatic)

A13	Nycti	120-143	-6.4	8	0.0-1.49	1	Moderate	6 (2 aromatic)
A14	MRV	120-143	-6.5	8	0.0-1.53	2	Moderate	8 (2 aromatic)
A15	Desmodus	120-143	-6.5	6	0.0-1.91	3	2 Moderate, 1 Weak	10 (1 aromatic)
A19	Nycti	120-143	-7.1	8	0.0-1.31	2	Moderate	8 (2 aromatic)
A21	Nishi	120-143	-6.4	5	0.0-1.58	0	-	12 (2 aromatic)
A13	Tadarida	166-197	-6.5	2	0.0-1.55	2	Moderate	7
A14	Tadarida	166-197	-6.7	4	0.0-0.73	3	Moderate	7
A19 – 1	Tadarida	166-197	-6.2	3	0.0-1.83	1	Moderate	9
A19 – 2	Tadarida	166-197	-6.2	3	0.0-1.83	3	Moderate	4
A5	DRV	217-240	-6.1	4	0.0-0.41	3	Weak	12 (3 aromatic)
A13	DRV	217-240	-6.7	5	0.0-1.93	2	1 Moderate, 1 Weak	9 (2 aromatic)
A14	DRV	217-240	-6.9	4	0.0-1.19	1	Moderate	8 (2 aromatic)
A19	DRV	217-240	-6.8	5	0.0-1.00	2	Moderate	13 (3 aromatic)
A21	DRV	217-240	-6.0	2	0.0-1.40	0	-	15 (4 aromatic)
A11	HEP	253-270	-6.0	3	0.0-0.68	1	Moderate	10 (1 aromatic)

A13	MRV	253-270	-6.7	4	0.0-1.96	4	1 Strong, 3 Moderate	4
A14	Tadarida	253-270	-6.7	2	0.0-1.30	None	-	8
A19	DRV	253-270	-6.9	6	0.0-1.34	None	-	12 (4 aromatic)
A21	HEP	253-270	-6.0	4	0.0-1.92	2	1 Moderate, 1 Weak	12 (2 aromatic)
A1	Desmodus	283-310	-6.5	3	0.0-1.75	2	Moderate	9 (1 aromatic)
A9	Callithrix	283-310	-7.1	1	0.0-1.08	1	Moderate	5
A11-1	Desmodus	283-310	-6.4	2	0.0-1.90	3	2 Moderate, 1 Weak	9 (1 aromatic)
A11-2	Desmodus	283-310	-6.4	2	0.0-1.90	3	Moderate	9 (1 aromatic)
A12	Silver	283-310	-6.4	8	0.0-1.40	1	Moderate	6
A14	Desmodus	283-310	-6.0	4	0.0-1.97	2	1 Strong, 1 Moderate	9 (1 aromatic)
A15	Callithrix	283-310	-6.5	1	0.0-1.25	2	Moderate	5
A19	Desmodus	283-310	-6.7	4	0.0-1.83	1	Moderate	13 (1 aromatic)
A21	Tadarida	283-310	-6.9	1	0.0-0.56	2	Moderate	6

Table 22 shows the results from active docking of glycoprotein with potential active sites and the ligands that achieved a value of -6.0 or lower.

Table 23 - Nucleoprotein Results from Active Site Docking

Ligand	Strain	Active site	Binding energy	Clusters	RMSD	Hydrogen bonding	Strength of H-bond	Hydrophobic Interactions
A5	MRV	349-403	-6.6	4	0.0-1.26	4	Moderate	4(1 aromatic)
A11	MRV	349-403	-6.6	6	0.0-0.9	4	Moderate-weak	3 (1 aromatic)
A19	DRV	349-403	-7.1	6	0.0-1.76	2	Moderate	9 (2 aromatic)
A21	Callithrix	349-403	-6.9	6	0.0-1.57	3	Moderate	3 (1 aromatic)
A5	Myotis	251-273	-6.9	3	0.0-0.34	2	Moderate-weak	9 (5 aromatic)
A5	Myotis	251-273	-6.9	3	0.0-0.34	2	Moderate	8 (4 aromatic)
A11	SAD	251-273	-8.0	1	0.0-1.99	4	Moderate-weak	7 (1 aromatic)
A15	Callithrix	251-273	-6.9	3	0.0-1.94	1	Moderate	7 (4 aromatic)
A19	HEP	251-273	-7.2	6	0.0-1.19	None	-	7 (3 aromatic)
A19	HEP	300-328	-7.5	2	0.0-1.85	2	Moderate-weak	11 (1 aromatic)
A19	Nycti	149-205	-6.9	6	0.0-1.97	3	Moderate	7 (1 aromatic)
A11	India	223-240	-7.2	1	0.08-0.74	2	Moderate	6
A19	SAD	223-240	-6.9	2	0.00-1.62	2	Moderate	8 (1 aromatic)
A21	ERA	223-240	-6.4	4	0.00-1.21	2	Moderate	13 (1 aromatic)
A5	MRV	27-38	-6.5	3	0.00-0.35	3	Moderate-weak	5 (1 aromatic)

A11	MRV	27-38	-6.7	1	0.32-1.98	None	-	11 (1 aromatic)
A19	Tadarida	27-38	-7.2	4	0.00-1.69	1	Moderate	11 (1 aromatic)
A21	Cerdocyon	27-38	-7.4	4	0.00-1.06	1	Moderate	7 (2 aromatic)

Table 23 shows the results from active docking of nucleoprotein with potential active sites and the ligands that achieved a value of -6.0 or lower.

Table 24 - Phosphoprotein Results from Active Site Docking

Ligand	Active Site	Binding Energy	Strain
A19	186-200	-4.8	-5.3 Eptesicus
A21	186-200	-5.0	-5.5 HEP
A19	265-290	-4.9	-7.6 Eptesicus
A21	265-290	-4.4	-6.1 Racoon
A19	240-260	-5.7	-7.8 PV
A21	240-260	-5.4	-6.7 PV
A19	223-240	-5.7	-6.7 Callithrix
A21	223-240	-5.5	-6.7 Desmodus

Table 24 shows the results from active docking of phosphoprotein with potential active sites and the ligands that achieved a value of -6.0 or lower. None of the results of this second round of docking achieved a value of -6.0 or lower, therefore none of the active sites or ligands were further analyzed.

Table 25 - Matrix Protein Results from Active Site Docking

Ligand	Strain	Active site	Binding energy	Clusters	RMSD	Hydrogen bonding	Strength of H-bond	Hydrophobic Interactions
A19	ERA	188-200	-6.5	2	0.2-1.32	2	Moderate	10
A19	DRV	172-188	-6.7	3	0.0-1.75	1	Moderate	11
A19	Cerdocyon	113-120	-7.7	5	0.0-1.95	2	1 moderate, 1 strong	11
A5	Nishi	113-120	-6.6	2	0.0-0.22	3	Moderate	9

A19	Nishi	88-100	-6.5	2	0.0-1.91	3-4	Moderate	8

Table 25 shows the results from active docking of matrix protein with potential active sites and the ligands that achieved a value of -6.0 or lower.

Table 26 - Large Protein Results from Active Site Docking

Ligand	Strain	Active site	Binding energy	Clusters	RMSD	Hydrogen bonding	Strength of H-bond	Hydrophobic Interactions
A1	Cerdocyon	521-582	-7.1	2	0.0-1.81	1	Moderate	10
A6	Racoon	521-582	-6.3	6	0.0-1.89	0	--	13 (2 aromatic)
A7	PV	521-582	-6.1	5	0.0-0.68	4	Moderate	10 (2 aromatic)
A8	SAD	521-582	-6.7	3	0.0-0.88	0	--	10 (1 aromatic)
A10	DRV	521-585	-6.6	3	0.0-1.86	3	3 Moderate, 1Weak	7 (2 aromatic)
A10	HEP	521-585	-6.6	5	0.0-1.83	1	1 Moderate, 1Weak	10 (2 aromatic)
A11	Callithrix	521-585	-6.9	3	0.0-1.62	1	Moderate	7 (1 aromatic)
A11	Nishi	521-585	-6.9	6	0.0-1.72	1	Moderate	8 (1 aromatic)
A11	Eptesicus	521-585	-6.9	4	0.0-1.79	2	Moderate	10 (2 aromatic)
A12	MRV	521-585	-6.6	8	0.0-1.37	2	Moderate	14 (3 aromatic)
A13	MRV	521-585	-7.4	6	0.0-1.86	1	Moderate	8 (2 aromatic)
A14	MRV	521-585	-7.3	4	0.0-1.64	2	Moderate	8 (2 aromatic)

A15	MRV	521-585	-6.3	8	0.0-0.58	1	Moderate	13 (2 aromatic)
A16	MRV	521-585	-6.2	4	0.0-1.8	1	Moderate	11 (1 aromatic)
A19	MRV	521-585	-7.7	4	0.0-1.8	0	--	13 (1 aromatic)
A21	MRV	521-585	-7.2	4	0.0-0.7	3	Moderate	11 (1 aromatic)
A19	SAD	620-700	-6.2	6	0.0-1.99	2	Moderate	9 (2 aromatic)
A1	Cerdocyon	1112-1286	-7.1	6	0.0-1.00	0	--	12 (2 aromatic)
A5	Eptesicus	1112-1286	-7.3	3	0.0-1.81	0	--	11 (3 aromatic)
A6	ERA	1112-1286	-6.6	9	0.00-0.99	1	Weak	12 (2 aromatic)
A7	ERA	1112-1286	-7.2	3	0.00-1.96	1	Moderate	12 (3 aromatic)
A8	Desmodus	1112-1286	-6.8	8	0.00-1.94	0	--	14 (3 aromatic)
A9	India	1112-1286	-6.2	7	0.0-0.72	2	Moderate	3 (1 aromatic)
A10	India	1112-1286	-6.3	8	0.0-1.72	3	2 Moderate, 1Weak	10 (1 aromatic)
A11	HEP	1112-1286	-7.1	8	0.00-0.69	0	--	8 (2 aromatic)
A12	SAD	1112-1286	-7.6	10	0.00-0.00	1	Moderate	13 (3 aromatic)
A13	India	1112-1286	-7.2	7	0.00-1.64	2	Moderate	7 (2 aromatic)
A14	Tadarida	1112-1286	-7.4	5	0.00-1.99	1	Moderate	8 (2 aromatic)

A15	India	1112-1286	-7.6	4	0.00-1.09	1	Moderate	9 (3 aromatic)
A16	Nishi	1112-1286	-6.5	9	0.00-1.44	2	Moderate	12 (4 aromatic)
A19	ERA	1112-1286	-8.4	7	0.00-1.83	3	Moderate	10 (1 aromatic)
A21	Nycti	1112-1286	-7.3	9	0.00-0.38	1	Moderate	9 (1 aromatic)
A22	Nycti	1112-1286	-7.1	4	0.00-1.67	2	Moderate	9 (4 aromatic)
A1	MRV	400-470	-7.0	6	0.00-1.37	--	--	9 (2 aromatic)
A5	PV	400-470	-7.0	4	0.00-1.04	2	Moderate	7 (2 aromatic)
A6	Vulpes	400-470	-6.8	5	0.00-1.95	1	Moderate	14 (1 aromatic)
A7	Racoon	400-470	-6.5	4	0.00-1.51	1	Moderate	11 (2 aromatic)
A8	Vulpes	400-470	-6.9	7	0.00-0.90	--	--	10 (2 aromatic)
A10	Nishi	400-470	-6.4	4	0.00-1.48	3	Moderate	8 (2 aromatic)
A11	Callithrix	400-470	-7.5	5	0.00-1.72	2	Moderate	9 (2 aromatic)
A12	Nycti	400-470	-7.2	7	0.00-1.78	--	--	14 (2 aromatic)
A13	Racoon	400-470	-6.9	5	0.00-1.60	2	1 Strong, 1 Moderate	7 (5 aromatic)
A14	Racoon	400-470	-7.0	8	0.00-1.24	1	Moderate	8 (1 aromatic)
A15	Racoon	400-470	-6.9	7	0.00-1.89	1	Moderate	9 (3 aromatic)

A16	Racoon	400-470	-6.7	6	0.00-1.89	--	--	12 (1 aromatic)
A19	SAD	400-470	-8.5	8	0.00-1.71	--	--	15 (2 aromatic)
A21	Cerdocyon	400-470	-7.4	8	0.00-1.02	2	Moderate	7 (4 aromatic)
A22	Racoon	400-470	-6.2	5	0.00-1.94	--	--	14 (2 aromatic)
A1	Tadarida	808-908	-6.8	3	0.00-1.26	--	--	17 (1 aromatic)
A5	PV	808-908	-6.5	2	0.00-1.97	1	Moderate	11
A6	HEP	808-908	-6.2	5	0.00-1.88	1	Moderate	16
A7	Nycti	808-908	-6.5	3	0.00-1.97	1	Moderate	12
A10	Vulpes	808-908	-6.2	4	0.00-1.92	2	Moderate	10
A10	SAD	808-908	-6.2	4	0.00-1.92	2	Moderate	12 (1 aromatic)
A11	SAD	808-908	-6.8	4	0.00-0.78	2	Moderate	12
A12	PV	808-908	-6.7	6	0.00-1.97	1	Moderate	12
A13	Tadarida	808-908	-6.7	3	0.00-1.89	1	Moderate	8 (2 aromatic)
A14	Tadarida	808-908	-6.9	4	0.00-1.71	--	--	10 (3 aromatic)
A15	PV	808-908	-6.6	5	0.00-1.95	--	--	13
A16	SAD	808-908	-6.2	6	0.00-1.84	1	Moderate	12

A19	SAD	808-908	-8.0	4	0.00-1.70	1	Moderate	14
A21	MRV	808-908	-7.2	1	0.00-1.66	1	Moderate	11

Table 26 shows the results from active docking of large protein with potential active sites and the ligands that achieved a value of -6.0 or lower.

The results indicating clusters and RMSD were taken from the .dlg file that is created post docking and includes details about the results of the process, as discussed in Materials and Methods. The results from hydrogen bonding, strength of H-bond and hydrophobic interactions were taken from the LigPlot (LASKOWSKI e SWINDELLS, 2011) program, as shown in Figure 7. With these results, a determination was made on which ligands were going to be visualized and analyzed further to discover how they could potentially hinder rabies proteins and stop their function in the host. The ligands were chosen based on how often they appeared with low binding energies, as well as other values shown in Tables 22 to 26. The chosen ligands are in Table 27.

Table 27 - Final Ligands Chosen to be Analyzed

A5
A11
A19
A21

Table 27 shows what ligands, from the 26 used, were chosen to be analyzed for their possible structural and pharmacological features against RABV proteins

Figure 7- Ligand A5 and Its Connections to Glycoprotein DRV Focusing on Active Site 43-69

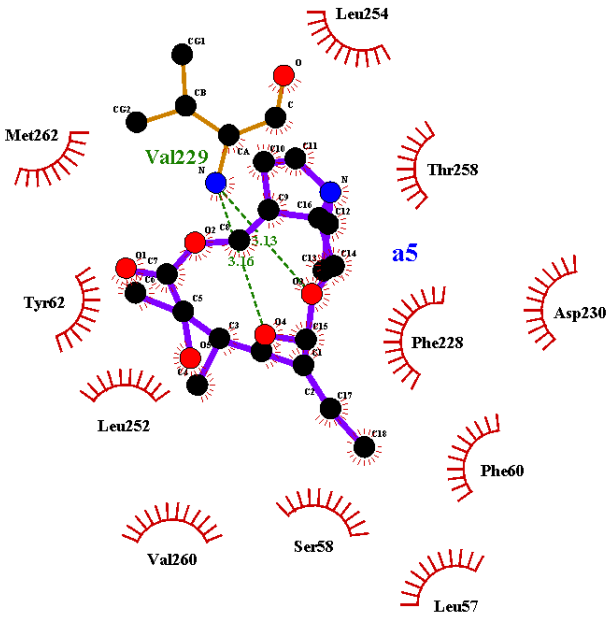


Figure 7 shows ligand A5 and where it bound when docking with DRV strain of glycoprotein. Picture shows it bound with Val229, making two hydrogen bonds of moderate strength, and shows the surrounding residues that formed hydrophobic interactions (the red "lashes"). Picture from LigPlot.

4.4 Active Sites

In order to evaluate the active sites, each ligand that got a binding energy of -6.0 or less in active docking was analyzed for where they actually bound during blind docking. Tables 28 to 48 will show the number of times a residue was employed for binding.

Table 28 - 188-200 Matrix Active Site Results

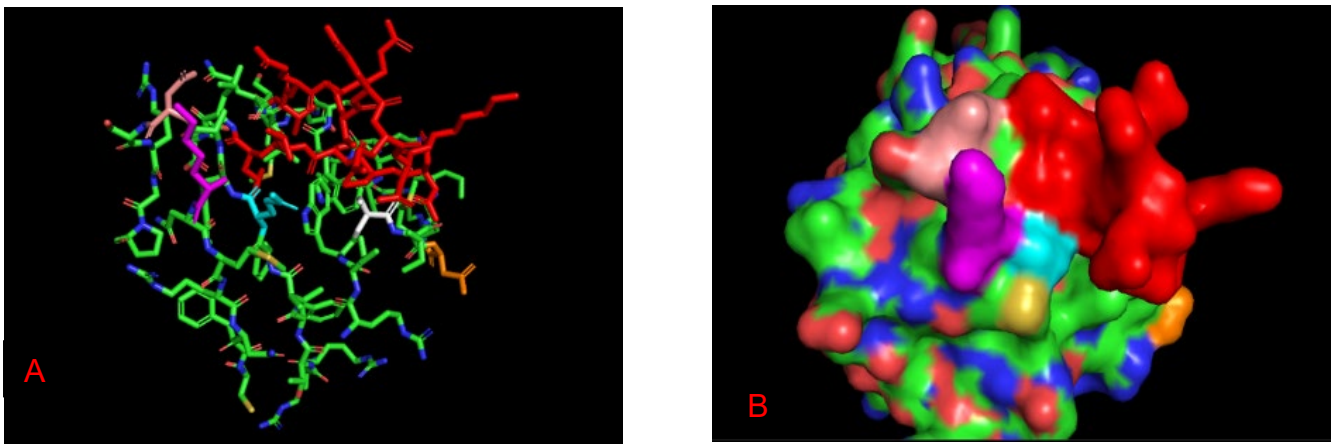
Residue	Number of times bound
Ser197	125
Gln123	73
Ser191	71
Glu192	51
Arg190	47
Ser198	28

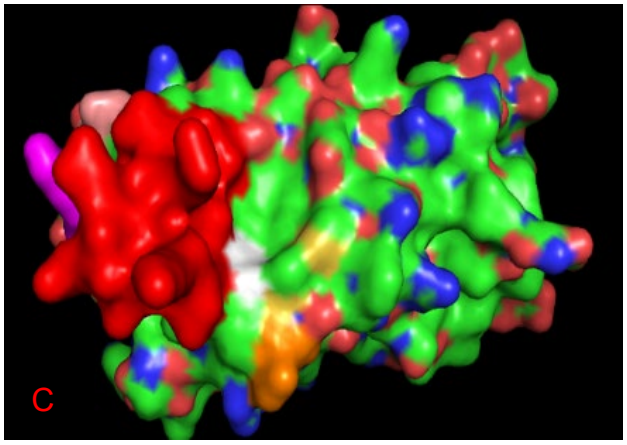
Lys48	24
Cys170	24
Thr188	22
Glu193	18
Asp196	15
Asp194	13
Lys195	13
Asn172	11
Asp126	10

Table 28 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The residues Gln123, Lys48, Cys170, Asn172, and Asp126 are not a part of the potential binding site 188-200. However, in Figure 8A it is possible to see that all these residues are in the general area of the 188-200 proposed potential active site, with Asn172 (in orange) being the one farthest away from the active site. Figure 8B shows the cavity that exists between residues 188-200 and Lys48, Asp126, and Gln123. While these last three residues were not the immediate residues proposed for this active site, they might complement the proposed active site. Figure 8C shows the cavity between residues 188-200 and Cys170 and Asn172. These seem to form less of a cavity than Figure 8B, with the orange residue being further away than it would make sense for an active site. 8D shows ligand A19 in the only position it acquired throughout this docking, in the cavity shown in picture 8C, touching the white residue.

Figure 8 - Residues 188-200 and Peripheral Residues of Active Site 188-200





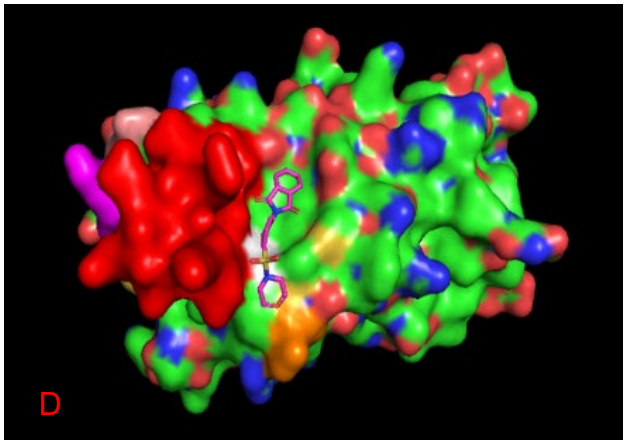


Figure 8 - Figures show active sites formed when using active docking 188-200 in the matrix protein. Residues 188-200 are in red, Lys48 in pink, Asn172 in orange, Cys170 in white, Asp126 in salmon, and Gln123 in light blue. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure A shows the location of each of these residues in comparison to the active site (ERA strain of matrix protein). B) Figure B shows that the red residues (188-200) formed a cavity with the light blue residue (Gln123), as well as possibly the pink residue (Lys48) and the salmon residue (Asp126). C) Figure C shows that a potential cavity was formed between 188-200 and Cys170 (white),but Asn172 (orange) was possibly too far from the residues used for potential active site. D) Figure D shows the position of ligand A19 when docked to the protein. This is the only pose it acquired, and it touches both the red potential active site, as well as the Cys170 (white) residue. Images made in PyMol.

Table 29 - 172-188 Matrix Active Site Results

Residue	Number of times bound
Ser198	129
Thr188	104
Ser182	89
Leu186	52
Cys170	43
Trp181	31
Asn172	17
Thr173	17
Asp196	14

Table 29 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The residues Ser198, Cys170, and Asp196 are not a part of this potential binding site (172-188). All these residues, however, are close to the potential binding site, as can be observed in Figure 9A. In figure 9B, specifically, it looks like all these extra residues, together with the proposed active site, form a binding cavity. This cavity was used by the ligand with lowest binding energy during active docking for binding, as figure 9C presents.

Figure 9 - Residues 172-188 and Peripheral Residues of Active Site 172-188

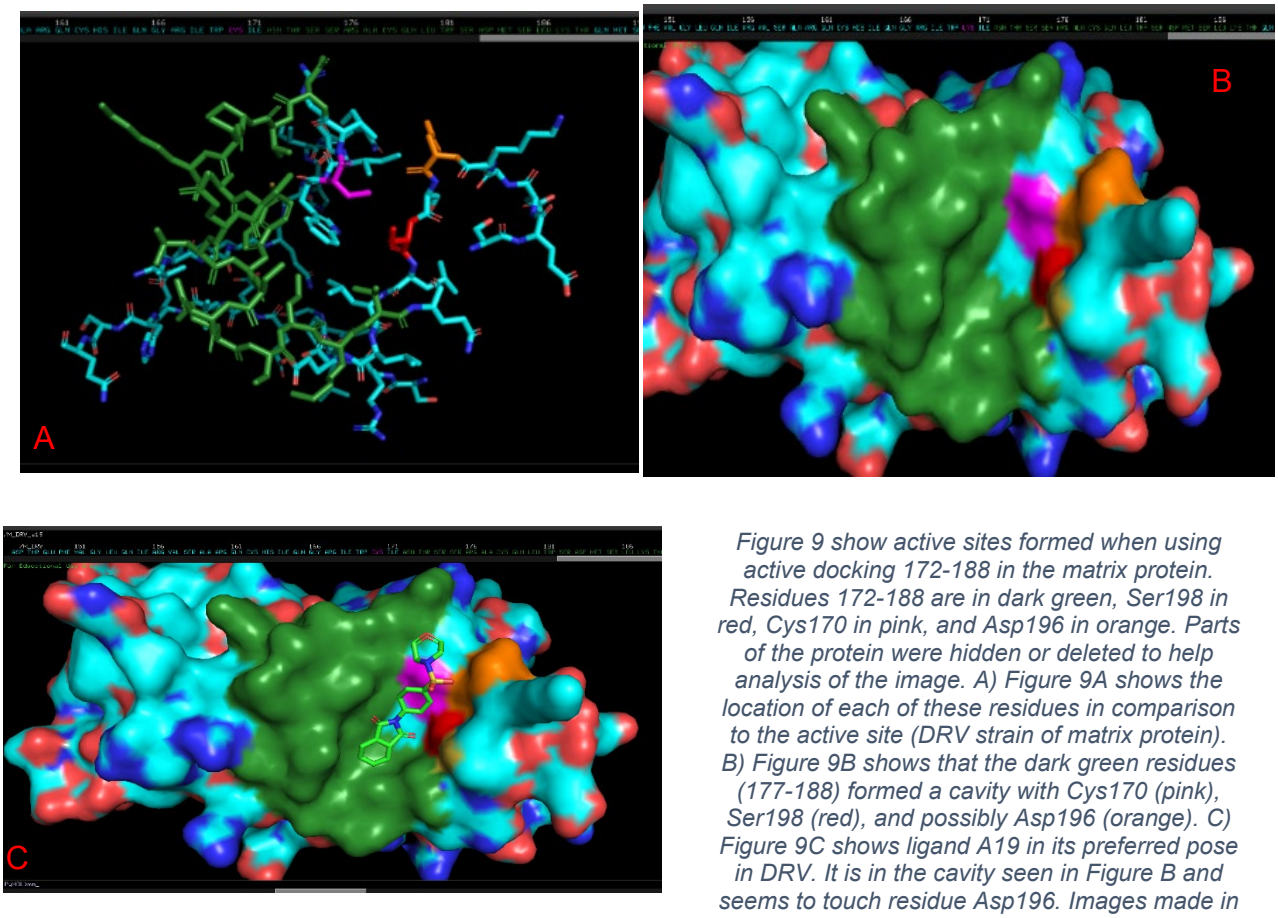


Figure 9 show active sites formed when using active docking 172-188 in the matrix protein. Residues 172-188 are in dark green, Ser198 in red, Cys170 in pink, and Asp196 in orange. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 9A shows the location of each of these residues in comparison to the active site (DRV strain of matrix protein). B) Figure 9B shows that the dark green residues (177-188) formed a cavity with Cys170 (pink), Ser198 (red), and possibly Asp196 (orange). C) Figure 9C shows ligand A19 in its preferred pose in DRV. It is in the cavity seen in Figure B and seems to touch residue Asp196. Images made in PyMol.

Table 30 - 113-120 Matrix Active Site Results

Residue	Number of times bound
Asn172	146
Arg118	139
Ser102	112
Asn174	86
Ile171	50
Lys115	45
Ala104	36
Arg167	32
Leu99	28
Val113	15
Asn86	11

Table 30 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The residues Asn172, Ser102, Asn174, Ile171, Ala104, Arg167, Leu99, and Asn86 are not a part of the potential binding site 113-120. Some of these peripheral residues, however, are close to and form cavities with the 113-120 area, as can be seen in Figure 10. Interestingly, residues Asn172, Ile171 and Asn174 are also seen binding to ligands in previous active sites mentioned (172-188 and 188-200), either because they were part of the intended binding site, or because they were peripheral residues. There might be a pattern in terms of the potential active site area when considering these three residues. Ligand A19, the one with the lowest binding site in this proposed active site, bound in the exact cavity made by the proposed active site and the peripheral residues, as can be seen in Figure 10D.

Figure 10 - Residues 113-120 and Peripheral Residues of Active Site 113-120

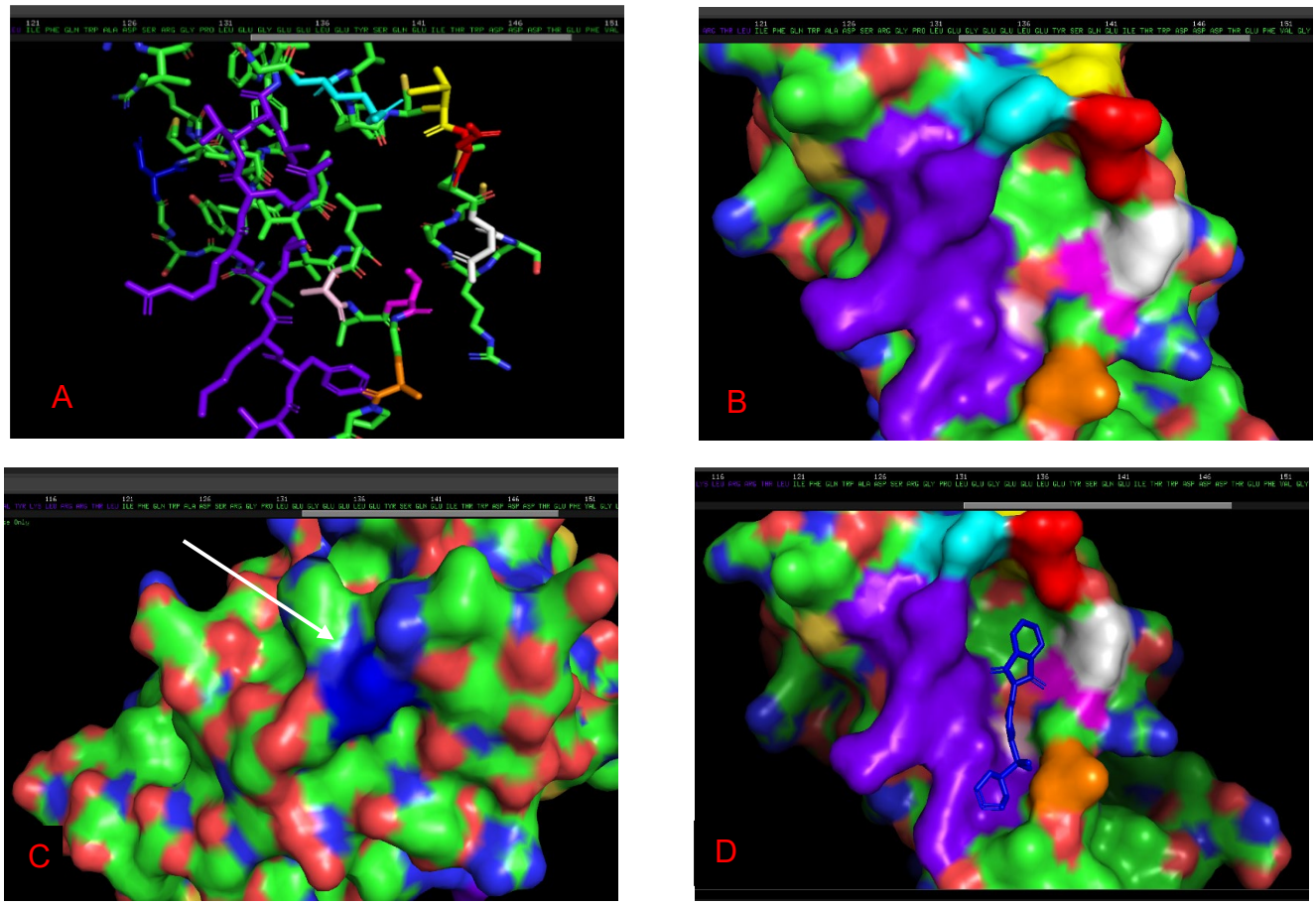


Figure 10 - Figures show active sites formed when using active docking 113-120 in the matrix protein. Residues 113-120 are in purple, Asn172 in red, Ser102 in pink, Ala104 in orange, Asn174 in white, Ile171 in yellow, Arg167 in light blue, Leu99 in light pink, and Asn86 in dark blue. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 10A shows the location of each of these residues in comparison to the active site (Cerdocyon strain of matrix protein). B) Figure 10B shows that the purple residues (113-120) formed a cavity Arg167 (light blue), Asn172 (red), Asn174 (white), Ser102 (pink), Leu99 (light pink). This cavity is also in close contact with Ile171 in yellow and Ala104 in orange. C) Figure 10C shows the opposite side of the protein when compared to Figure B. There, it's clear that residue Asn86 (dark blue, shown by white arrow) is on the opposite side of both the potential active site and most peripheral residues that were used for binding in this active docking. D) Figure 10D shows ligand A19 in its preferred pose in Cerdocyon. It is in the cavity seen in Figure B and seems to be touching Ala104 (in orange). Figures made in PyMol.

Table 31 - 84-100 Matrix Active Site Results

Residue	Number of times bound
Ser198	133
Thr188	127
Cys170	63
Ser182	52
Asp196	32
Cys178	31
Leu186	25
Trp181	23
Asn172	12

Table 31 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

None of the residues that the ligand bound to in this particular active site were a part of the proposed active sites. In Figure 11 it is clear that the residues 84-100 are not on the surface of the protein, but deeper into it, which might be why it wasn't available to the ligand (and probably wouldn't be available for binding or as an active site). Of note is that residues such as Ser198, Thr188, Cys170, Asp196, Leu186, Trp181, and Asn172 have appeared at least once in other binding sites, even when they were not the focus of the potential active site.

Figure 11 - Residues 88-100 and Peripheral Residues of Active Site 88-100

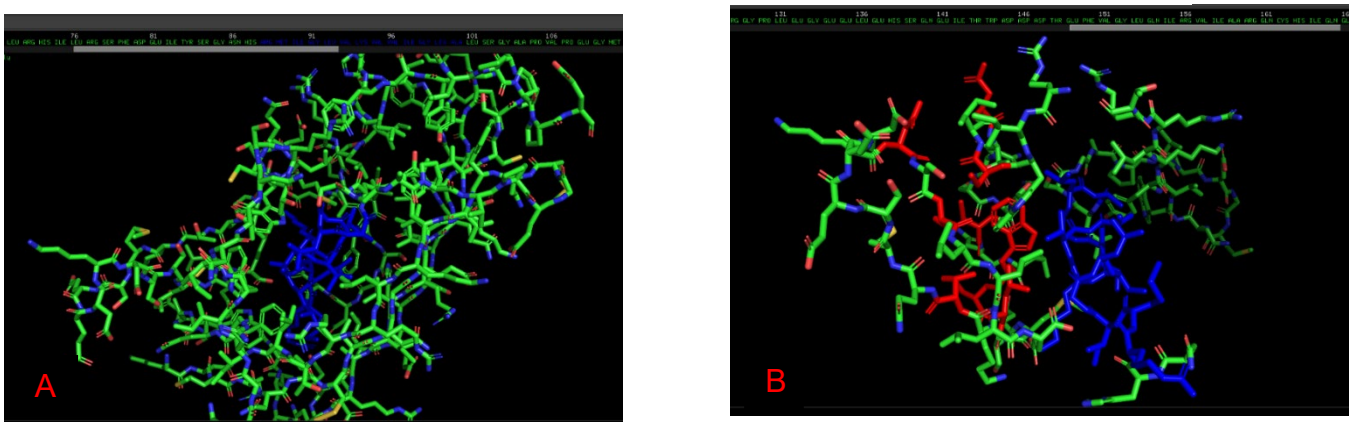


Figure 11 shows active sites formed when using active docking 88-100 in the matrix protein. Residues 88-100 are in dark blue, while all the other residues that appeared in Table 30 are in red. A) Figure 11A shows the location of the potential active site proposed, in dark blue. There are many residues on top and on the sides of 88-100, which might mean it isn't very likely of being available for binding to ligands (Nishigahara strain of matrix protein). B) Figure 11B shows, in red, the peripheral residues where A19 actually bound to in this active site docking. Figures made in PyMol.

Table 32 - 349-403 Nucleoprotein Active Site Results

Residue	Number of times bound
Arg400	161
Arg358	120
Phe359	112
Arg357	109
Thr354	105
Leu251	92
Glu356	45
Thr252	45
Lys352	25
Phe355	23
Thr399	22
Gly353	16
Tyr259	14
Glu370	11
Glu398	10

Table 32 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The residues Leu251, Thr252, and Tyr259 are not a part of the potential binding site 349-403 for glycoprotein. Figure12 shows both the potential active site residues as well as these peripheral residues. Some of these residues that were not intended as the potential active site did form a cavity that could potentially be a binding site. In Figure 12A it is visible that the blue residues (Leu251, Thr252, and Tyr259) are somewhat close to the potential active site (in red). In Figure 12B, a cavity between the potential binding site and the extra residues can be seen, showing that these extra residues might be a part of cavity binding. Ligand A19 (Figure 12C) shows a pose where it binds to the cavity made between the potential active site and the extra residues, although others poses were seen where it just bound to the potential active site previously predicted (in red).

Figure 12 - Residues 349-403 and Peripheral Residues of Active Site 349-403

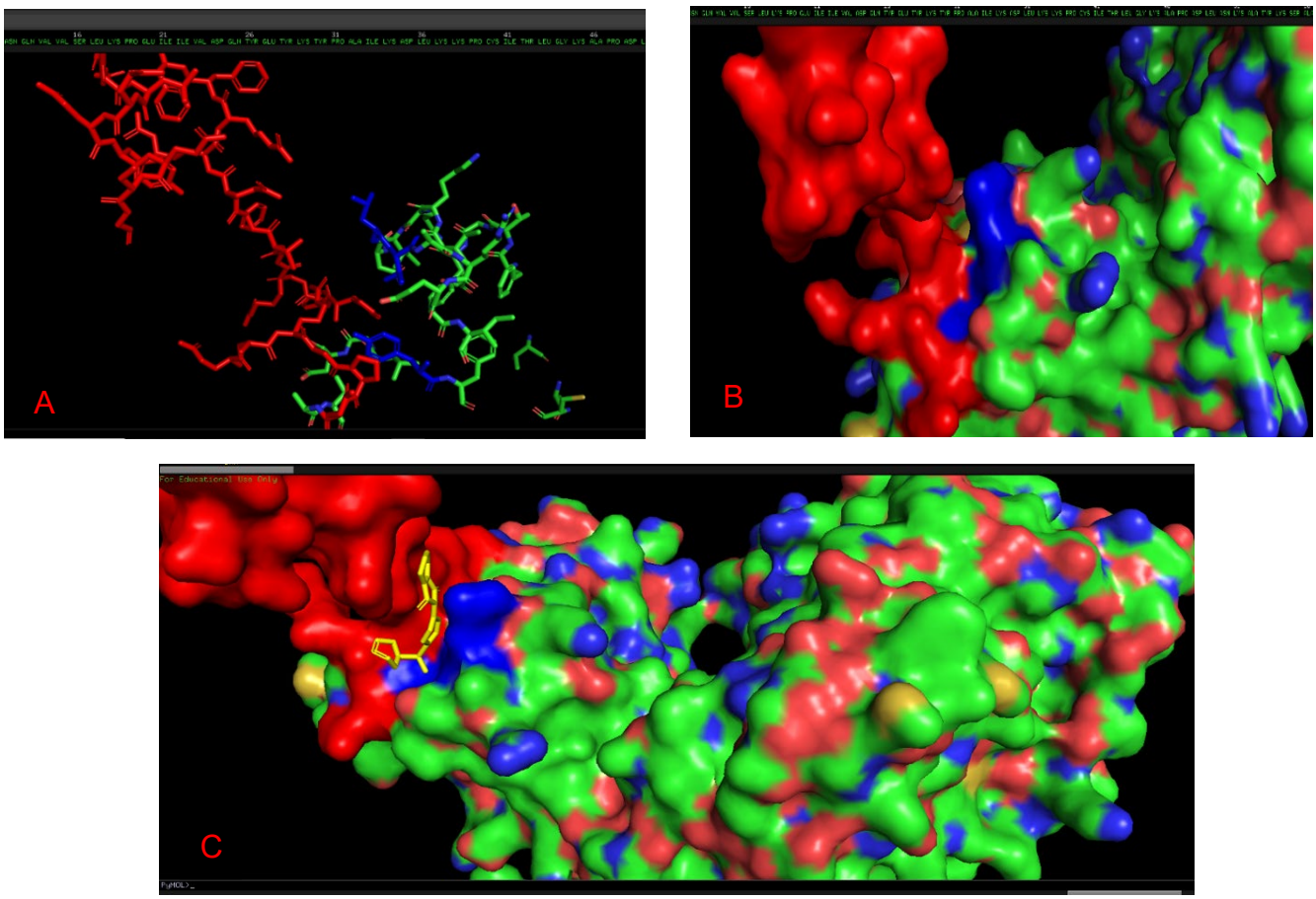


Figure 12 shows active sites formed when using active docking 349-403 in the nucleoprotein. Residues 349-403 are in red, while residues Leu251, Thr252, and Tyr259 are in blue. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 12A shows the location of each of these residues in comparison to the active site (DRV strain of nucleoprotein). B) Figure 12B shows that the three peripheral residues did form a sort of cavity with the predicted binding site, in blue. C) Figure C shows ligand A19 in one of its poses during binding, resting in the cavity made between the predicted binding site and the peripheral residues. Image made in PyMol.

Table 33 - 300-328 Nucleoprotein Active Site Results

Residue	Number of times bound
Arg271	52
Phe245	46
Tyr28	41
Lys247	38
Asn303	36
His307	24
Ala304	18
Arg323	13
Val240	10

Table 33 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Over half of the residues that were used for binding in this active site were not the ones predicted as potential active sites. Residues Arg271, Phe245, Tyr28, Lys247, and Val240 were all outside of the potential active site. Figure 13 shows the position of these residues. Figure 13B shows the position of these residues on the surface of the protein, with the white, pink, and orange residues close to each other. Furthermore, it's possible to see a potential cavity that is formed between the orange and white residues. Figure 13C shows the ligand which got the lowest binding value in the active docking in its preferred pose throughout docking. It does fit in between the white and orange residues.

Figure 13 - Residues 300-328 and Peripheral Residues of Active Site 300-328

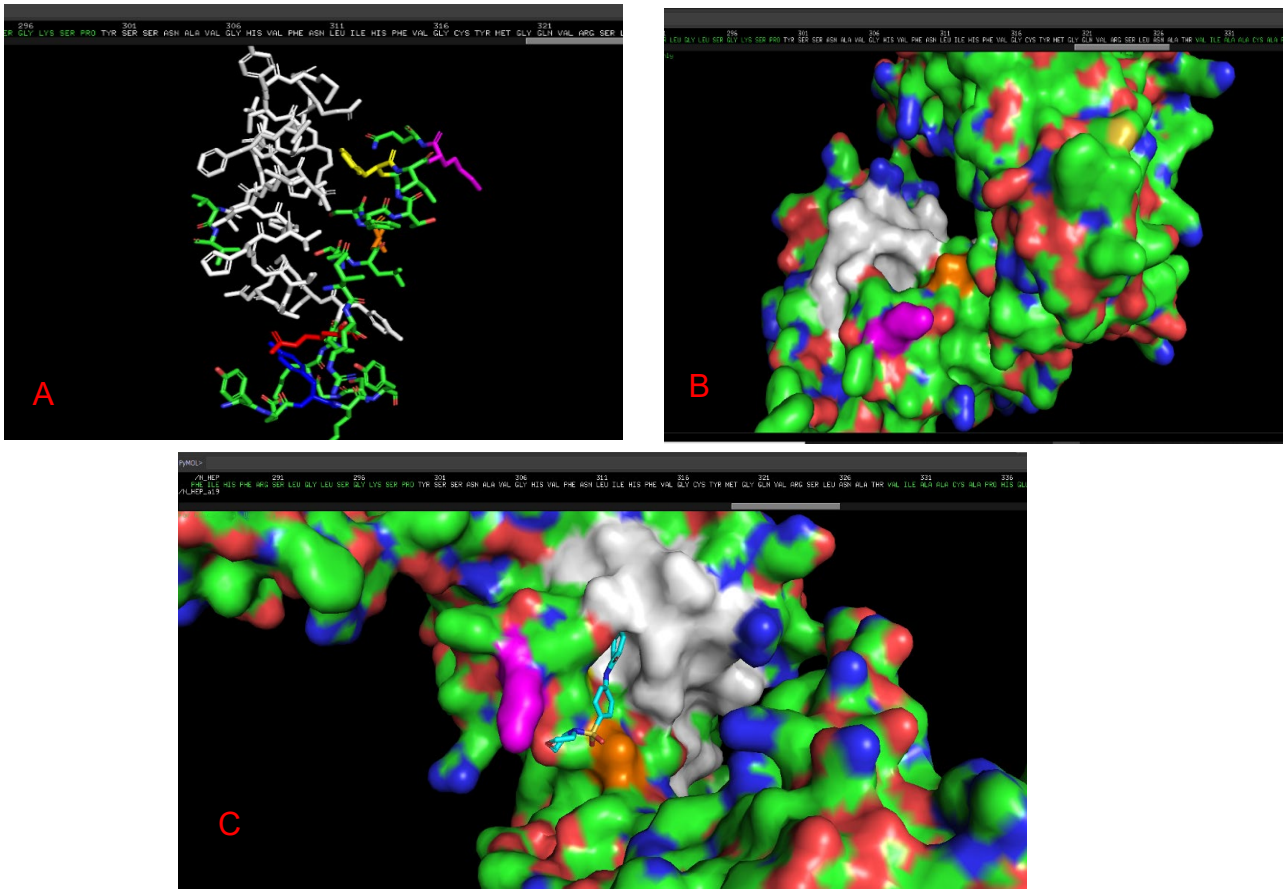


Figure 13 shows active sites formed when using active docking 300-328 in the nucleoprotein. Residues 300-328 are in white, while residue Arg271 is in red, Phe245 in yellow, Tyr28 in blue, Lys247 in pink, and Val240 in orange. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 13A shows the location of each of these residues in comparison to the active site (HEP strain of nucleoprotein). B) In figure 13B, it's possible to see that both orange (Val240) and pink (Lys247) residues are close to/touch the potential active site in white. However, no cavity is formed between them. C) Figure 13C shows ligand a19 binding in the cavity of the 300-328 potential active site (in white), touching the orange residue (Val240). Images made on PyMol.

Table 34 - 149-205 Nucleoprotein Active Site Results

Residue	Number of times bound
Ser222	22
Val226	19
Asn196	17
His219	17
Trp197	17
Arg290	17
Arg225	15
Lys162	15
Lys192	14
Ala223	14
Met60	11
Asn61	11
Ser198	11
Lys38	11
His184	11
Leu292	11
Ala62	10
Asn157	10

Table 34 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The residues Ser222, Val226, His219, Arg290, Arg225, Ala223, Met60, Asn61, Lys38, Leu292, and Ala62 were all outside of the potential binding site. Figure 14A shows the position of these residues on Nucleoprotein strain Nyctinomops. Figures 14B and 14C show different cavities that were formed by the determined active site and different extra residues that were bound to the ligands. It is clear in those pictures that it would be impossible for one ligand to connect to all these residues, especially considering that some residues lie on opposite sides of the protein. Figure 14D shows the ligand which got the lowest value in active docking in this proposed active site. The blue and pink residues were not used by this ligand, and instead it connected to residues in red (proposed active site), yellow, and white.

Figure 14 - Residues 149-205 and Peripheral Residues of Active Site 149-205

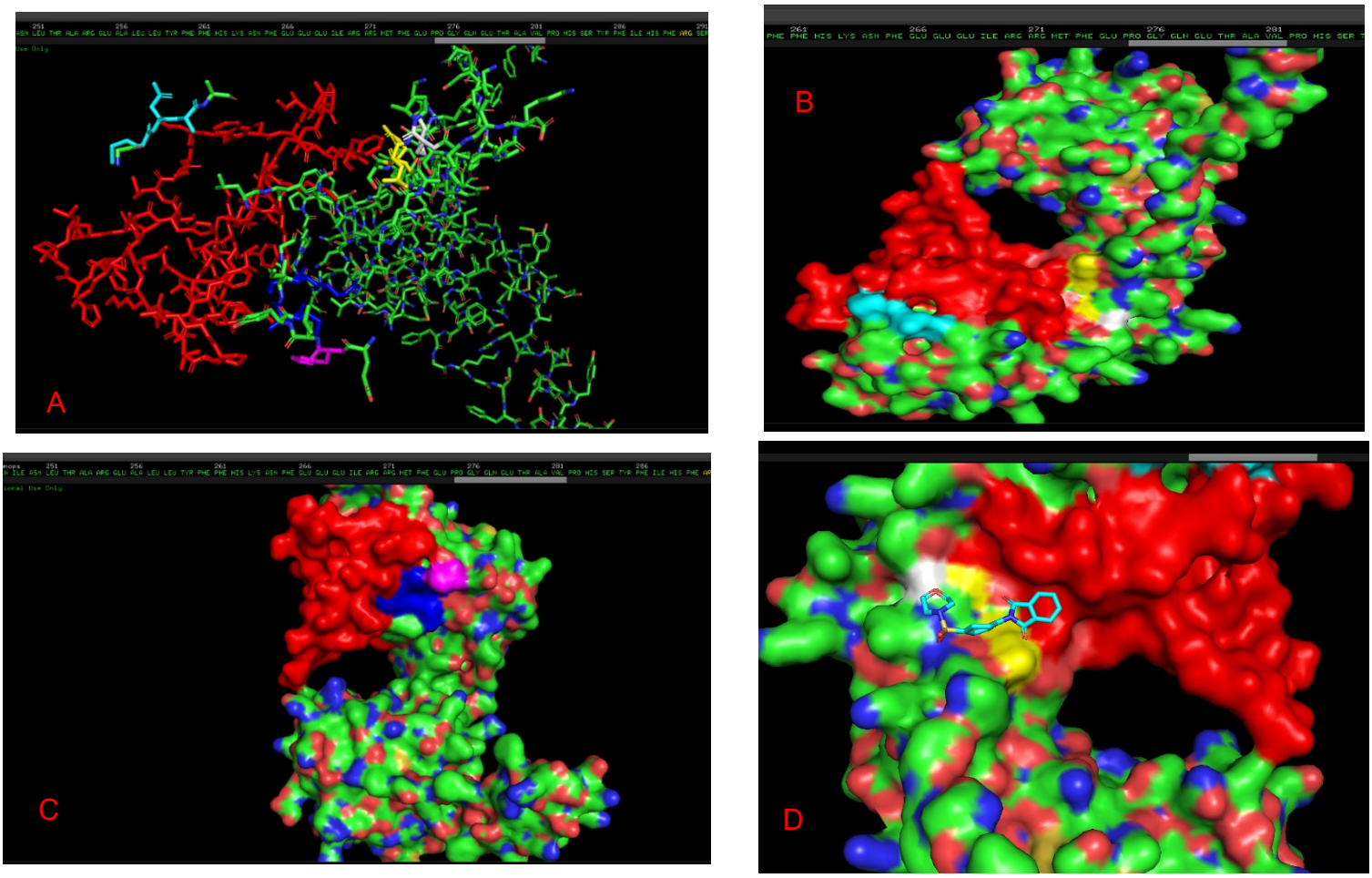


Figure 14 shows active sites formed when using active docking 149-205 in the nucleoprotein. Residues 149-205 are in red, while extra residues 222, 223, 225, and 226 are in blue, 290 and 292 in yellow, 60, 61, and 62 in light blue, Lys38 in white and His219 in pink. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 14A shows the location of each of these residues in comparison to the active site (Nyctinomops strain of nucleoprotein). B) In figure 14B, it's possible to see that residues in white, yellow and light blue are close to the potential binding site in red, with the light blue residues forming a possible cavity with the red residues C) Figure 14C shows another side of the protein (the "top"), where blue and pink residues are next to the potential active site but no actual cavity is formed. D) Figure 14D shows one of the poses ligand a19 formed when docking to this active site. It does not bind to any particular cavity, but it does seem to connect to red residues (predicted active site), as well as yellow (290 and 292) and white (Lys38) residues. Image made in PyMol.

Table 35 - 251-273 Nucleoprotein Active Site Results

Residue	Number of times bound
Phe260	109
Glu266	91
Arg270	67
Arg400	44

Leu251	43
Ile257	33
Thr354	28
Arg254	24
Leu258	24
Leu257	21
Ile269	20
His262	20
Thr243	19
Arg271	12

Table 35 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Arg400, Thr354, and Thr 243 were not a part of the potential active site 251-273. As can be seen in Figure 15, the extra residues were not too far from the predicted potential active site, however, as seen in figure 15B only residues Thr354 and Thr243 formed a cavity with the potential active site. This cavity, as noticed in figure 15C, is not used by ligand A11, however.

Figure 15 - Residues 251-273 and Peripheral Residues of Active Site 251-273

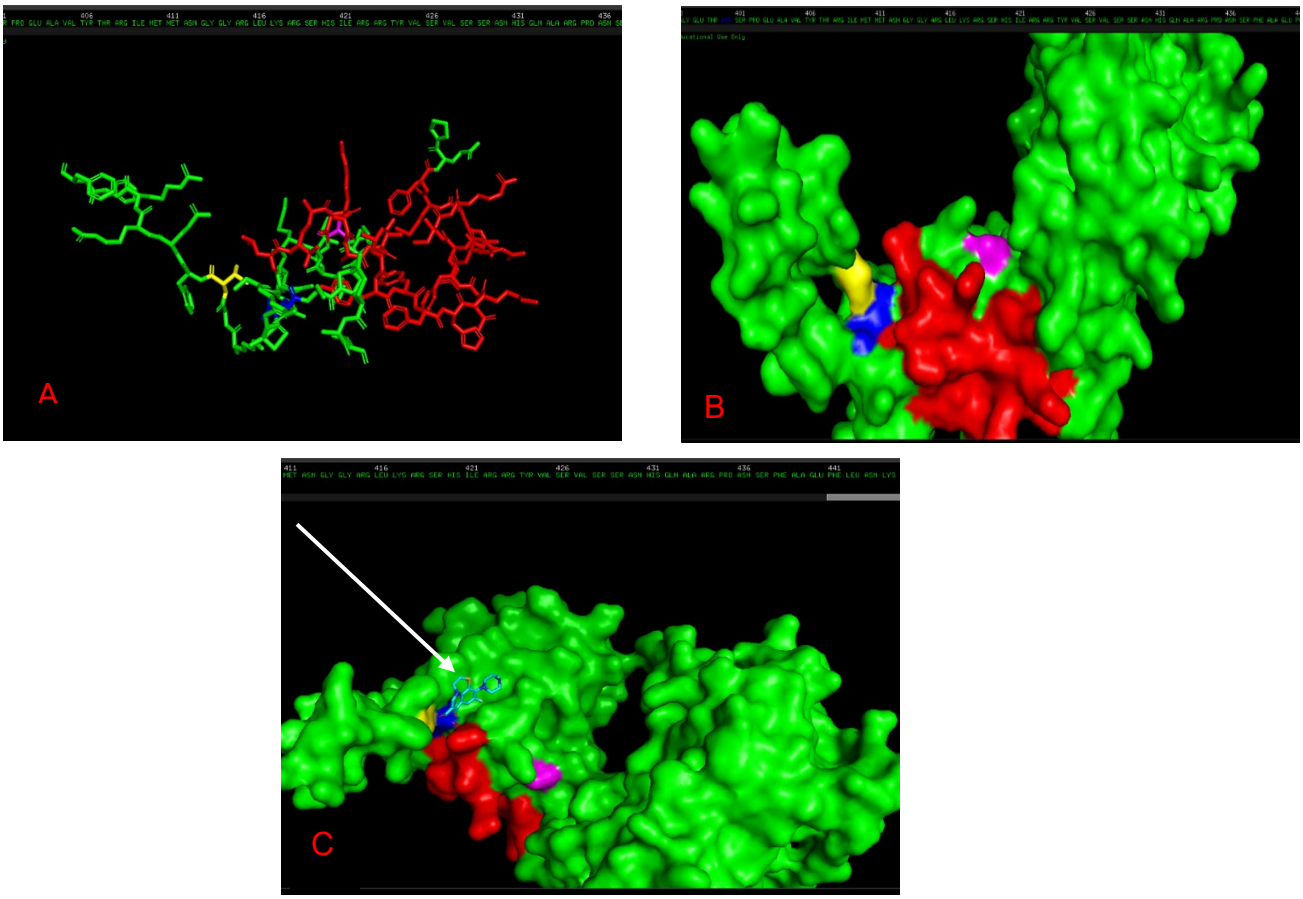


Figure 15 shows active sites formed when using active docking 251-273 in the nucleoprotein. Residues 251-273 are in red, while Arg400 is in blue, Thr354 in yellow, and Thr243 in pink. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 15A shows the location of each of these residues in comparison to the active site (SAD strain of nucleoprotein). B) In figure 15B, it's possible to see that residues in blue and yellow are somewhat close to the red residues, while pink is farther away. Blue and yellow residues form a potential small cavity with the red potential binding site. C) Figure 15C shows ligand A11 (in light blue, marked by arrow) in the only pose it acquired throughout this active docking, touching both blue (Arg400) and red residues. In no moment did it lay in the potential active site that can be seen in picture 15B. Image made in PyMol.

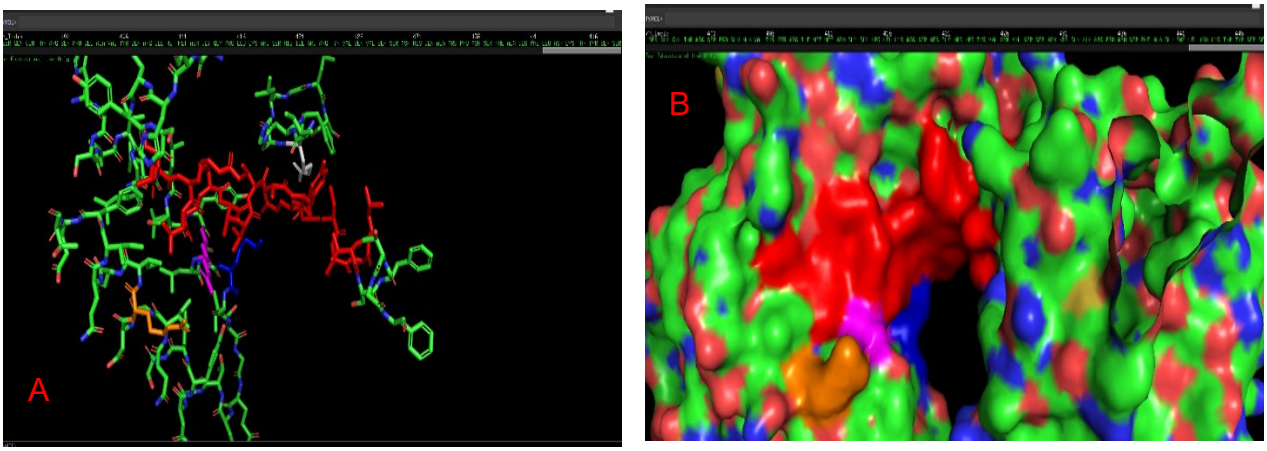
Table 36 - 223-240 Nucleoprotein Active Site Results

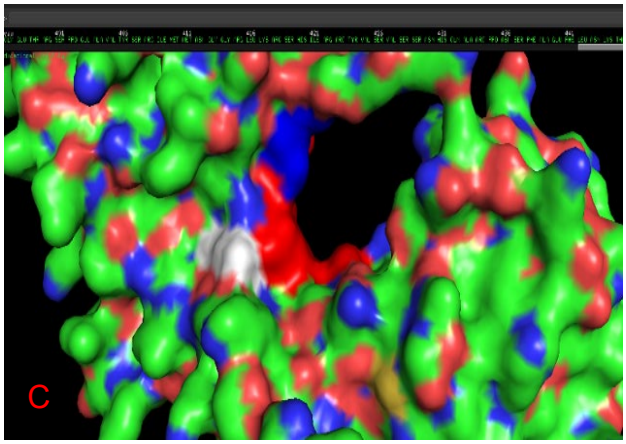
Residue	Number of times bound
Arg149	186
Arg168	178
Lys152	61
Asp235	40
Arg290	19

Table 36 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

All residues that bound more than 10 times to the ligands were outside of the predicted active site, expect for Asp235. Despite that, Figure 16A shows that some (pink and blue) residuesare close to the predicted active site. However, as seen in figure 16B only residues Arg149 and Arg152 formed a cavity with the potential active site. Figure 16C also shows that residue Arg290 is on the opposite side of the protein when compared to the other peripheral residues, and it didn't form a cavity with the predicted potential active site. The cavity seen in 16B is not used by ligand A11, as can be seen in picture 16D.

Figure 16 - Residues 223-240 and Peripheral Residues of Active Site 251-273





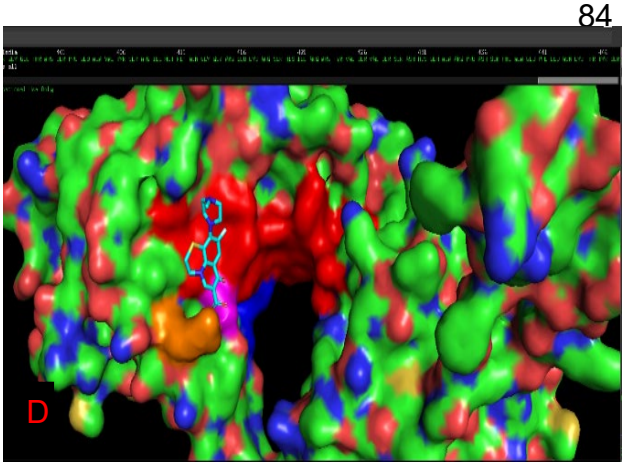


Figure 16 show active sites formed when using active docking 223-240 in the nucleoprotein. Residues 223-240 are in red, while Arg149 is in pink, Arg168 in orange, Arg152 in blue, and Arg290 in white. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 16A shows the location of each of these residues in comparison to the active site (India strain of nucleoprotein). B) In figure 16B, it's possible to see that residues in blue, orange and pink are somewhat close to the red residues, but no binding cavity seems to be made. C) Figure 16C shows the other side of the molecule, where residue Arg290, in white, residues, as well as a portion of red and blue residues. No cavity is made by the closeness of the red and white residues. D) Figure 16D shows ligand A11 (in light blue) in the only pose it acquired throughout this active docking, touching both pink (Arg149) and red residues. It doesn't exactly law in the cavity shown in picture 16B. Image made in PyMol.

Table 37 - 27-38 Nucleoprotein Site Results

Residue	Number of times bound
Lys38	128
Lys29	106
Arg290	87
Pro275	76
Ser295	60
Leu294	55
Asp35	44
Tyr28	39
Phe205	35
Lys34	29
Gly296	29
Tyr30	27
Arg204	23
Ser291	20
Gly276	18
Ile41	16
Asn202	14

Pro201	14
Gln277	14
Ala32	11

Table 37 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Arg290, Pro275, Ser295, Leu294, Phe205, Gly296, Arg204, Ser291, Gly276, Ile41, Asn202, Pro201, and Gln277 are outside of the predicted potential active site. While Ile41 is incredibly close to the proposed active site, the other peripheral residues are farther away. Figures 17A, 17B, and 17C show that even the peripheral residues are generally close to the predicted potential active site. Pictures 17D and 17E show the two poses ligand A11 acquired in this blind docking.

Figure 17 - Residues 27-38 and Peripheral Residues of Active Site 27-38

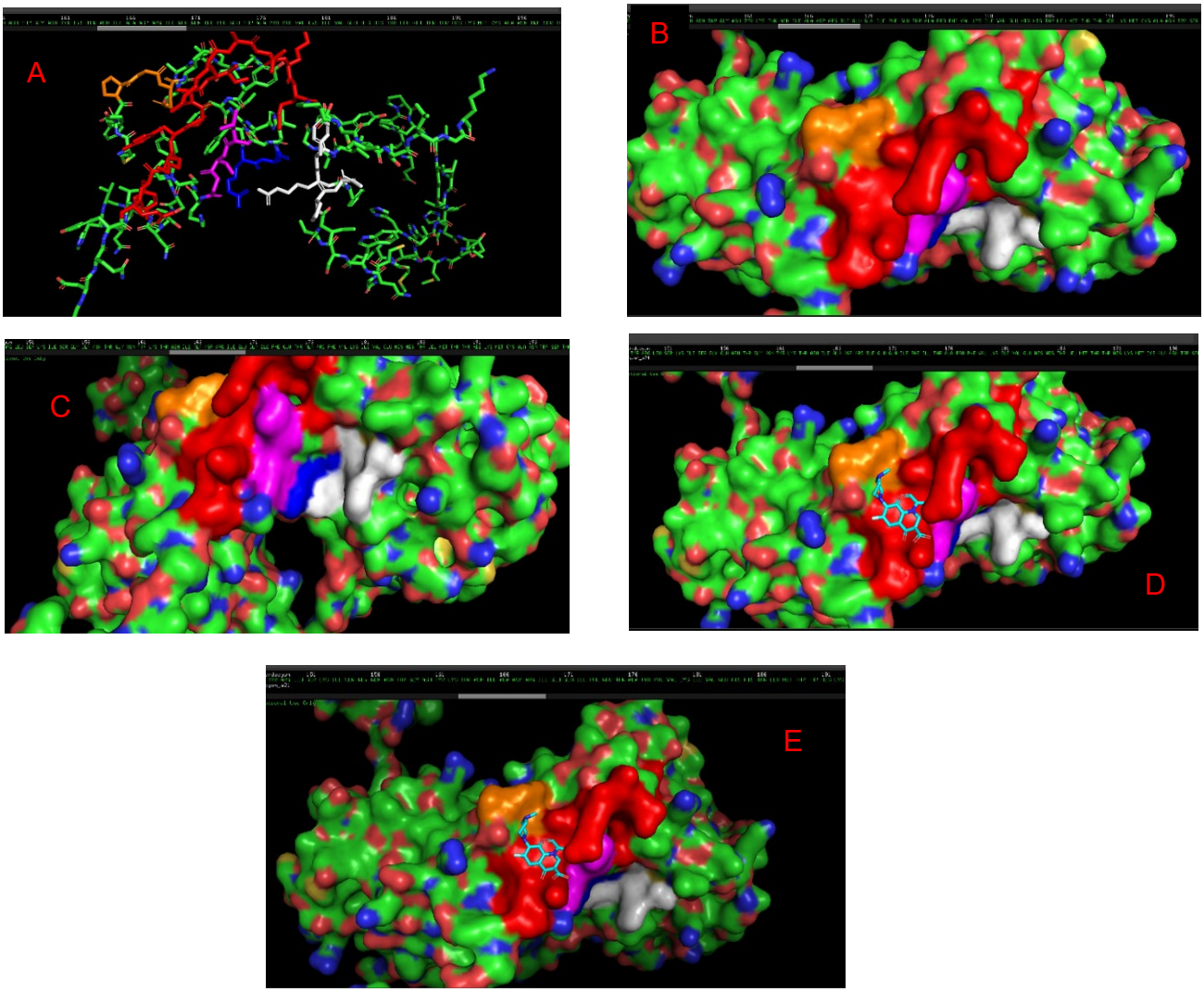


Figure 17 show active sites formed when using active docking 27-38 in the nucleoprotein. Residues 27-38 are in red, while Arg290 and Ser291 are in blue; Pro275, Gly276, and Gln277 are in orange; Ser295, Leu294, and Gly296 are in pink; and Pro201, Asn202, Arg204, and Phe205 are in white. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 17A shows the location of each of these residues in comparison to the active site (Cerdocyon strain of nucleoprotein). B) In figure 17B, it's possible to see that residues orange are close to the predicted active sites, forming a possible cavity. C) Figure 17C shows the other side of the predicted potential site, where pink, blue, and white residues lie, all close to the red residues, with pink, blue, and white form a possible big cavity with the red residues D) Figure 17D shows ligand A11 (in light blue) in one of the poses it acquired throughout this active docking, touching both orange (Gln277) and red residues, as well as within the possible cavity seen in picture B. E) Figure 17E shows the other pose ligand A11 acquired throughout this active docking, in the cavity of the red residues, but touching both orange (Gln277) and pink (Gly29) residues. Image made in PyMol.

Table 38 - 43-69 Glycoprotein Active Site Results

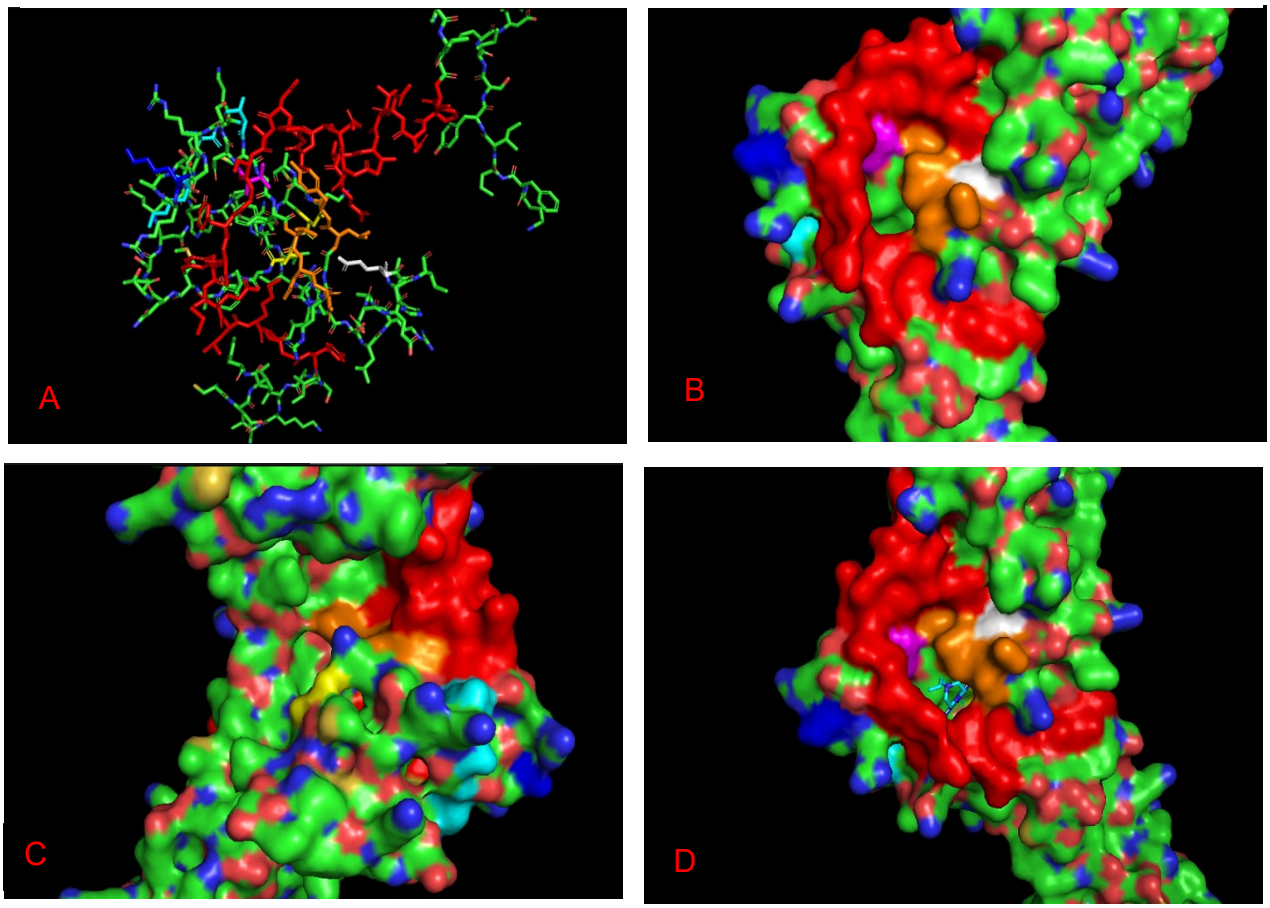
Residue	Number of times bound
Lys217	275
Glu50	193
Leu57	175
Gly53	159
Thr55	120
Asp51	101
Thr212	80
Val49	63
Phe60	58
Val229	52
Glu52	51
Asp230	48
Glu231	32
Glu288	29
Asn56	27
Gly59	27
Cys54	24
Tyr62	21
Thr225	20
Ser58	18
Tyr235	18
Glu64	17

Lys221	16
Asn55	16
Ser56	14
Ser55	13
Asp256	12
Thr258	11
Ser61	11
Leu234	10
Asn223	10

Table 38 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Lys217, Thr212, Val229, Asp230, Glu231, Glu288, Thr225, Tyr235, Lys221, Asp256, Thr258, Leu234, and Asn223 are outside of the predicted potential active site. In Figure 18B it's visible that residues pink, orange and white are close to the projected active site, with orange, pink, and red seeming to form a binding cavity. On the other side of the protein, as can be seen from Figure 18C, residues in yellow and light blue are near the projected binding site, but no cavities are formed. In Figures 18D and 18E, ligand A9 binds to the cavity visible in Figure 18B, even if it

Figure 18 - Residues 43-69 and Peripheral Residues of Active Site 43-69



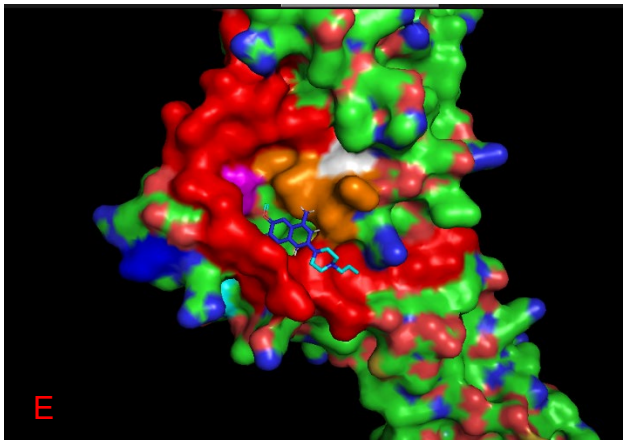


Figure 18 show active sites formed when using active docking 43-69 in the nucleoprotein. Residues 43-69 are in red, while Lys217 is in blue, Thr212 in pink; Val229, Asp230, Glu231, Leu234, and Tyr235 are in orange; Lys221, Asn223, and Thr225 are in light blue; Glu288 in white, and Asp256 andThr258 in yellow. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 18A shows the location of each of these residues in comparison to the active site (Nishigahara strain of glycoprotein). B) In figure 18B, it's possible to see that residues orange and pink are close to the predicted active sites, with orange seeming to form a cavity with the red residues. C) Figure 18C shows the other side of the predicted potential site, where yellow residues and parts of light blue and orange residues lie, all but yellow close to the red residues. None seem to form a cavity, however. D) Figure 18D shows ligand A9 (in light blue) in one of the poses it acquired throughout this active docking, touching both orange and red residues, as well as lying in the cavity they form together. E) Figure 18E shows the other pose ligand A9 acquired throughout this active docking, touching both red and orange residues. Image made in PyMol.

Table 39 - 120-143 Glycoprotein Active Site Results

Residues	Number of times bound
Lys148	97
Ser150	95
Tyr127	93
Asn133	88
Lys104	74
Ile152	72
Tyr135	65
Val152	60
Tyr138	59
Glu129	58
Glu128	56
Glu149	54
Leu141	45
Leu131	37

Pro136	35
Ile154	33
Trp140	29
Asp137	24
Lys120	23
Ser130	20
Arg126	19
Ala112	19
His132	17
Thr91	17
Thr146	16
Lys145	16
Ala116	16
Glu86	15
Pro134	13
Arg107	11
Arg103	11
Phe93	10
Gly95	10

Table 39 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Lys148, Ser150, Lys104, Val/Ile152, Glu149, Ile154, Ala112, Thr91, Thr146, Lys145, Ala116, Glu86, Arg107, Arg103, Phe93, and Gly95 are outside of the predicted potential active site. In Figure 19B it's visible that residues yellow and pink are close to the proposed active site (blue), and Figure 19C shows that the blue and pink residues do form a cavity. In terms of binding, Figures 19D and 19E show that ligand A19 had two different poses, one that lays in the cavity seen in Figure 19C and one in Figure 19D.

Figure 19 - Residues 120-143 and Peripheral Residues of Active Site 120-143

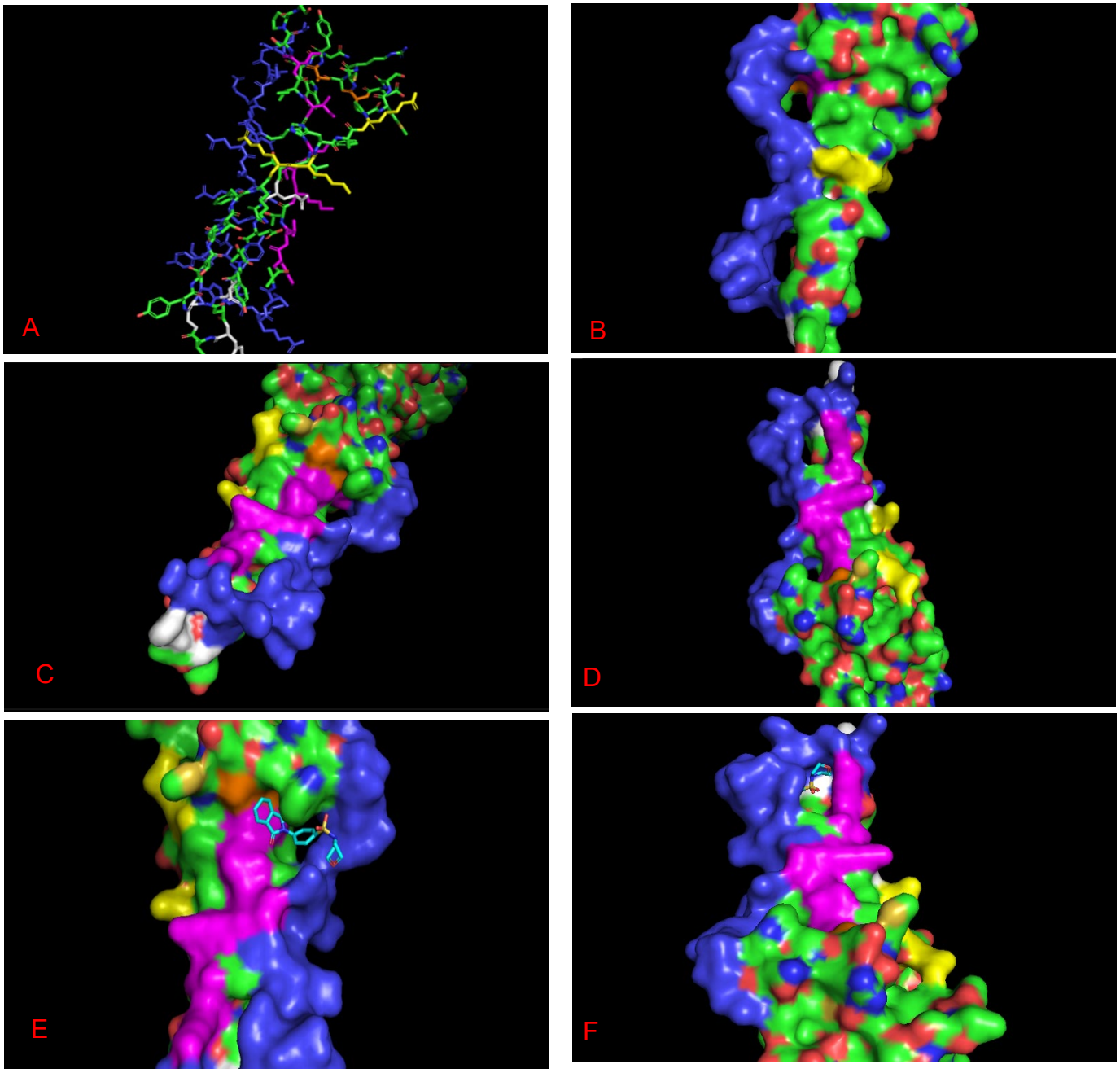


Figure 19 show active sites formed when using active docking 12-143 in the glycoprotein. Residues 120-143 are in blue, while Lys145, Thr146, Lys148, Glu149, Ser150, Val/Ile152, and Ile154 are in pink; Lys104, Arg103, and Arg107 in yellow; Glu86, Thr91, Phe93, and Gly95 are in white; Ala112 and Ala116 are in orange. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 19A shows the location of each of these residues in comparison to the active site (Nyctinomops strain of glycoprotein). B) In figure 19B, it's possible to see that residues yellow and pink are close to the predicted active sites and, in picture 19C, it's possible to see that these residues (orange and pink) form a cavity with the blue residues. D) Figure 19D shows a different cavity, one between the blue residues, pink residues and white residues at the end of the protein. E) Figure 19E shows ligand A19 (in light blue) in one of the poses it acquired throughout this active docking lying in the cavity seen in Figure C E) Figure 19E shows the other pose ligand A9 acquired throughout this active docking, now in the cavity seen in Figure 19D, touching the white residues. Image made in PyMol.

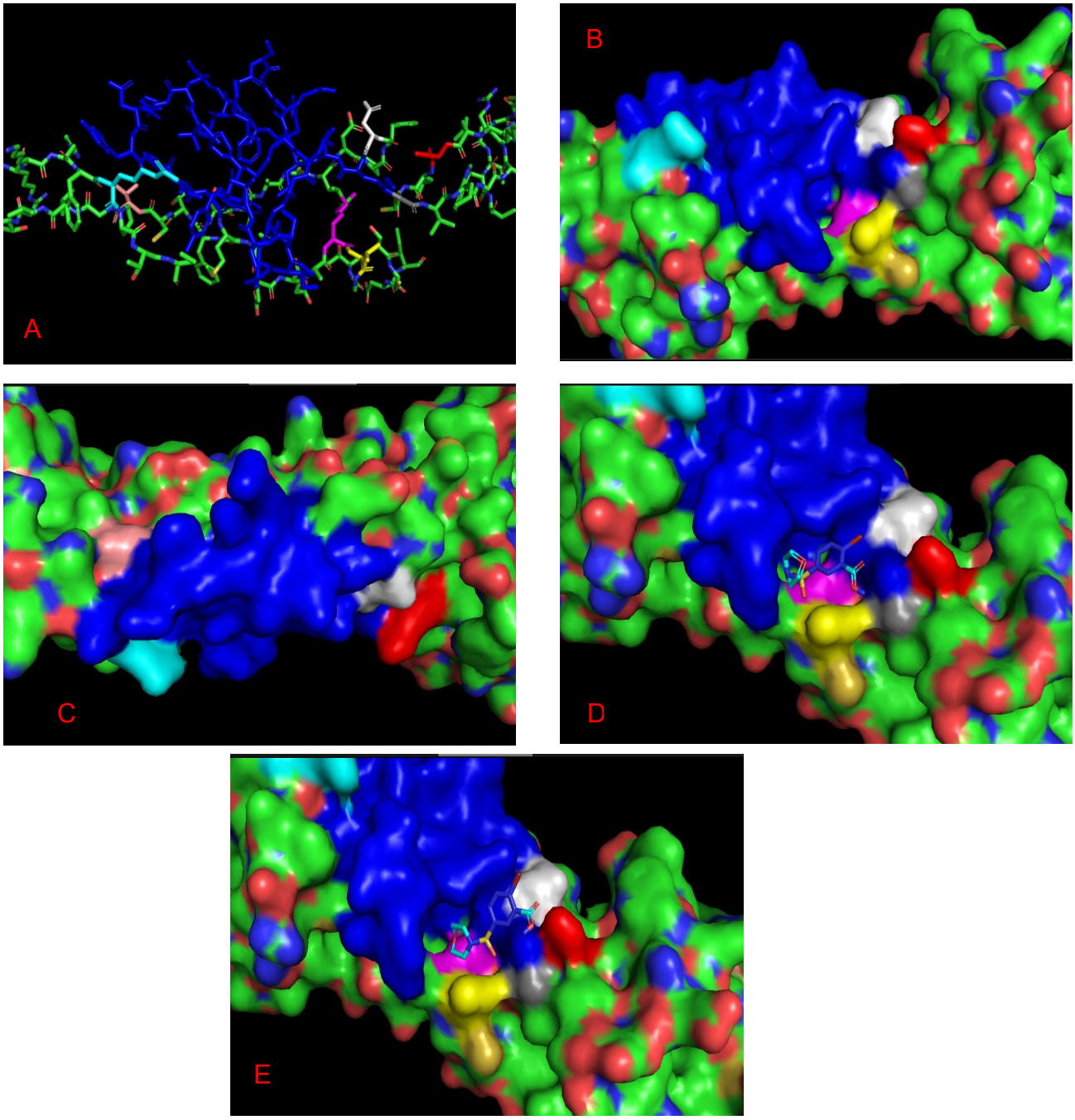
Table 40 - 166-197 Glycoprotein Active Site Results

Residue	Number of times bound
Lys166	113
Arg203	98
Ser167	61
Leu204	44
Lys177	41
Thr195	38
Ser185	33
Ser179	31
Asp165	31
Lys245	29
Thr190	26
Gly180	26
Cys178	25
His192	25
Tyr194	25
Leu179	24
Thr81	21
Arg166	19
Ser184	18
Thr179	17
Asp193	16
Tyr187	15
Gly248	13
Ser189	13
Thr186	11
Pro174	11
Arg107	11

Table 40 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Arg203, Leu204, Asp165, Lys245, Thr81, Gly228, and Arg107 are outside of the predicted potential active site. In Figure 20B it's visible that residues Arg203, Leu204, Asp165, and Arg107 are near the proposed active site (in dark blue), but only Arg203 and Leu204 seem to form a binding cavity. Figure 20C shows the other side of the protein, and the salmon residue near the blue residues. It does not form a binding cavity, however. Figures 20D and 20E show ligand A14 in the cavity shown in Figure 20B.

Figure 20 - Residues 166-197 and Peripheral Residues of Active Site 166-197



Figures 20 show active sites formed when using active docking 166-197 in the glycoprotein. Residues 166-197 are in blue, while Arg203 is in pink, Leu204 in yellow, Asp165 in white, Lys245 in red, Thr81 in salmon, Gly248 in gray, and Arg107 in light blue. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 20A shows the location of each of these residues in comparison to the active site (Tadarida strain of glycoprotein). B) In figure 20B, it's possible to see that residues yellow, white, pink, and light blue are near the potential active site in dark blue, but only the pink residue forms a cavity with the dark blue residues. C) Figure 20C shows a different side of the protein where the salmon residue is near the blue residues. No cavity is formed, however. D) Figure 20D shows ligand A14 (in light blue) in the most common of the poses it acquired throughout this active docking lying in the cavity seen in Figure 19B. E) Figure 20E shows the other pose ligand A9 acquired throughout this active docking, in the same cavity, but also touching the white residue. Image made in PyMol.

Table 41 - 217-240 Glycoprotein Active Site Results

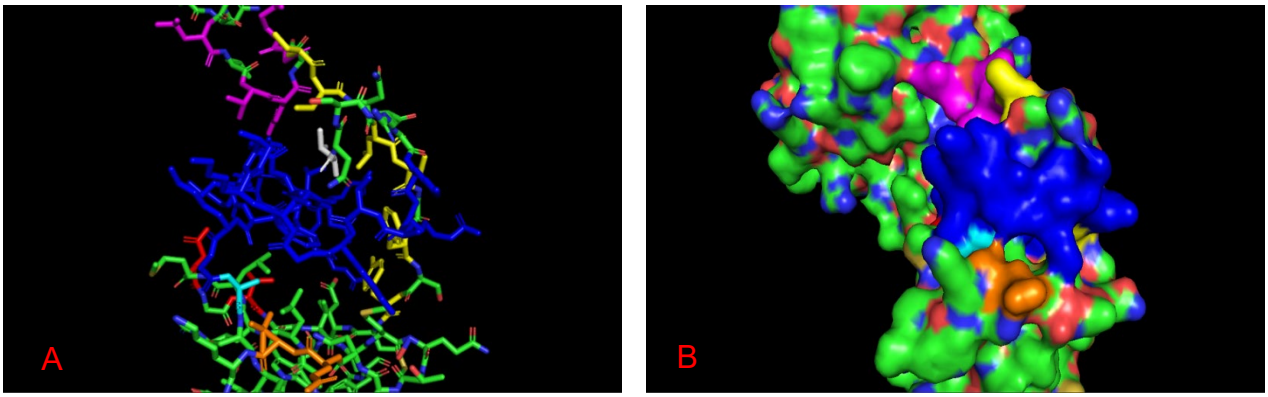
Residue	Number of times bound
Tyr235	74
Lys236	69
Thr225	55
Ser237	50
Val229	50
Val49	49
Cys54	47
Glu50	45
Leu57	39
Arg218	37
Lys217	35
Leu238	34
Thr55	33
Tyr62	32
Cys226	31
Leu47	31
Phe60	31
Thr212	28
Lys218	27
Gly53	23
Ala241	20
Lys221	20
Gly240	19
Leu234	18

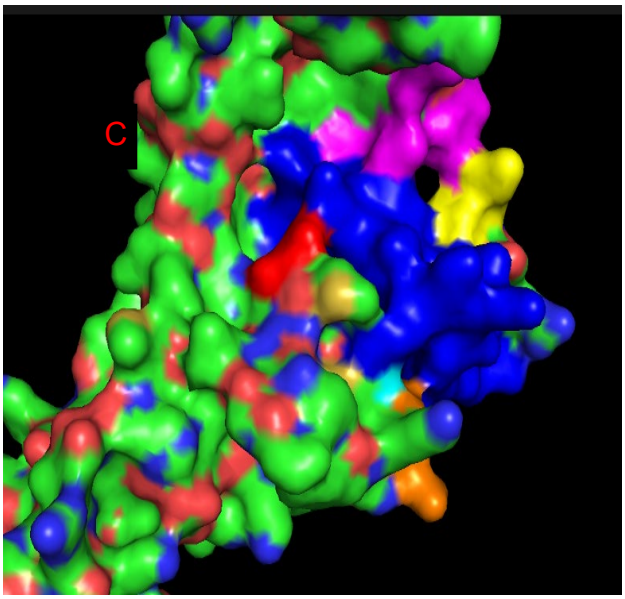
Lys239	17
Thr258	15
Asp256	15
Thr268	13
Ser58	12
Glu52	12
Asp267	11
Lys224	10
Gly227	10
Glu267	10
Asn223	10

Table 41 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Arg203, Leu204, Asp165, Lys245, Thr81, Gly228, and Arg107 are outside of the predicted potential active site. In Figures 21B, 21C, and 21D it's visible that residues in pink, yellow, white, orange, and red are near to the blue proposed binding site. Some also form cavities with the proposed active site. Figure 21D shows the ligand, in all its poses in this active docking, was laying in the active cavity formed between the blue residues, together with yellow, white, and pink.

Figure 21 - Residues 217-240 and Peripheral Residues of Active Site 217-240





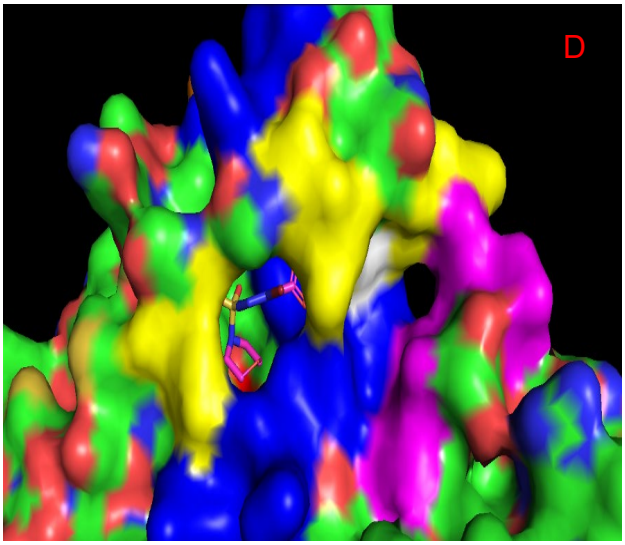


Figure 21 show active sites formed when using active docking 217-240 in the glycoprotein. Residues 217-240 are in blue, while Leu47, Val49, Glu50, Glu52, and Gly53 are in pink; Cys54, Thr55, Leu57, Ser58, Phe60, and Tyr62 in yellow; Thr212 is in white, Ala241 in light blue, Asp256, and Thr258 in red, and Asp/Glu267 and Thr268 in orange. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 21A shows the location of each of these residues in comparison to the active site (DRV strain of glycoprotein). B) In figure 21B, it's possible to see that residues pink, orange and light blue are near the proposed active site, and there's one possible cavity formed between orange and blue residues. C) Figure 21C shows a different side of the protein where yellow, white, and pink residues are near the blue residues. Two cavities are formed, one between yellow and blue, and one between yellow, white, pink, and blue. D) Figure 21D shows the only pose ligand A14 acquired during this active docking. It is in the binding cavity made by the blue, white, and yellow residues shown in Image 21C. Image made in PyMol.

Table 42 - 253-270 Glycoprotein Active Site Results

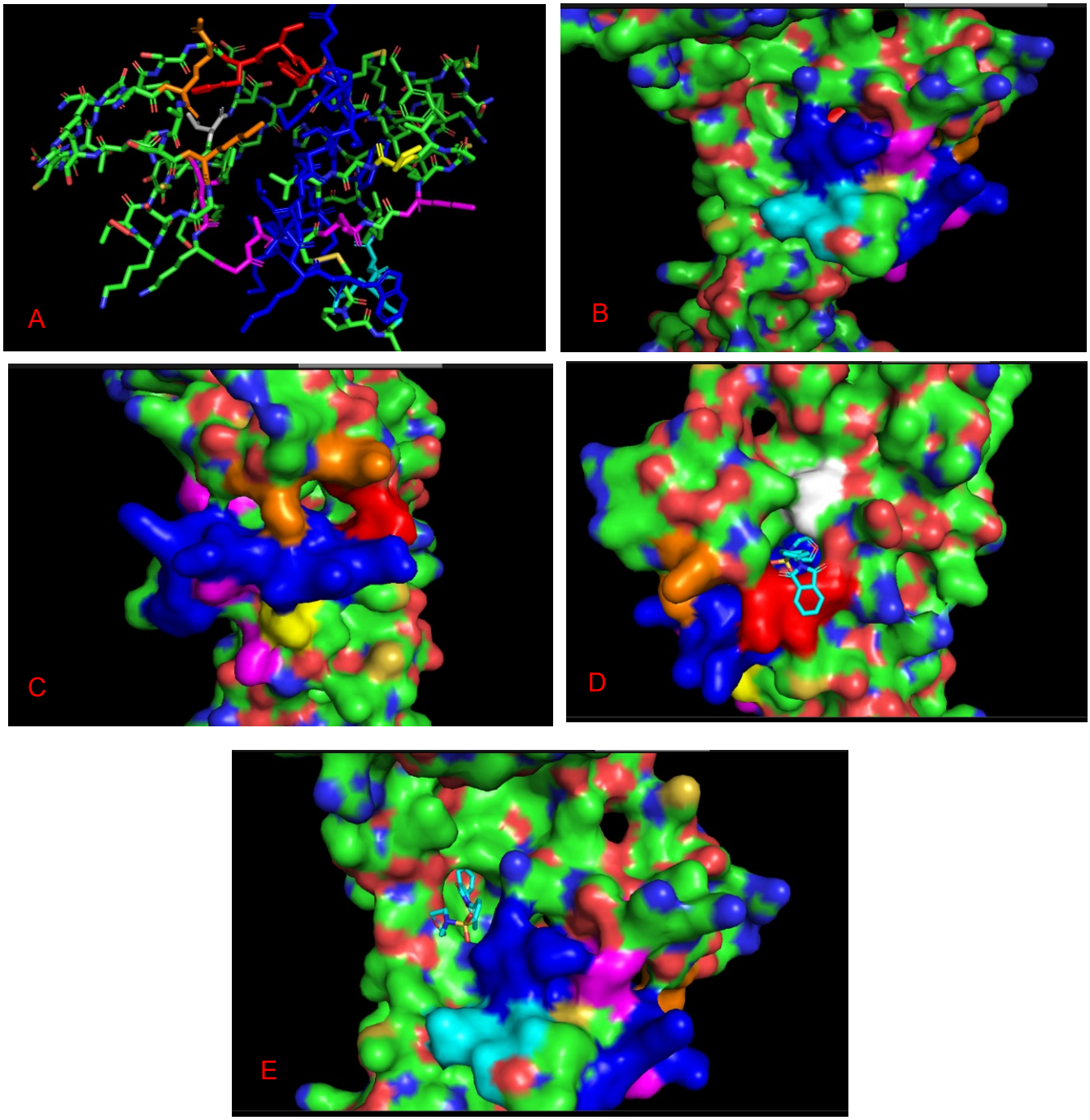
Residue	Number of times bound
Arg218	82
Gly240	67
Lys269	62
Ala241	61
Glu267	59
Trp270	53
Thr268	50
Lys243	49
Tyr62	45
Ser265	41
Lys245	40

Asp266	37
Lys221	31
Ser61	29
Arg253	29
Ser250	25
Gln263	24
Thr264	24
Asp256	17
Leu238	17
Gln275	16
Arg243	16
Asp274	14
Lys218	14
Val262	13
Val229	13
Ser261	12
Leu276	12
Phe60	10

Table 42 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Phe60, Ser61, Tyr62, Arg/Lys218, Lys221, Leu238, Gly240, Ala241, Lys/Arg243, Lys245, Val229, Ser250, Asp274, Gln275, and Leu276 are outside of the predicted potential active site. In Figures 22A, 22B, and 22C, it's visible that residues in red, pink, cyan, orange, and yellow are near to the blue proposed binding site. Cyan forms a cavity with the blue residues (Figure 22B), and the same can be said for red and blue (Figure 22C). Figures 22D and 22E shows the ligand in the two different poses it acquired during docking. In Figure 22D, it is in a cavity made between residues in red and dark blue, and in Figure 22E it is only in a cavity made by the blue residues with other extra residues.

Figure 22 - Residues 253-270 and Peripheral Residues of Active Site 253-270



Figures 22 show active sites formed when using active docking 253-270 in the glycoprotein. Residues 253-270 are in blue, while Phe60, Ser61, and Tyr62 are in red; Arg/Lys218 and Lys221 in orange; Leu238, Gly240, Ala241, and Lys/Arg243 in pink; Val229 in white; Ser250 in yellow; and Asp274, Gln275, and Leu276 in cyan. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 22A shows the location of each of these residues in comparison to the active site (DRV strain of glycoprotein). B) In figure 22B, it's possible to see that residues in pink, and cyan are near the proposed active site, with dark blue and cyan forming a possible cavity. C) Figure 22C shows a different side of the protein where yellow, orange, and red residues are near the blue residues. Cavities can be seen formed between blue, orange, and red residues. D) Figure 22D shows ligand A19 in one of its conformations, in the cavity formed between red and blue residues. E) Figure 22E shows the other pose ligand A19 acquired during this active docking, in a cavity formed between the blue residues and extra residues of the protein. Image made in PyMol.

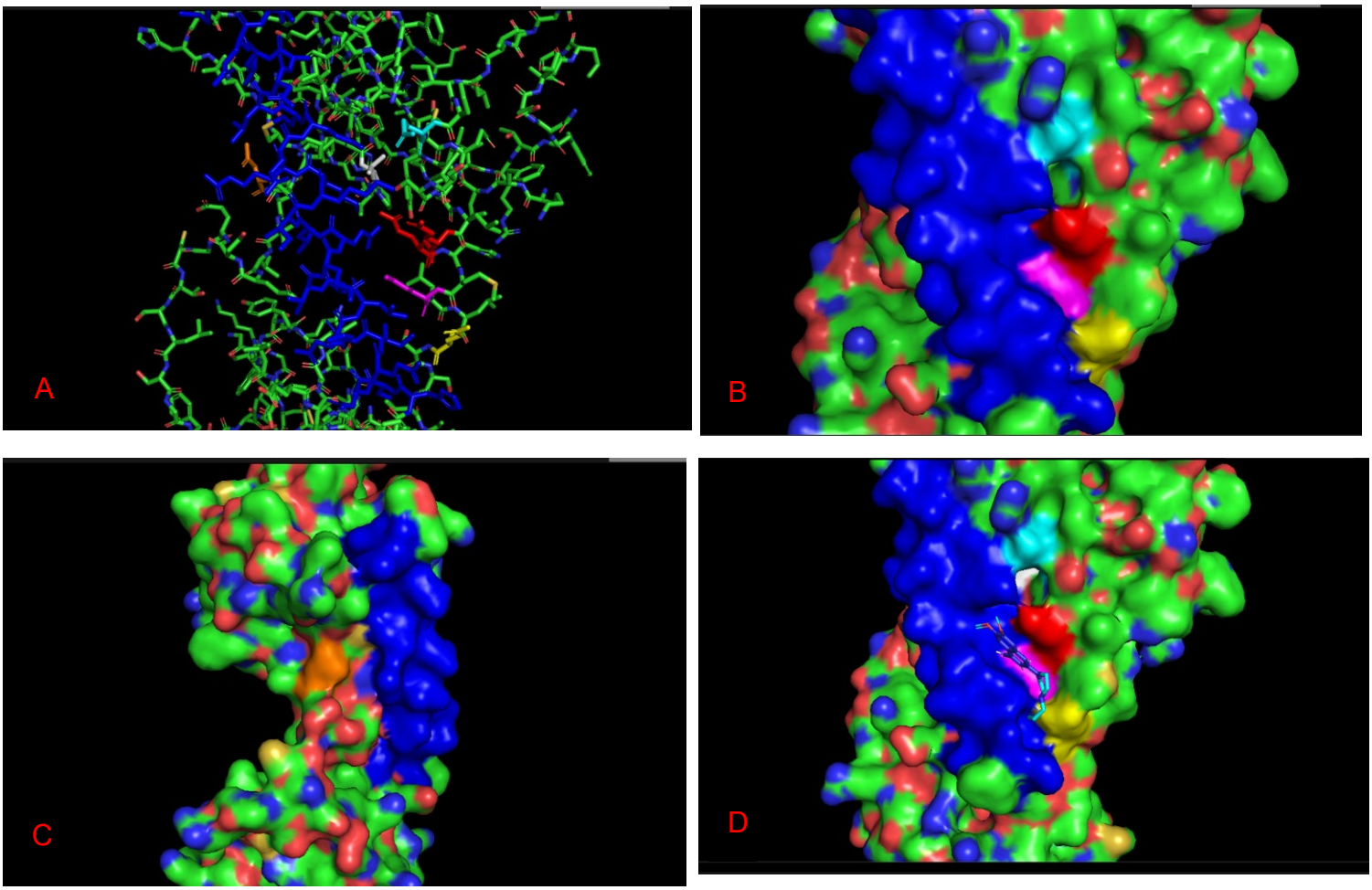
Table 43 - 283-310 Glycoprotein Active Site Results

Residue	Number of times bound
Lys297	289
Gln401	273
Glu405	144
Glu294	102
Glu408	95
Leu290	95
Lys298	72
Glu293	44
Arg299	43
Leu326	29
Gln402	22
Arg297	21
Glu307	12
Asn46	12
Glu394	11
Asp304	11

Table 43 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Gln401, Glu405, Glu408, Leu326, Gln402, Asn46, and Glu394 are outside of the predicted potential active site. In Figures 23B, and 23C it's visible that residues in pink, yellow, red, orange, and light blue are near to the blue proposed binding site. Some also form cavities with the proposed active site. Figure 23D shows the ligand, in all its poses in this active docking, was laying in the active cavity formed between the blue residues, together with red and pink.

Figure 23 - Residues 283-310 and Peripheral Residues of Active Site 283-310



Figures 23 show active sites formed when using active docking 283-310 in the glycoprotein. Residues 283-310 are in blue, while Gln401 and Gln402 are in red; Glu405 is in pink, Glu408 in yellow, Leu326 is in white, Asn46 in orange, and Glu394 in cyan. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 23A shows the location of each of these residues in comparison to the active site (Callithrix strain of glycoprotein). B) In figure 23B, it's possible to see residues in pink, red, and cyan are near the proposed active site, with dark blue, pink, and red forming a possible cavity. C) Figure 23C shows a different side of the protein where orange can be seen near the proposed active site, with a possibility cavity formed. D) Figure 23D shows ligand A9 in the only conformation it acquired, in the cavity formed between red, pink, and blue residues. Image made in PyMol.

Table 44 - 620-700 Large Protein Active Site Results

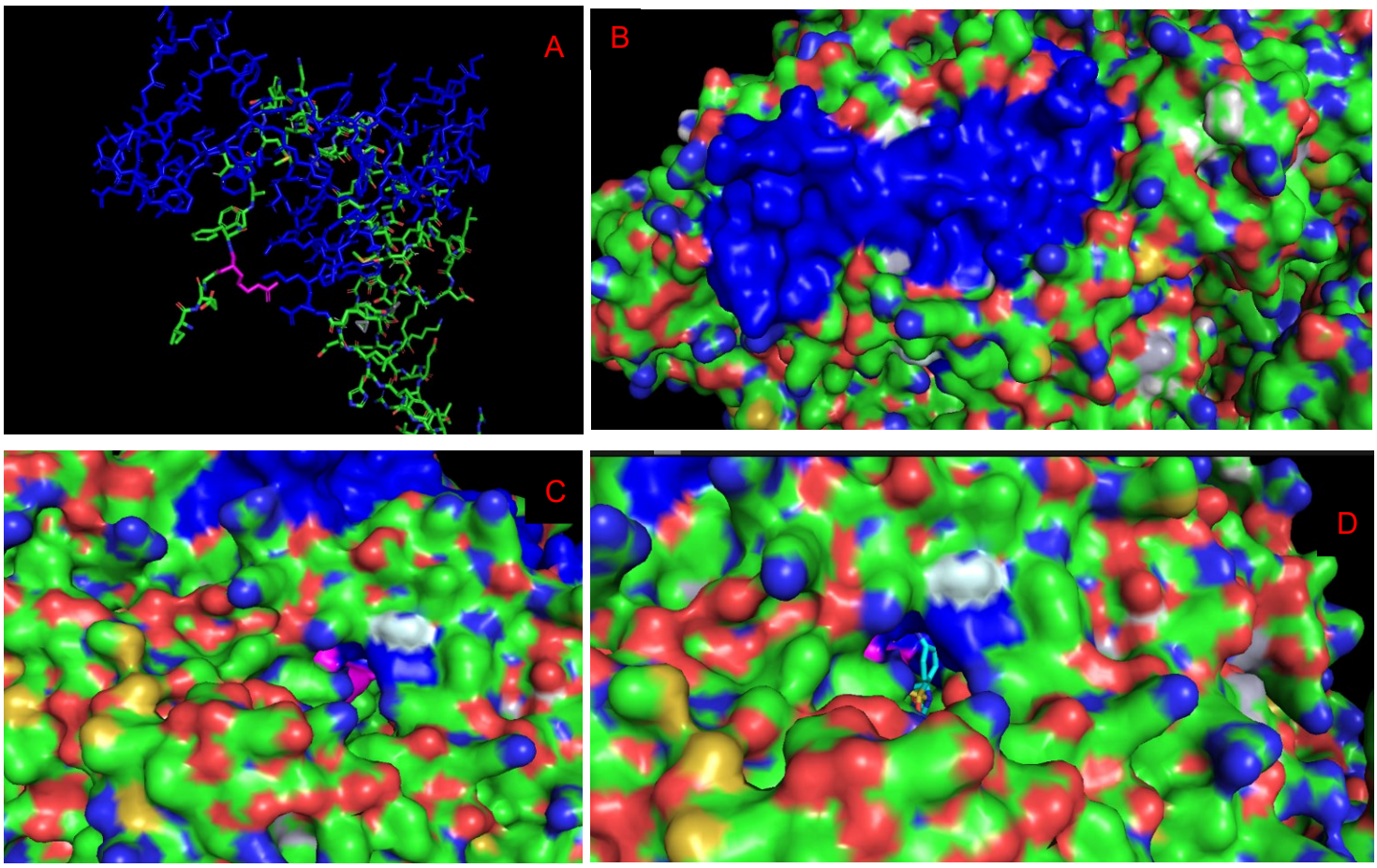
Residue	Number of times bound
Asn689	16
Ser649	14
Arg650	12

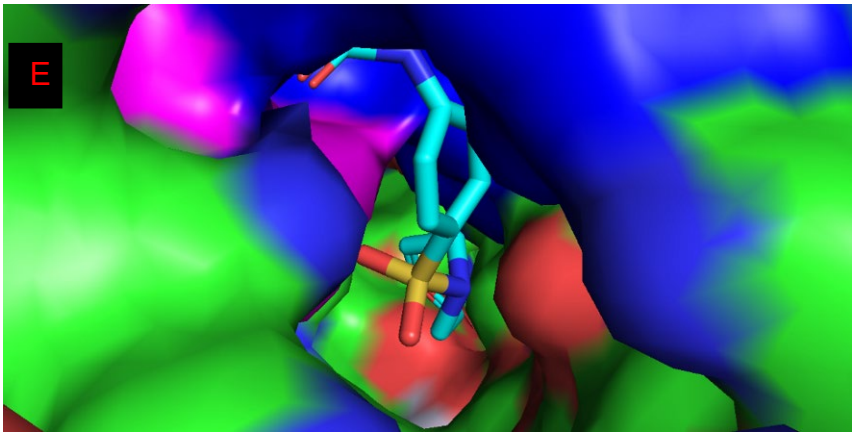
Gln656	11
Lys657	11
Arg552	10
Asn623	10

Table 44 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

The only residue outside of the potential active site that bound to ligand A19 is Arg552. In Figure 24C it's possible to see the extra residue formed a cavity with the blue residues (shown by arrow). The figures make it harder to notice because the residue in pink is "inside" the protein instead of on its surface (like the blue residues). In Figures 24D and 24E it's possible to see the ligand, in all its poses in this active docking, was laying in the cavity shown in figure 24C, touching blue and pink residues.

Figure 24 - Residues 620-700 and Peripheral Residues of Active Site 620-700





Figures 24 show active sites formed when using active docking 620-700 in the large protein. Residues 620-700 are in blue, while Arg552 is in pink. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 24A shows the location of each of these residues in comparison to the active site (SAD strain of large protein). B) In figure 24B, it's possible to see that only the blue residues (the proposed active site) were on the surface of the protein. The other residues were all on the inside of the protein, making visualization harder. C) Figure 24C shows a small cavity where the pink residue is. It's possible to see it is forming a cavity with the blue residues. An arrow points to the cavity. D) Figure 24D shows ligand A19 in its only conformation, in the cavity formed between the residues. An arrow points to it. E) Figure 24E shows a closer view of the ligand in the cavity. Image made in PyMol.

Table 45 - 521-585 Large Protein Active Site Results

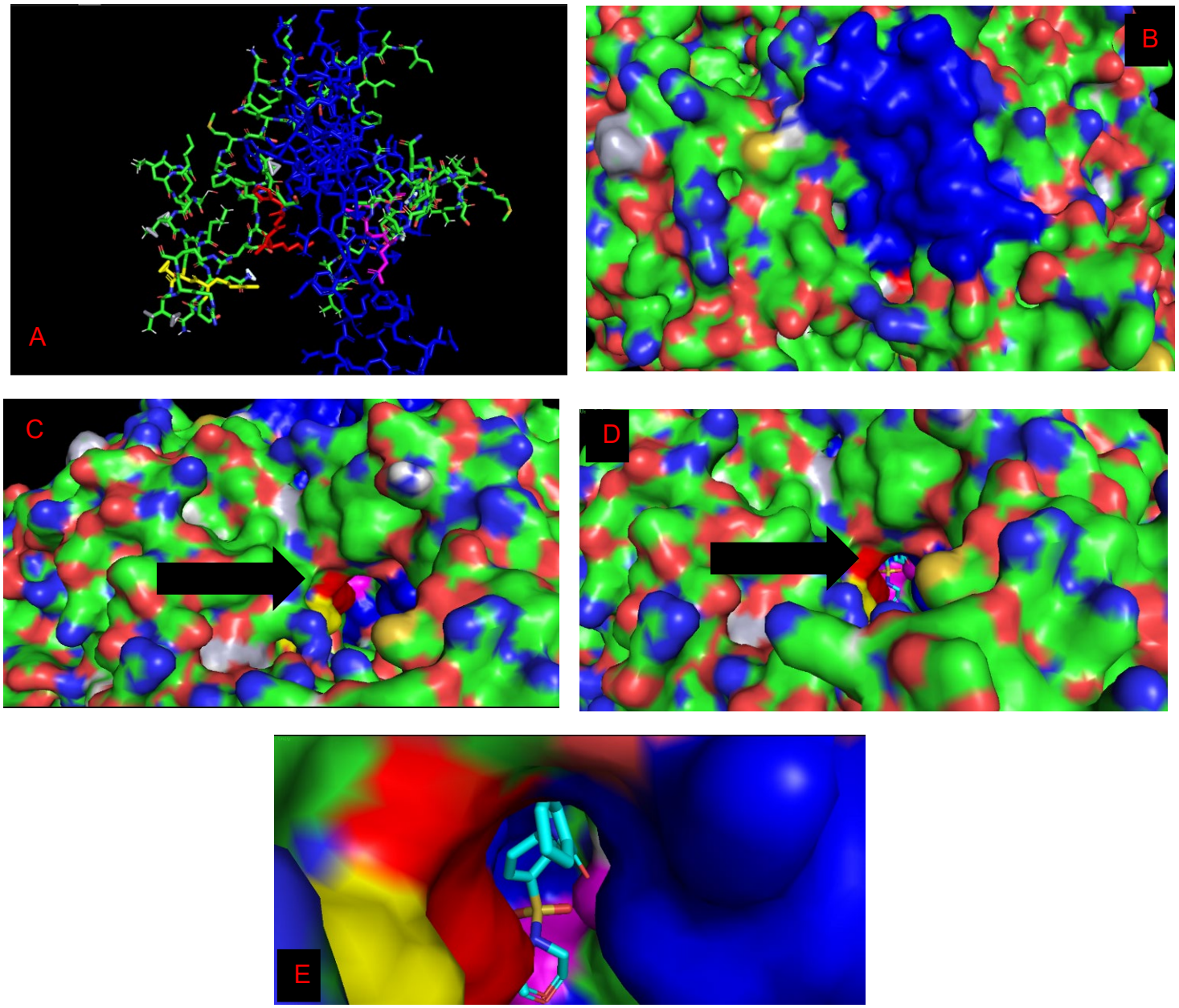
Residue	Number of times bound
Ser482	152
Arg562	139
Arg1051	125
Ala555	75
Val566	72
Lys481	57
Gly697	55
Met557	36
Glu696	21
Lys543	19
Arg563	19
Arg1052	17
Glu569	15
Lys701	14
Ala556	14
Glu544	13

Ser483	11

Table 45 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Ser482, Arg1051, Lys481, Gly697, Glu696, Arg1052, Lys701, and Ser483are outside the potential active site. In Figure 25C it's possible to see that residues pink, red, and yellow form a cavity with the blue residues (on the top). The figures make it harder to notice because the residues in pink, red, and yellow are "inside" the protein instead of on its surface (like the blue residues). In Figures 25D and 25E it's possible to see the ligand, in all its poses in this active docking, was laying in the cavity shown in figure 25C, touching blue, pink, and red residues.

Figure 25 - Residues 521-585 and Peripheral Residues of Active Site 521-585



Figures 25 show active sites formed when using active docking 521-585 in the large protein. Residues 521-585 are in blue, while Ser482, Lys481, and Ser483 are in red; Arg1051 and Arg1052 in yellow; and Gly697, Glu696 and Lys701 in pink. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 25A shows the location of each of these residues in comparison to the active site (MRV strain of large protein). B) In figure 25B, it's possible to see that only the blue residues (the proposed active site) were on the surface of the protein. The other residues were all on the inside of the protein, making visualization harder. C) Figure 25C shows a small cavity where the pink, red and yellow residues are. It's possible to see they are forming a cavity with the blue residues. An arrow points to them. D) Figure 25D shows ligand A19 in its only conformation, in the cavity formed between all the residues. An arrow points to it. E) Figure 25E shows a closer view of the ligand in the cavity. Image made in PyMol.

Table 46 - 1112-1285 Large Protein Active Site Results

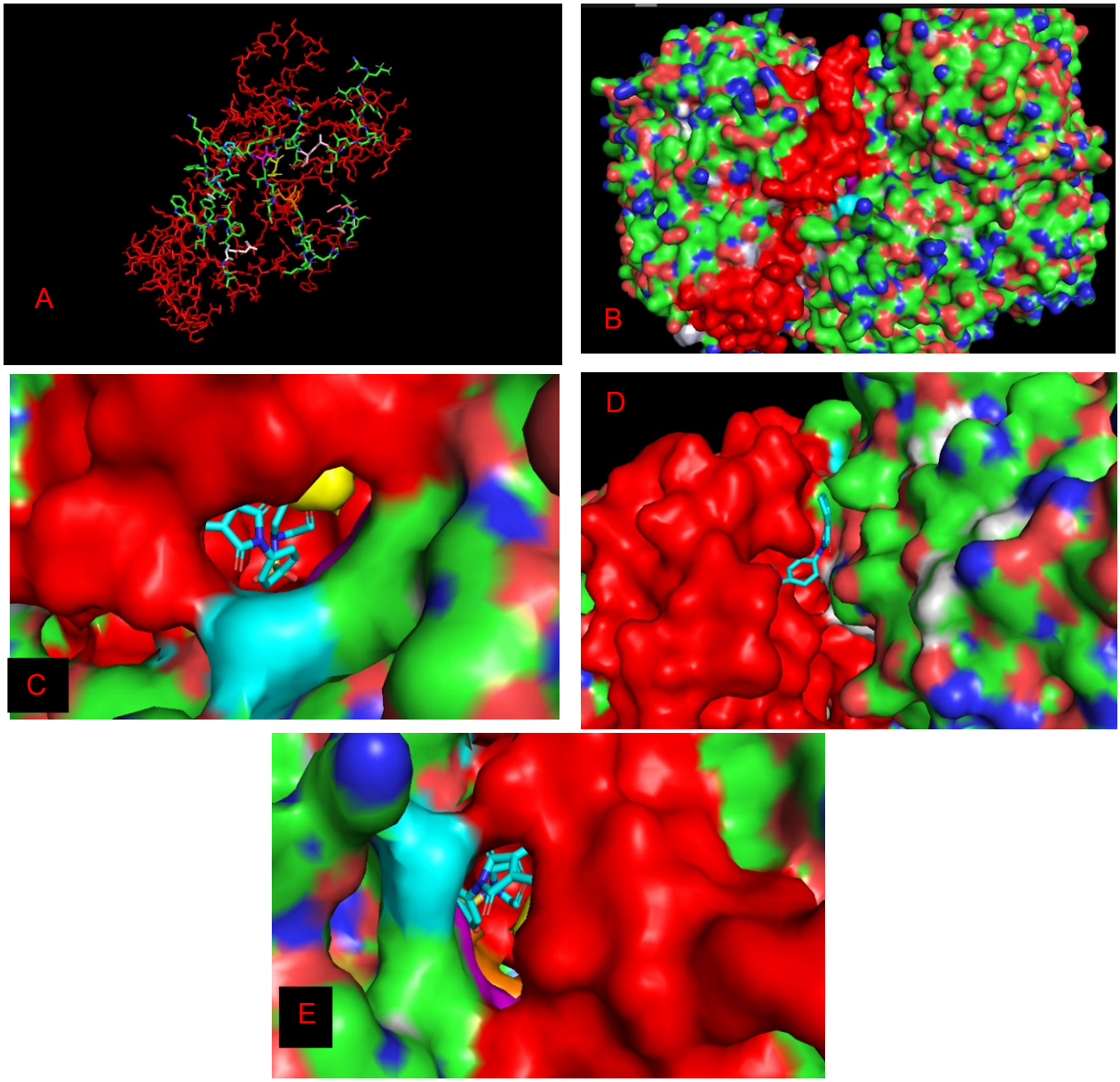
Residue	Number of times bound
Asn1280	72
Ile1404	62
Arg1235	61
Arg1242	50
Thr1268	48
Lys1244	46
Asn1408	44
Tyr1399	42
Pro1406	37
Asp1277	35
Arg1533	30
Ser1194	28
Ala1246	24
Asn1400	23
Ser1249	23
Glu1197	23
Ser1737	21
Ser1198	20
Ser1272	18
Lys594	17
Gly1164	16
Ala1736	16
Lys591	15
Met1271	14

Thr1170	13
Ser1169	11

Table 46 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Ile1404, Thr1286, Asn1408, Tyr1399, Pro1406, Arg1533, Asn1400, Ser1737, Lys594, Ala1736, and Lys59 are outside of the predicted potential active site. In Figure 26B it's visible that the red residues are the only on the surface of the molecule, making it harder to visualize the extra residues and their position in reference to the potential active site. Figures 26C, 26D and 26E show the different poses the ligand acquired and it was laying in cavities where it touched red, yellow, cyan, orange, and pink residues.

Figure 26- Residues 1112-1285 and Peripheral Residues of Active Site 1112-1285



Figures 26 show active sites formed when using active docking 1112-1285 in the large protein. Residues 1112-1285 are in red, while Ile1404 and Arg1533 is in pink, Thr1286 in yellow; Asn1408 and Pro1406 in cyan; Tyr1399 and Asn1400 in orange; Ser1737 and Ala1736 is in light pink; Lys594 is in salmon, and Lys59 in teal. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 26A shows the location of each of these residues in comparison to the active site (ERA strain of large protein). B) In figure 26B, it's possible to see that only the red residues (the proposed active site) were on the surface of the protein. The other residues were all on the inside of the protein, making visualization harder. C) Figure 26C shows a small cavity where the red, cyan and yellow residues are. It's possible to see they are forming a cavity, and one of the poses of ligand A19 lays in that cavity. D) Figure 26D shows ligand A19 in another of its poses, touching the red residues but none of the other extra residues. E) Figure 26E shows the last cavity where ligand A19 bound, a cavity formed by red, cyan, pink, and orange residues. Image made in PyMol.

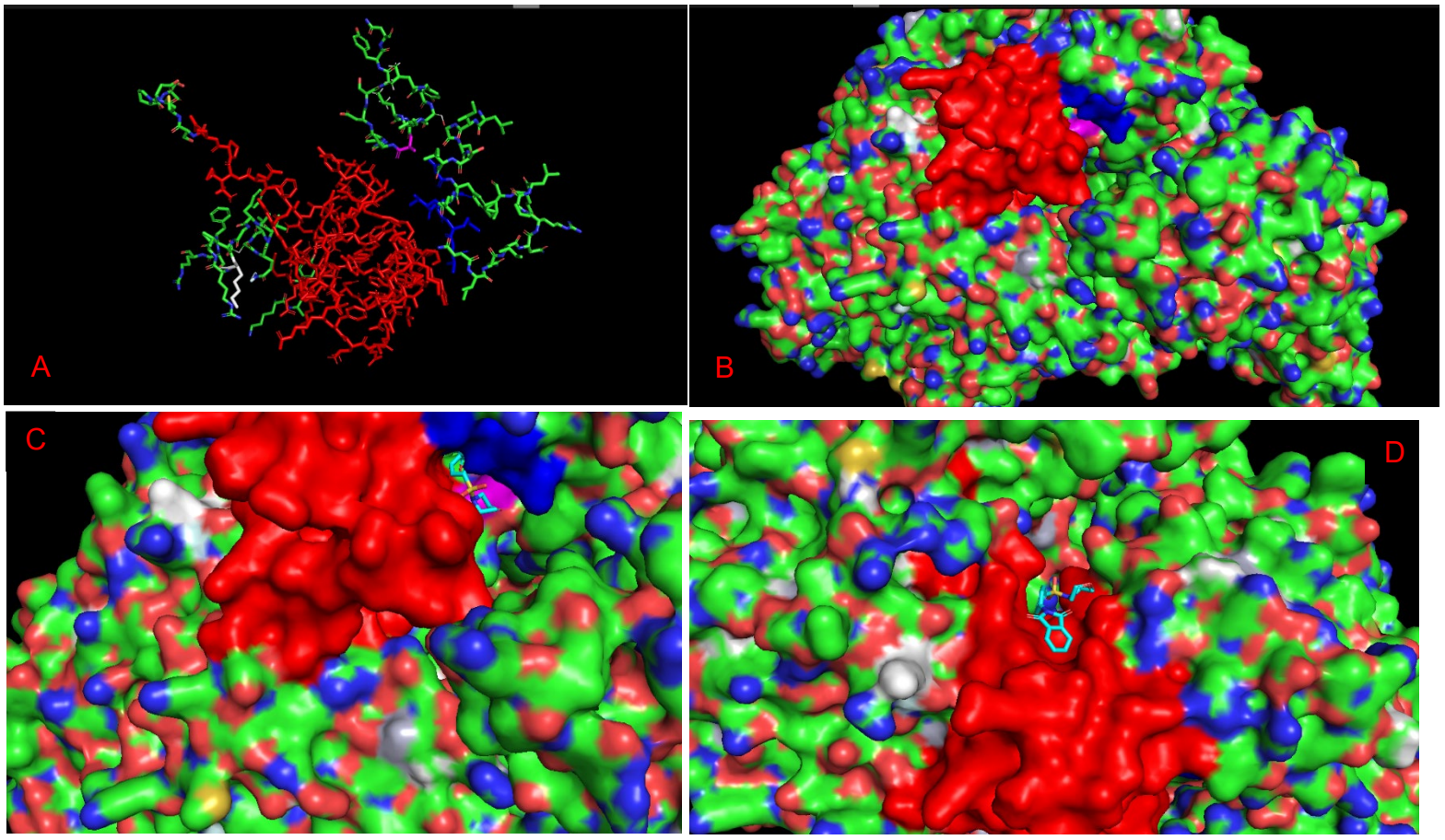
Table 47 - 400-470 Large Protein Active Site Results

Residue	Number of times bound
Arg1437	49
Ser420	49
Asp417	43
Lys418	36
Glu467	36
Trp444	31
Thr1436	22
Asp432	17
Ile465	17
Arg1438	17
Ala1393	16
Tyr423	14
Gln442	14
Thr1439	14
Arg427	13
Trp445	12
Ser421	12
Lys645	11
Cys1433	11
Asp418	11
Gln443	11
Thr1437	11

Table 47 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Arg/Thr1437, Thr1436, Arg1438, Ala1393, Thr1439, Lys645, and Cys1433 are outside of the predicted potential active site. Residues Arg1437 and Thr1437 were treated as only one for the purpose of this active site analysis, as they are only different because of point mutations throughout strains. In Figure 27B it's visible that the red, blue, and and pink residues form a cavity, while red and white residues form another, with all of them being close together. Figure 27C shows a cavity formed between the red, pink, and blue residues where ligand A19 bound to the most times. Figure 27D shows another cavity ligand A19 used, this one formed only by the red residues.

Figure 27 - Residues 400-470 and Peripheral Residues of Active Site 400-470



Figures 27 show active sites formed when using active docking 400-470 in the large protein. Residues 400-470 are in red, while Arg1437, Thr1436, Arg1438, Thr1439, and Cys1433 are in blue, Ala1393 in pink, and Lys645 in white. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 27A shows the location of each of these residues in comparison to the active site (SAD strain of large protein). B) In figure 27B, it's possible to see that all the residues are close to each other, all on the surface of the protein. It's also possible to see that the residues in red, pink,and blue form a cavity, while the residues in white and red form another. C) Figure 27C shows ligand A19 in the cavity between blue, pink, and red. This was the pose it was found most times during this docking. D) Figure 27D shows ligand A19 in another of its poses, only touching the red residues and in a cavity formed only by the potential active site. Image made in PyMol.

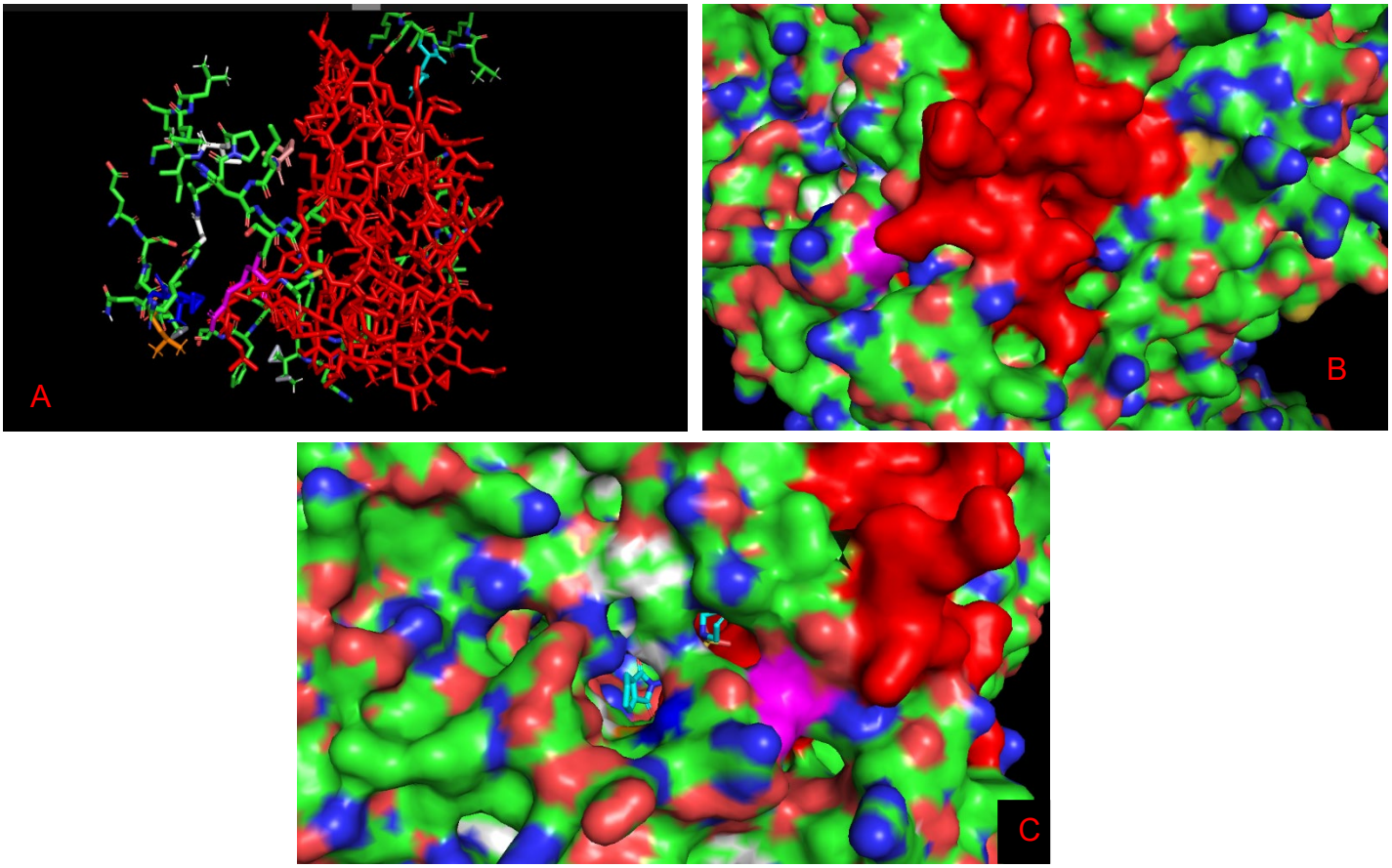
Table 48 - 808-908 Large Protein Active Site Results

Residue	Number of times bound
Val901	58
Ile969	37
Arg1079	36
Arg811	31
Asn833	30
Thr832	29
Leu967	26
Gly972	24
Thr976	20
Thr829	19
Arg808	18
Arg545	18
Phe884	18
Gly899	17
Pro896	17
Asn883	16
Glu1073	14
Tyr874	13
Ser828	11
Arg890	11
Ser883	10
Ser816	10

Table 48 shows residues that were accessed at least 10 times when evaluating each of the ligands that achieved a -6.0 or less in this particular active site.

Residues Ile969, Arg1079, Leu967, Gly972, Thr976, Arg545, and Glu1073 are outside of the predicted potential active site. In Figure 28B it's visible that the red and pink residues are the only on the surface of the molecule, making it harder to visualize the extra residues and their position in reference to the potential active site. Figure 28C shows the only cavity formed between the red residues and extra residues, as well as the ligand A19 laying in this cavity (where it was found in every single of its 10 poses). Although not pictured, the white residue was "inside" this cavity, touching the ligand as well.

Figure 28 - Residues 808-908 and Peripheral Residues of Active Site 808-909



Figures 28 show active sites formed when using active docking 808-908 in the large protein. Residues 808-908 are in red, while Ile969 is in blue, Arg1079 in pink, Leu967 in orange, Gly972 and Thr976 in white, Arg545 in cyan, and Glu1073 in salmon. Parts of the protein were hidden or deleted to help analysis of the image. A) Figure 28A shows the location of each of these residues in comparison to the active site (SAD strain of large protein). B) In figure 28B, it's possible to see that only the red residues (the proposed active site), as well as the pink and (Arg1079) were on the surface of the protein. The other residues were all on the inside of the protein, making visualization harder. C) Figure 28C shows a small cavity where the pink, red and white (inside the protein and not shown) residues are. It's possible to see they are forming a cavity with the blue residues. This cavity is the only place ligand A19 bound to in this protein throughout this docking, and it is shown in the Figure. Image made in PyMol.

Considering these results, the ligands that appeared most often in both blind and active docking, and had had lower energies compared to others were: A5, A11, A19, A21. Before diving into each ligand, Table 49 shows some properties of each ligand that might be relevant for both its functioning as well as possible blood-brain barrier (BBB) passage. Things like molecular weight, solubility and pKa may be of utmost importance for BBB passage, and all these ligands seem to be within the range of what's needed to cross the BBB. The elimination half-life, metabolization pathway, and bioavailability are important to measure for how long these drugs would stay in the body and how well they would be excreted were they to be used as medicine. Finally,

both commercialization and suggested doses may be important for practical use of these ligands as medicine, as well as possible extra research and drug trials.

Table 49 - Information on Chosen Ligands After Active Docking

	A5	A11	A19	A21
IUPAC Name	(6S,11S,1S,7S,17R)-4-ethylidene-7-hydroxy-6,7-dimethyl-2,9-dioxa-14-azatricyclo[9.5.1.0<14,17>]heptadecane-3,8-dione	7-fluoro-6-(4-methylpiperazin-1-yl)-10-oxo-4-thia-1-azatricyclo[7.3.1.$0^{5,13}$]trideca-5(13),6,8,11-tetraene-11-carboxylic acid	2-[4-(4-Morpholinylsulfonyl)phenyl]-1H-isoindole-1,3(2H)-dione	7-fluoro-2-methyl-6-(4-methylpiperazin-1-yl)-10-oxo-4-oxa-1-azatricyclo[7.3.1.05,13]trideca-2,5(13),6,8,11-pentaene-11-carboxylic acid
Drug Name	Platyphylline (stereoisomer)	Rufloxacin	N/A	2,3-dehydroofloxacin
Molecular Formula	C18H27NO5	C17H18FN3O3S	C18H16N2O5S	C18H18FN304
Molecular Weight	337.416	363.414	372.40	359.357
Solubility (in water)	135.9mg/ml	0.96 mg/ml	N/A	1.44 mg/ml
Elimination Half-Life	N/A	Over 30 hours (RIMOLDI, FIORETTI, *et al.*, 1992)	N/A	5 – 8 hours (STEIN, FLOR e BEALS, 1991)

Metabolization Pathway	Hepatic (RUAN, LIAO e LIN, 2014)	Renal (PERRY, MANT, *et al.*, 1993)	N/A	Renal (STEIN, FLOR e BEALS, 1991)
pKa	9.4 (in contact with tissue) (POMEROY e RAPER, 1971)	pKa1 – 5.4 pKa2 – 6.09	N/A	pKa1 – 6.05 pKa2 – 8.11 (BABIC, HORVAT, *et al.*, 2007)
Commercialization	16 vendors	38 vendors	12 vendors	5 vendors
Bioavailability	N/A	60 – 70% (in animals) (RIMOLDI, FIORETTI, *et al.*, 1992)	N/A	95 – 100% (STEIN, FLOR e BEALS, 1991)
Suggested Doses	N/A	400 mg	N/A	400mg

Table 49 shows biological information of the potential ligands that could have an effect against RABV. It wasn't possible to find information for all the ligands. All information was taken from Zinc15 (STERLING e IRWIN, 2015) *and DrugBank* (DRUGBANK)*, unless otherwise specified.*

5.0 DISCUSSION

When choosing an active site, it is almost impossible to pick one where only the relevant residues will be available for binding. In AutoDock, a “box” is created over the area where the user wants the active site to be used for docking. This means that, thanks to the tertiary structure of proteins, many of the potential active sites that were used resulted in ligands binding to residues outside the potential active site. Naturally, the active site of any given protein is going to be a cavity that can include both residues that are bound to the ligand and those that are not. Each of the potential binding sites that achieved an average binding energy of -6.0 or less were analyzed to see where exactly most of the bindings happened. It is with this information, plus the average binding energy information, that it was possible to come up with a result for what might be the active sites of each of these proteins.

This research aimed to gather more knowledge of RABV and its five proteins. This is imperative if the scientific community wants to find a way to defeat rabies on a global scale that does not limit itself to vaccination resources and knowledge that are not common in countries in development.

The results point out to specific areas of each of the proteins as active sites that can be used by ligands to block certain functions of the virus. Likewise, the results also show that the ligands that had the best results both during blind and active docking for all proteins were ligands A5, A9, A11, and A21.

5.1 Ligands

5.1.1 Ligand A5

Ligand A5 (ZINC649 (STERLING e IRWIN, 2015)) is a cyclic molecule, with a 12-membered ring attached to a pyrrolizidine. It is a stereoisomer of platyphylline, a member of a class of natural products that are used as antibiotics called macrolides. In general, macrolides inhibit bacterial protein synthesis, and can either kill bacteria or inhibit its growth (BEST PRACTICE MANUAL, 2012). Platyphylline is also a pyrrolizidine alkaloid, a group of alkaloids in plants that can have toxic or pharmacological properties (MOREIRA, PEREIRA, *et al.*, 2018). Platyphylline, specifically, has been used and researched mostly in its cholinergic antagonist abilities, and not in its potential antibacterial effects.

5.1.1.1 Pharmacology

Platyphylline is an alkaloid found in the *Senecio platyphylus* plant. It is an M-cholinoblocker, a cholinergic antagonist, being able to block neurotransmission in muscarinic synapses in both sympathetic and parasympathetic nerves (BOBYROV, VAZHNICHA, *et al.*, 2018). It is similar to atropine (BOBYROV, VAZHNICHA, *et al.*, 2018) (CHEN, HARRIS e ROSE, 1940), which competes with acetylcholine by binding to receptors and reversibly interrupting the binding of the neurotransmitter, both in central and peripheral muscarinic synapses.

Most pyrrolizidine alkaloids are considered to be hepatotoxic. However, platyphylline, due to not having a saturated C1-C2 bond in its pyrrolizidine ring might not be toxic (BOBYROV, VAZHNICHA, *et al.*, 2018). As other cholinergic antagonists, it can be used for its sedative effects, as well as for spasms of cerebral vessels and hypertension. In large doses, it can lead to excitation and hallucination. It can also lead to dilated pupils, dry mouth, and tachycardia. Platyphylline, however, has a diminished central action when compared to atropine (BOBYROV, VAZHNICHA, *et al.*, 2018), and it is not as prescribed or used as medication in comparison to atropine, being mostly used in Eastern European countries.

Atropine can be used to stop severe spasming (BAWASKAR, BAWASKAR e BAWASKAR, 2017) and to avoid bradycardia, which are both symptoms characteristic of rabies. It's possible that platyphylline could have a similar effect with less toxicity in a rabies patient in terms of possible pharmacological effects. This would need to be checked with *in vitro* experiments.

5.1.1.2 Structure

The ligand in question is actually a stereoisomer of platyphylline. The ligand A5 seems to be the S enantiomer, while platyphylline is the R enantiomer. No research seems to have been made in terms of whether the S enantiomer is active. Ibuprofen, for example, an anti-inflammatory drug, is made of a racemic mixture of the S(+) and the R(-) enantiomer. While they are similar in terms of properties such as solubility, boiling point, among others, S+ ibuprofen is the only stereoisomer that has an antiinflamatory effect and inhibits cyclooxygenase and reduces inflammation, while R-ibuprofen shows no antiinflamatory effect (only when it is in the body does it transform into the S+ enantiomer and has an effect). It is, then, possible that the stereoisomer of platyphylline would have different effect in the body, or even no pharmacological effect.

With that in mind, the structure of ligand A5 is incredibly important as it might be the only method by which it inhibits RABV proteins and diminishes their pathological effects in the host. Figure 29 shows ligand A5 and platyphylline side by side.

One characteristic of RABV is that it can bind to nicotinic acetylcholine receptors (nAChR) in muscles, and potentially in the CNS (HUEFFER, KHATRI, *et al.*, 2017). However, there has been some research to suggest that parasympathetic muscarinic receptors in the brain could have a function when it comes to the infection and its subsequent wasting of the immune system (DUMRONGPHOL, SRIKIATKHACHOM, *et al.*, 1996). If that is true, platyphylline could potentially block the receptors where a (or more) RABV protein(s) binds, inhibiting its effect, besides its aforementioned effects on RABV proteins themselves.

Figure 29 - Platyphylline (left) and Its Stereoisomer (Ligand A5 - right)

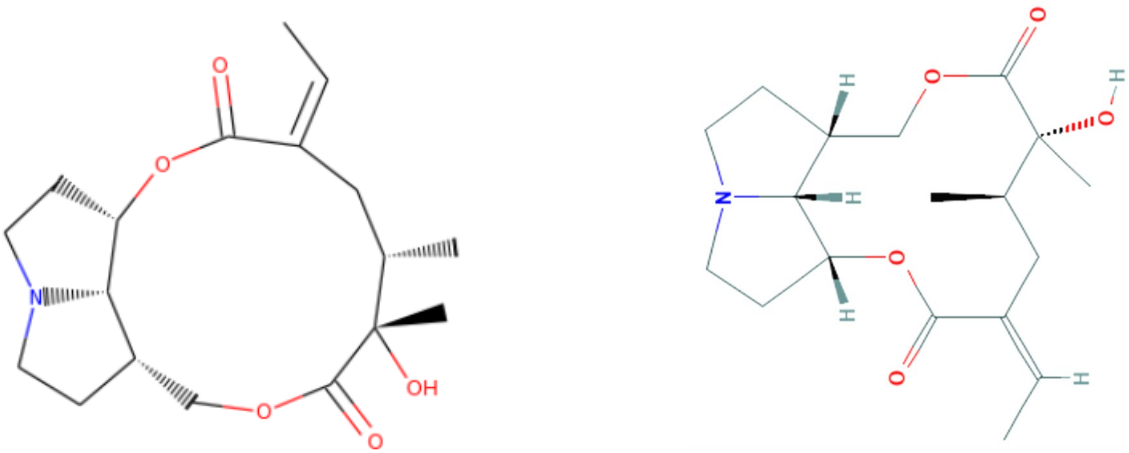


Figure 29 shows side-by-side images of platyphylline and ligand A5 to show they are non-superimposable mirror images of each other. Images from PubChem website (NCBI RESOURCE COORDINATORS, 2018).

To further differentiate platyphylline from its stereoisomer Ligand A5, one blind docking of platyphylline (3D downloaded from PubChem, with PubChem CID 5281742 (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION) with each of the five RABV proteins was conducted. The areas where platyphylline bound were similar to the areas where A5 bound, but the binding energy was lower for all proteins when compared to A5, which might be because of its different orientation.

Platyphilline had results of -6.0 or less in blind docking to all RABV proteins except glycoprotein. When looking at the atoms most accessed by the proteins, the atoms highlighted in Figure 30 were accessed more than 10 times, with the oxygen

atom highlighted in red being accessed 22 times. Most of the protein residues that were used for these bindings in the matrix protein were Asn172, Asn174, Ile171, Ser198 and Thr177; in the nucleoprotein were Trp197, Gln156, Asn154, Thr158, Gly125, Phe359, Arg358, Arg357, Glu356, Thr104, Glu266, Glu337, Glu127, Arg400, and Ser58; in the glycoprotein they were Lys339, Thr312, Ser308, Leu306, Val229, Trp270, Leu341, Lys245, Asp274, Thr25, Val49, Lys298, Lys339, and Gly68; in the large protein, they were Asp1752, Ser1645, Asn111, Met738, Asn140, Asn518, Thr1268, Asn1280, and Arg1143.

Figure 30 - Highlighted Atoms of Ligand A5

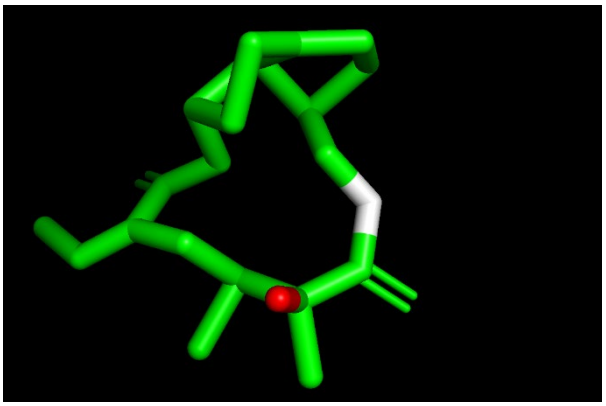


Figure 30 shows the atoms most accessed during docking of ligand A5. Image made using PyMol.

5.1.2 Ligand A11

Ligand A11 (ZINC538328 (STERLING e IRWIN, 2015)), of molecular formula C17H18FN3O3S, is also known as Rufloxacin. Rufloxacin is a type of quinolone antibiotic, more specifically a type of fluoroquinolone, as it exhibits a fluorine atom in its structure. Quinolones are known to inhibit the activity of DNA gyrase and IV topoisomerase II, interfering with DNA replication, and fluoroquinolones can be used both against Gram-negative and Gram-positive bacteria (OLIPHANT e GREEN, 2002). It can be used in infections such as those of the urinary tract, as well as possible respiratory infections (OLIPHANT e GREEN, 2002). Rufloxacin is one of the least known and used fluoroquinolones, with its sibling, Ofloxacin, which will be discussed later, being much more utilized.

5.1.2.1 Pharmacology

Studies (DALHOFF, 2014) mention the pleiotropic nature of fluoroquinolones, or their ability to be used against different types of pathogens (bacteria, viruses, and parasites). This ability can be due to different mechanisms of action that the drug can

use. In the case of viral infections, some studies have shown that fluoroquinolones act as antioxidants against reactive oxygen species (ROS) and nitric oxide (NO) in influenza infection (ENOKI, ISHIMA, *et al.*, 2015). While rabies has different infection stages in comparison to influenza, it has been shown that both ROS and NOs are produced during infection and are one of the causes for immunosuppression in the host (KAMMOUNI, WOOD, *et al.*, 2015) (NAKAMICHI, INOUE, *et al.*, 2020) (MADHU, SINGH, *et al.*, 2016).

Other studies in different viruses have shown that it is possible that fluoroquinolones inhibit helicase activity, diminish viral load, inhibit viral replication *in vitro* and, in the case of HIV, inhibit *in vitro* reverse transcriptase (DALHOFF, 2014). Also, it has been proposed that the antibacterial effect of quinolones is related to the ability of the molecule to bind to double-stranded DNA (RICHTER, PAROLIN, *et al.*, 2004) (VALISENA, PALUMBO, *et al.*, 1990), facilitated by magnesium ions that could be coordinated by a free carboxylic group. If that is the case, there is reason to believe that fluoroquinolones could do the same to viral nucleic acids (RICHTER, PAROLIN, *et al.*, 2004). While no research has been done using both Rufloxacin and RABV, previous research between its class of antibiotics and others virus may indicate that Rufloxacin (as well as other fluoroquinolones) could have interesting effects in controlling rabies infection.

5.1.2.2 Structure

In terms of its structure, Rufloxacin contains a methyl-piperazine (highlighted in red in Figure 31B), a fluorine (in blue in figure 31B), a thia-azatricyclo group with a ketone (in black in figure 31B), and a carboxylic acid (in green in figure 31B). In reference pH (approximately 7.4), the carboxylic acid is ionized, with a negative charge on the oxygen instead of the double oxygen, so it is considered a carboxylate. Under physiological conditions, the intracellular brain pH is approximately 7.2 (ORLOWSKI, CHAPPELL, *et al.*, 2011), and it is possible that the negative charge on the molecule would complicate its passage through the BBB (KASINATHAN, JAGANI, *et al.*, 2014). On the other hand, the structure of Rufloxacin gives it a pKa1 of 5.4 and a pKa2 of 6.09 (THE HUMAN METABOLOME DATABASE), which is within the range for entrance to the BBB, thanks to its basicity. The structure also has a fluorine and a number of cyclic groups, which could help BBB absorption.

One study showed that Rufloxacin is able to cross the blood-brain barrier in patients with both inflamed and uninflamed meninges and that, because of its long half-life, the drug would maintain high concentrations in the CSF for a longer time than other fluoroquinolones (MORETTI, PAULUZZI e CESANA, 2020).

Figure 31 - Structure of Ligand A11 and Relevant Functional Groups

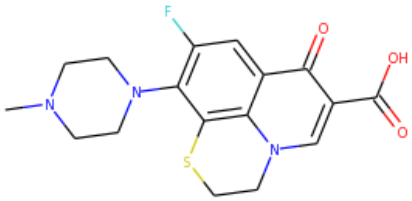


Figure 31A - structure of Rufloxacin (ligand A11). Image from Zinc15.

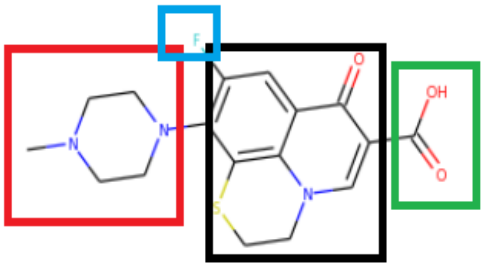


Figure 31B - functional groups of Rufloxacin (ligand A11). Image edited from Zinc15.

When analyzing the binding with glycoprotein and matrix protein, binding was not stable, with residues not matching when comparing strains. Further, the position of binding in glycoprotein with A11 was not stable, with binding happening in all three nitrogens and double bonds depending on the strain and residue. Nucleoprotein binding was more stable, with binding happening mostly to Arg400, Leu251, Arg149, and Arg290, matching a potential active site (see 5.2.2). In the large protein, the binding was also stable, accessing mostly potential active site residues (see 5.2.4), such as Ser482, Ser483, Gly697, Arg562, Gly698, Met557, Arg563, and Lys482. In terms of the ligand, mostly the double O's were used in binding, as well as the O- in the carboxylic acid. The stability of binding residues, both of the protein and the ligand, might indicate both a stronger interaction between them and a stronger attraction.

5.1.3 Ligand A19

Ligand A19 (ZINC591669 (STERLING e IRWIN, 2015)), of molecular formula C18H16N2O5S, has no previous proven use in drug discovery or medicine. The ligand has been researched for antimalarial activity, but it wasn't found to have effects (GAMO, SANZ, *et al.*, 2010). Since there has been no proved activity of this ligand in any type of research, this molecule was separated in its components in order to find potential antiviral effects.

5.1.3.1 Pharmacology

Figure 32 breaks down the structures of ligand A19. In black, a morpholine is highlighted, in red we can see two benzenes (which are probably useful in the molecule for its inherent stability), in pink a sulfone, and in green a phthalimide (a benzene connected to a succinimide, also called an isoindole).

While this ligand specifically hasn't been the focus of many articles and research, these functional groups have been researched in terms of their potential antiviral activity. Morpholine (in black in Figure 32), for example, is a molecule used in many different drugs, and useful for improving the aqueous solubility of drugs (KUMARI e SINGH, 2020). In terms of antiviral activity, research has shown that morpholine may suppress activity of hepatitis C virus (HCV) by using its oxygen and binding to HCV's binding site, and some research has been done about the antiviral activity of morpholine in HIV protease (KUMARI e SINGH, 2020).

Different isoindoles have been researched for potential antiviral activity from different perspectives, such as inhibition of virus entry and uncoating, to protease inhibition. A reason why isoindoles may have these specific activities is its lipophilic and aromatic character, which can carry different bioactive side chains (CSENDE e PORKOLÁB, 2020).

A ligand very similar to ligand A19 in its isoindole character (but with an ethyl phenyl ring instead of morpholine) showed inhibitory activity against DENV2 NS2B–NS3 protease. In molecular docking, this ligand had many interactions, a lot of them strong, with the active site of DENV2, showing a possible structural inhibition (CSENDE e PORKOLÁB, 2020).

Figure 32 - Ligand A19

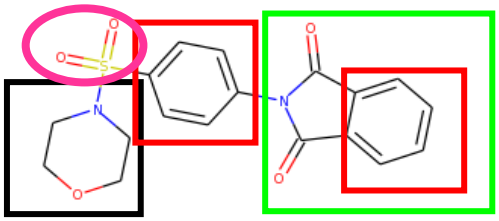


Figure 32 – Ligand A11 with highlighted functional groups. Image from Zinc15

Compounds with sulfur have been used in medicine for a long time (FENG, TANG, *et al.*, 2016). Sulfones, specifically, are used in drugs such as dapsone (for

tuberculosis), and for the treatment of some types of malaria (MITCHARD, 1988). It works by inhibiting folic acid synthesis, which can lower the amount of folate, an important tool in immune response (REVELL, O'DOHERTY, *et al.*, 1991) (MIKKELSEN e V., 2019).

5.1.3.2 Structure

When observing the connections ligand A19 made with all of the five proteins of RABV, it is clear that it made a great number of connections and interactions. Besides conventional hydrogen bonds and carbon-carbon bonds, it also a number of pi-alkyl interactions, pi-pi, pi-sulfur, pi-sigma, and pi-amide interactions. The atoms most accessed of the ligand were the double bond oxygens attached to the sulfur atom as well as the double oxygens in the phtalimide. Sulfur was also accessed a number of times.

When looking at the protein residues that accessed these atoms, the most accessed in the glycoprotein were Tyr235, Val315, Lys236, Lys245, Leu341, Lys120, Ile36, and Ile154. In the large protein, they were Thr1268, Pro1474, Ile969, Asn1280, Asp417, and Ser1194. In the matrix protein, they were Thr188, Ser198, Ser182, Asn172, Lys115, Cys178, Arg118, and Ser139. Regardless of the protein, A19 made more connections (especially of the weaker pi types) than any other ligand.

A19 was the best ligand overall during active and blind docking. It had the best results across the board. While this might mean ligand A19 has potential to block RABV structures, it also shows little specificity of this ligand, which might in turn mean that it would have no realistic effect *in vivo*.

5.1.4 Ligand A21

Ligand A21 (ZINC596282 (STERLING e IRWIN, 2015)), of molecular formula $C_{18}H_{18}FN_3O_4$, is also known as 2,3-dehydroofloxacin. It is almost identical to Ofloxacin, with 2,3-dehydroofloxacin being a dehydrogenated version of Ofloxacin (having two less hydrogens due to an extra bond). Ofloxacin, like Rufloxacin, is a type of quinolone antibiotic, more specifically a type of fluoroquinolone, as it exhibits a fluorine atom in its structure. Quinolones are known to inhibit the activity of DNA gyrase and IV topoisomerase II, interfering with DNA replication, and fluoroquinolones can be used both against Gram-negative and Gram-positive bacteria (OLIPHANT e GREEN,

2002). It can be used in infections such as those of the urinary tract, as well as possible respiratory infections (OLIPHANT e GREEN, 2002).

5.1.4.1 Pharmacology

Both Ligands A11 and A21 have similar pharmacology, due to them being in the same class of antibiotics. With that said, there is a lot more research involving Ofloxacin (which is very similar to ligand A21) than Rufloxacin (ligand A11). Studies have shown that Ofloxacin could potentially have antiviral activity, with a mechanism involving inhibition of topoisomerase and hindering both viral DNA and viral RNA syntheses (IKEDA, YAZAWA e NISHIMURA, 1987). No research has been made in terms of the antiviral capability of 2,3-dehydroofloxacin, although it has been shown that it has a higher minimum inhibitory concentration needed to be used as an antibiotic in comparison to ofloxacin (AUGERI, FRAY e KLEINMAN, 1990).

In terms of its functional groups, 2,3-dehydroofloxacin has a piperazine (highlighted in blue in Figure33C), a ozaxine (highlighted in green in Figure 33C), and a benzene. The ozaxine with the benzene could also be called a benzoxazine (highlighted in red in Figure 33C), and with the nitrogen from the ozaxine the functional group in pink in Figure 33C is a 4-oxoquinolone.

Although no specific effect or mechanism of action has been established, different studies have found that ozaxines and modified benzoxazines may be useful in a myriad of ways for drug design, whether for antibacterial, antiviral, or even drugs for types of cancers (KOZLOVSKAYAA, ANDREI, *et al.*, 2019) (ZINAD, MAHAL, *et al.*, 2019).

Similarly to ozaxines, there are no mechanisms of action already established for oxoquinolones, however research has been conducted to find potential antiviral effects of the compound. Oxoquinolones and some derivatives have shown to be possible drugs for HIV, working against HIV-1 reverse transcriptase (SANTOS, ABREU, *et al.*, 2009). Studies have also mentioned the capability of oxoquinolone derivatives to potentially inhibit the NS5B polymerase of human cytomegalovirus (BATALHA, FOREZI, *et al.*, 2020).

5.1.4.2 Structure

Figures of ofloxacin and 2,3-dehydroofloxacin (Ligand A21) are below (Figure 33). 2,3-dehydroofloxacin has an extra bond in its ozaxine ring, therefore being more saturated than ofloxacin. Usually, extra unsaturation makes a molecule more reactive, making it possible that it is more available for binding than its original molecule (ofloxacin). The ligand acquired values of -6.0 or lower in all proteins but the matrix protein. Through evaluation of both these ligands (Ofloxacin was also used in this research as Ligand 1, and it only acquired values of -6.0 or lower in the large protein and glycoprotein), it was seen that A21 created more interactions than A1, often pi-sigma or pi-alkyl interactions. Even though these interactions are not as strong as covalent bonds, the simple ability of the ligand to bind more times and in different positions to the protein may explain why A21 had better results than A1.

Ofloxacin has a pKa1 of 5.45 and a pKa2 of 6.2, and these numbers should be similar for Ligand A21 (DRUGBANK, 2021). In terms of BBB passage, it is similar to Rufloxacin and would probably make it through. Again, like Rufloxacin, it is possible that the negative charge on the oxygen of the carboxylic acid could complicate its passage through the BBB as, in Reference pH, it is ionized.

In terms of most accessed residues, in the nucleoprotein A21 accessed residues like Arg400, Leu251, Lys51, Thr354, Lys29, Ser55, Glu127, and Arg149 the most. For glycoprotein, they were Lys245, Lys104, Lys298, Lys148, Lys236, Lys74, Leu57, and Ile36. In the large protein, the residues were Lys1687, Glu1473, Lys418, Ser1475, Ser420, Ser1414, Thr1268, and Asn1400. For the matrix protein, the bindings were not stable and not many residues appeared a number of times.

Figure 33 - Ligand A21

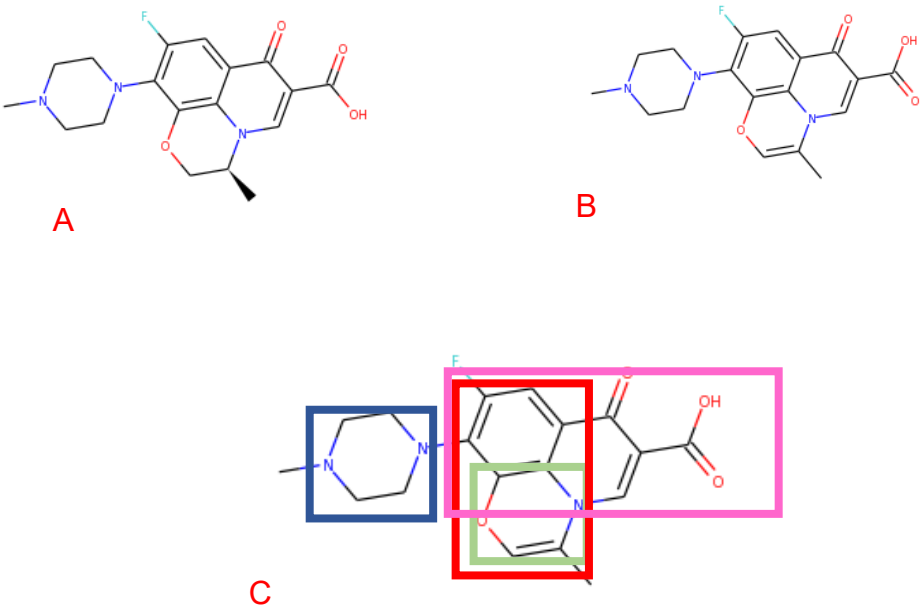


Figure 33A shows Ofloxacin, a known drug similar to Ligand A21, if not for the lack of an extra double bond in its ozaxine ring. Figure 33B shows Ligand A21. Figure 33C shows Ligand A21 with some of its functional groups highlighted. Piperazine is highlighted in blue, ozaxine in green, a bezonxazine in red, and an oxoquinolone in pink. All images are from Zinc15.

5.2 Active Sites

The structure of proteins makes it so choosing a specific active site does not necessarily mean that only those residues will be accessed by the ligand. Thanks to the folding of proteins, residues numbered far from the potential active site can be close to the potential active site. Considering this, it was important to analyze exactly which residues the ligands bound to so as to come to a conclusion on what the active site of each RABV protein might be.

5.2.1 Matrix Protein Active Sites

Four potential active sites were analyzed for the matrix protein.

The first potential active site encompassed residues 188-200 (12 residues) of the matrix protein. 15 residues of the protein bound 10 or more times to the relevant

ligands (the ones that got -6.0 or lower during active site docking). Out of these 15 residues, 10 were a part of the potential active site (67.7%), and 5 (Gln123, Lys48, Cys170, Asn172, and Asp126) were outside. Out of these 5, 3 (Gln123, Lys48, Asp126) formed a cavity with the residues 188-200.

The second potential active site encompassed residues 172-188 (16 residues) of the matrix protein. 9 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 9, 6 were a part of the potential active site (67.7%), and 3 (Ser198, Cys170, and Asp196) were outside. Out of these, 2 (Ser 198 and Cys170) formed a cavity with the residues 172-188.

The third potential active site encompassed residues 113-120 (7 residues) of the matrix protein. 11 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 11, 3 were a part of the potential active site (27.3%), and 8 (Asn172, Ser102, Asn174, Ile171, Ala104, Arg167, Leu99, and Asn86) were outside. Out of these 8, 5 (Asn172, Ser102, Asn174, Arg167, and Leu99) formed a cavity with the residues 113-120.

The fourth, and last, potential active site encompassed residues 84-100 (16 residues) of the matrix protein. 9 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). However, none of the residues that bound to the potential ligands were residues within this potential active site. The residues that did bound were: Ser198, Thr188, Cys170, Ser182, Asp196, Cys178, Leu186, Trp181, Asn172.

The last two potential active sites showed worse results in terms of relevant ligands binding to the potential active sites. In the last case, this is likely because the residues 84-100 are deeper into the protein in its folded shape, leaving other (more superficial) residues more available for binding. The third case (residues 113-120) may just not be a potential active site, or possibly the small number of residues (only 7) made it so a higher number of residues outside the potential active site needed to be accessed by the ligands.

Considering, then, both the active sites that had an approximately 67% of connection between the potential active sites and the residues that actually bound to the ligands, it’s possible that these two are closer to, or are, the active site of the matrix

protein. Furthermore, some of the extra residues (residues that were not in the potential active site but were found to bind to the relevant ligands), such as Cys170, Asn172, Ser198, and Asp196 were found to bind in more than one potential active site, even when these residues were not a part of the potential active site.

With this information, it is possible to say residues Cys170 to Leu200 are in the active site of the matrix protein of RABV. These are not residues that have been found in literature to be the possible active site of the matrix protein before. Considering this potential active site, the ligand that fit best would be A19.

5.2.2 Nucleoprotein Active Sites

Six potential active sites were analyzed for the nucleoprotein.

The first potential active site encompassed residues 349-403 (54 residues) of the matrix protein. 15 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 15 residues, 12 were a part of the potential active site (80%), and 3 (Leu251, Thr252, and Tyr259) were outside. Out of these 3, they all formed a cavity with the residues 349-403.

The second potential active site encompassed residues 300-328 (28 residues) of the nucleoprotein. 9 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 9, 4 were a part of the potential active site (44.4%), and 5 (Arg271, Phe245, Tyr28, Lys247, and Val240) were outside. Out of these 5, 1 (Val240) formed a cavity with the residues 300-328.

The third potential active site encompassed residues 149-205 (56 residues) of the nucleoprotein. 18 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 18, 7 were a part of the potential active site (38.9%), and 11 (Ser222, Val226, His219, Arg290, Arg225, Ala223, Met60, Asn61, Lys38, Leu292, Ala62) were outside. Out of these 11, 5 (Arg290, Met60, Asn61, Leu292, and Ala62) formed a cavity with the residues 149-205.

The fourth potential active site encompassed residues 251-273 (22 residues) of the nucleoprotein. 14 residues of the protein bound 10 or more times to the relevant

ligands (the ones that got -6.0 or lower during active site docking). Out of these 14, 11 were a part of the potential active site (78.6%), and 3 (Arg400, Thr354, and Thr243) were outside. Out of these 3, 2 (Arg400 and Thr354) formed a cavity with the residues 251-273.

The fifth potential active site encompassed residues 223-240 (17 residues) of the nucleoprotein. 5 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these, 0 were a part of the potential active site, and 5 (Arg149, Arg168, Lys152, Asp235, Arg290) were outside. Out of these 5, none seemed to form a cavity with the residues 223-240.

The sixth, and last, potential active site encompassed residues 27-38 (11 residues) of the nucleoprotein. 20 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 20, 7 were a part of the potential active site (35%), and 13 (Arg290, Pro275, Ser295, Leu294, Phe205, Gly296, Arg204, Ser291, Gly276, Ile41, Asn202, Pro201, Gln277) were outside. Out of these 13, 5 residues (Arg290, Pro275, Gly296, Gly276, and Gln277) formed a binding cavity with the residues 27-38.

The second, third, fifth and sixth potential active sites showed worse results in terms of relevant ligands binding to the potential active sites. The second (300-328), fifth (223-240), and sixth (27-38) active sites were the ones that had the least number of residues in them, which may have required the ligands to search for residues outside of the proposed site for binding to be the strongest possible. The third active site (149-205) had a big number of residues in it, but still got a low percentage of residues binding to the ligands. Even the extra residues that bound in this particular active site were not a part of other active sites that got better results and were not found as extra residues in other active sites. This may indicate that this area of the protein is not an active site.

Considering, then, the active sites that had the best results (349-403 with 80%, 251-273 with 78.6%), it is possible that these are closer to, or are, the active site of the nucleoprotein. Furthermore, some of the extra residues (residues that were not in the potential active site but were found to bind to the relevant ligands), such as Thr252 and Arg290 were found more than once in and outside of their potential active sites.

With this information, it is possible to narrow down potential active sites of the nucleoprotein to residues 251-403. From literature, it has been shown that residues

273 and 394 of nucleoprotein were important for virus evasion, and that residues 355-372 could be a potential active site for N-N interaction (RIEDEL, VASISHTAN, *et al.*, 2019). Residues 273 and 394, as well as residues ranging from 355-372 are within the 251-403 range, and it is possible that these residues are, indeed, the active sites of nucleoprotein. Considering this potential active site, the ligand that fit best would be A5.

5.2.3 Glycoprotein Active Sites

Six potential active sites were analyzed for the glycoprotein.

The first potential active site encompassed residues 43-69 (26 residues) of the glycoprotein. 31 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 31 residues, 18 were a part of the potential active site (58.1%), and 13 (Lys217, Thr212, Val229, Asp230, Glu231, Glu288, Thr25, Tyr235, Lys221, Asp256, Thr258, Leu234, and Asn223) were outside. Out of these 13, 5 (Val229, Asp230, Glu231, Tyr235, and Leu234) formed a cavity with the residues 43-69.

The second potential active site encompassed residues 120-143 (23 residues) of the glycoprotein. 32 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 32, 16 were a part of the potential active site (50%), and 16 (Lys148, Ser150, Lys104, Ile/Val152, Glu149, Ile154, Ala112, Thr91, Thr146, Lys145, Ala116, Glu86, Arg107, Arg103, Phe93, and Gly95) were outside. Out of these, 13 (Lys148, Ser150, Lys104, Val/Ile152, Glu149, Ile154, Ala112, Thr91, Thr146, Lys145, Ala116, Glu86, Phe93, and Gly95) formed a cavity with the residues 120-143.

The third potential active site encompassed residues 166-197 (31 residues) of the glycoprotein. 27 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 27, 20 were a part of the potential active site (74.1%), and 7 (Arg203, Leu204, Asp165, Lys245, Thr81, Gly248, and Arg107) were outside. Out of these 7, 2 (Arg203 and Leu204) formed a cavity with the residues 166-197.

The fourth potential active site encompassed residues 217-240 (23 residues) of the glycoprotein. 35 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 34, 17

were a part of the potential active site (50%), and 17 (Val49, Cys54, Glu50, Leu57, Thr55, Tyr62, Leu47, Phe60, Thr212, Gly53, Ala241, Thr258, Asp256, Thr268, Ser58, Glu52, and Asp/Glu267) were outside. Out of these 17, 11 (Val49, Cys54, Glu50, Leu57, Thr55, Tyr62, Leu47, Phe60, Gly53, Ser58, and Glu52) formed a cavity with the residues 217-240.

The fifth potential active site encompassed residues 253-270 (17 residues) of the glycoprotein. 27 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 27, 12 were a part of the potential active site (44%), and 15 (Arg/Lys218, Glu240, Ala241, Lys/Arg243, Tyr62, Lys245, Lys221, Ser61, Ser250, Leu238, Gln275, Asp274, Val229, Leu276, and Phe60) were outside. Out of these 15, 6 (Arg/Lys218, Tyr62, Lys221, Ser61, Leu276, and Phe60) formed a cavity with the residues 253-270.

The sixth, and last, potential active site encompassed residues 283-310 (27 residues) of the glycoprotein. 16 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 16, 9 were a part of the potential active site (56.3%), and 7 (Gln401, Glu405, Glu408, Leu326, Gln402, Asn46, and Glu394) were outside. Out of these 7, 4 (Gln401, Glu405, Gln402, and Asn46) formed a cavity with the residues 283-310.

The second, fourth and fifth potential active sites got the worse results in terms of relevant ligands binding to the potential active sites. These binding sites were the ones that had the least number of residues in them, which may have required the ligands to search for residues outside of the proposed site for binding to be the strongest possible. Although showing only 50% of accuracy (50% of the residues used for binding were in the proposed active site), the extra residues that bound to the proposed active site 217-240 were residues found in other potential active sites (namely Arg/Lys218, Lys221, Val229, Asp230, Glu231, Leu234, and Tyr235).

Considering, then, the active sites that had the best results (43-69 with 58.1%, 166-197 with 74.1%, and the 283-310 with 56.3%), it’s possible that these are closer to, or are, the active site of the glycoprotein. Furthermore, some of the extra residues (residues that were not in the potential active site but were found to bind to the relevant ligands), such as Arg/Lys218-Tyr235, Asn46, Leu47, Val49, Glu50, Glu52, Gly53,

Cys54, Thr55, Leu57, Ser58, Phe60, Ser61, and Tyr62 were found more than once in and outside of their potential active sites.

With this information, it is possible to narrow down potential active sites of the glycoprotein, such as residues 43-69, 166-197 and/or 217-310. These last two match what has been found by previous research (mentioned in section 4.2.1), making them highly likely to be active sites of the protein. Considering these active sites, there is not one ligand that seems more likely to work with the glycoprotein, as they all seemed to fit in the active site somewhat.

5.2.4 Large Protein Active Sites

Five potential active sites were analyzed for the large protein.

The first potential active site encompassed residues 620-700 (80 residues) of the large protein. 7 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 7 residues, 6 were a part of the potential active site (85.7%), and 1 (Arg552) was outside. Arg552 did form a cavity with the potential active site residues (620-700).

The second potential active site encompassed residues 521-585 (64 residues) of the large protein. 17 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 17, 9 were a part of the potential active site (52.9%), and 8 (Ser482, Arg1051, Lys481, Gly697, Glu696, Arg1052, Lys701, Ser483) were outside. All of these formed a cavity with the potential active site residues 521-585.

The third potential active site encompassed residues 1112-1286 (174 residues) of the large protein. 26 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 26, 15 were a part of the potential active site (57.6%), and 11 (Ile1404, Thr1286, Asn1408, Tyr1399, Pro1406, Arg1533, Asn1400, Ser1737, Lys594, Ala1736, and Lys59) were outside. Out of these 11, 7 (Arg1533, Ser1737, Lys594, and Lys59) formed a cavity with the residues 1112-1286.

The fourth potential active site encompassed residues 400-470 (70 residues) of the large protein. 21 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 21, 14

were a part of the potential active site 66.7%), and 7 (Arg/Thr1437, Thr1436, Arg1438, Ala1393, Thr1439, Lys645, Cys1433) were outside. Out of these 7, all formed a cavity with the residues 400-470.

The fifth, and last, potential active site encompassed residues 808-908 (100 residues) of the large protein. 22 residues of the protein bound 10 or more times to the relevant ligands (the ones that got -6.0 or lower during active site docking). Out of these 22, 15 were a part of the potential active site (68.1%) and 7 (Ile969, Arg1079, Leu967, Gly972, Thr976, Arg545, Glu1073) were outside. Out of these 15, 3 (Arg1079, Gly972, Thr967) formed a cavity with the residues 808-908.

The second, and third potential active sites got the worse results in terms of relevant ligands binding to the potential active sites. These binding sites were the ones that had the least and the most number of residues in them, respectively. The lack of residues may have required the ligands to search for residues outside of the proposed site for binding to be the strongest possible, and the big number of residues may have made the docking less specific.

Considering, then, the active sites that had the best results (620-700 with 85.7%, 400-470 with 66.7%, and the 808-908 with 68.1%), it's possible that these are closer to, or are, the active site of the glycoprotein. Furthermore, some of the extra residues (residues that were not in the potential active site but were found to bind to the relevant ligands), such as Lys481, Ser482, Ser483, Arg545, Arg552, Gly696, and Gly697 were found more than once in and outside of their potential active sites.

With this information, it is possible to narrow down potential active sites of the large protein to somewhere in between residues 400-700. This range could possibly match previous molecular docking results, which gave residues M585, E620, K621, W622, N623, E696, L698, A726, and K778 as potential active sites of the large protein (KIRIWAN e CHOOWONGKOMON, 2021). Considering this potential active site, the ligand that fit best would be A11.

5.2.5 Phosphoprotein Active Sites

No proposed active sites of phosphoprotein obtained results with ligands that were equal to or lower than -6.0. This made it impossible for this research to observe and analyze any potential active sites of this protein, as well as come up with a conclusion on it.

6.0 CONCLUSION

1 Accurate 3D-structure models of each of the five RABV proteins were obtained to which previously unknown active sites were mapped.

2. Through molecular docking and analysis of docking and homology results, four ligands were considered to yield the best results and could block RABV life cycle. In terms of active site, each protein (except for phosphoprotein) was analyzed and a range of residues were given as the possible active sites of each protein.

7.0 REFERENCES

ADEBAMBO, K.; GUNARATNAM, S. Molecular Docking Investigation of New Inhibitors of Falciparum vivax. **Computational Molecular Bioscience**, v. 8, n. 2, p. 43-67, Jun 2018.

AGOSTINHO, W. C.; BRANDAO, P. E. Intracerebral transfection of anti-rabies virus antibodies is an effective therapy for rabies. **Journal of Neurovirology**, v. 26, p. 764-768, July 2020.

ALBERTINI, A. A. V.; RUIGROK, R. W. H.; BLONDEL, D. Rabies Virus Transcription and Replication. **Research Advances in Rabies**, p. 1-22, 2011.

APPOLINARIO, C. M.; JACKSON, A. C. Antiviral therapy for human rabies. **Antiviral Therapy**, v. 20, n. 1, p. 1-10, 2014.

AUGERI, J. D.; FRAY, H. D.; KLEINMAN, F. E. Synthesis and Antibacterial Activity of 2, 3-Deydroofloxacin. **J. Heterocyclic Chem**, n. 27, 1990.

AUTODOCK. **AutoDock Scripps**, 2006. Disponivel em: <http://autodock.scripps.edu/wiki/AutoDock>. Acesso em: 20 jul. 2021.

BABIC, S. et al. Determination of pKa values of active pharmaceutical ingredients. **Trends in Analytical Chemistry**, v. 26, n. 11, p. 1043-1061, 2007.

BASTIDA-GONZÁLEZ, F. et al. Predicted 3D Model of the Rabies Virus Glycoprotein Trimer. **Biomed Res Int**, 2016.

BATALHA, N. P. et al. 4-Oxoquinoline Derivatives as Antivirals: A Ten Years Overview. **Current Topics in Medicinal Chemistry**, v. 20, p. 224-255, 2020.

BAWASKAR, S. H.; BAWASKAR, H. P.; BAWASKAR, H. P. Rabies: A Novel Clinical Presentation. **Indian J Crit Care Med**, v. 21, n. 12, p. 872-874, Dec 2017.

BEN KHALIFA, Y. et al. The matrix protein of rabies virus binds to RelAp43 to modulate NF-kappaB-dependent gene expression related to innate immunity. **Scientific Reports**, 2016.

BERMAN, M. H. et al. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235-242, 2000.

BEST PRACTICE MANUAL. The appropriate use of macrolides. **Best Practice Journal**, v. 44, p. 32-38, 2012.

BOBYROV, M. . V. et al. Pharmacology - Textbook for students of medical higher educational institutions. Vinnytsia: Nova Knyha, 2018. p. 55-62.

BRANDÃO, E. P. et al. Short-Interfering RNAs as Antivirals Against Rabies. **The Brazil Journal of Infectious Diseases** , v. 11, n. 2, p. 224-225, 2007.

CENTER FOR DISEASE CONTROL AND PREVENTION. Cost of Rabies Prevention. **Centers for Disease Control and Prevention**, 2019. Disponivel em: <https://www.cdc.gov/rabies/location/usa/cost.html>. Acesso em: June 2021.

CENTER FOR DISEASE CONTROL AND PREVENTION. Neglected Tropical Diseases. **Centers for Disease Control and Prevention**, 202. Disponivel em: <https://www.cdc.gov/globalhealth/ntd/diseases/index.html>. Acesso em: Jun 2021.

CENTER FOR DISEASE CONTROL AND PREVENTION. Rabies Around the World. **Centers for Disease Control and Prevention**, 2020. Disponivel em: <https://www.cdc.gov/rabies/location/world/index.html>. Acesso em: June 2021.

CHEN, K. . K.; HARRIS, N. . P.; ROSE, L. . C. The Action and Toxicity of Platyphilline and Seneciphylline. **Journal of Pharmacology and Experimental Therapeutics**, v. 68, n. 1, p. 130-140, 1940.

CHENIK, M. et al. In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. **Journal of General Virology**, v. 75, p. 2889-2896, 1994.

CONSORTIUM, T. U. UniProt: a worldwide hub of protein knowledge. **Nucleic Acid Res.**, v. 47, n. D1, p. D506–D515, 2019.

COORDINATORS, N. R. Database resources of the National Center for Biotechnology Information. **Nucleic Acid Res.**, v. 46, p. D8-D13, Jan 2018.

CSENDE, F.; PORKOLÁB, A. Antiviral activity of isoindole derivatives. **Journal of Medicinal and Chemical Sciences**, 2020.

DALHOFF, A. Antiviral, antifungal, and antiparasitic activities of fluoroquinolones optimized for treatment of bacterial infections: a puzzling paradox or a logical

consequence of their mode of action? **European Journal of Clinical Microbiology & Infectious Diseases**, v. 34, p. 661-668, Dec 2014.

DALLAYKAN, S.; OLSON, A. J. Small-Molecule Library Screening by Docking with PyRx. **Methods Mol Biol**, p. 243-250, 2015.

DASSAULT, S. BIOVIA Discovery Studio 2020. **2021**, San Diego.

DRUGBANK. Ofloxacin. **DrugBank**, 2021. Disponivel em: <https://go.drugbank.com/drugs/DB01165>. Acesso em: 2021.

DRUGBANK. Search over 500,000 drugs & drug products on DrugBank Online. **DrugBank**. Disponivel em: <https://go.drugbank.com/>. Acesso em: 2021.

DU PONT, V.; PLEMPER, K. R.; SCHNELL, J. M. Status of antiviral therapeutis against rabies virus and related emerging lyssavirus. **Current Opinion in Virology**, v. 35, p. 1-13, 2019.

DUMRONGPHOL, H. et al. Alteration of muscarinic acetylcholine receptors in rabies viral-infected dog brains. **Journal of the Neurological Sciences**, 1996.

ENGLES, D.; ZHOU, X. Neglected tropical diseases: an effective global response to local poverty-related disease priorities. **Infectious Diseases of Poverty**, 2020. Disponivel em: <https://idpjournal.biomedcentral.com/articles/10.1186/s40249-020-0630-9>. Acesso em: 15 maio 2021.

ENOKI, Y. et al. Pleiotropic Effects of Levofloxacin, Fluoroquinolone Antibiotics, against Influenza Virus-Induced Lung Injury. **Plos one**, June 2015.

FENG, M. et al. Sulfur Containing Scaffolds in Drugs: Synthesis and Application in Medicinal Chemistry. **Current Topics in Medicinal Chemistry**, v. 16, n. 11, p. 1200-1216, 2016.

FINKE, S.; MUELLER-WALDECK, R.; CONZELMANN, K. K. Rabies virus matrix protein regulates the balance of virus transcription and replication. **Journal of General Virology**, v. 84, p. 1613-1621, 2003.

FISHER, R. C.; STREICKER, G. D.; SCHNELL, J. M. The spread and evolution of rabies virus: conquering new frontiers. **Nat Rev Microbiol**, v. 16, n. 4, p. 241-255, April 2018.

GAMO, F. J. et al. Thousands of chemical starting points for antimalarial lead identification. **Nature**, v. 465, p. 305-310, March 2010.

GUEX, N.; PEITSCH, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. **Electrophoresis**, n. 18, p. 274-2723, 1997.

GUPTA, A. et al. The Phosphoprotein of Rabies Virus Is Phosphorylated by aUnique Cellular Protein Kinase and Specific Isomersof Protein Kinase C. **Journal of Virology**, v. 74, n. 1, p. 91-98, Jan 2000.

HANWELL, D. M. et al. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. **Journal of Cheminformatics**, v. 4, n. 12, p. 1-17, 2012.

HUEFFER, K. et al. Rabies virus modifies host behaviour through a snake-toxin like region of its glycoprotein that inhibits neurotransmitter receptors in the CNS. **Scientific Reports**, 2017.

IKEDA, S.; YAZAWA, M.; NISHIMURA, C. Antiviral activity and inhibition of topoisomerase by ofloxacin, a new quinolone •derivative. **Antiviral Research**, n. 8, p. 103-113, 1987.

JOCHMANS, D.; NEYTS, J. The path towards effective antivirals against rabies. **Vaccine**, v. 37, p. 4660-4662, 2019.

KAMMOUNI, W. et al. Rabies virus phosphoprotein interacts with mitochondrial Complex I and induces mitochondrial dysfunction and oxidative stress. **J Neurovirol**, v. 21, n. 4, p. 370-82, Aug 2015.

KASINATHAN, N. et al. Strategies for drug delivery to the central nervous system by systemic route. **Drug Delivery**, v. 22, n. 3, p. 243-257, June 2014.

KELLEY LA, E. A. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, v. 10, p. 845-858, 2015.

KIRIWAN, D.; CHOOWONGKOMON, K. In silico structural elucidation of the rabies RNA-dependent RNA polymerase (RdRp) toward the identification of potential rabies virus inhibitors. **J Mol Model**, v. 27, n. 6, p. 183, 2021.

KOUZNETZOFF, A.; BUCKLE, M.; TORDO, N. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. **J Gen Virol**, v. 79, p. 1005-1013, May 1998.

KOZAKOV, D. et al. Optimal clustering for detecting near-native conformations in protein docking.. **Biophysical journal**, 2005. 867-875.

KOZLOVSKAYAA, I. L. et al. Antiviral activity spectrum of phenoxazine nucleoside derivatives. **Antiviral Research**, v. 163, p. 117-124, 2019.

KUMARI, A.; SINGH, K. R. Morpholine as ubiquitous pharmacophore in medicinal chemistry: Deep insight into the structure-activity relationship (SAR). **Bioorganic Chemistry**, 2020.

LASKOWSKI, R. A.; SWINDELLS, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. **J. Chem. Inf. Model**, v. 51, p. 2778-2786, 2011.

LENTZ, T. L. Rabies virus binding to an acetylcholine receptor alpha-subunit peptide. **J Mol Recognit**, 1990. 82-88.

LENTZ, T. L. et al. Is the acetylcholine receptor a rabies virus receptor? **Science**, 1982. 4529.

LIPINSKI, A. C. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 23, p. 3-25, Aug 1996.

MADHU, B. P. et al. Correlation of inducible nitric oxide synthase (iNOS) inhibition with TNF-α, caspase-1, FasL and TLR-3 in pathogenesis of rabies in mouse model. **Virus Genes**, v. 52, n. 1, p. 61-70, Feb 2016.

MASATANI, T. et al. Rabies virus nucleoprotein functions to evade activation of the RIG-I-mediated antiviral response. **Journal of Virology**, v. 84, p. 4002-4012, 2010.

MASATANI, T. et al. Amino acids at positions 273 and 394 in rabies virus nucleoprotein are important for both evasion of host RIG-I-mediated antiviral response and pathogenicity. **Virus Research**, v. 155, n. 1, p. 168-174, Jan 2011.

MEBATSION, T.; WEILAND, F.; CONZELMANN, K.-K. Matrix Protein of Rabies Virus Is Responsible for the Assembly and Budding of Bullet-Shaped Particles and Interacts

with the Transmembrane Spike Glycoprotein G. **J Virol**, v. 73, n. 1, p. 242-250, Jan 1999.

MIKITSH, J. L.; CHACKO, A.-M. Pathways for Small Molecule Delivery to the Central Nervous System Across the Blood-Brain Barrier. **Perspect Medicin Chem**, v. 6, p. 11-24, 2014.

MIKKELSEN, K.; V., A. Vitamin B12, Folic Acid, and the Immune System. In: MAHMOUDI, M.; REZAEI, N. **Nutrition and Immunity**. [S.l.]: [s.n.], 2019. p. 103-114.

MITCHARD, M. Sulphur Compounds Used in Medicine. **Drug Metabolism and Drug Interactions**, v. 6, n. 3/4, 1988.

MOREIRA, R. et al. Pyrrolizidine Alkaloids: Chemistry, Pharmacology, Toxicology and Food Safety. **Int J Mol Sci**, v. 19, n. 6, 2018.

MORETTI, V. M.; PAULUZZI, S.; CESANA, M. Penetration of Rufloxacin into the Cerebrospinal Fluid in Patients with Inflamed and Uninflamed Meninges. **Antimicrobial Agents and Chemotherapy**, Dec 2020.

MORRIS, G. M. et al. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. **J. Computational Chemistry**, v. 30, n. 16, p. 2785-91, 2009.

MORRIS, M. G. et al. **Automated Docking of Flexible Ligands to Flexible Receptors**. [S.l.]. 2012.

NAKAGAWA, K. et al. Molecular Function Analysis of Rabies Virus RNA Polymerase L Protein by Using an L Gene-Deficient Virus. **Journal of Virology**, v. 91, n. 20, Sep 2017.

NAKAMICHI, K. et al. Rabies Virus Stimulates Nitric Oxide Production and CXC Chemokine Ligand 10 Expression in Macrophages through Activation of Extracellular Signal-Regulated Kinases 1 and 2. **Journal of Virology**, Dec 2020.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **PubChem Compound Summary for CID 5281742, Platyphylline**. Disponivel em: <https://pubchem.ncbi.nlm.nih.gov/compound/Platyphylline.>. Acesso em: Jun 2021.

NCBI RESOURCE COORDINATORS. Database resources of the National Center for Biotechnology Information. **Nucleic Acid Res**, v. 46, n. D1, p. D8-D13, 2018.

NIGG, J. A.; WALKER, L. P. Overview, Prevention, and Treatment of Rabies. **Pharmacotherapy**, v. 29, n. 10, p. 1182-1195, 2009.

OGINO, M. et al. The rabies virus L protein catalyzes mRNA capping with GDP Polyribonucleotidyltransferase Activity. **Viruses**, v. 8, n. 5, p. 144, 2016.

OLIPHANT, M. C.; GREEN, M. G. Quinolones: A Comprehensive Review. **Am Fam Physician**, v. 65, n. 3, p. 455-465, Feb 2002.

ORLOWSKI, P. et al. Modelling of pH dynamics in brain cells after stroke. **Interface Focus**, March 2011.

PAGADALA, S. N.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a review. **Biophys Rev.**, v. 9, n. 2, p. 91-102, April 2017.

PERRY, G. et al. Pharmacokinetics of rufloxacin in patients with impaired renal function. **Antimicrobial Agents and Chemotherapy**, v. 37, n. 4, p. 637-641, Apr 1993.

PHILIPS, M. A.; STEWART, M. A. . W. D. L. A. X. Z. R. Has Molecular Docking Ever Brought Us a Medicine? In: VLACHAKIS, D. **Molecular Docking**. [S.l.]: IntechOpen, 2018. Cap. 8, p. 141-178.

POMEROY, R. A.; RAPER, C. Pyrrolizidine alkaloids: actions on muscarinic receptors in the guinea-pig ileum. **Br J Pharmacy**, v. 41, p. 683-690, 1971.

REVELL, P. et al. Folic acid absorption in patients infected with the human immunodeficiency virus. **J Intern Med**, v. 230, n. 3, p. 227-31, 1991.

RICHTER, S. et al. Antiviral Properties of Quinolone-based Drugs. **Current Drug Targets - Infectious Disorders** , v. 4, n. 2, p. 111-116, 2004.

RIEDEL, C. et al. Cryo EM structure of the rabies virus ribonucleoprotein complex. **Scientific reports**, v. 9, n. 1, 2019.

RIMOLDI, R. et al. Pharmacokinetics of rufloxacin once daily in patients with lower respiratory tract infections. **Infection**, v. 20, n. 2, p. 89-93, 1992.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein. **Nature Protocols**, v. 5, p. 725-738, 2010.

RUAN, J.; LIAO, C.; LIN, G. Lack of metabolic activation and predominant formation of an excreted metabolite of nontoxic platynecine-type pyrrolizidine alkaloids. **Clinical Research in Toxicology**, v. 27, n. 1, p. 7-16, Jan 2014.

SANTOS, F. et al. Synthesis, antiviral activity and molecular modeling of oxoquinoline derivatives. **Bioorganic & Medicinal Chemistry**, v. 17, p. 5476-5481, 2009.

SCHNEIDER, G. **Prediction of Drug-Like Properties**. Austin. 2000-2013.

SCHRÖDINGER, LLC. The PyMOL Molecular Graphics System.

SHITYAKOV, S.; FORSTER, C. In silico predictive model to determine vector-mediated transport properties for the blood-brain barrier choline transporter. **Advances and Applications in Bioinformatics and Chemistry**, v. 2014, n. 7, p. 1-14, Aug 2014.

SONTHONNAX, F. et al. Lyssavirus matrix protein cooperates with phosphoprotein to modulate the Jak-Stat pathway. **Sci Rep**, v. 9, n. 1, p. 12171, 2019.

STEIN, E. G.; FLOR, C. S.; BEALS, S. B. Safety of multiple doses of ofloxacin in healthy volunteers. **Drugs Exp Clin Res**, v. 17, n. 10-11, p. 525-529, 1991.

STERLING, T.; IRWIN, J. J. ZINC 15 – Ligand Discovery for Everyone. **J. Chem. Inf. Model.**, v. 55, n. 11, p. 2324-2337, 2015.

THE HUMAN METABOLOME DATABASE. Metabocard for Rufloxacin. **Human Metabolome Database**. Disponivel em: <https://hmdb.ca/metabolites/HMDB0042009>. Acesso em: 2021.

TOMAR, N. R. et al. Molecular Docking Studies with Rabies Virus Glycoprotein to Design Viiral Therapeutics. **Indian J Pharm Sci**, v. 72, n. 4, p. 486-90, 2010.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455-461, 2010.

UNWIN, N. Refined Structure of the Nicotinic Acetylcholine Receptor at 4A Resolution. **J Mol Biol**, v. 346, n. 4, p. 967-989, 2005.

VALISENA, S. et al. Relevance of Ionic Effects on Norfloxacin Uptake by Eschericia Coli. **Biochemical Pharmacology**, v. 40, n. 3, p. 431-436, 1990.

VIDY, A.; CHELBI-ALIX, M.; BLONDEL, D. Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. **Journal of Virology**, v. 79, n. 22, p. 14411-14420, Nov 2005.

WHO. Neglected Tropical Diseases. **World Health Organization**. Disponivel em: <https://www.who.int/neglected_diseases/diseases/en/>. Acesso em: 13 abr. 2020.

WORLD HEALTH ORGANIZATION. Rabies. **World Health Organization**, 2021. Disponivel em: <https://www.who.int/news-room/fact-sheets/detail/rabies>. Acesso em: Jun 2021.

XIA, X. Bioinformatics and Drug Discovery. **Curr Top Med Chem**, v. 17, n. 15, p. 1709-1726, Jun 2017.

YANG, F. et al. Structural Analysis of Rabies Virus Glycoprotein Reveals pH-Dependent Conformational Changes and Interactions with a Neutralizing Antibody. **Cell Host & Microbe**, v. 27, n. 3, p. 441-453, 2020.

YOUSAF, Z. M. et al. Rabies molecular virology, diagnosis, prevention, and treatment. **Virology Journal**, v. 9, n. 50, 2012.

ZAN, J. et al. Rabies virus matrix protein induces apoptosis by targeting mitochondria. **Exp Cell Res**, v. 347, n. 1, p. 83-94, 2016.

ZHANG, L. et al. Screening of Potent Inhibitor of H1N1 Influenza NSI CPSF30 Binding Pocket by Molecular Docking. **Advances in Infectious Diseases**, v. 2, n. 4, p. 92-96, 2021.

ZINAD, S. D. et al. Medicinal chemistry of oxazines as promising agents in drug discovery. **Chem Biol Drug Des**, p. 1-32, 2019.