COMPUTATIONAL INVESTIGATION OF ALLOSTERY IN NANOBODY – ANTIGEN COMPLEXES

A Thesis Submitted to
the Graduate School of Engineering and Science of
Izmir Institute of Technology
in Partial Fulfilment of the Requirements for the Degree of

MASTER OF SCIENCE

in Biotechnology

by
Nurşah HALİSDEMİR

July 2024
İZMİR

We approved the thesis of **Nurşah HALİSDEMİR**

Examining Committee Members:

Assist. Prof. Arzu UYAR
Department of Bioengineering, Izmir Institute of Technology

Assoc. Prof. Nur Başak SÜRMELİ ERALTUĞ
Department of Bioengineering, Izmir Institute of Technology

Assist. Prof. Dr. Deniz KARATAŞ
Department of Bioengineering, Celal Bayar University

10 July 2024

Assist. Prof. Arzu UYAR
Supervisor, Department of Bioengineering, Izmir Institute of Technology

Assist. Prof. Hümeyra TAŞKENT SEZGİN
Co-Supervisor, Department of Bioengineering, Izmir Institute of Technology

Assoc. Prof. Ali Oğuz BÜYÜKKİLECİ
Head of the Department of Biotechnology and Bioengineering Graduate Program

Prof. Dr. Mehtap EANES
Dean of the Graduate School of Engineering & Science

ACKNOWLEDGMENTS

I would like to acknowledge the IZTECH Scientific Research Projects Coordinator Unit (project no: AUDP-2022-İYTE-3-0041) for financial support.

First and foremost, I would like to express my heartfelt appreciation to my beloved thesis supervisor Assist. Prof. Arzu UYAR, and co-supervisor Assist. Prof. Hümeyra TAŞKENT SEZGİN, who provided me with the chance to research this subject, supported me in every way, showed understanding, and passed on their experience to me.

I would also like to express my unending gratitude to all of my instructors at the Izmir Institute of Technology, who attended or did not attend my lecture, who provided the most accurate answers to every question I posed, who demonstrated the importance of questioning to obtain the correct information, and who always guided me in the right direction. I want to convey my unending love, thanks, and gratitude to my wonderful instructors at Erciyes University Faculty of Pharmacy, who have given me their support in every aspect since the beginning of my university studies and inspired me to pursue a career in pharmacy.

I would like to express my unending love and gratitude to my best friend and source of motivation, my cat Beta, whom I adopted on February 5, 2023, and who is more than a pet for me, and to my beloved family, who have provided me with material and moral support at every stage of my life, as well as their shadow and presence when I was helpless. I want to convey my heartfelt appreciation to all of my friends, to my UBioModel office team who are like a family to me, to my boyfriend who softens me on his chest when I am stressed, and to colleagues who helped me up when I fell, who were always there for me no matter what, and who made me feel their presence from the other side of the world.

ABSTRACT

COMPUTATIONAL INVESTIGATION OF ALLOSTERY IN NANOBODY-ANTIGEN COMPLEXES

Functions of biomolecules can be regulated by allostery, where a molecule binds to a site far from the active site of the structure. Recently, nanobodies are promising for disease treatment and imaging. Exploration of the existence of allostery and new potential sites in nanobodies have not yet been studied in detail, and computational approaches help speed up research on allostery in nanobodies. For this aim, three nanobody-antigen complexes, namely Caplacizumab - von Willebrand factor (vWF), Nanobody 87 - NTCP, and Nanobody 2-67 - SARS-CoV-2 Spike, were computationally examined using Essential Site Scanning Analysis (ESSA) method to predict potential binding sites. The default cutoff distance value of ESSA (10 Å) is replaced with a lower (7.3 Å) and a higher (13 Å) cutoff value where the total number of modes is also increased to 20 in the new ESSA. The old and new ESSA are applied to all structures, and the results are compared to each other to better understand the effect of new parameters on the success of ESSA. It is observed that ESSA is improved with the new parameters, and more successful results are obtained when the cutoff value is 7.3 Å, especially in finding essential residues in complementarity-determining regions (CDRs) for the nanobodies studied. The improved ESSA also revealed some other binding sites in the antigens studied where they interact with other proteins and ligands. The improved method in this thesis might help develop antigen-specific new-generation therapeutics and better diagnostic tools.

ÖZET

NANOKOR-ANTİJEN KOMPLEKSLERİNDEKİ ALOSTERİNİN HESAPLAMALI İNCELENMESİ

Gelişen, değişen ve dijitalleşen dünyada hesaplamalı biyokimya alanının sunduğu avantajlar birçok araştırmanın bilgisayar ortamına taşınmasını sağlamıştır. Kümülatif bilgi birikimi ile büyüyen bilim dünyasında nanobodilerin tanı ve tedavide sağladığı faydalar ile nanokor-antijen komplekslerinin araştırılması son zamanlarda giderek artmıştır. Bu bağlamda bu çalışmada, nanokor-antijen komplekslerindeki bağlanma bölgeleri dışında kalan önemli bölgeler araştırılarak yapıların iç dinamiklerinde meydana gelen alosterik değişimler ve bu değişimlerin yapılara yansıması yorumlanmıştır. Çalışmada Caplacizumab – von Willebrand Faktör (vWF), Nanokor 87 – NTCP ve Nanokor 2-67 – SARS-CoV-2 Spike protein kompleksleri incelenmiştir. Bu yapılara ek olarak bu kompleks yapıların bir bileşeni olan antijenlere bağlanan diğer moleküller de çalışmanın konuşu olmuştur. Bu incelemelerde bir potansiyel bağlanma bölgesi tahmin uygulaması olan Essential Site Scanning Analysis (ESSA) uygulaması kullanılmıştır. Yapıların bağlanma bölgelerinin analizi z-skor sonuçlarına göre yapılmış olup önemli bölgeler 5'in üzerinde z-skor sonucuna sahip bölgeler olarak belirlenmiştir. Bu çalışmadaki önemli bulgulardan biri de incelenen nanokor, antijen yapılarının birbirleri ile bağlanma bölgeleri dışında da önemli bölgelere sahip olmasıdır. Bağlanma bölgesi dışında bulunan ve moleküllerin iç dinamiklerinde meydana gelen alosterik değişimlerden kaynaklandığı düşünülen bu bölgelerin, gelecekte nanokor – antijen komplekslerinin bağlanma mekanizmalarının açıklanmasına yeni bir bakış açısı sunacağı ve kapsamlı yapısal analizlerin önemini vurgulamaktadır.

To all the females. To the children who are hopeful for their future,

To my motivational source, my lovely cat Beta…

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

Ab	Antibody
Ag	Antigen
ANM	Anisotropic Network Model
C	Constant
CDR	Complementarity Determining Region
ENM	Elastic Network Model
ESSA	Essential Site Scanning Analysis
Fab	Fragment of Antigen Binding
Fc	Fragment of Crystallizable
GNM	Gaussian Network Model
GPCR	G-Protein Coupled Receptor
HBV	Hepatitis B Virus
HDV	Hepatitis D Virus
Ig	Immunoglobulin
kDa	Kilodalton
mAb	Monoclonal Antibody
Nb	Nanobody
NBC	Non-Bounded Contact
nm	Nanometer
NTCP	Sodium Taurocholate Cell Polypeptide
PDB	Protein Data Bank
pH	Potential of Hydrogen
RBD	Receptor Binding Domain
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
sdAbs	Single-domain Antibodies
TTP	Thrombotic Thrombocytopenic Purpura
V	Variable
VHH	Variable Domain of Heavy Chain-only Antibody
VUB	Vrije University of Brussel

vWD	von Willebrand Disease
vWF	von Willebrand Factor
β2AR	β2 Adrenergic Receptor

CHAPTER 1

INTRODUCTION

With the development of technology from the past to the present, many different and innovative approaches have been developed in science and the health sector. One of the prominent areas in the developing, changing, and digitalizing new world order has been the field of computer-aided research. While computers have made users' lives easier in many areas, they have also enabled much research to be transferred to the computer environment. With the advantages offered by computer-aided research, processing, grouping, comparing, storing, and using data has become safer. For all these reasons, the importance and use of computer-aided research continue to increase (Stoumpos, Kitsios and Talias 2023).

Throughout history, human beings have discovered many diseases and found treatment methods for these diseases. Experimental studies carried out in the past are not as cheap, and easy as they used to be due to the excessive cost of living, and time is a valuable concept. Nowadays, these experimental studies are supported by studies conducted in computer environments. Computational sciences are instrumental in the discovery of new molecules, alternative binding sites, the examination of changes in the internal dynamics of molecules, organisms that cause diseases, and the development of treatment methods with innovative approaches to these organisms (Tulchinsky and Varavikova 2014).

A disease is a condition that affects daily life and is caused by all kinds of molecules that are foreign to the body and come from outside to infect the body. Diseases or illnesses are caused only by foreign substances coming from outside. A disease or illness can be caused by the entry of foreign substances into the body which disrupts the body's internal balance (Nicholson 2016). Any foreign molecule that comes to the body from outside, is foreign to the body, stimulates the living system, and is not the living system itself is called an antigen, in other words, an immunogen. The body shows some reactions to these substances when they enter the body. These reactions can mostly

manifest as pain, fever, cough, sneezing, headache, and infection. The body has a defense mechanism to fight against foreign substances. This defense mechanism is called the 'immune system' (Beisel 1999).

The immune system is the entirety of the processes that protect living things against all kinds of foreign molecules and recognize and destroy these substances. In other words, the immune system is defined as the body's resistance to diseases, especially infectious diseases. This system, whose main task is to destroy all kinds of substances and organisms that are foreign to the body, stands out with its functions of resistance to infectious agents, protection against foreign antigens, and resistance to tumor development (Rao 2018; Abbas, Lichtman and Pillai 2020).

1.1. From Antibodies to Nanobodies

Antigens are recognized by the immune system when they enter the body. As a result of recognition, the production of antibodies against antigens begins in the body. The discovery of antibodies, in other words, immunoglobulins (Igs), dates back many years. The discovery of antibodies during serum treatment and vaccine development studies against the diphtheria pathogen earned German physiologist Emil von Behring the Nobel Prize in Physiology and Medicine in 1901 (Kaufmann 2017).

Antibodies, which are of great importance for the immune system, are molecules that are specific to their antigens thanks to their particular Y-shaped structure, consisting of glycoprotein structures and heavy and light protein chains, as shown in Figure 1 (Janeway, et al. 2001). However, antibodies contain Fab (Fab: Fragment of antigen binding) and Fc (Fragment of crystallizable) regions. The two N-terminal fragments are called the Fab region, and the C-terminal fragment is called the Fc region. While the Fab region is responsible for antigen binding, the Fc region is responsible for biological activity. Each antibody molecule contains two Fab regions and one Fc region. The flexible region between the Fab and Fc regions is called the "hinge region". This region allows the arms of the molecule to recognize more molecules. In addition, variable (V) regions on the heavy and light chains of antibodies, are involved in antigen recognition; There are

constant (C) regions that provide structural integrity and undertake effector functions (Nicholson 2016; Abbas, Lichtman and Pillai 2020; Tsumoto and Kuroda 2022).

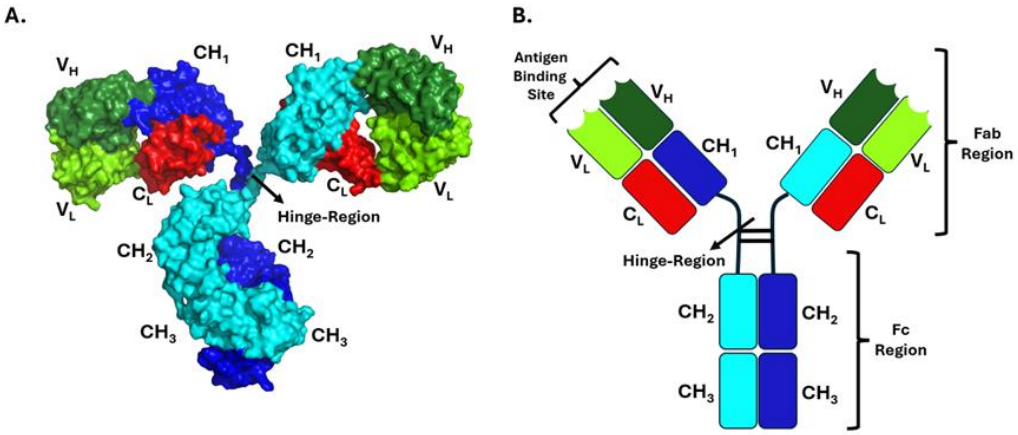


Figure 1. The representation of an antibody structure (PDB ID: 1IGT) A: Antibody 3D structure, B: Antibody structure schematized in the form of Y (Harris, et al. 1997; Berman, et al. 2000).

Antibodies, whose discovery has paved the way for many studies in the world of science and the field of immunology, are stable molecules that are easily produced and have the potential to bind to the same or different target proteins. The size of antibodies that stand out with their high antigen binding capacity is approximately 150 kDa. The fact that antibodies are tiny provides them with a significant advantage in targeted treatments and delivery to hard-to-reach body parts, and they are also used in medical diagnosis, treatment, and imaging. Antibodies are also used in biotechnological approaches developed to reduce the possible side effects of conventional drugs. Antibodies specific to antigens that cause complications in the body also provide advantages to scientists in targeted, personalized treatments. Although antibodies have various advantages, smaller molecules called “nanobodies” with similar properties have been discovered over time (Rapley 1995).

Nanobodies, in other words, single-domain antibodies (sdAbs) or VHHs, are molecules discovered by chance by a group of students working in the laboratory of

Professor Raymond Hamers at the Vrije University of Brussel (VUB) to develop a test for a type of infection (Arbabi-Ghahroudi 2017). The term VHH, used to refer to nanobodies, was first used by the VUB team in 1993 to refer to a VH domain derived from camelid heavy chain antibodies (Bo-Kyung, et al. 2023). In one of the earlier studies in this field, Sally Ward and colleagues published a report explaining the concept of single-domain antibodies (sdAbs) in 1989. In the historical process of fragments derived from camelids called VHH, medical development and use of these structures appear to have consisted of many important stages (Arbabi-Ghahroudi 2017).

During the ten-year period between 1993 and 2003, called the discovery phase, nanobodies were defined, and their gene regions were mapped. These structures, whose sequence analysis was performed, and crystal structures were elucidated, were also isolated with enzyme inhibitors and antigen-specific nanobodies in this period (Mitchell and Colwell 2018). The discovery period also coincides with the establishment of the Ablynx company as a subsidiary of the Vlaams Interuniversitair Instituut voor Biotechnologie in December 2001. The second phase of developments regarding nanobodies covers the period between 2003 and 2013 (Arbabi-Ghahroudi 2017). During these years, more than fifty researchers from various parts of the world have conducted many studies on the subject, and articles on these studies have been published. Researchers in this research group are responsible for investigating the uses of nanobodies in biotechnology and medicine. As a result of the research, it has been discovered that nanobodies are used as imaging agents in the control of platelet aggregation, treatment of viral infections, treatment of rheumatoid arthritis, and imaging of some tumor types. The studies carried out since 2014 constitute the third phase of the discovery of nanobodies. At this stage, clinical experiments with nanobodies were also conducted, and patent applications were made in this field by various companies (such as Merck, Boehringer Ingelheim, Sanofi, and Ablynx) in the summer of 2014 in Europe and in the summer of 2017 in the USA. Caplacizumab (bivalent anti-vWF nanobody for the treatment of rare blood clotting disorders), the first nanobody-based drug, took its place on the market in 2019 (Hassanzadeh-Ghassabeh, et al. 2013).

Nanobodies have many advantages. These structures, which have high affinity, good resolution, and specificity, are used in various fields due to their small size (2.5 nm diameter; ~ 4 nm length; ~ 15 kDa), low production costs, low immunogenicity, and high binding potential to tissues (Wang, et al. 2016). In addition, nanobodies, which are matured and strengthened *in vivo* with minimal modification potential, also contribute to

targeted immune system stimulation and the creation of phage display libraries, thanks to their small size and high stability (Franzel, Schirrmann and Hust 2016). Nanobodies, which generally have short half-lives, can be used as imaging agents in tumor imaging and help clear molecules such as radioisotopes in the blood, as they are structures that can extend their half-lives with their pharmacokinetic properties (Zheng, et al. 2022). Since 2016, nanobodies have been isolated against more than 120 therapeutic targets in diseases related to oncology, *in vivo* imaging, hematology, infectious diseases, neurological and inflammatory disorders, and some of these therapeutics have progressed to advanced stages of clinical trials (Arbabi-Ghahroudi 2017).

The VHH region, which provides nanobodies with many advantages and corresponds to the variable regions located at the C-terminal end of the heavy chains of antibodies, is a region that shows high stability and high resolution in addition to the advantages of specificity and high affinity. The antigen binding regions in VHHs, each with different conformations and different amino acids, are called "hypervariable regions (complementarity determining region) (CDR region). There are three CDR (CDR-1, 2, 3) regions here (Figure 2). Thanks to the CDR regions, the basic structure of the molecule is preserved, and maximum variability is achieved. Among these, the region with more antigen diversity than other CDR regions is mostly CDR-3 regions (Mitchell and Colwell 2018). In a study conducted by Mitchell *et al.* examining 156 nanobody antigen complexes, it was concluded that the solubility of nanobodies depends on the length of the CDR-3 regions (Mitchell and Colwell 2018).

As a result of many studies conducted on the structures available on the RCSB Protein Data Bank (RCSB PDB) (Berman, et al. 2000) web server, some conformations have been discovered for the CDR-3 regions, as shown in Figure 3. These conformations are divided into three different structural classes. These are helical bending, non-helical folded conformations, and β-sheet-like extended structures. Helical bend conformations are divided into four subclasses depending on the position of the helices. Coiled conformations are also divided into two subclasses (Tsumoto and Kuroda 2022; Oresta, Ametrano and Coscia 2021).

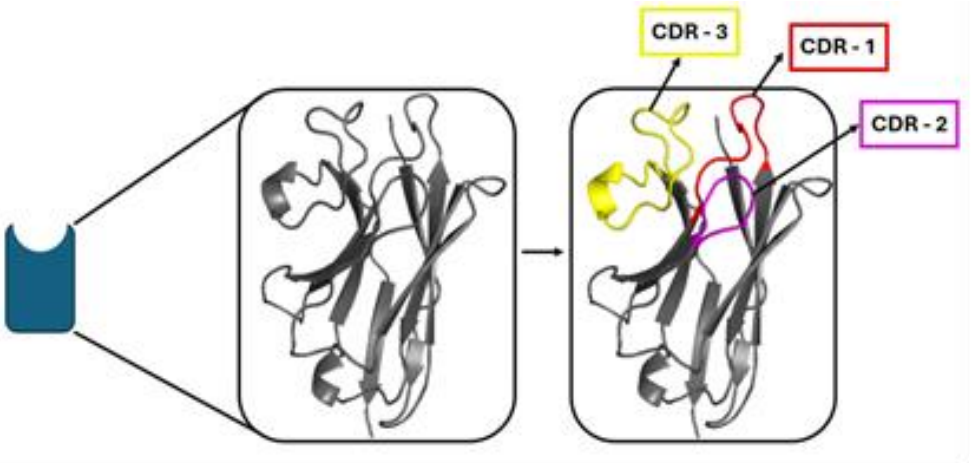


Figure 2. Representation of regions corresponding to CDR regions in a nanobody (PDB ID: 7EOW) (Lee, et al. 2021; Berman, et al. 2000).

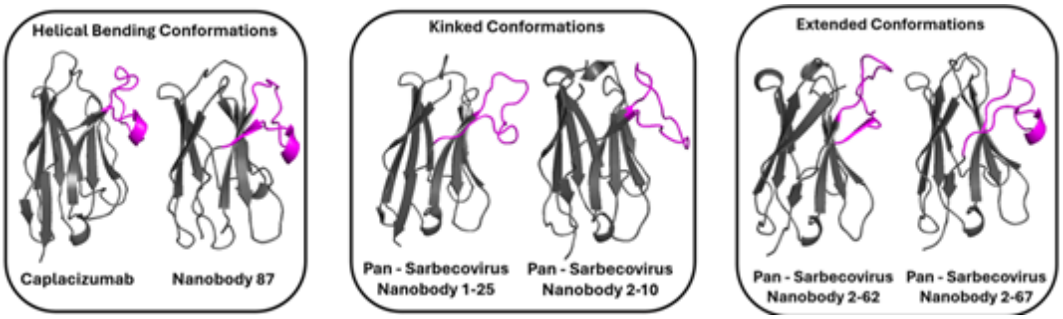


Figure 3. The representation of three different conformations of CDR3 found in VHH (Lee, et al. 2021; Goutam, et al. 2022; Xing, et al. 2022; Berman, et al. 2022).

Unlike classical monoclonal antibodies (mAb), nanobodies can reach hard-to-reach body parts (for example, intracellular targets) more easily. In addition, their small molecular size, target specificity, and long CDR-3 regions provide nanobodies with ease of use in targeted treatments, unlike classical drugs. As a result of the research by Jovčevska and Muyldermans, it has been observed that, unlike antibodies, these structures, which contain a single disulfide bond, have higher intracellular expression (Jovčevska and Muyldermans 2020).

While nanobodies have advantages, they also have disadvantages. Stability problems are observed in these structures, which can be easily affected by environmental factors such as elevated temperature and low pH. These small structures are also used in detecting environmental chemicals due to their physicochemical properties. However, for these uses, excellent insulation is required. This is a factor that limits the use of nanobodies in this field. Due to their unique structure, their use in targeted therapies makes nanobodies significant in the diagnosis and treatment of infectious diseases (Bo-Kyung, et al. 2023).

Just like antibodies, nanobodies specific to their antigens also have particular regions called paratopes that are responsible for antigen recognition. Paratopes are regions that add specificity to nanobodies. Therefore, nanobodies can recognize and bind to their antigens through paratopes located in a specific region at the junction between the two light chains and the heavy chains of the nanobodies. The corresponding region of this region in the antigen molecule is called the epitope region. The epitopes in antigens recognize and bind to their specific nanobodies. These regions are on the surface of the antigen and shaped according to the paratope of the nanobody (Figure 4) (Weller 2018).

In a study conducted by Mitchell *et al.* by examining 156 nanobody and antigen complex structures, it was concluded that the paratope region of the nanobody that interacts with the epitope region of the antigen is mostly the CDR-3 region (Mitchell and Colwell 2018). Antigens have many different epitope regions and stimulate the immune system with these regions, resulting in an immune response in the body (Schroeder Jr and Cavacini 2010).

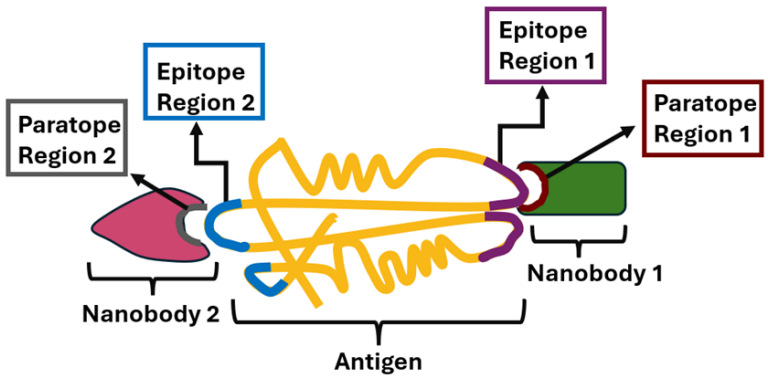


Figure 4. Representation of different nanobodies recognized by different epitope regions on the antigen and paratope regions found in nanobodies.

1.2. Allostery in Nanobody-Antigen Complexes

Proteins are molecules that are very important for every living organism. Proteins appear in almost every metabolic event that occurs in an organism. Thanks to their three-dimensional structure, unique surface properties, and internal dynamics, proteins undertake essential tasks such as intracellular signal transmission, regulation of enzyme activities, and fulfillment of metabolic activities. Diseases can be treated and prevented by understanding and explaining the changes in the internal dynamics of proteins (Albert, Johnson and Lewis 2002).

As is known, proteins composed of amino acids are molecules that are in motion and can open and close, and these movements are also reflected in their internal dynamics. The term allostery, which emerges in the internal dynamics of constantly moving proteins and forms the basis of many mechanisms in the body, was first introduced in 1961 by Jacques Monod and Francois Jacob when they realized that proteins interacting with each other had different structures (Liu and Nussinov 2016).

Much research has been conducted on the allostery over time. As a result of this research, the concept of allostery, in its most general definition, is all the changes that occur in proteins due to the binding of a molecule to a place far from the active site of the protein. Although the mechanism underlying the concept of allostery is not fully explained, it is estimated to arise from the physical properties of macromolecules (Wodak, et al. 2019).

Allostery was defined in the 1980s (Cooper and Dryden 1984) as the set of changes that occur in the protein system as a result of an effect applied to the protein structure. Various conferences and workshops have been organized to explain this concept. In the congresses and workshops organized, the results from computational and experimental analyses were evaluated, and the importance of allosteric mechanisms was emphasized (Wodak, et al. 2019).

Nanobodies specific to their antigens bind to their antigens with high affinity. As a result of this binding, nanobody–antigen complexes are formed. The binding of antigens to nanobodies may cause some changes in these protein molecules (Figure 5), and allosteric interactions might explain these changes. These may affect the binding affinity

of the molecule, as well as modulate biological activities through conformational changes (Mitchell and Colwell 2018).

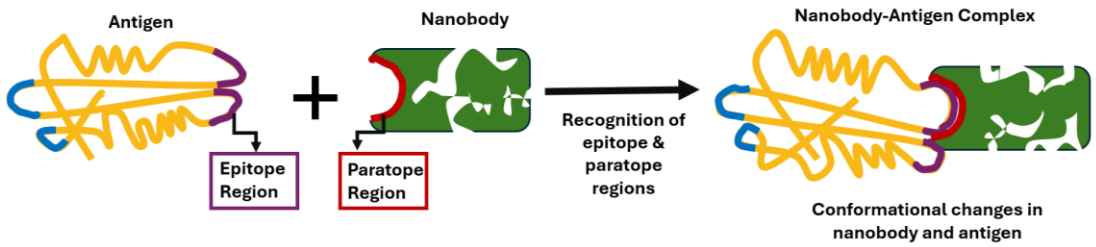


Figure 5. Schematic representation of the epitope and paratope regions of nanobody and antigen molecules recognizing each other and forming complexes and allostery.

A few research has been conducted to better understand and explain the concept of allostery in nanobody-antigen complexes. In a study by Staus *et al.* on allosteric nanobodies, receptor behavior, and allosteric changes were examined for two nanobodies (Nb60 and Nb80) directed to the G-protein coupled receptor (GPCR). In this study, it was observed that a series of conformational changes occurred in the receptor by binding Nb60 to the inner part of the receptor. When these conformational changes were tested with a ligand binding to this receptor, it was observed that the affinity of the ligand increased in the presence of Nb60 (Status, et al. 2016). The results obtained in this mostly experimental study coincide with the results found by Dror *et al.* by simulations (Dror, et al. 2009).

Another study by Chen *et al.* examined the allosteric effects that occur when a nanobody 80 binds to the β2 adrenergic receptor (β2AR), In another study by Chen *et al.* they examined the allosteric effects and conformational changes that occur in the receptor when a nanobody 80 binds to the β2 adrenergic receptor (β2AR), regardless of the ligand. The results obtained concluded that the nanobody80 (Nb80) binding to the receptor makes conformational changes in the receptor and supports active site-like conformations in the binding of various ligands, independent of the ligand. In addition, Chen and his colleagues noted in their study that they observed structural changes in the receptor's signaling pathway, due to the nanobody80 (Nb80) that was attached to the receptor (Chen, et al. 2021).

In a study both experimentally and computationally conducted by Yang *et al.* allosteric regulation through designed nanobodies was used to inhibit the interaction between the receptor and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Researchers stated that especially the CDR-2 and CDR-3 regions played an important role in the 16 designed nanobodies, where they bind to three different binding sites on SARS-CoV-2 (Yang, et al. 2021).

As seen in many studies, computational approaches have played an important role in understanding the conformational changes that occur in the internal dynamics of molecules and the process that develops because of these changes. Similarly, the discovery and understanding of the concept of allostery has formed the basis of many studies and enabled the understanding of many biological mechanisms occurring in the body. In addition, nanobodies have provided various advantages to researchers with their unique properties. At the same time, including allostery in research on nanobody antigen complexes forms the basis of many future studies (Wodak, et al. 2019).

This thesis focuses on possible allosteric changes that might exist in nanobody and antigen structures and their complexes.

Firstly, in the introduction (Chapter 1), the importance of computational approaches in current research is emphasized, and the structures of nanobodies derived from antibodies, what allostery is, the importance of allosteric mechanisms, and research on this subject are mentioned. In the Materials and Methods sections (Chapter 2), detailed information is given about the nanobody-antigen complex systems used in this study and the computational methods used to examine the changes occurring in these complex systems. Later, in the Result and Discussion (Chapter 3) section, the results of this study and the findings are presented. This section examines the possible effects of changes in the internal dynamics of proteins in terms of allostery. Finally, in the Conclusion section (Chapter 4), the general significance of the results obtained and suggestions for future research are presented.

CHAPTER 2

MATERIALS AND METHODS

In science, knowledge progresses cumulatively. In computational sciences, the currently known information has been reached through the methods developed and theories presented by many scientists from the past to the present. This situation subjects scientific studies to a multifaceted evaluation. For this reason, thinking and interpreting from multiple perspectives is an important criterion when evaluating the results of a research or study.

Proteins serve many critical roles in organisms. Although the mechanisms of some proteins have been elucidated, the working mechanisms of many proteins have not yet been elucidated. Nanobodies are structures composed of amino acids, just like proteins (see section 1.1 for the nanobody structure). Amino acids have various functions in biological systems, and they also form the building blocks of structures composed of amino acids (such as nanobodies and antigens). There are 20 essential amino acids in the protein structures found in organisms. Each amino acid contains an amino group (NH2), a carboxyl group (COOH), a hydrogen atom (H), and a changing side chain (R) in its structure. This side chain not only determines the chemical properties of amino acids but also plays a critical role in the shape and functions of the structures in which they are found (Schumacher, et al. 2018). Some amino acids in nanobodies interact with antigens through interactions, such as hydrogen bonding, non-bonded contact (NBC), and salt bridge. In this thesis, three nanobody–antigen structures selected from various classes were examined. The main features of these structures examined are given in Table 1.

Table 1. Properties of the nanobody-antigen complex structures used in this thesis study (Berman et al. 2000).

PDB ID	Nb & Ag Names	Mutation(s)	Classifications	References
7EOW	Caplacizumab	No	Structural Protein	(Lee, et al. 2021)
	von Willebrand Factor (vWF)	No		
7PQG	Nanobody 87	No	Membrane Protein	(Goutam, et al. 2022)
	NTCP	Yes		
8CYJ	Nanobody 2-67	No	Viral Protein	(Xiang, et al. 2022)
	SARS-CoV-2 Spike Protein	No		

2.1. Nanobody-Antigen Complex Structures

2.1.1. Caplacizumab-von Willebrand Factor (vWF) Complex

There may be many different factors (genetic factors, environmental factors, and external antigens are some of these factors) underlying the development of diseases. One of them is blood clotting problems caused by von Willebrand Factor (vWF) deficiency, which is an autoimmune disorder.

The von Willebrand Disease (vWD) is a disease that occurs because of a mutation of the gene that controls the von Willebrand Factor (vWF), a glycoprotein structure that is effective in blood clotting, or the deficiency of this factor. Due to the deficiency in this protein, platelet cells, effective in blood clotting, cannot stick together or connect to blood vessels. This condition is medically called Thrombotic thrombocytopenic purpura (TTP). In this disease, prolonged clotting time or sometimes uncontrolled bleeding can be observed in case of an injury. Large bruises on the body, bleeding gums, heavy and prolonged bleeding from cuts and injuries, bleeding during and after birth, and frequent and prolonged nose bleeding are among the symptoms of this disease. The disease can be

detected by blood clotting tests. Caplacizumab, an FDA-approved drug, is a nanobody drug used against vWF in treating TTP (Lee, et al. 2021).

Caplacizumab works by targeting the A1 domain of the von Willebrand Factor (vWF), inhibiting the interaction between platelets, and stopping microthrombosis. Caplacizumab, which shows highly specific binding to von Willebrand Factor (vWF), does not prevent interaction with other coagulation elements. It does not cross-react with other blood elements in the environment, erythrocytes, and platelets. Thanks to this high specificity, off-target effects were not observed in pre- and post-clinical studies (Sargentini-Maier, et al. 2019).

Figure 6 shows the interaction sites of Caplacizumab and von Willebrand Factor (vWF). Similarly, Table 2 shows the nanobody and antigen interaction sites retrieved from the PDBsum web server. In Figure 7, the residues in the interaction regions given on the PDBsum web server are shown on the complex structure.

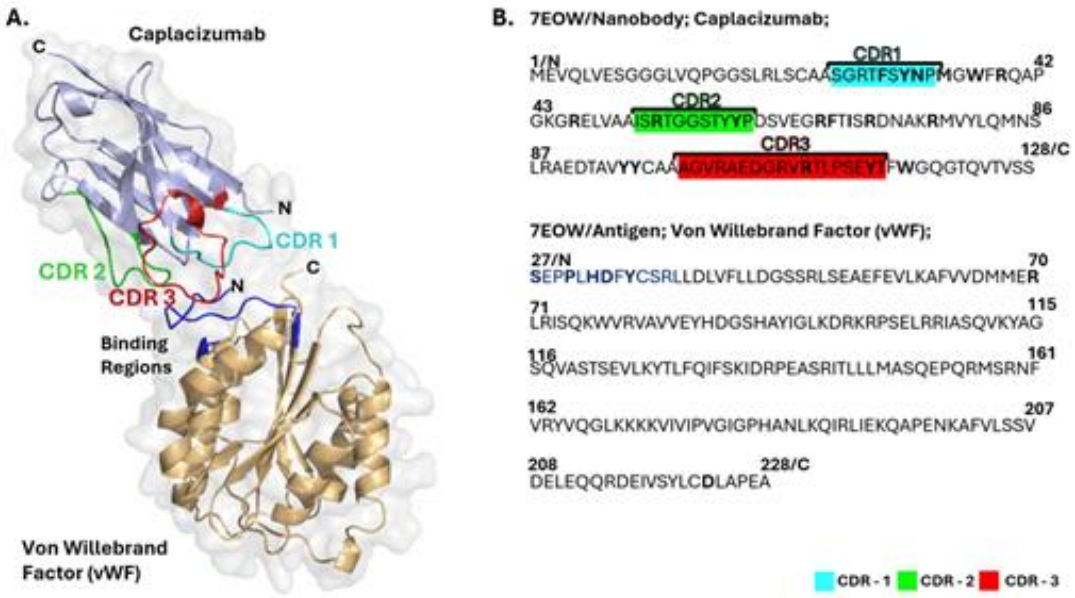


Figure 6. Crystal structure of Caplacizumab and von Willebrand Factor (vWF) complex (PDB ID: 7EOW). A: Representation of the Caplacizumab in light blue color and vWF interaction interface in gold color (The interaction region in the antigen is dark blue.) B: Primary structures of Caplacizumab and von vWF. (Boldface expressions in the sequence indicate the amino acids in the interaction regions) (Lee, et al. 2021; Berman, et al. 2000).

Table 2. Table of the amino acids that interact between the Caplacizumab nanobody and von Willebrand Factor (vWF) and the bond type between them (PDBsum).

Caplacizumab-von Willebrand Factor (vWF) (PDB ID: 7EOW)		
Caplacizumab (Nb)	**Bond Types**	**von Willebrand Factor (vWF) (Ag)**
Residue names & numbers		**Residue names & numbers**
Arg54	Hydrogen Bond	Asp33
Thr55		
Gly107		
Glu105		Tyr35
		Arg38
Arg78	Non–bonded contact (NBC)	Ser27
Asn75		Pro29
Arg54		Pro30
Tyr32		
Thr55		His32
Arg54		
Thr55		Asp33
Gly107		
Arg54		
Glu105		
Ser53		
Glu105		Phe34
Arg54		Tyr35
Glu105		
Ser31		
Tyr32		
Ser31		Ser37
Arg103		Arg38
Glu105		
Phe30		Arg72
Arg28		
Arg103		Asp223
Glu105	Salt Bridge	Arg38
Arg103		Asp223

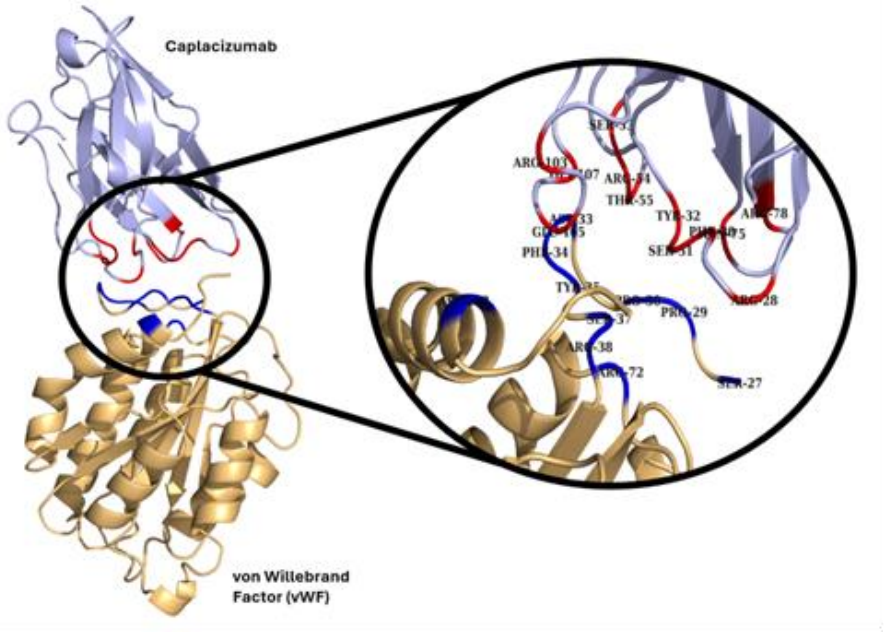


Figure 7. Names of amino acids interacting between Caplacizumab nanobody and von Willebrand Factor (vWF) and representation of the interaction site (PDBsum).

2.1.2. Nanobody 87-NTCP Complex

The liver is a vital organ that produces bile by removing bile salts from the blood, absorbing fat-based nutrients, and excreting metabolites and drugs. Bile salts, which enable the fats that come with food to become smaller fat drops, are taken into the cells by polypeptide (NCTP) receptors carried together with Na^{+-} taurocholate in the liver. This receptor undergoes a conformational change to transport bile salts into the cell, enabling transmembrane passage, as well as providing the emergence of the key determining residues for the binding of hepatitis B and D (HBV/HDV) viruses to the outside of the cell, and has a fundamental role in liver pathology as the cellular entry receptor of these viruses. Chronic hepatitis B (HBV) infection is an important cause of hepatocellular carcinoma and liver cirrhosis, affecting approximately 250 million people globally (Goutam, et al. 2022).

Related to this issue, Goutam and his colleagues studied a nanobody that would inhibit the entry mechanism of hepatitis B and D viruses into the liver cells. Scientists

have worked with nanobody 87 (Nb87) and 91 (Nb91) structures. These studied structures recognize the epitope regions of hepatitis B and D virus (HBV and HDV) antigens and prevent the entry of these antigens into the cell by closing the open passage pore in NTCP. In this study, it was shown that nanobody 87 (Nb87) has a stronger inhibitory effect than nanobody 91 (Nb91). The binding of nanobody 87 (Nb87) to the receptor caused allosteric changes in inward conformations. Researchers noted that thanks to these conformational changes in the structure, the entry of the virus antigen into the cell was prevented. However, researchers also noted that binding nanobodies to the structure did not prevent bile salt entry into the cell. Nanobody 87 has an inhibitory effect by selectively controlling the entry of hepatitis B and D antigen fragments into the cell (Goutam, et al. 2022). Figure 8 shows nanobody 87 (Nb87) and NTCP (Ag) interaction sites. Similarly, Table 3 shows the nanobody and antigen interaction sites retrieved from the PDBsum web server. In Figure 9, the residues in the interaction regions given on the PDBsum web server are shown on the complex structure.

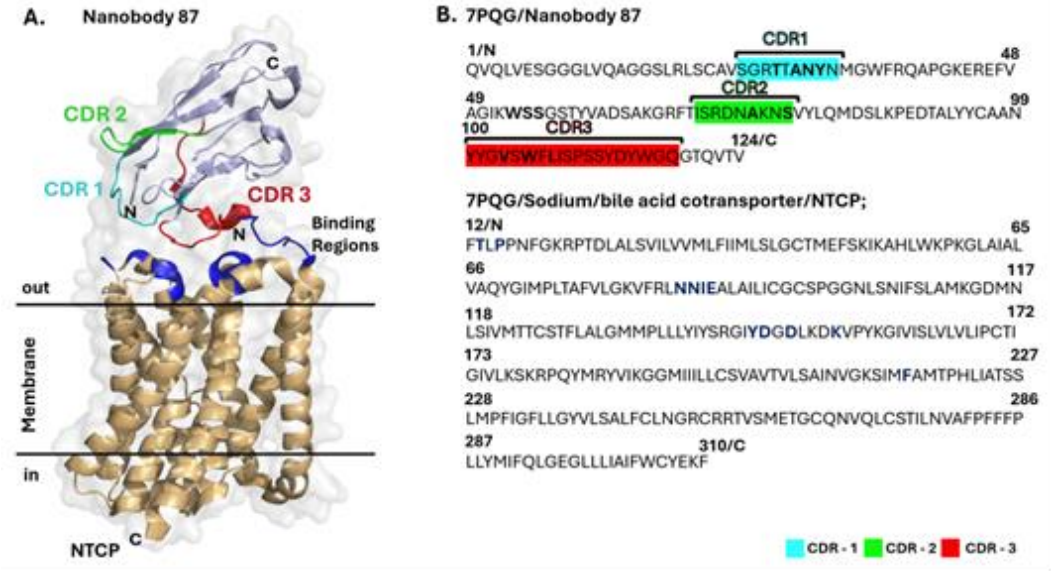


Figure 8. Representation of the interaction sites of Nanobody 87 and the NTCP receptor (PDB ID: 7PQG). A: Nanobody 87 interaction interface in light Blue and NTCP in gold (The NTCP interaction regions are dark blue.) B: Representation of the sequence of Nanobody 87 and NTCP. (Boldface expressions in the sequence indicate the amino acids in the interaction regions) (Goutam, et al. 2022; Berman, et al. 2000).

Table 3. Table of the amino acids that interact between the nanobody 87 (Nb) and NTCP (Ag) and the bond type between them (PDBsum).

Nanobody 87-NTCP (PDB ID: 7PQG)		
Nanobody 87 (Nb)	**Bond Types**	**NTCP (Ag)**
Residue names & numbers		**Residue names & number**
Asn31	Hydrogen Bond	Asn86
Tyr100		Asn87
Gly102		
Ser109		Glu277
Asn31	Non-bonded Contacts (NBC)	Asn86
		Asn87
		Tyr146
Tyr59		Thr13
		Pro15
Tyr100		Asn87
		Asp149
Tyr101		Asn87
		Ile88
		Lys153
Gly102		Asn87
		Ile88
		Asn86
		Glu89
Val103		Asn86
		Ile88
Trp105		Phe216
		Glu277
		Pro281
		Val278
		Ile279
		Gly280
Phe106		Pro15
		Thr13
Leu107		Glu277
Ile108		Pro15
Ser109		Glu277
Arg27	Salt Bridge	Asp147

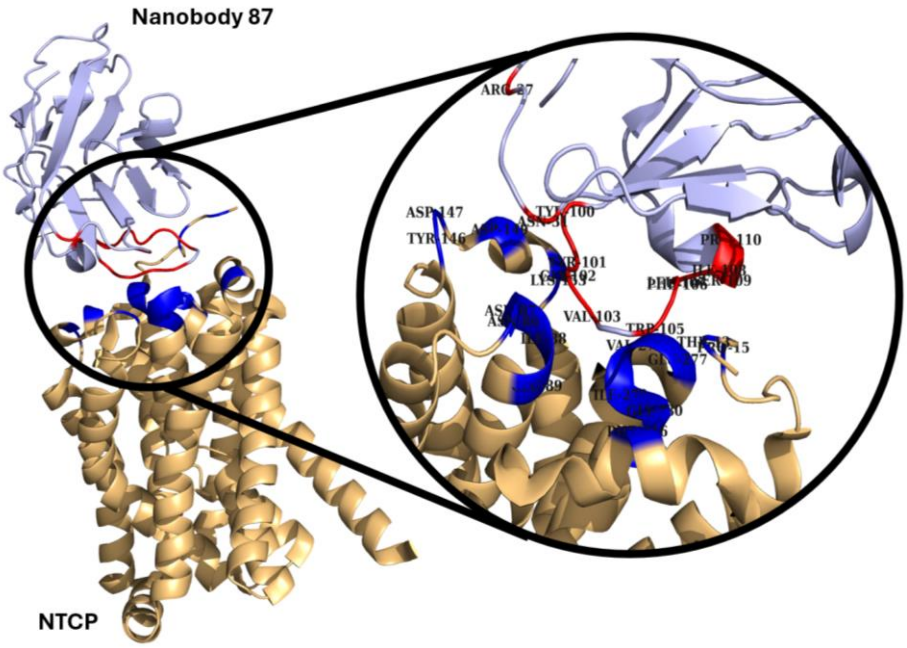


Figure 9. Names of amino acids interacting between Nanobody 87 and NTCP and representation of the interaction site (PDBsum).

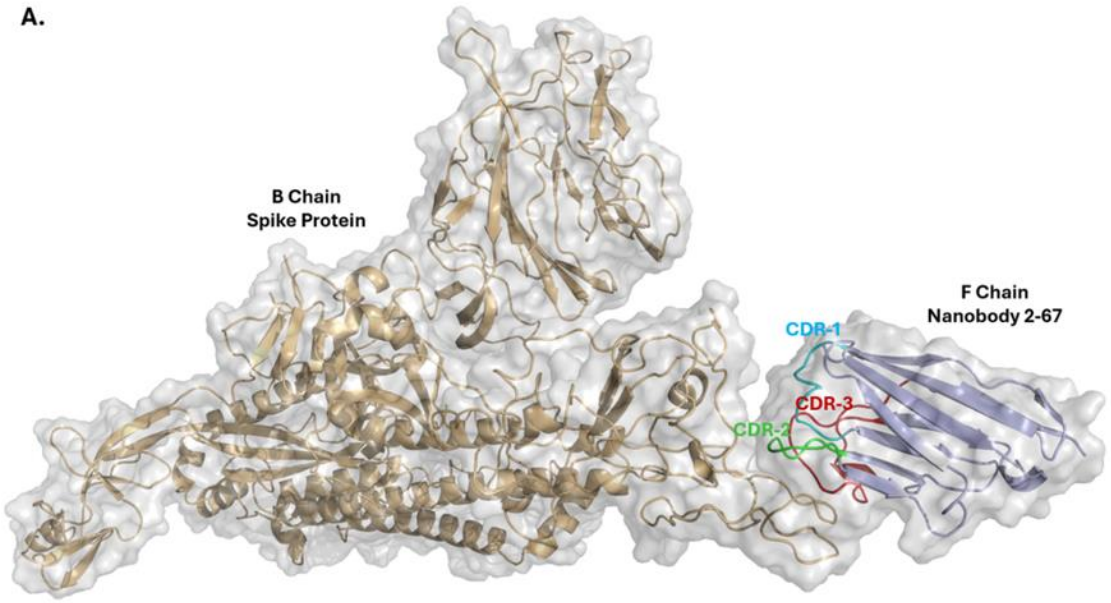
2.1.3. Nanobody 2-67 – SARS-CoV-2 Spike Protein Complex

Recently, a pandemic has occurred all over the world, characterized by cough and high fever that affects the respiratory tract in people and life in many ways. The family of viruses related to SARS-CoV and SARS-CoV-2, which caused this pandemic, are called coronaviruses (sarbecovirus). This virus family has attracted the attention of researchers due to its high genetic diversity, high recombination abilities, and rapid proliferation. Researchers are conducting various studies on this subject to prevent future dangers. The research conducted mainly targets the spike protein of the SARS-CoV-2 virus and aims to understand the mechanism. Nanobodies have become the focus of researchers in this field due to their unique properties. One of them, which was conducted by Xiang and colleagues, has focused on nanobodies that potently neutralize the virus and can bind to the receptor binding domain (RBD) with high affinity (Xiang, et al. 2022).

Xiang and colleagues examined pan-sarbecovirus nanobodies (psNbs), which bind strongly and specifically with high affinity to all known sarbecovirus species. As a

result of this review, they listed 13 pan-sarbecovirus nanobodies (psNbs) that can bind to the spike protein or receptor binding domain (RBD) of the SARS-CoV-2 virus. These listed nanobodies target five different epitope regions that are small, flat, and flexible, containing more than 75% of the conserved surface residues in the RBD domain. Due to the examinations, the listed nanobodies were structurally divided into 4 different classes. Nanobodies divided into groups are distinguished from each other by their various properties. For example, the average neutralization potential of group A (Nb 2-67) is more potent than that of group C (Nb 1-22, Nb 2-10, Nb 2-45). Group B nanobodies show relatively higher potency. Nb 2-62, one of the nanobodies examined, binds tightly to the RBD but is a weaker neutralizer than other nanobodies because it cannot bind to the spike protein before the effect begins. Although Nb 2-63 and Nb 2-28 in Group D are in the same group and neutralize the virus at the same rate, they are nanobodies with different properties (Xiang, et al. 2022).

As a result of the research, the total structure showing the interaction areas with four different nanobodies and spike proteins taken from the PDBsum web server is shown in Figure 10. The interaction zones taken from this PDBsum web server are given in Table 4. In Figure 11, the interaction zones taken from the PDBsum web server are shown on the complex structure.



B. 8CYA/ Antigen; SARS-CoV-2 Spike Protein

15 CVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFAS 94
95 TEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQP 174
175 FLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSS 256
257 GWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPF 338
339 GEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIA 419
420 DYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQ 498
499 PTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRD 578
579 PQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNS 659
660 YECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYIC 743
744 GDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNK 825
826 VTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF 906
907 NGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPP 987
988 EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTY 1067
1068 VPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDS 1147

8CYA/Nanobody; Nanobody 2-67

1 EVQLVESGGGLVQTGGSLRLSCALSGYTFSIFPTAWFRQAPGKEREFVAGIRWNGSTRDYTEYADFVKGRFTISRDNAKN 80
81 MVYLQMISLKPEDTALYYCAASDGVIDGTNANAYRYWGQGTQVTVSS 127

Figure 10. Representation of Nanobody 2-67 bound to the spike protein of the SARS-CoV-2 virus (PDB ID: 8CYA). A: spike protein is gold colored and nanobody 2-67 is lightblue colored (CDR-1 is blue, CDR-2 is green, CDR-3 is red, Spike protein interaction region is in dark blue). B: Representation of sequences of spike protein and nanobody 2-67. (Boldface expressions in the sequence indicate the amino acids in the interaction regions) (Xiang, et al. 2022; Berman, et al. 2000).

Table 4. Table of the amino acids that interact between the nanobody 2-67 the spike protein of the SARS-CoV-2 virus (Ag) the bond type between them (PDBsum).

Nanobody 2-67 - The spike protein of the SARS-CoV-2 virus		
Nanobody 2-67 (Nb)	**Bond Types**	**Spike Protein (Ag)**
Residue names & numbers		**Residue names & numbers**
Asp107	Hydrogen Bond	Gly476
Asn54		Val483
Arg52		Gly485
Asp107		Asn487
Ile106		Tyr489
Gly108		
Val105	Non-bonded Contacts (NBC)	Phe456
Asp107		Ala475
		Gly476
		Ser477
Asn54		Val483
Arg52		Glu484
Asn54		
Tyr60		
Thr57		
Arg52		Glu485
Tyr60		
Tyr60		Phe486
Glu62		
Arg52		Asn487
Glu62		
Asp107		
Gly108		
Arg52		Cys488
Val105		Tyr489
Arg52		
Ile106		
Asp107		
Gly108		

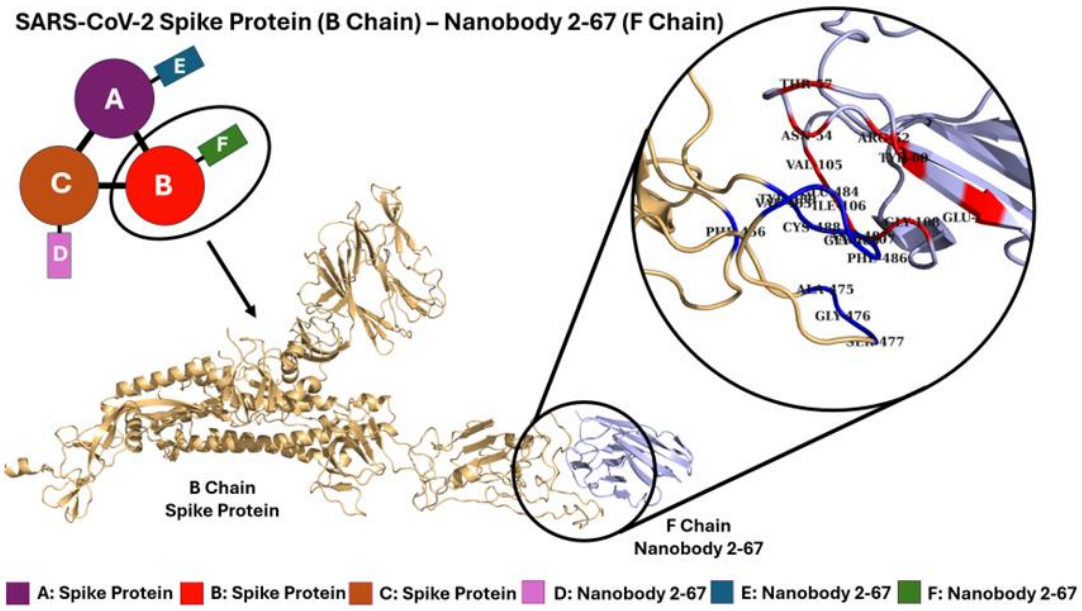


Figure 11. Names of amino acids interacting between Nanobody 2-67 and spike protein of the SARS-CoV-2 Virus (Ag) and representation of the interaction site (PDBsum).

Various computational methods enable the analysis and interpretation of the interaction regions of nanobody–antigen complexes. These methods also help understand the appearance of molecules in space. Elastic Network Models (ENM), a coarse-grained normal mode analysis, is one of the methods used to explore protein dynamics and allostery. The following section explains the ENM methodology in detail.

2.2. Elastic Network Model (ENM)

Proteins are essential in biological systems and play significant roles in cellular activities, therefore, they have been the subject of many studies due to their three-dimensional, flexible, and dynamically mobile structures. Molecular modeling applications, an important tool for understanding the movements and dynamics of proteins exhibiting allosteric behavior, form the basis of many future studies. In this context, the Elastic Network Model (ENM) helps researchers make sense of the

movements of proteins with the simple network model it creates (López-Blanco and Chacón 2016; Bahar, Atilgan and Erman 1997).

Elastic Network Model (ENM) provides information about possible molecule binding sites, allosteric behaviors, movements, functions, and internal dynamics of molecules in the modeling of biological molecules with a wide variety of data (such as atomic coordinates, B factor) received from RCSB (Protein Data Bank). It is a computational molecular modeling method (López-Blanco and Chacón 2016; Bahar, Atilgan and Erman 1997).

Proteins consist of amino acids, each with a carbon-alpha atom. The working principle of this analysis method, which is based on the carbon-alpha atom of proteins and includes two different applications [Anisotropic Network Model (ANM) and Gaussian Network Model (GNM)], is as follows. A cutoff value is determined to be applied to each carbon-alpha atom, with the central position being the carbon-alpha of the amino acids. This determined cutoff value refers to the radius of the spheres whose centers are carbon-alphas. Spheres with radius cutoff values are drawn separately for each carbon-alpha atom. It is accepted that all the atoms inside these drawn spheres interact with the carbon-alpha atom at the center of the sphere. All atoms interacting with each other are shown with a network model (Figure 12) (López-Blanco and Chacón 2016; Atilgan, et al. 2001).

The Gaussian Network Model (GNM), a 1D ENM, operates based on the carbon-alpha atoms of the amino acids that make up proteins. This analysis method defines the relationship of amino acids in the protein structure with other carbon-alpha atoms located at a certain cutoff distance from the carbon-alpha atoms with springs. In this method, which is considered as if there is a spring (representation) between each carbon-alpha atom in contact, it is concluded that the proteins make a harmonic movement with the potential energy accumulated in the spring between the atoms as a result of the movements of the proteins. In other words, in a structure containing N residues, there are $(N - 1)$ modes (motions) that can be obtained from. Each harmonic mode obtained from GNM has a frequency. The change in frequency depending on the stable state of the examined structure provides information about the internal dynamics of the molecule, the flexibility of the structure, and molecular motion (Bahar, Atilgan and Erman 1997; Atilgan, et al. 2001).

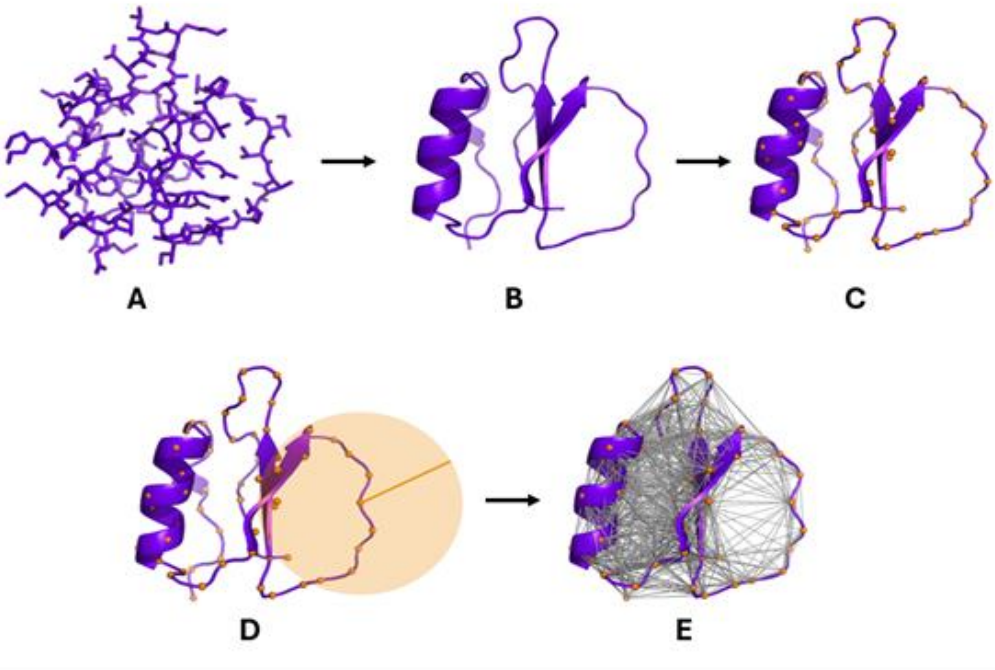


Figure 12. The details of the Elastic Network Model (ENM) process steps are as follows. A: Protein structure consisting of stick-like amino acids, B: Protein structure consisting of alpha helix and β-sheets, C: Displaying/selecting the carbon-alphas for each amino acid that makes up the protein (Orange dots represent carbon-alphas) D: Drawing spheres at the distance value determined for each selected carbon-alpha. E: Creation of the network structure of atoms located in the sphere whose center position is carbon-alphas and participating in the interaction (Animation Credit: Assoc. Dr. Özge Kurkcuoglu Levitas).

Another analysis method based on ENM is called Essential Site Scanning Analysis (ESSA), which helps to compare protein dynamics between pre-and post-binding states. The next section explains the ESSA method in detail.

2.3. Essential Site Scanning Analysis (ESSA)

The Essential Site Scanning Analysis (ESSA) method, based on the Elastic Network Model (ENM), detects regions in proteins essential for biological activity.

Examples of these regions in proteins that are essential for biological activity are responsible for catalytic activity, potential molecule binding regions, and allosteric regions. One of the main purposes of this application, which analyses proteins based on their amino acid sequences and three-dimensional structures, is to determine the possible functional regions of proteins (Kaynak, Bahar and Doruker 2020).

In ESSA, like in the elastic network model (ENM), spheres with a radius at a certain cutoff value are drawn by considering the carbon-alpha atoms of amino acids in a protein structure as the center. A network model is created by assuming interaction with all atoms within the cutoff radius value of each carbon-alpha-centered sphere drawn. However, unlike ENM, ESSA draws spheres with radius at the same cutoff value separately for each side chain to consider the side chains of amino acids (Figure 13). In this way, ESSA, which looks at the effect of side chains on protein internal dynamics, is an analysis method suitable for large systems and can quickly analyze the molecule in bound and unbound states (Kaynak, Bahar and Doruker 2020).

In this analysis method, a z-score is calculated to determine how likely each amino acid is to be a binding site. Thanks to this scoring process, the contribution of the position of amino acids to functional activity can be calculated. Thus, potential binding regions in proteins are determined. First, the frequency shift for each residue is calculated using Eq. 2.1 (Kaynak, Bahar and Doruker 2020).

$$\Delta\lambda_k^{(i)}(\%) = \frac{(\lambda_k^{(i)} - \lambda_k)}{\lambda_k} \times 100 \quad \text{(Equation 2.1.)}$$

Here, *i* is the selected residue, *k* is the mode, lambda *(λ)* is the frequency, λ_k represents the frequency when the carbon-alpha atoms considered and $\lambda_k^{(i)}$ represents the frequency when the side chains for each carbon-alpha atom considered one by one, and delta *(Δ)* is the difference process. Here, the movement of each residue in each mode was examined and the change in eigenvalues was evaluated as a % value (Kaynak, Bahar and Doruker 2020). Thus, the calculated percentage change in the eigenvalues in the k^{th} mode or *k* modes is a response corresponding to the crowding around the selected residue *i*.

Changing the order of modes that proteins display or keeping them the same causes differences with respect to the reference and resulting network model. Mode data

has been reordered to achieve maximum efficiency. A scoring is defined with these listed modes. The z-scores obtained as a result of Eq. 2.2 given below (Kaynak, Bahar and Doruker 2020).

$$z_i = \frac{<\Delta\lambda_{1-10}^{(i)}> - \mu}{\sigma} \text{ (Equation 2.2.)}$$

Here, to increase the robustness of the results, defined an effective shift in the global mode frequencies for each residue in the $< \Delta\lambda_{1-10}^{(i)} >$ over the first ten modes [if protein movements are better observed ($1 \leq k \leq 10$)]. As a result of this calculation, a z-score is assigned. This z-score describes the dynamics of molecule binding sites close to residue *i*. In addition, a color scale is used to better visualize the z-score values obtained. In addition, the reason for choosing the first ten modes here they are mostly related to the function of the proteins (Kaynak, Bahar and Doruker 2020).

The PyMOL application was used to display the results, with z-score values fixed between 1 and 7 for each residue (Figure 14). In order to better interpret the results, colors were used, and the results were colored according to the B-factor in the PyMOL application. After this rescoring, predicted essential residues with a z-score value above 5 appear in red and orange as shown in Figure 14. These regions represent potential binding sites.

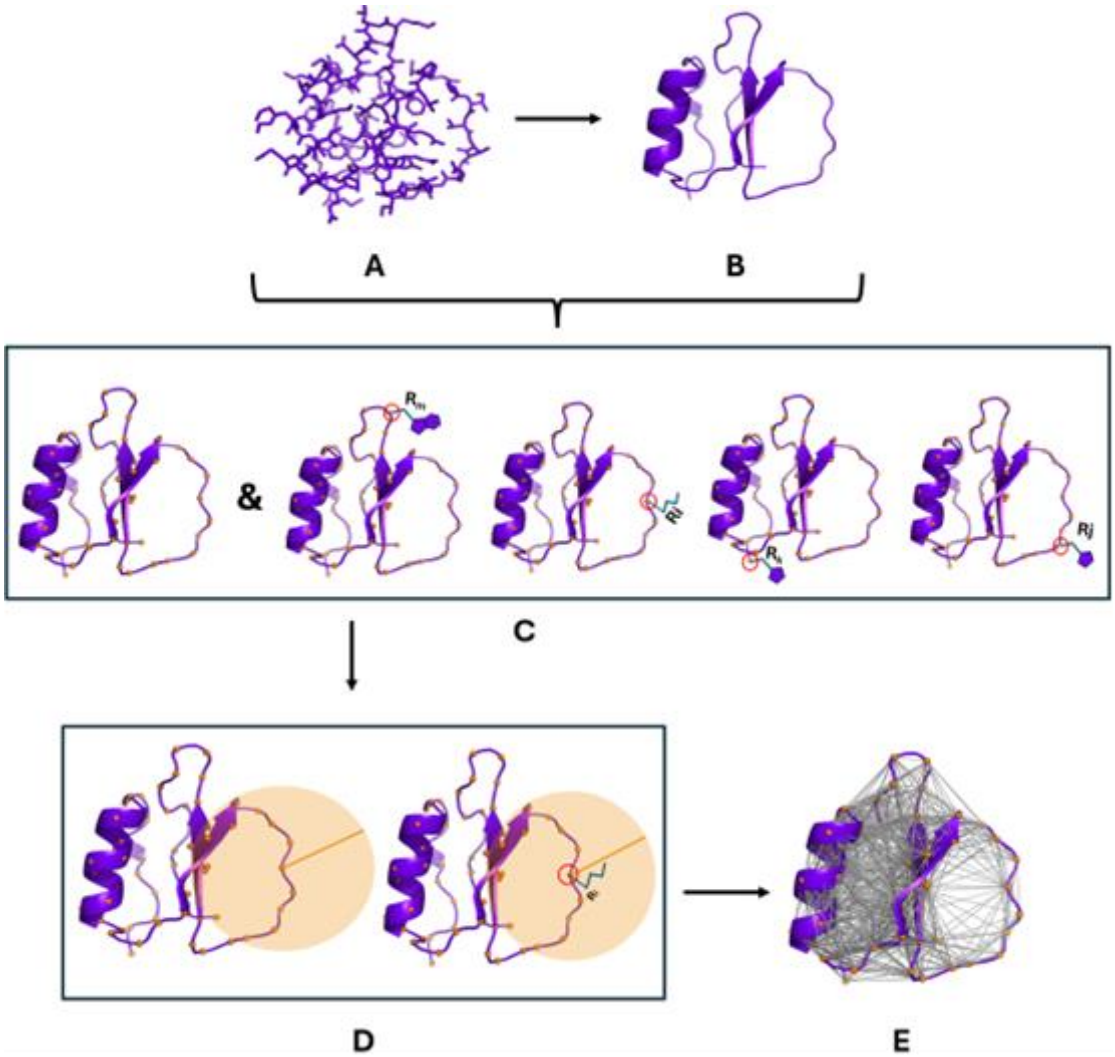


Figure 13. The details of the Essential Site Scanning Analysis (ESSA) process steps are as follows. A: Protein structure consisting of stick-like amino acids, B: Protein structure consisting of alpha helix and β-sheets, C: Showing/selecting the carbon-alphas and side chain of different amino acids that make up the protein (Orange dots represent carbon alphas) D: Drawing spheres at the distance value determined for each selected carbon-alpha and side chain, E: Creation of the network structure of the atoms located within the drawn sphere and involved in the interaction (Animation Credit: Assoc. Dr. Özge Kurkcuoglu Levitas and Assist Prof. Arzu Uyar).

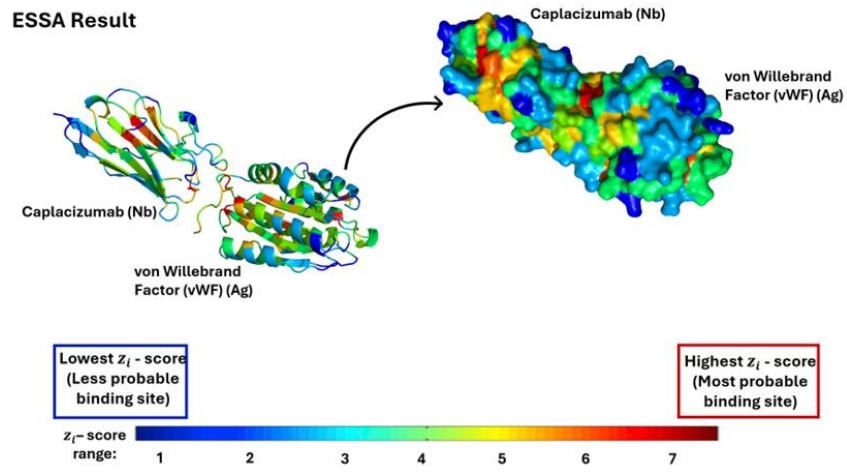


Figure 14. Representation of the color scale in the data obtained as a result of ESSA performed on the Caplacizumab and von Willebrand Factor (vWF) complex (PDB ID: 7EOW) (Lee, et al. 2021) structure. Regions expressed z-score value above 5 appear (in red or orange color) indicate potential binding sites (A: representation of the structure in cartoon form, B: surface form).

The default parameter values in ESSA are as follows: the mode number is 10, and the cutoff value is 10 Å. In this thesis, three different cutoff values (7.3 Å, 10 Å, and 13 Å) are tested, and the total number of modes is taken as 20 in ESSA calculations for the studied nanobody-antigen systems. The results obtained using each cutoff value are shown separately. In addition, combined results were reported and named “combined-ESSA". Here, a result is obtained for each selected cutoff distance value in the results obtained separately from three separate cutoff distance values, at whichever cutoff value the highest z-score result is obtained for each residue, the corresponding color for that residue at that cutoff value appears in the combined-ESSA result. For example, for the arginine residue located at position 73 in the Caplacizumab nanobody; Arg73, ESSA cutoff distance: 7.3 Å, color: red, Arg73, ESSA cutoff distance: 10 Å, color: yellow, Arg73, ESSA cutoff distance: 13 Å, color: green. Since the color with the highest z-score value among the results is red, Arg73 appears in red in the combined-ESSA result. Considering all this information, in this thesis study, three complex structures with PDB IDs 7EOW, 7PQG, and 8CYJ and the components forming these complex structures were evaluated separately in the result and discussion section.

CHAPTER 3

RESULTS AND DISCUSSIONS

In science, knowledge progresses cumulatively. In computational sciences, the currently known information has been reached through the methods developed and theories presented by many scientists from the past to the present. This situation subjects scientific studies to a multifaceted evaluation. For this reason, thinking and interpreting from multiple perspectives is an important criterion when evaluating the results of a research or study. The data obtained from the analyses carried out in this thesis study were interpreted based on this criterion.

In this thesis, Essential Site Scanning Analysis (ESSA) was applied to the nanobody and antigen structures that form three complex structures and the complex structures themselves, and the results are evaluated in this chapter.

The main purpose of the method (ESSA) is to find interaction sites or predict potential binding sites in the examined structures. Therefore, the questions that need to be answered and the main purpose when starting this thesis are as follows: Were the binding site found even when the nanobody and antigen structures were not in the complex form? What changed when the nanobody and antigen structures formed a complex? In this thesis, which sought answers to these questions, the PyMOL application was used to evaluate the ESSA results and z-score values calculated in ESSA are used to color each structure in PyMOL. The residues with z-score values above 5 were considered essential for binding. Moreover, one of the basic criteria that plays a critical role in defining a specific interaction site in protein structure analysis is the cutoff value. In this thesis study, three different cutoff values (7.3 Å, 10 Å, and 13 Å) were used for each structure. Afterward, the results for each cutoff and combined-ESSA results are given, and the intersection between the interacting residues in complex structures obtained from PDBSum and ESSA results are emphasized in tables.

In addition, some expressions have been used in this thesis study to describe some residues. The first of these expressions is 'sequence neighbor'. Sequence neighbor was

used to identify residues with a z-score value above 5 that were up to 2 or 3 residues away from the positions of the CDR regions in the primary sequences of the structures. Another expression used is 'spatial neighbor. Those are the residues with a high z-score value positionally adjacent in space and determined by calculating the 6-7 Å (Cα-Cα distance) away from the CDR regions in structures.

3.1. Caplacizumab-von Willebrand Factor (vWF) Complex

3.1.1. Caplacizumab (Nanobody) Structure

Caplacizumab is an FDA-approved nanobody drug used in Thrombotic thrombocytopenic purpura (TTP), which occurs because of deficiency or mutation of the glycoprotein von Willebrand Factor (vWF). There are 128 amino acids in the crystal structure of this structure, which is in nanobody form (Lee, et al. 2021).

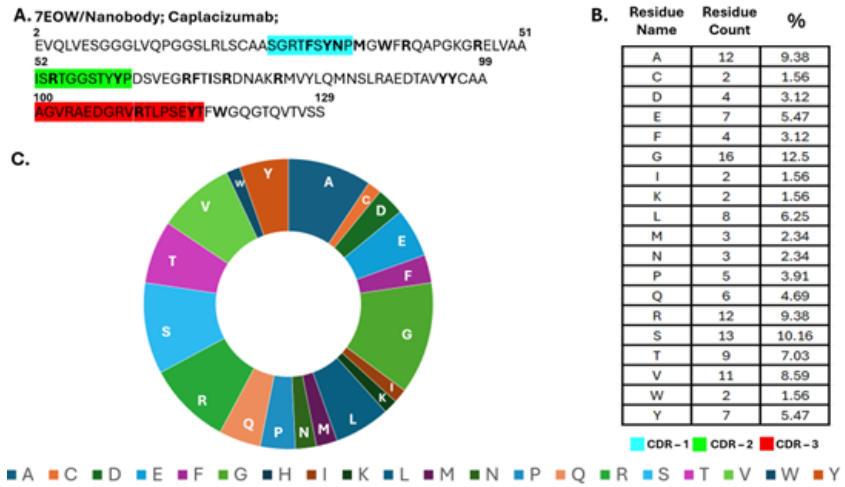


Figure 15. Representation of the primary sequence data regarding the Caplacizumab nanobody. A: Caplacizumab sequence [PDB ID: 7EOW, (Lee, et al. 2021)] B: Count and percentage values for each amino acid in Caplacizumab, C: Pie-chart representation of the amino acid distribution in Caplacizumab.

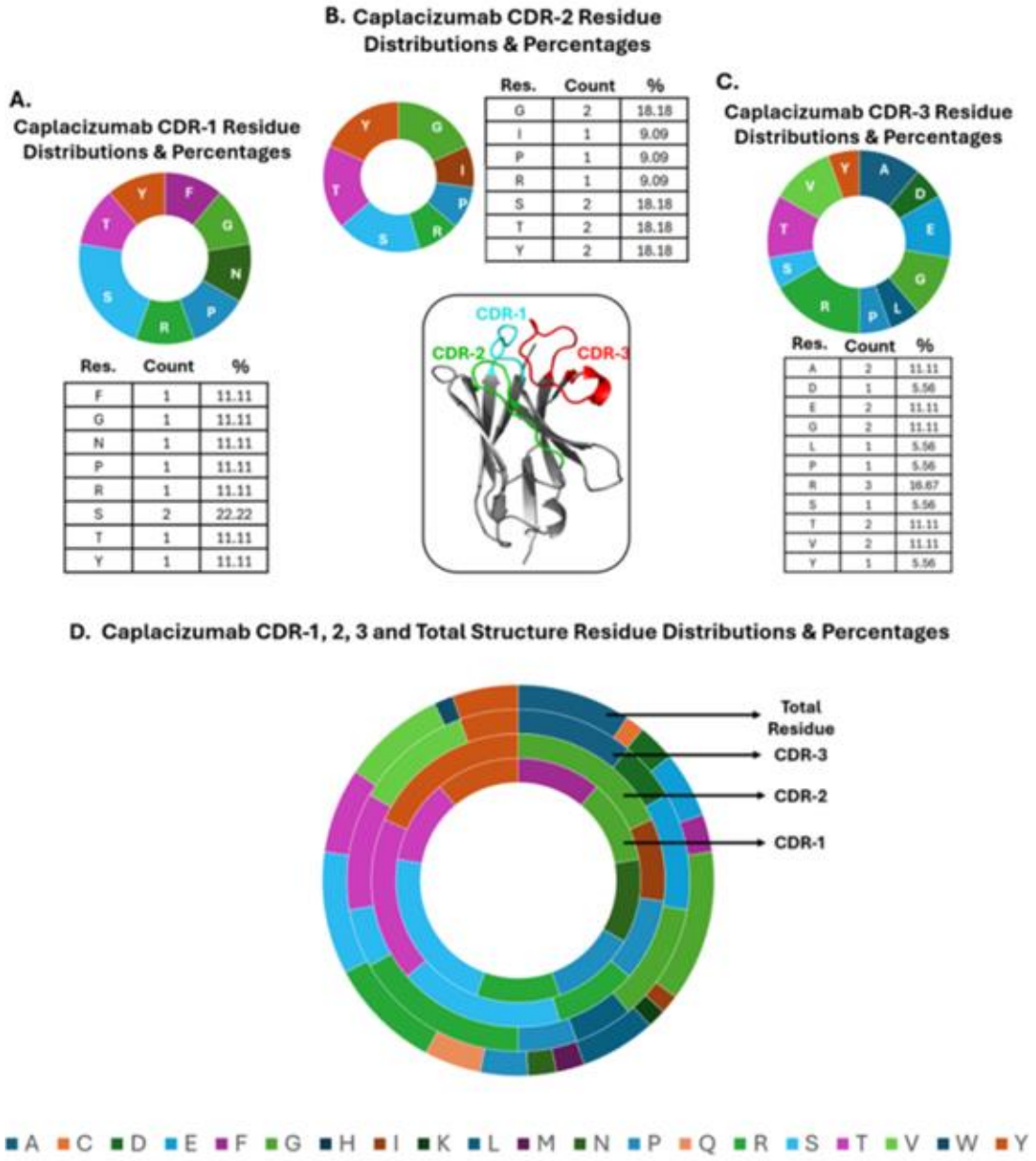


Figure 16. Representation of the primary sequence data of CDR regions located in the Caplacizumab nanobody [RCSB PDB ID: 7EOW, (Lee, et al. 2021)] and percentage distributions of amino acids in A: the CDR-1 region, B: the CDR-2 region, and C: the CDR-3 region. D: All amino acids in the Caplacizumab nanobody pie-chart of amino acids found in acids and CDR regions.

The sequence of this structure, which does not have any mutations, and the distribution of amino acid diversity and counts are shown in Figure 15. When Figure 15 is examined, looking at the residue percentage distribution in this structure containing

128 residues, serine and glycine residues are above 10%. In addition, the most abundant residues in the structure are glycine, serine, arginine, alanine, and valine, respectively.

The location of the CDR regions in the Caplacizumab nanobody structure, their amino acid counts, and the diversity of amino acids in these regions are shown in Figure 16. When Figure 16 is examined, each CDR region has different percentages and counts of residues. These residues, which are different in number and arrangement, distinguish these regions from each other and make them special. The number of residues in the CDR-3 region is higher than in other CDR regions. In Figure 16, the comparison of the total number of residues in the nanobody structure and the number of residues in the CDR regions is also shown as a pie chart.

The results from ESSA, which was performed by assuming the carbon-alpha atom and side chains as the central position for each residue and based on varying cutoff distance values (7.3 Å, 10 Å, and 13 Å), are colored according to the z-score values in PyMOL. The figures of the structures are presented from different angles (0°, 90°, 180°, 270°, top and bottom side) in Figure 17. Here, red and orange-colored regions represent potential binding and interaction sites. As can be seen in these results, some of the important regions show interaction points, CDR regions, or the regions located in the inner region (framework) of the structure.

However, in the ESSA result obtained using separate cutoff distance values are merged, thus a combined result was also obtained. These results from different angles (0°, 90°, 180°, 270°, top and bottom side) are given Figure 18. While creating the combined-ESSA result, for each residue, the highest z-score result among the three cutoff distance values was taken. For example, if only one residue appears red or orange at one cutoff distance value taken separately, this residue is shown as red or orange in the combined result. The representation of the red and orange residues on the structure in the analysis results for different cutoff distance values is shown in Figure 19. The table of red and orange residues as a result of the analysis performed on the Caplacizumab nanobody at different cutoff distance values (7.3 Å, 10 Å, and 13 Å) is shown in Table 5.

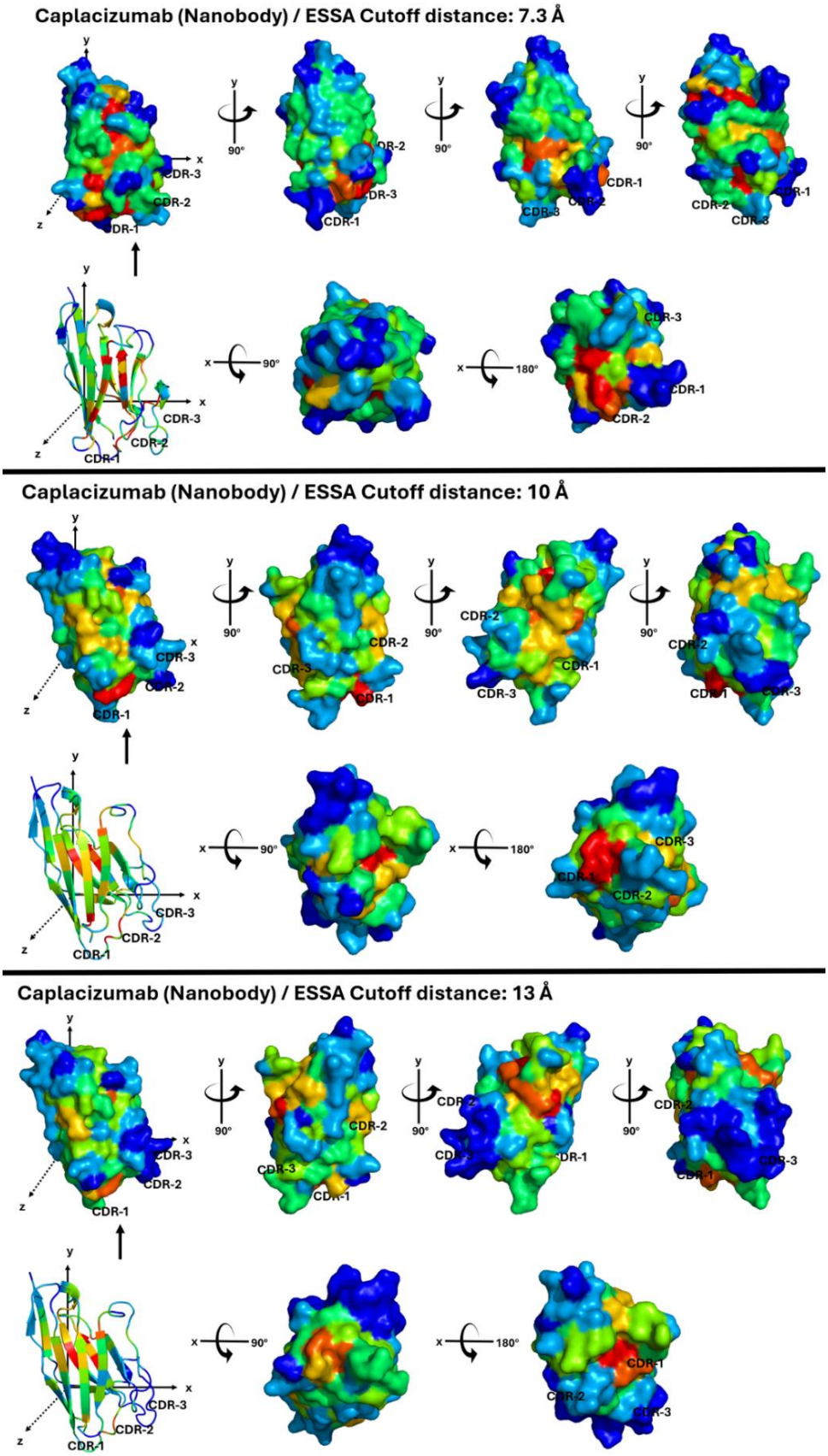


Figure 17. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the Caplacizumab nanobody.

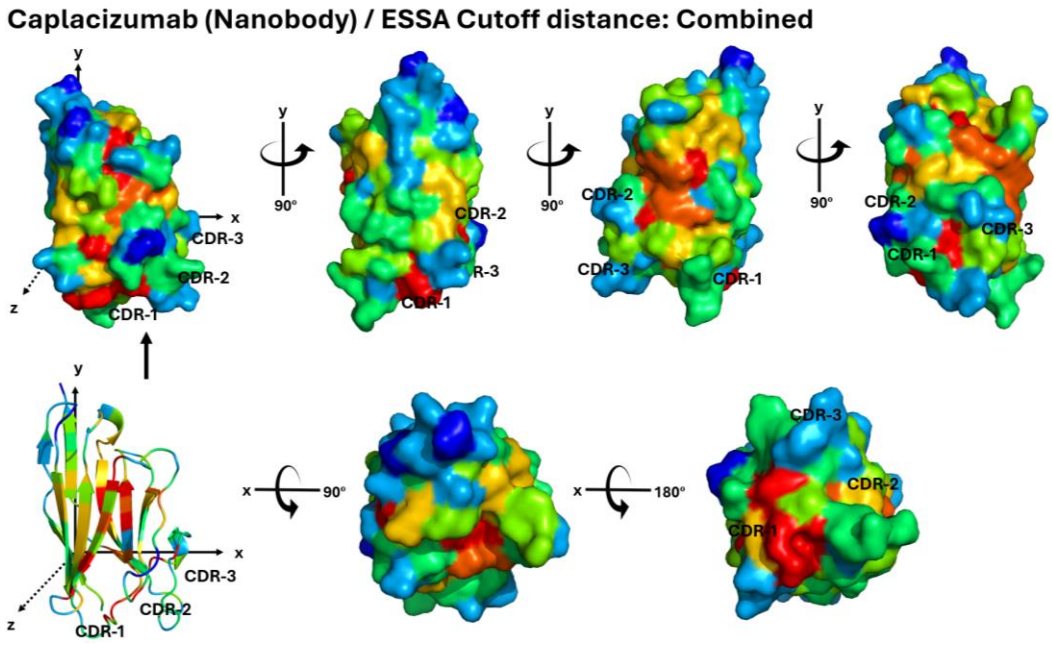


Figure 18. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the Caplacizumab nanobody.

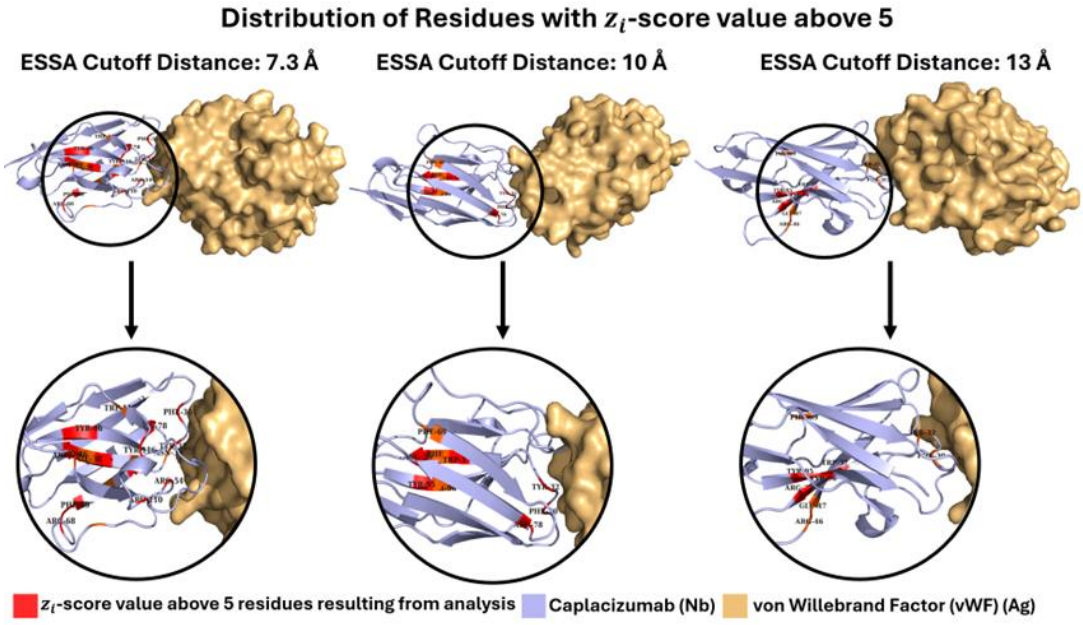


Figure 19. The most essential residues on the Caplacizumab structure determined using different cutoff distance values in ESSA.

Table 5. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the Caplacizumab nanobody (Only high z-score values are reported in the table. *: Residues specified as interaction sites in PDBsum are indicated with an asterisk).

ESSA, Cutoff Distance: 7.3 Å			**ESSA, Cutoff Distance: 10 Å**		**ESSA, Cutoff Distance: 13 Å**	
Caplacizumab		**vWF**	**Caplacizumab**	**vWF**	**Caplacizumab**	**vWF**
Phe30*(CDR1)		Arg72*	Phe30*(CDR1)	Arg72*	Phe30*(CDR1)	Arg72*
Try32*(CDR1)		Pro30*	Tyr32*(CDR1)	Pro30*	Tyr32*(CDR1)	Pro30*
		Tyr35*		Tyr35*		Tyr35*
Asn33	Met35		Trp37		Trp37	
Trp37	Arg39		Phe38		Arg39	
Arg54*(CDR2)		Asn33*	Arg39		Arg46	
		Pro30*	Phe69		Glu47	
		His32*	Arg78*	Ser27*	Phe69	
		Asn33*	Tyr95		Tyr95	
		Tyr35*	Tyr96		Tyr96	
Tyr61(CDR2)						
Arg68	Phe69					
Ile71	Arg73					
Arg78*		Ser27*				
Tyr95						
Arg110(CDR3)						
Tyr116(CDR3)						
Trp119						
Total: 17		**Total: 9**	**Total: 9**	**Total: 4**	**Total: 9**	**Total: 3**

As a result of the analysis performed on the Caplacizumab nanobody at different cutoff distance values (7.3 Å, 10 Å, and 13 Å), residues with z-score values above 5 are listed in Table 5. A total of 17 residues were found as a result of ESSA analysis, with the cutoff distance value being 7.3 Å. According to PDBsum data, the four residues marked

with asterisks in Table 5 (Phe30, Tyr32, Arg54, and Arg78) interact with antigen vWF. These residues are also in the CDR region (CDR-1 and CDR-2).

An equal number of residues (9 residues) were found in the analysis performed by taking the cutoff distance values as 10 Å and 13 Å. The number of residues found at these two cutoff distance values is almost half of the analysis result by taking the cutoff distance value as 7.3 Å. Tyr96 was seen as essential at both cutoff distance values. As a result of the analysis with a cutoff distance value of 10 Å, Phe38 and Arg73 residues were found, unlike the analysis result with a cutoff distance value of 13 Å. The residues detected as essential only in one cutoff distance value are as follows: Phe38 by taking the cutoff distance value as 10 Å; Arg46 and Glu47 by taking the cutoff distance value as 13 Å. When the data in Table 5 were compared, it was seen that Phe30 and Tyr32 residues, which are in CDR-1, were common.

The CDR-1 in the Caplacizumab nanobody is the region between Ser26 - Pro34 residues. The residues close to this region in sequence (referred to as “sequence neighbors”) are Met35 and Trp37. Interestingly, the z-score values of these residues are above 5 when the cutoff value is 7.3 Å (as well as cutoff 10 Å and 13 Å for Trp37). The CDR-2 region includes residues between Ile52 - Pro62. There have not found relatively close to this site that are considered “sequence neighbors.” In the Caplacizumab nanobody, the CDR-3 region starts with residue Ala100 and ends with residue Thr117. One residue (Tpr119) is sequence neighbors to this region. On the other hand, four residues, namely Arg68, Phe69, Arg73, and Arg78, are “positionally adjacent/spatial neighbors” to the CDR-2 in the Caplacizumab nanobody.

Additionally, in a study by Kelow S. P., *et al.* a loop region, which is close to the CDRs, was the main focus of the study is on antibody-antigen binding. Kelow S. P. and his colleagues, who referred to this loop region as the DE loop, explained that this region is of critical importance for antibodies (Kelow, Adolf-Bryfogle and Dunbrack 2020). The DE loop region is the region between the 70th and 75th residues in the sequence. Caplacizumab have a residue (Arg78) in a loop corresponding to the DE loop with a z-score value above 5 (ESSA cutoff distance: 7.3 Å). The Arg78 residue is adjacent to CDR-1 and CDR-2.

3.1.2. The von Willebrand Factor (vWF) (Antigen) Structure

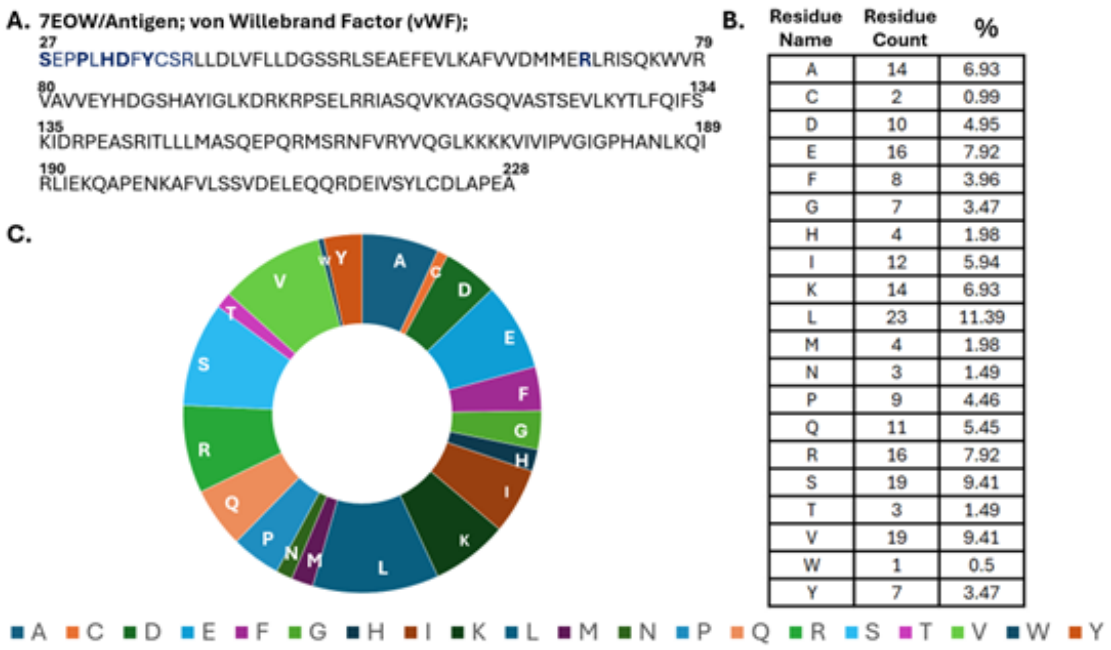
The von Willebrand Factor (vWF), which has a glycoprotein structure, is a molecule necessary to stabilize the circulatory factor that forms a platelet plug in the damaged area as a result of damage to the vessels in the organism. Due to this molecule's deficiency or mutation, Thrombotic thrombocytopenic purpura (TTP) and von Willebrand disease (vWD) occur. There are 194 amino acids in the crystal structure of vWF, which do not have any mutations (Lee, et al. 2021).

The sequence of this structure, and the distribution of amino acid diversity and counts are shown in Figure 20. When Figure 20 is examined, looking at the residue percentage distribution in this structure containing 194 residues, leucine residue is above 10%. In addition, the amino acids with a percentage of 5% or more in the structure are alanine, aspartate, glutamate, isoleucine, lysine, leucine, glutamine, arginine, serine, and valine.

The same analysis explained in the previous section was done for this structure. The results are given from various angles (0°, 90°, 180°, 270°, top and bottom side) in Figure 21. As seen in these results, some of the important residues are in the interaction region, while some are located outside the interaction region.

However, the analysis results obtained using separate cutoff distance values are combined, then a single combined-ESSA result that includes the highest value in three cutoff values for each amino acid was also obtained. These results are presented from different angles (0°, 90°, 180°, 270°, top and bottom side) in Figure 22. Residues with z-score values above 5 obtained from ESSA using different cutoff values are shown on the structure in Figure 23.

The residues with z-score value above 5 as a result of the ESSA analysis performed on the von Willebrand Factor (vWF) (Ag) at different cutoff distance values (7.3 Å, 10 Å, and 13 Å) are listed in Table 6. A total of 19 residues were found when the cutoff distance value was used as 7.3 Å. Among these residues, the two residues shown in asterisk in the table (His32 and Tyr35) correspond to the interaction region (also including CDR-1 and CDR-2) between the von Willebrand Factor (vWF) antigen and the Caplacizumab nanobody as given on the PDBsum web server.



Residue Name	Residue Count	%
A	14	6.93
C	2	0.99
D	10	4.95
E	16	7.92
F	8	3.96
G	7	3.47
H	4	1.98
I	12	5.94
K	14	6.93
L	23	11.39
M	4	1.98
N	3	1.49
P	9	4.46
Q	11	5.45
R	16	7.92
S	19	9.41
T	3	1.49
V	19	9.41
W	1	0.5
Y	7	3.47

Figure 20. Representation of the primary sequence data regarding the von Willebrand Factor (vWF) (Antigen). A: von Willebrand Factor (vWF) sequence [PDB ID: 7EOW, (Lee, et al. 2021)] B: Count and percentage values for each amino acid in vWF, C: Pie-chart representation of the amino acid distribution in vWF.

The total number of essential residues increased, and 24 residues were obtained when the cutoff distance value was used as 10 Å (Table 6). However, a smaller number of residues (16 in total) appeared essential when a higher cutoff value (13 Å) was used. The ESSA detected residues Tyr35 and Arg72 at the antigen-nanobody interaction site using both cutoff values. All three cutoff values point to the residue Tyr35, which interacts with both CDR-1 and CDR-2. In addition, cutoff 10 Å detected one more residue (Phe34), which interacts with the non-CDR region of the nanobody.

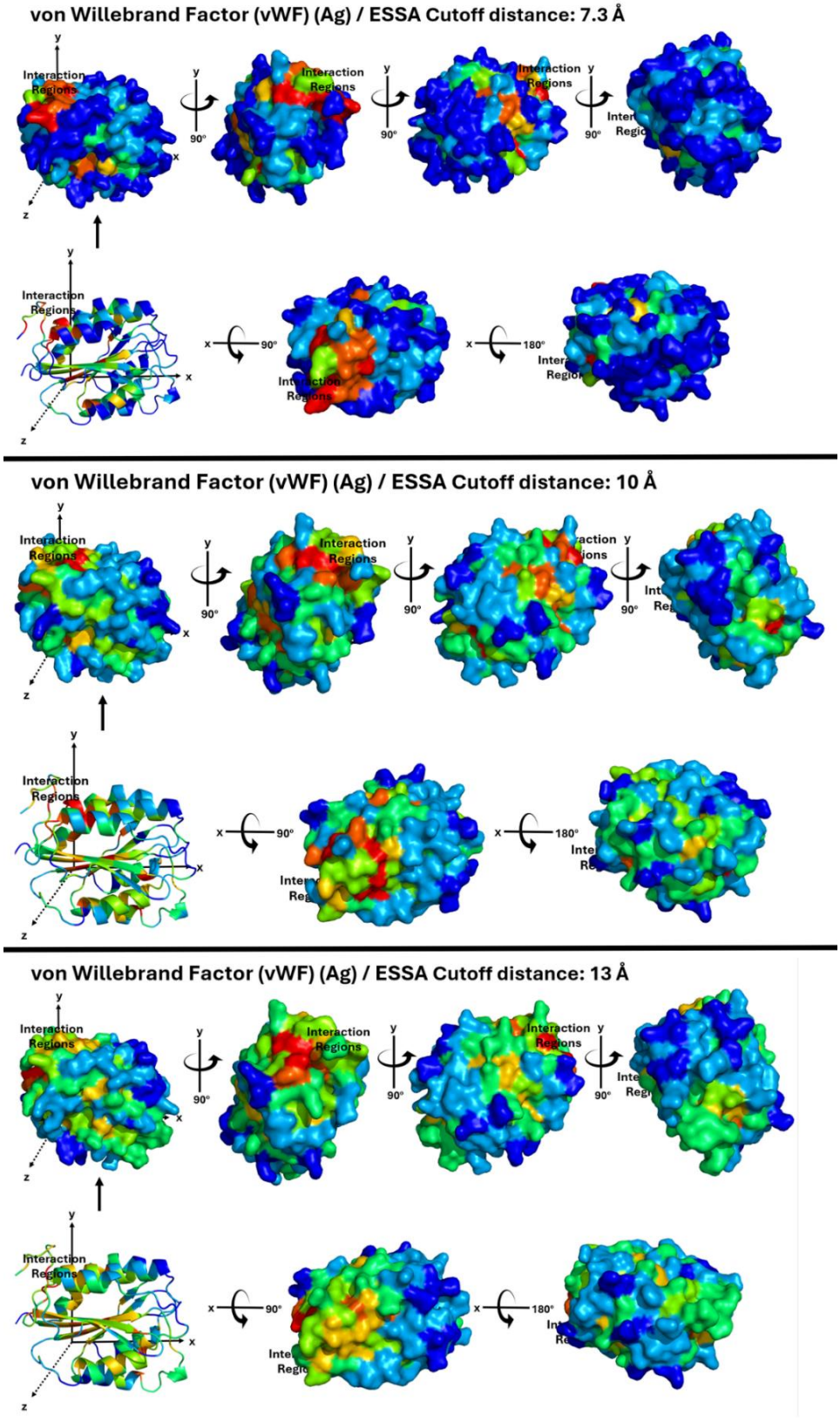


Figure 21. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the von Willebrand Factor (vWF) antigen.

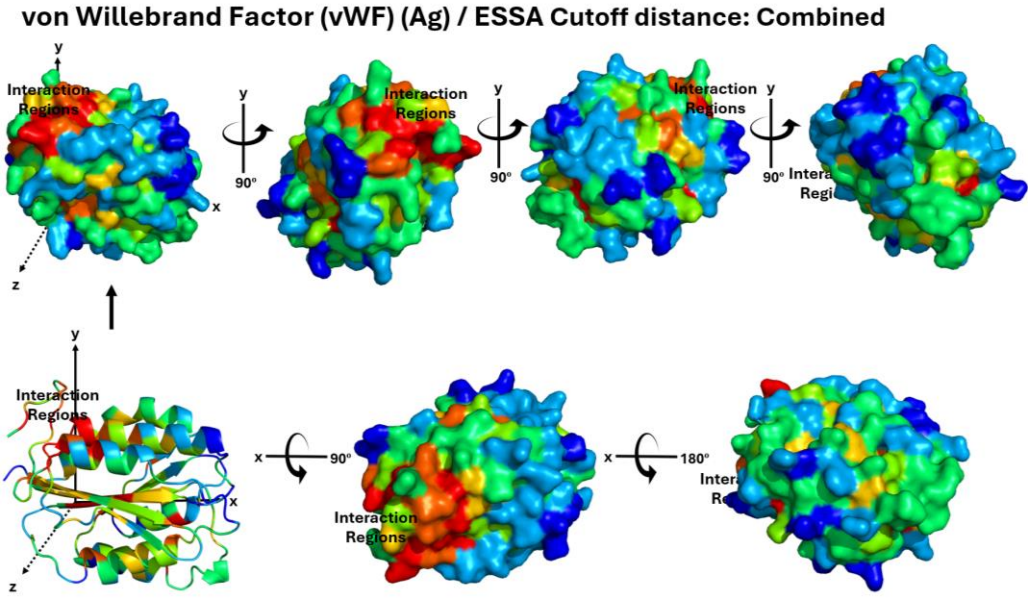


Figure 22. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the von Willebrand Factor (vWF) antigen.

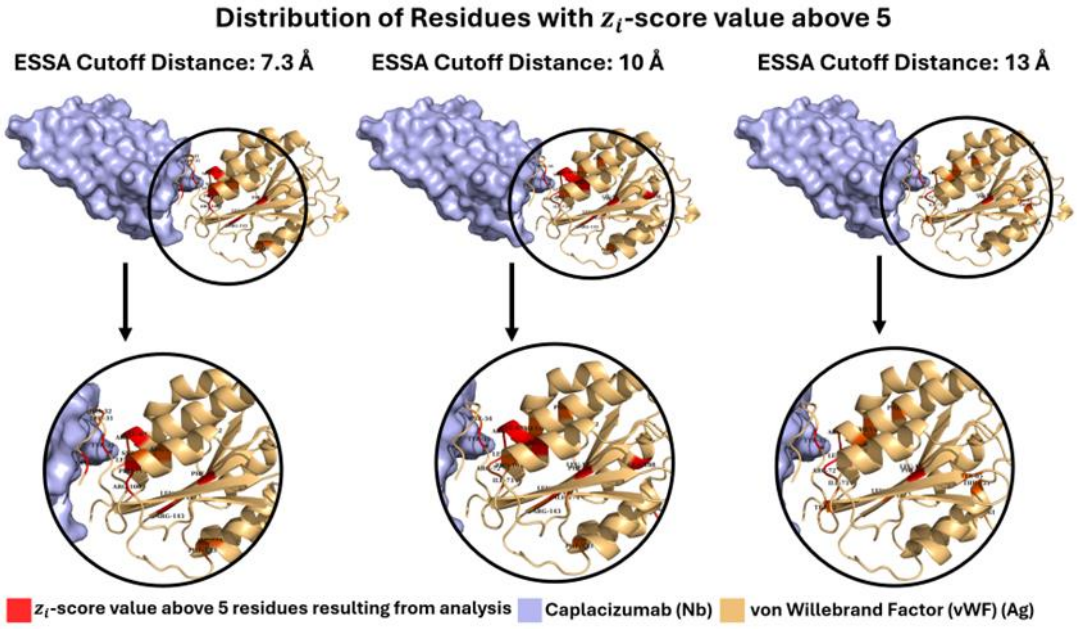


Figure 23. The most essential residues on the von Willebrand Factor (vWF) (Ag) structure determined using different cutoff distance values in ESSA.

Table 6. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the von Willebrand Factor (vWF) (Only high z-score values are reported in this table. *: Residues specified as interaction sites in PDBsum are indicated with).

ESSA, Cutoff Distance: 7.3 Å		**ESSA, Cutoff Distance: 10 Å**		**ESSA, Cutoff Distance: 13 Å**	
vWF	**Caplacizumab**	**vWF**	**Caplacizumab**	**vWF**	**Caplacizumab**
Glu28		Phe34*	Glu105*	Tyr35*	Ser31*
Leu31		Tyr35*	Ser31*		Tyr32*(CDR1)
His32*	Thr55*		Tyr32*(CDR1)		Arg54*(CDR2)
	Arg54*(CDR2)		Arg54*(CDR2)		Glu105*
Tyr35*	Ser31*		Glu105*	Leu42	
	Tyr32*(CDR1)	Leu42		Phe44	
	Arg54*(CDR2)	Phe44		Phe63	
	Glu105*	Phe63		Met67	
Leu42		Met67		Met68	
Phe44		Met68		Arg70	
Glu69		Glu69		Leu71	
Arg70		Arg70		Arg72*	Arg28*
Leu71		Leu71			Phe30*(CDR1)
Arg100		Arg72*	Arg28*	Ile73	
Pro101			Phe30*(CDR1)	Trp77	
Ser102		Ile73		Tyr85	
Arg105		Pro101		Thr121	
Ile132		Arg105		Ile144	
Phe133		Phe133		Leu146	
Arg143		Arg143		Phe161	
Ile144		Ile144			
Phe202		Leu146			
Tyr220		Phe161			
		Ile174			
		Gln188			
		Phe202			
		Tyr220			
		Leu221			
Total: 19	**Total: 6**	**Total: 24**	**Total: 7**	**Total: 16**	**Total: 6**

Interestingly, two residues in vWF antigen (Phe161 and Gln188 underlined in Table 6), which do not interact with the Caplacizumab found using the cutoff distance value of 10 Å and 13 Å, were found to be in the interaction region between the vWF and an antibody called NMC-4 IGG-1 as shown in Figure 24 and 25. There are two available crystal structures for the vWF in complex with NMC-4 IGG-1 (PDB IDs: 1OAK and 1FNS-mutant) (Celikel, et al. 1998; Celikel, Ruggeri and Varughese 2000). For both crystal structures of the vWF in complex with antibody NMC-4 IGG-1, the antigen part of the antibody-antigen and the nanobody-antigen (PDB ID: 7EOW) complex were aligned, then a merged binding pose for the PDB ID: 1OAK and 7EOW was obtained (Figure 24), and for the PDB ID: 1FNS and 7EOW was obtained (Figure 25). In both figures, the antigen is in surface representation and colored according to the ESSA results obtained using cutoff values 10 Å and 13 Å. Phe161 in the vWF-Caplacizumab complex (PDB ID: 7EOW) corresponds to Phe634 in the vWF-NMC-4 IGG-1 complex (PDB ID: 1OAK), where it is interacting with the antibody. Similarly, Gln188 in PDB ID: 7EOW corresponds to Gln661 in PDB ID: 1FNS. This difference in residue index is due to the difference in the residue numbering in the corresponding PDB files.

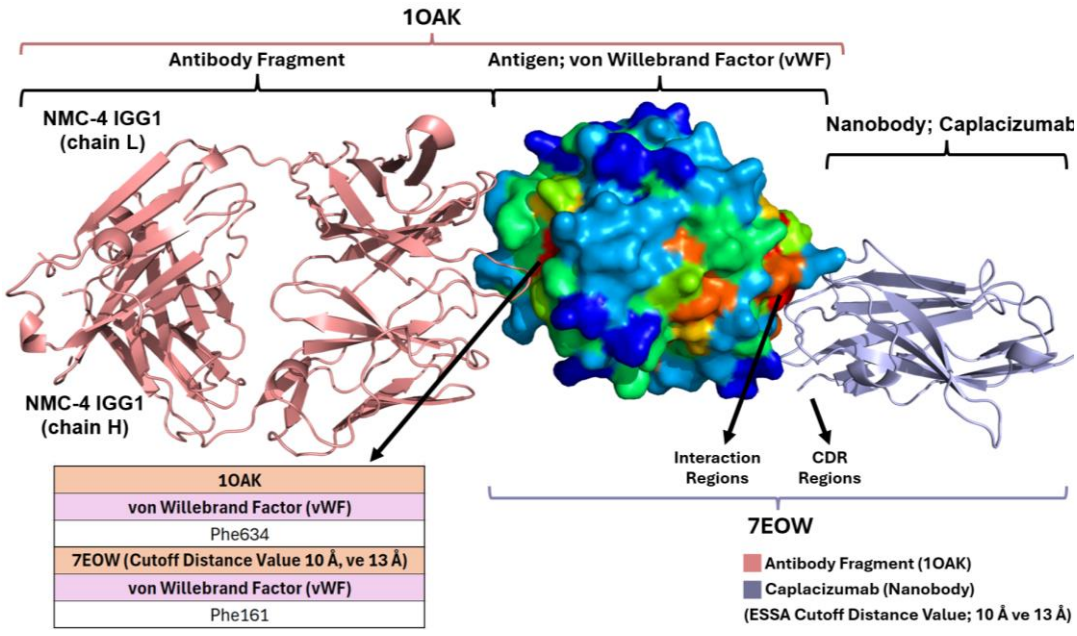


Figure 24. Representation of the merged binding poses for the vWF antigen in complex with the nanobody Caplacizumab (7EOW) and antibody NMC-4 IGG-1 (1OAK) (Lee, et al. 2021; Celikel, et al. 1998).

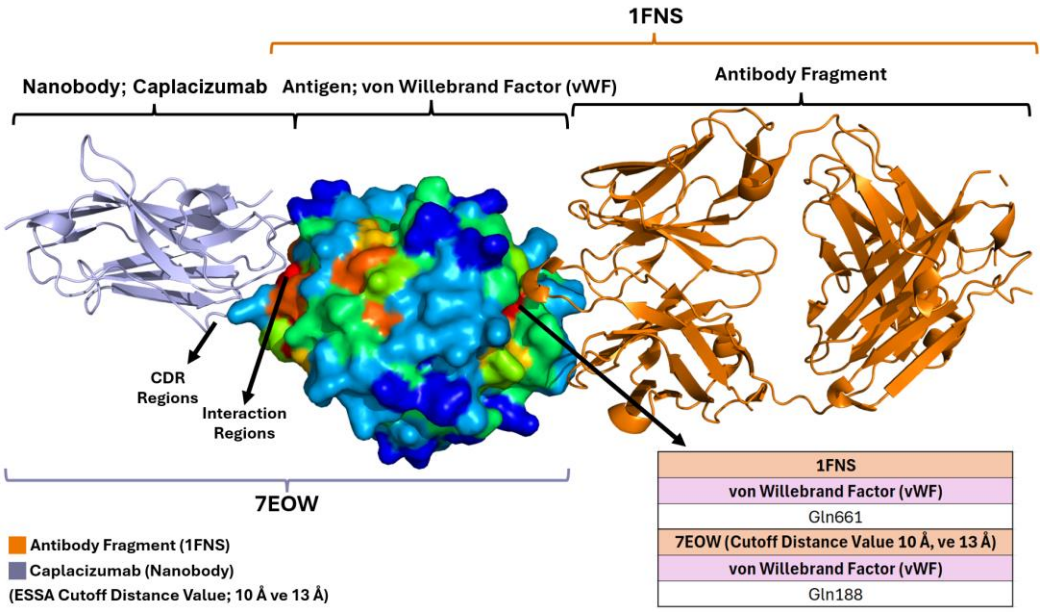


Figure 25. Representation of the merged binding poses for the vWF antigen in complex with the nanobody Caplacizumab (7EOW) and antibody NMC-4 IGG-1 (1FNS) (Lee, et al. 2021; Celikel, Ruggeri and Varughese, 2000).

3.1.3. Caplacizumab-von Willebrand Factor (vWF) Complex Structure

There are a total of 322 residues in the complex crystal structure of Caplacizumab and von Willebrand Factor, whose nanobody and antigen structures were examined separately. The numbers and percentage distributions of these residues are shown in Figure 26.

The residues with a total number of counts over 20 in the complex structure are serine, leucine, valine, arginine, alanine, glutamate, and glycine. The residues with a total number of less than 5 in the complex structure are cysteine, histidine, and tryptophan. The highest percentile is serine residue with 9.7%, while the lowest percentile is tryptophan residue with 0.91% (Figure 26).

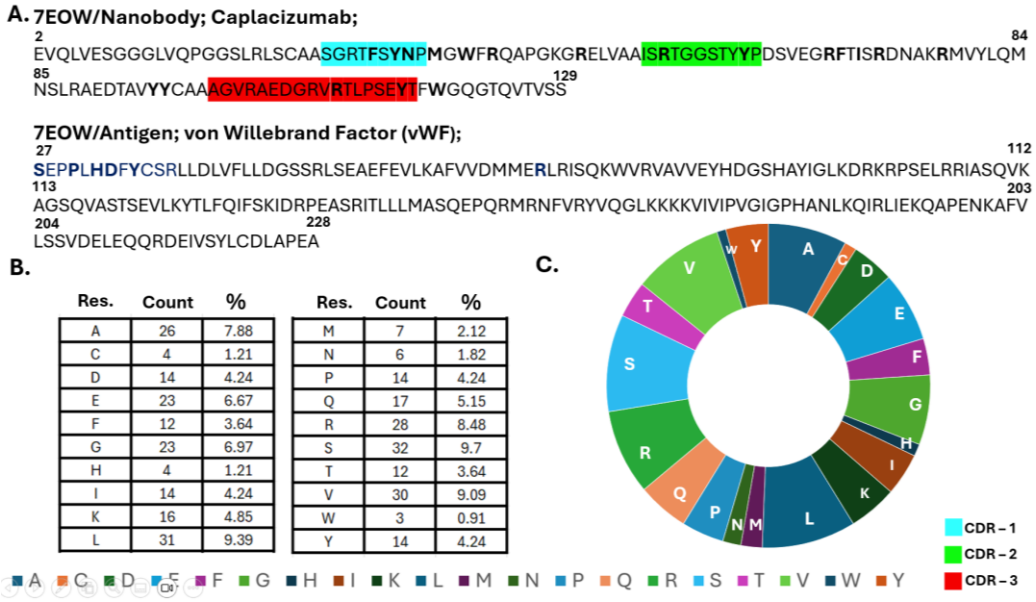


Figure 26. Representation of the primary sequence data regarding the Caplacizumab (Nanobody) and von Willebrand Factor (vWF) (Antigen) complex. A: Complex structure sequence [PDB ID: 7EOW, (Lee, et al. 2021)] B: Count and percentage values for each residue in the complex structure, C: Pie-chart representation of the residue distribution in the complex structure.

The ESSA results using varying cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the complex structure are given in Figure 27 from different angles (0°, 90°, 180°, 270°, top, and bottom side). The z-score values obtained using varying cutoff distance values are used to color the structures in the PyMOL. The residues with z-scores above 5 represent essential sites for binding. As seen in these results, some of the important residues are within the interaction region, while the others are located outside the interaction region. The display of residues with z-score values above 5 on the structure in the analysis results for different cutoff distance values is shown in Figure 29.

The combined-ESSA results are also given from different angles (0°, 90°, 180°, 270°, top, and bottom side) in Figure 28. For each amino acid, whichever of the three cutoff values had the highest z-score value was used or taken as the z-score.

As a consequence of the ESSA analysis performed on the von Willebrand Factor (vWF) (Ag) at different cutoff distance values (7.3 Å, 10 Å, and 13 Å), the residues with a z-score value above 5 are listed in Table 7. The total number of predicted residues using

three different cutoffs are as follows: 12 residues at cutoff 7.3 Å; 18 residues at cutoff 10 Å, and 17 residues at cutoff 13 Å. As a result of the analysis performed for all three cutoff distance values, the Tyr32 residue was found to be a common essential amino acid in the nanobody. When the results of the complex nanobody with cutoff distance values of 10 Å and 13 Å were examined, it was seen that Trp37 was a sequence neighbor. It was observed that Arg78 was positionally adjacent to CDR-1 and CDR-2 at a cutoff value of 7.3 Å.

In the free from of von Willebrand Factor (vWF) antigen results, two residues (Phe161 and Gln188) located in the antigen connection region of the two structures with PDB IDs 1FNS and 1OAK were found only in the ESSA results, which were done by taking the cutoff distance value as 10 Å. Although the antigen was in complex with the Caplacizumab nanobody, these two residues still showed a high z-score value above 5.

Table 7. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the complex structure.

ESSA, Cutoff Distance: 7.3 Å		**ESSA, Cutoff Distance: 10 Å**		**ESSA, Cutoff Distance: 13 Å**	
Caplacizumab	**vWF**	**Caplacizumab**	**vWF**	**Caplacizumab**	**vWF**
Tyr32(CDR1)	Glu28	Phe30(CDR1)	Phe34	Phe30(CDR1)	His32
Arg78	Phe34	Tyr32(CDR1)	Try35	Tyr32(CDR1)	Phe34
Glu105	Arg72	Trp37	Arg70	Trp37	Tyr35
	Trp77	Phe38	Arg72	Arg39	Phe44
	Arg100	Arg39	Trp77	Arg46	Arg70
	Pro101	Glu47	Phe133	Arg54	Arg72
	Phe133	Arg68	Arg143	Tyr95	Trp77
	Arg143	Phe69	Phe161	Tyr96	
	Glu277	Tyr95	Gln188	Arg103	
				Glu105	
Total: 3	**Total: 9**	**Total: 9**	**Total: 9**	**Total: 10**	**Total: 7**

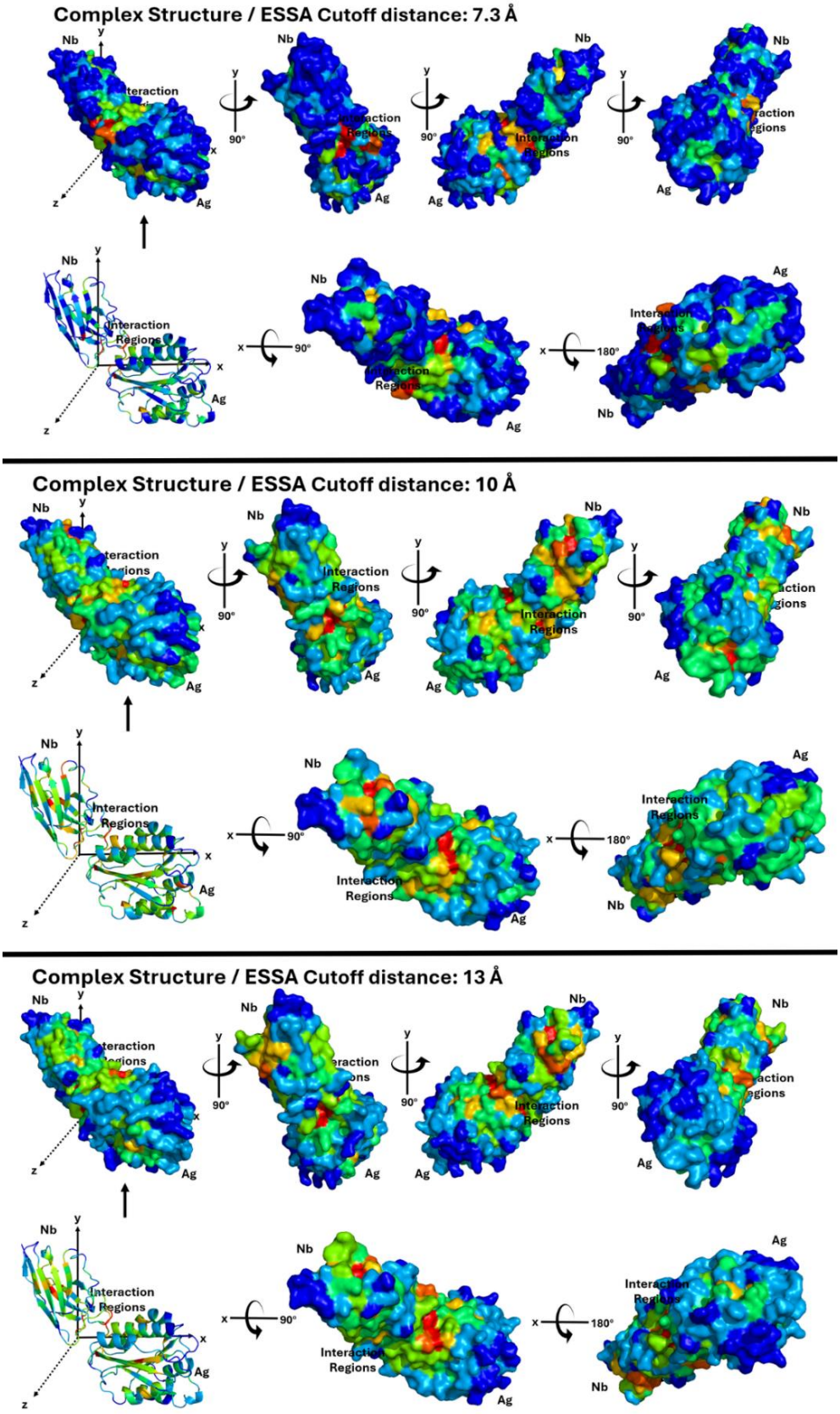


Figure 27. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

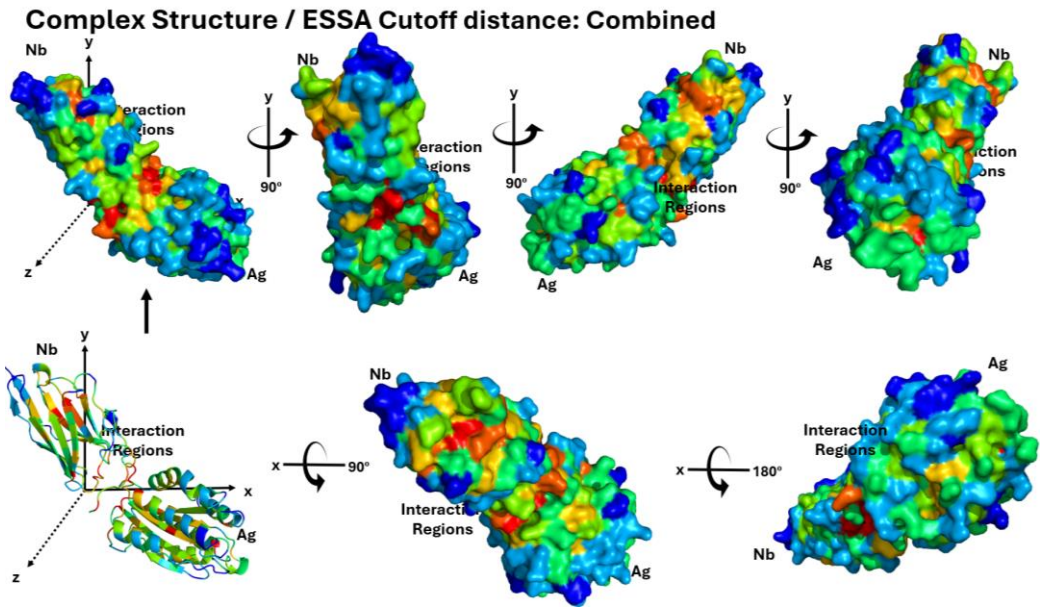


Figure 28. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

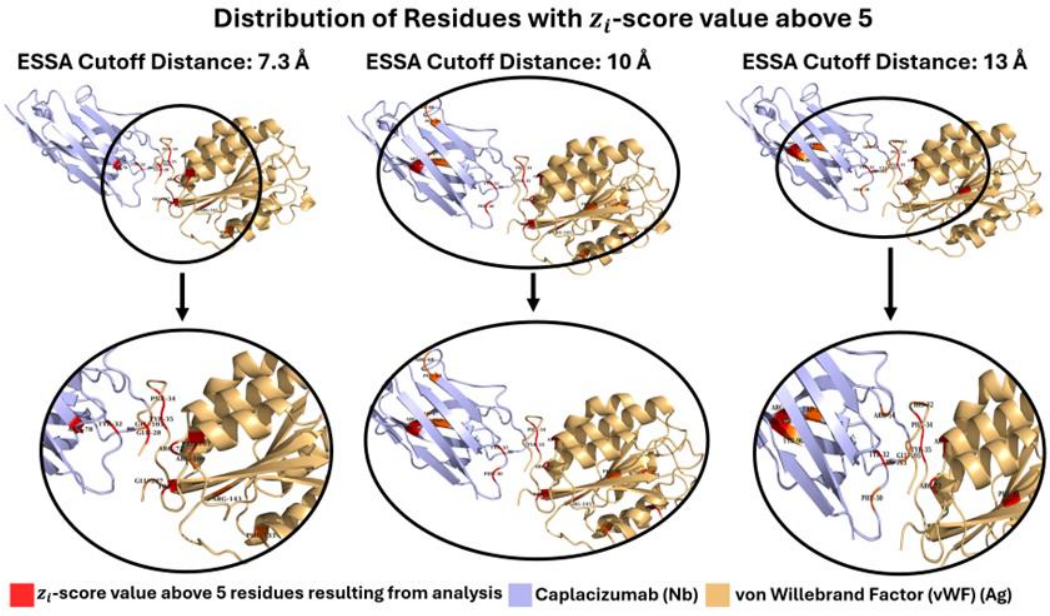


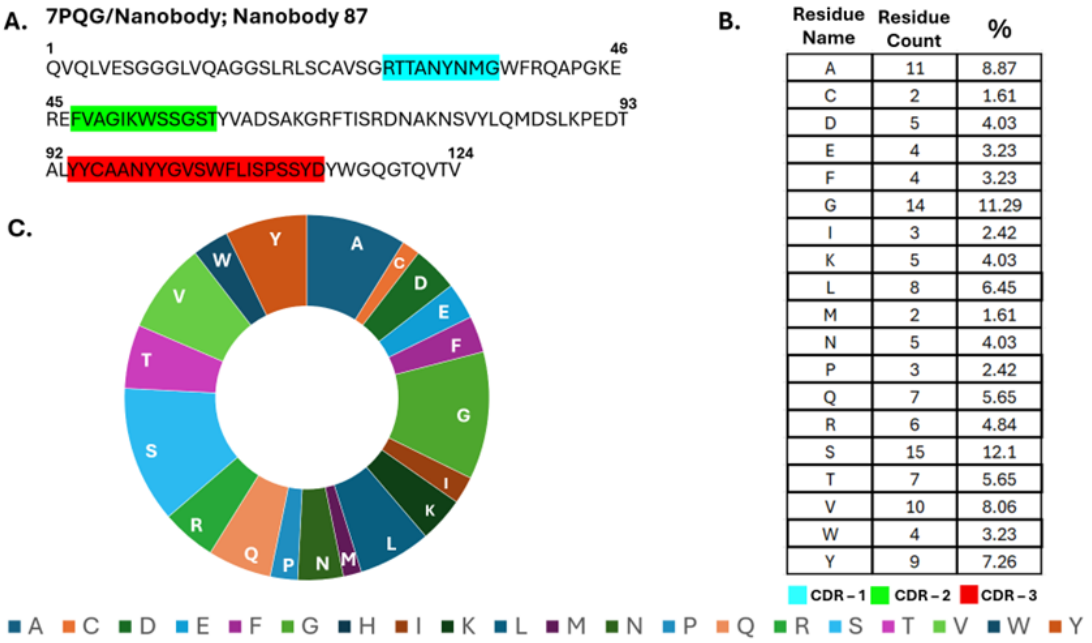
Figure 29. The most essential residues on the complex structure determined using different cutoff distance values in ESSA for the complex structure.

3.2. Nanobody 87-NTCP Complex

3.2.1. Nanobody 87 Structure

Nanobody 87, which contains 124 residues in its crystal structure, binds to NTCP in the liver cell and selectively blocks the entry of Hepatitis B and D (HBV and HDV) virus antigens into the cell (Goutam, et al. 2022).

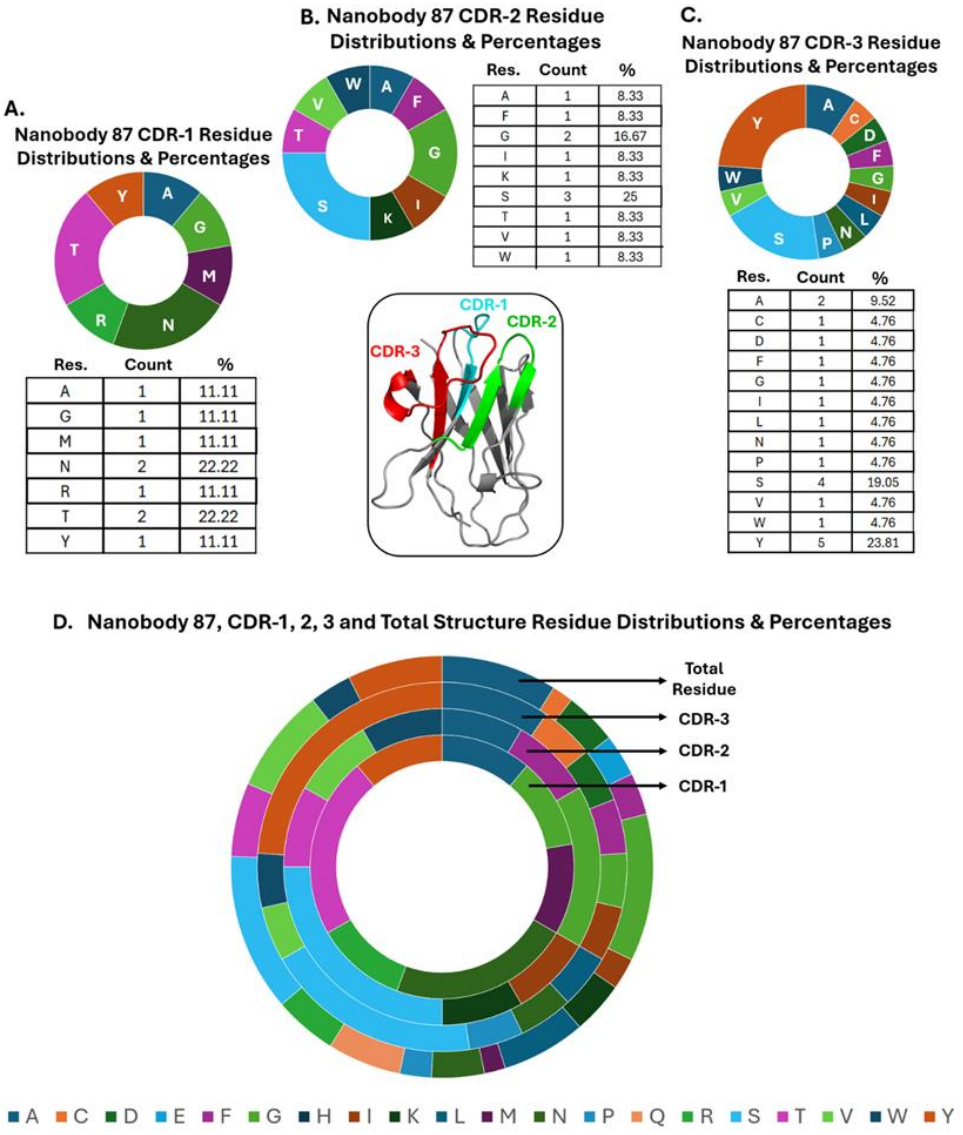
The most abundant residues in this structure, whose sequence and percentage distributions of the residues are given in Figure 30, are serine, alanine, valine, and glycine. Data regarding the CDR regions of this structure, which contains counts and percentages of residues in the CDR regions, are given in Figure 31.



Residue Name	Residue Count	%
A	11	8.87
C	2	1.61
D	5	4.03
E	4	3.23
F	4	3.23
G	14	11.29
I	3	2.42
K	5	4.03
L	8	6.45
M	2	1.61
N	5	4.03
P	3	2.42
Q	7	5.65
R	6	4.84
S	15	12.1
T	7	5.65
V	10	8.06
W	4	3.23
Y	9	7.26

CDR-1 CDR-2 CDR-3

Figure 30. Representation of the primary sequence data regarding Nanobody 87. A: Nanobody 87 sequence [PDB ID: 7PQG, (Goutam, et al. 2022)]. B: Count and percentage values for each residue in Nanobody 87, C: Pie-chart representation of the residue distribution in Nanobody 87.



Res.	Count	%
A	1	11.11
G	1	11.11
M	1	11.11
N	2	22.22
R	1	11.11
T	2	22.22
Y	1	11.11

Res.	Count	%
A	1	8.33
F	1	8.33
G	2	16.67
I	1	8.33
K	1	8.33
S	3	25
T	1	8.33
V	1	8.33
W	1	8.33

Res.	Count	%
A	2	9.52
C	1	4.76
D	1	4.76
F	1	4.76
G	1	4.76
I	1	4.76
L	1	4.76
N	1	4.76
P	1	4.76
S	4	19.05
V	1	4.76
W	1	4.76
Y	5	23.81

Figure 31. Representation of the primary sequence data of CDR regions located in the Nanobody 87 [PDB ID: 7PQG, (Goutam, et al. 2022)]. Types and percentage distributions of residues in A: the CDR-1 region, B: the CDR-2 region, and C: the CDR-3 region. D: Pie-chart representation of the residue distribution in Nanobody 87 CDR regions.

The ESSA results for three different cutoff distances are presented from different angles (0°, 90°, 180°, and 270°) through 3D structures in Figure 32. The combined result for Nanobody 87 from different angles is shown in Figure 33. The essential residues with z-score values above 5 obtained using different cutoff distance values are shown in Figure 34 and provided as a list in Table 8.

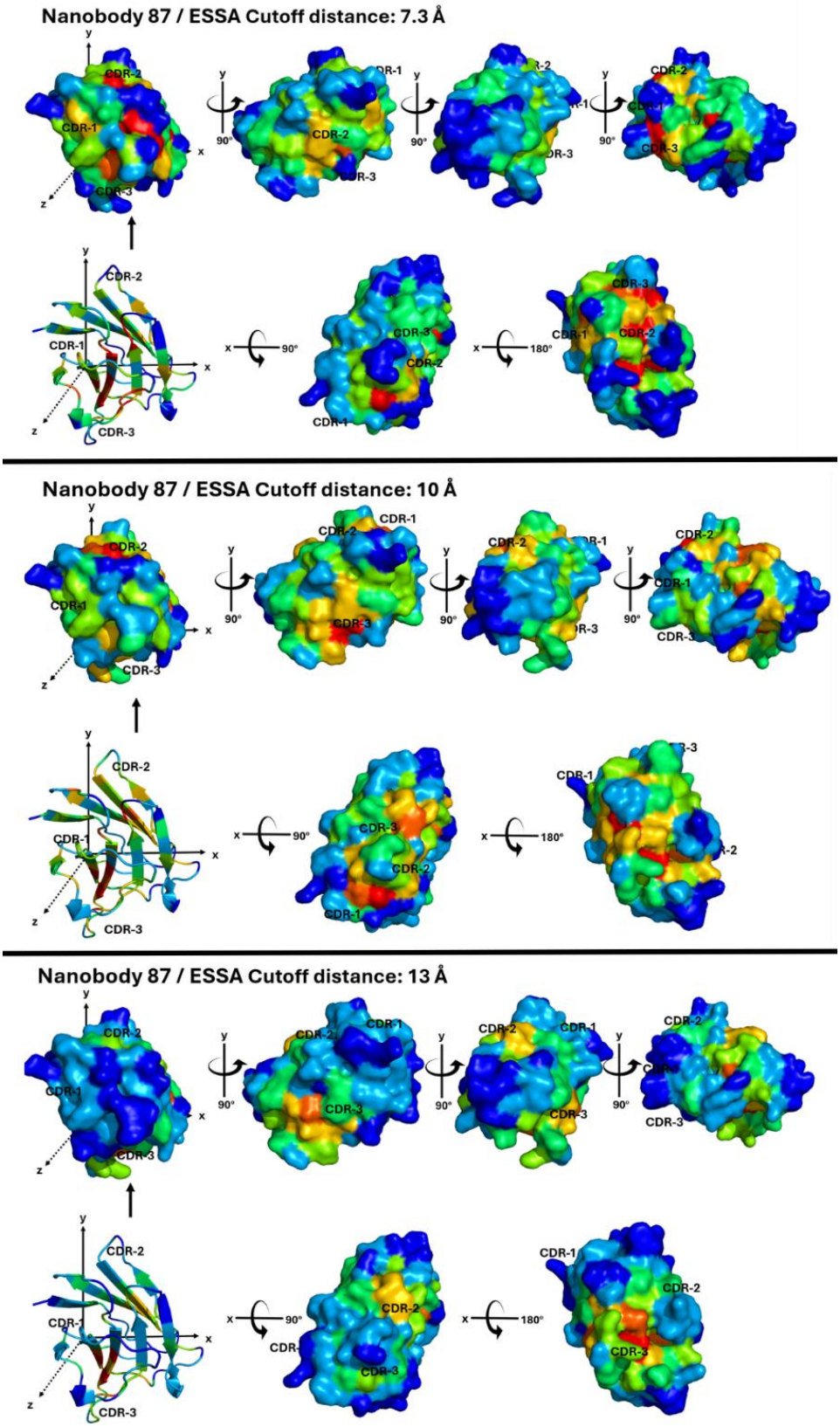


Figure 32. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff distance values for the Nanobody 87.

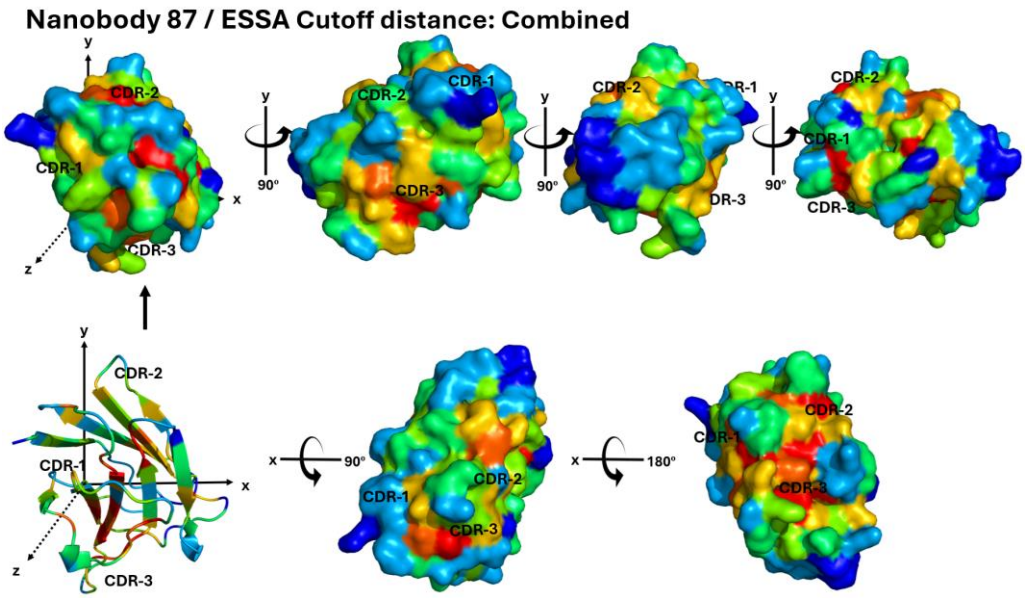


Figure 33. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the Nanobody 87.

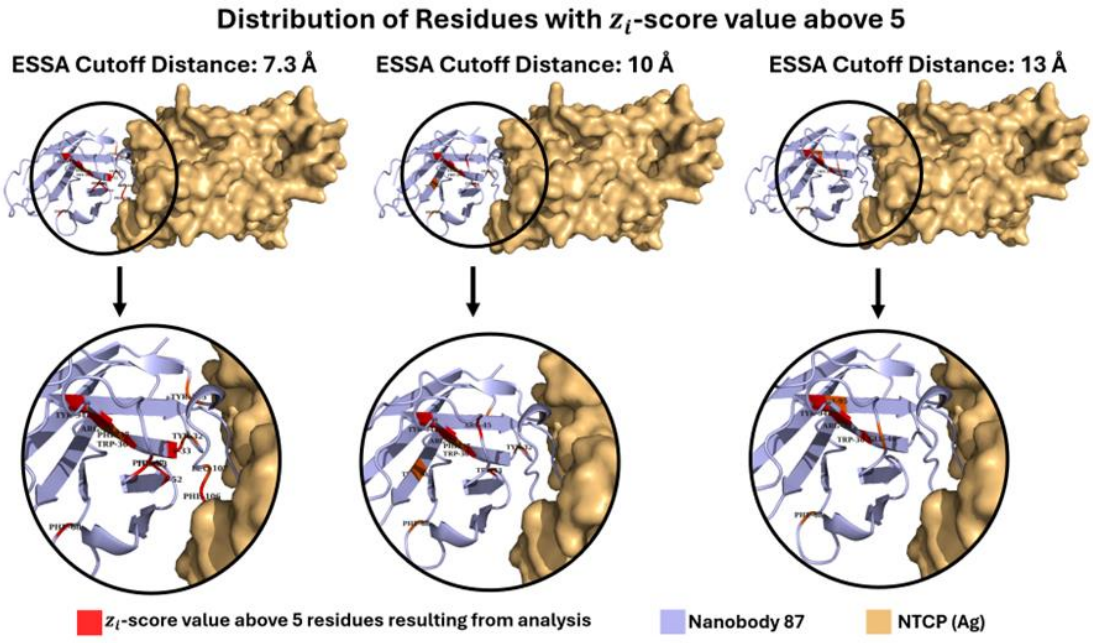


Figure 34. The most essential residues on the Nanobody 87 structure determined using different cutoff distance values in ESSA.

Table 8. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the Nanobody 87 (Only high z-score values are reported in the table. *: Residues specified as interaction sites in PDBsum are indicated with an asterisk. Ag: Antigen, Nb: Nanobody).

ESSA, Cutoff Distance: 7.3 Å		**ESSA, Cutoff Distance: 10 Å**		**ESSA, Cutoff Distance: 13 Å**	
Nanobody87		**Nanobody87**		**Nanobody87**	
Tyr32(CDR1)	Asn33(CDR1)	Thr28(CDR1)	Tyr32(CDR1)	Trp36	Arg38
Trp36	Phe37	Trp36	Phe37	Glu46	Phe68
Arg38	Phe47	Arg38	Arg45	Tyr94(CDR3)	Tyr95
Lys52(CDR2)	Trp53(CDR2)	Trp53(CDR2)	Phe68		
Phe68	Tyr94(CDR3)	Tyr80	Tyr94(CDR3)		
Phe106*(CDR2) – Ag: Thr13*					
Phe106*(CDR2) – Ag: Thr15*					
Leu107*(CDR3) – Ag: Glu277*					
Tyr113*(CDR3)					
Total: 13		**Total: 10**		**Total: 6**	

In terms of the total number of residues, the highest number of residues (13 essential residues) was found as a result of the analysis performed at cutoff 7.3 Å. In the analysis performed with the cutoff distance value of 10 Å, 10 essential residues were identified. Interestingly, with a higher cutoff distance value (13 Å), a smaller number of residues (6) were detected. Arg38, Phe68, and Trp36 residues were common at all three cutoff distance values. On the other hand, there are some residues that are only found in ESSA with a specific cutoff distance value; Phe106 and Leu107 in the CDR-3 region of nanobody 87 (marked with an asterisk in Table 8) are in the interaction region with the antigen in the PDBsum database in ESSA with a cutoff distance of 7.3 Å. Glu46 and Tyr95 in ESSA with a cutoff distance of 13 Å, Thr28, Arg45, and Tyr80 with 10 Å were found. In the analysis results with cutoff distance values of 7.3 Å and 10 Å, Trp36, Phe37, and Arg38 residues are neighbors of the CDR-1 region. Glu46 residue, which is adjacent to CDR-2, was observed as an essential residue in the analysis with a cutoff distance value

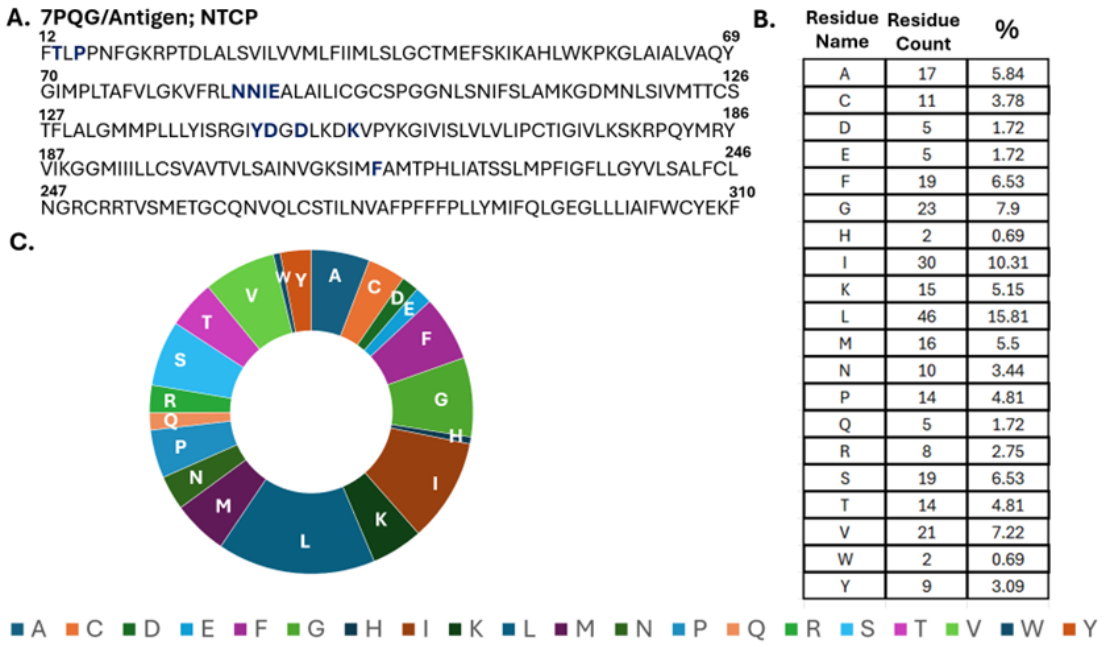
of 13 Å. Another residue adjacent to CDR-2 (Arg45) was found essential in the analysis with a cutoff distance value of 10 Å.

3.2.2. NTCP (Antigen) Structure

Bile salts, which enable the fats that come with food to become smaller fat drops, are taken into the cells by polypeptide (NCTP) receptors carried together with Na^{+-} taurocholate in the liver. This receptor undergoes a conformational change to transport bile salts into the cell, enabling transmembrane passage, as well as providing the emergence of the key determining residues for the binding of hepatitis B and D (HBV/HDV) viruses to the outside of the cell, and has a fundamental role in liver pathology as the cellular entry receptor of these viruses (Goutam, et al. 2022).

The total number of mutations in the NTCP antigen, which has 298 residues in its crystal structure, is 9. Although there is no explanation for these mutations in the crystal structure's article, in this thesis study, analyses were done on the sequence taken from PDB. The percentage distribution of the remains found in the structure is shown in Figure 35. As can be seen from here, there are different numbers and types of residues in the NTCP antigen. The most abundant residue among these residues is Leucine. Additionally, the regions shown in bold and dark blue in the figure are the residues in the interaction regions according to the PDBsum database. The least abundant residues are tryptophan and histidine. The analysis results based on different cutoff distance values are shown in Figure 36 from various angles. The combined ESSA result of cutoff distance values is given in Figure 37. In addition, the residues with a z-score value above 5 from the ESSA analyses performed for different cutoff distance values (7.3 Å, 10 Å, and 13 Å) are shown in Figure 38.

The results shown in Table 9 are the residues with a z-score value above 5 in the ESSA analysis results done by taking different cutoff distance values (7.3 Å, 10 Å, and 13 Å).



Residue Name	Residue Count	%
A	17	5.84
C	11	3.78
D	5	1.72
E	5	1.72
F	19	6.53
G	23	7.9
H	2	0.69
I	30	10.31
K	15	5.15
L	46	15.81
M	16	5.5
N	10	3.44
P	14	4.81
Q	5	1.72
R	8	2.75
S	19	6.53
T	14	4.81
V	21	7.22
W	2	0.69
Y	9	3.09

Figure 35. Representation of the primary sequence data regarding the NTCP (Antigen). A: NTCP sequence [PDB ID: 7PQG, (Goutam, et al. 2022)]. B: Count and percentage values for each residue in NTCP, C: Pie-chart representation of the residue distribution in NTCP.

When looking at the results, according to the results of the ESSA analysis performed with the cutoff distance value as 7.3 Å, the phenylalanine residue in the 216th position among the residues with a z-score value above 5 was found in the interaction region taken from the PDBsum web server (marked with an asterisk in Table 9). Using higher cutoff distance values (10 Å and 13 Å), the interaction zones given in the PDBsum web server could not be found. Among the residues with z-score values above 5, five common residues were found in the analyses performed for each cutoff value. These residues are Phe108, Gln293, Glu296, Ile301, and Trp305.

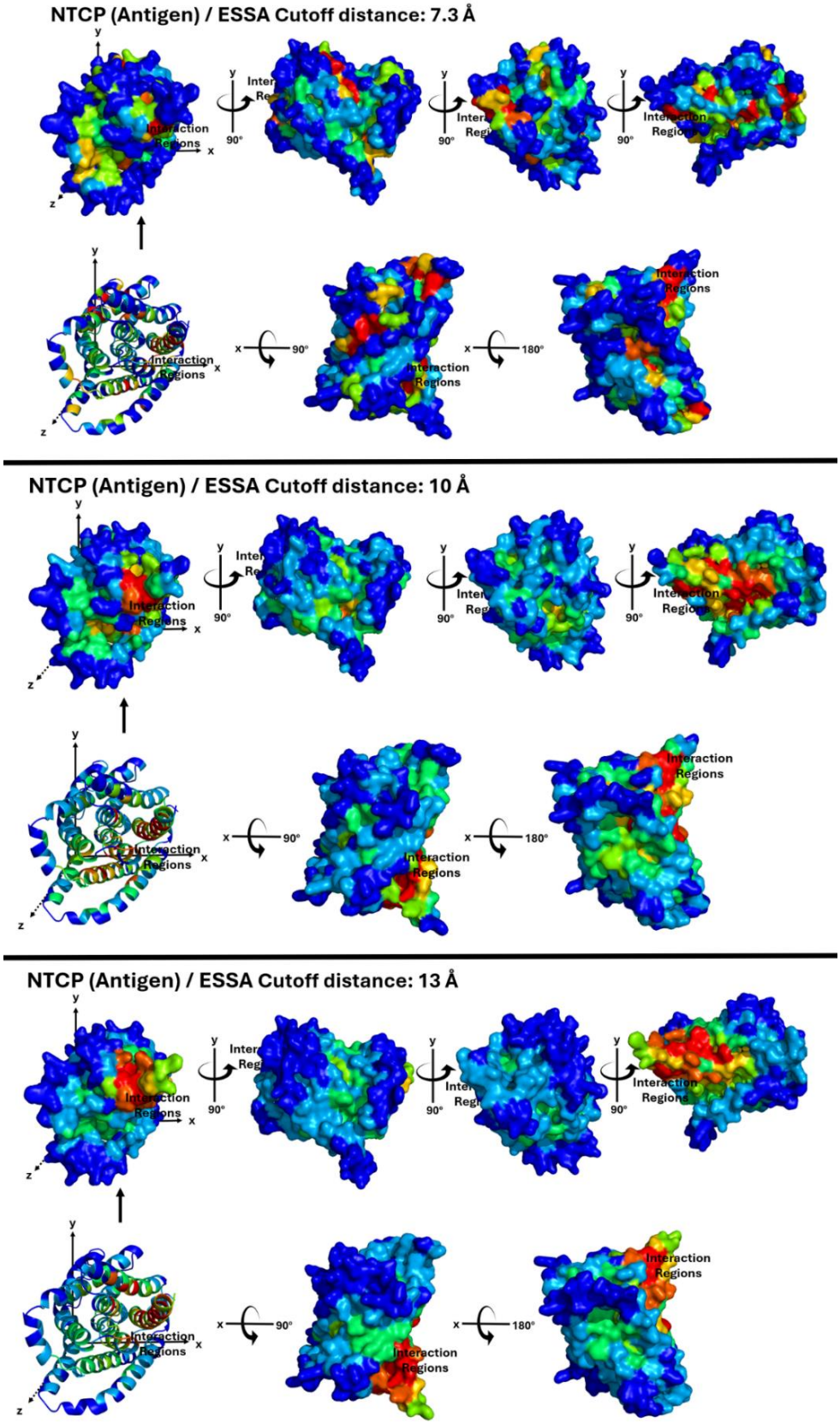


Figure 36. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the NTCP antigen.

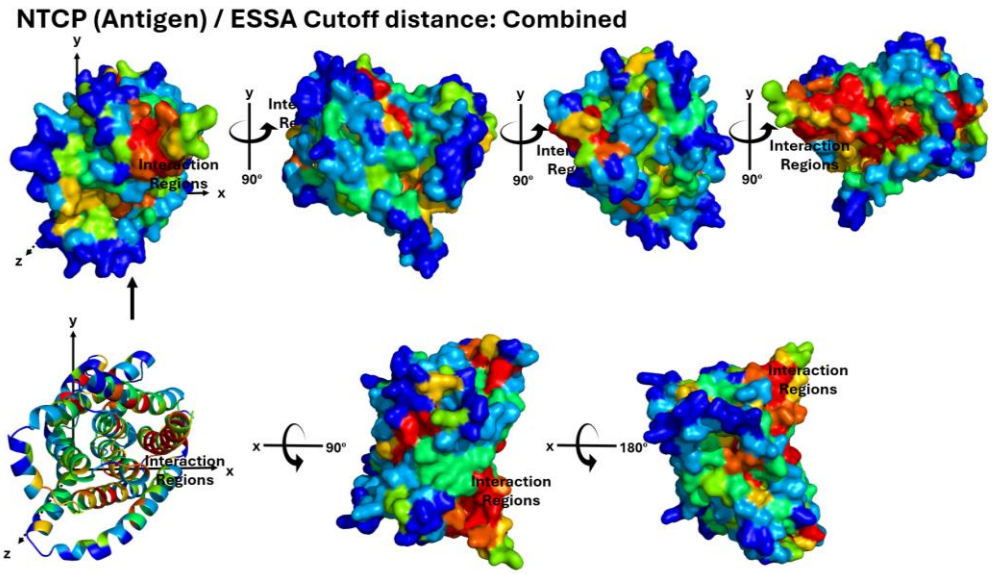


Figure 37. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the NTCP antigen.

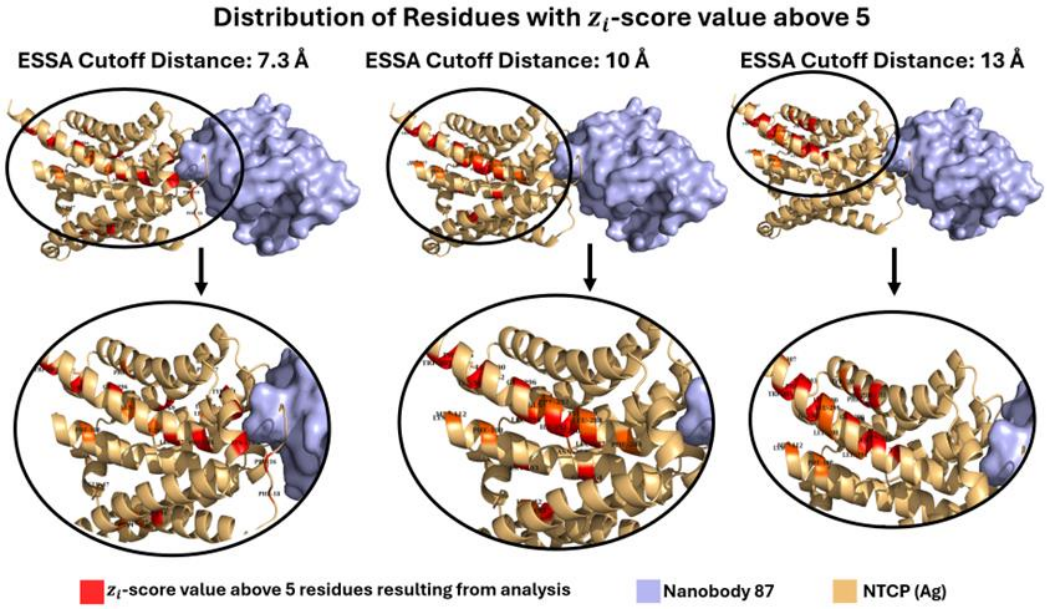


Figure 38. The most essential residues on the NTCP (Ag) structure determined using different cutoff distance values in ESSA.

Table 9. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the NTCP (Only high z-score values are reported in this table. *: Residues specified as interaction sites in PDBsum are indicated with. Ag: Antigen, Nb: Nanobody).

ESSA, Cutoff Distance: 7.3 Å		**ESSA, Cutoff Distance: 10 Å**		**ESSA, Cutoff Distance: 13 Å**	
NTCP		**NTCP**		**NTCP**	
Pro16	Phe18	Ile38	Leu42	Phe108	Met112
Met34	Ile37	Asn103	Phe108	Lys113	Phe231
Glu47	Tyr69	Met112	Lys113	Phe234	Tyr238
Phe77	Leu91	Ile195	Arg252	Arg252	Tyr289
Leu94	Phe108	Asn262	Gln264	Phe292	Gln293
Leu137	Tyr141	Phe284	Leu287	Leu294	Glu296
Ile194	Lys212	Leu288	Tyr289	Leu298	Leu299
Phe216* - Nb: Trp105*(CDR3)		Met290	Ile291	Leu300	Ile301
Phe234	Phe284	Phe292	Gln293	Ile303	Phe304
Leu287	Met290	Leu294	Glu296	Trp305	Tyr307
Gln293	Glu296	Leu300	Ile301		
Ile301	Trp305	Phe304	Trp305		
Total: 23		**Total: 24**		**Total: 20**	

Interestingly, residues Lys212 and Phe216 found essential as a result of ESSA analysis with a cutoff distance value of 7.3 Å, are also the binding site of Fab regions complexed with the NTCP. These two residues interact with Fab sites found in two structures with PDB IDs 7ZYI and 7FCI (Liu, et al. 2022; Park, et al. 2022). Among these residues, which correspond to the same position in the antigen sequence, the Lys212 residue interacts with the Tyr104 residue in the H-chain located in the Fab region of the 7FCI structure. Phe216 residue interacts with Tyr101 and Ala103 residues in the H-chain and Tyr31 residue in the L-chain. These interaction regions are shown in Figure 39.

Similarly, the Lys212 residue interacts with Tyr100H located in the H-chain of the complex NTCP structure with YN69083Fab with PDB ID 7ZYI, and the Phe216 residue interacts with Tyr100H and Arg100F residues in the H-chain and Tyr92 residue located in the L-chain. The interaction region is shown in Figure 40.

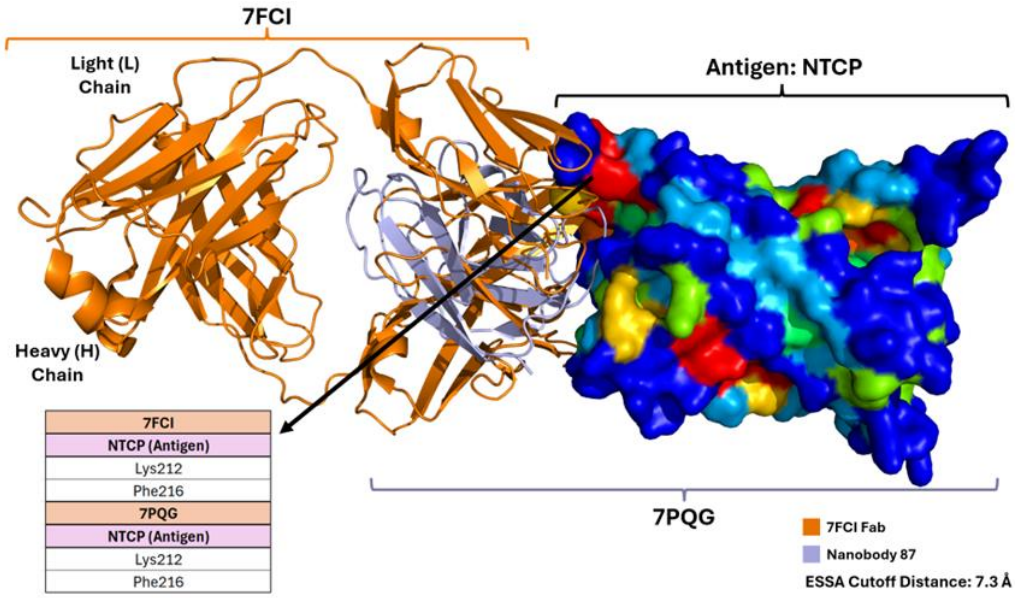


Figure 39. Representation of the merged binding poses for the NTCP antigen in complex with the Nanobody 87 (7PQG) and Fab domain 7FCI (Goutam, et al. 2022; Park, et al. 2022).

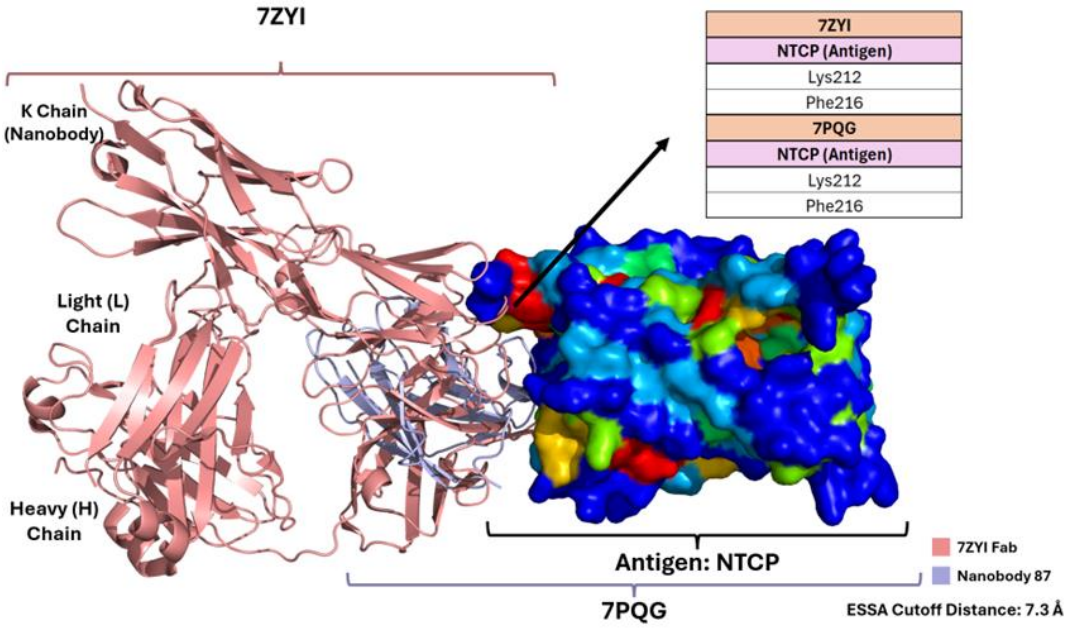
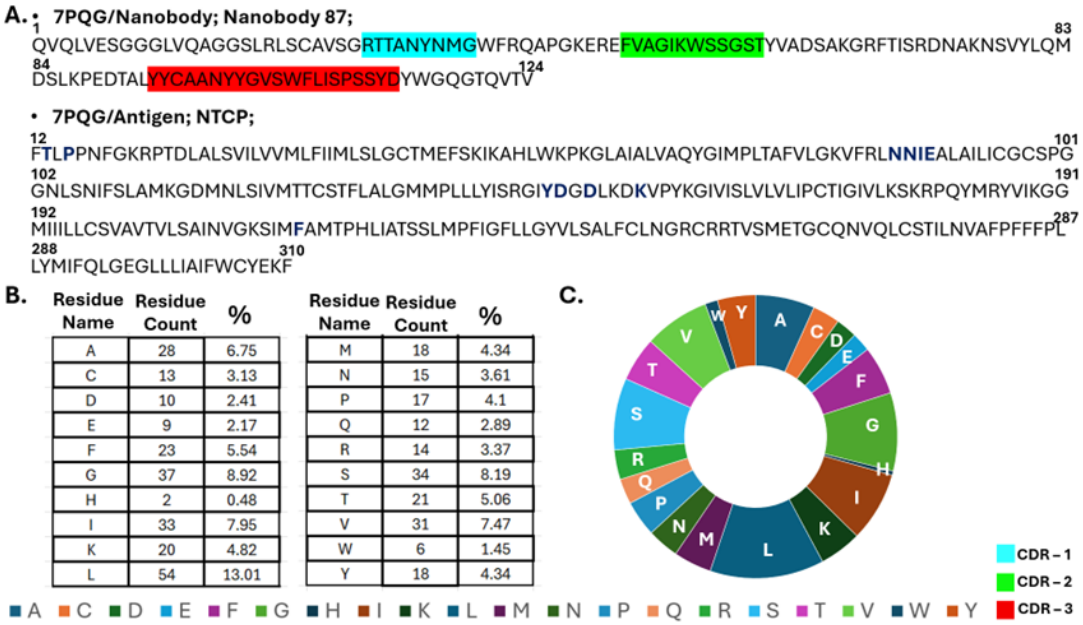


Figure 40. Representation of the merged binding poses for the NTCP antigen in complex with the Nanobody 87 (7PQG) and YN69083 Fab domain 7ZYI (Liu, et al. 2022; Goutam, et al. 2022).

3.2.3. Nanobody 87-NTCP (Antigen) Complex Structure

There are 422 residues in the crystal structure of the Nanobody 87 - NTCP complex. The number and percentage distribution of these residues are shown in Figure 41. When the crystal structure is examined (Figure 41), the residue found in the highest count and percentage is leucine. The least abundant residue is histidine.

The view of the complex structure colored according to z-score values obtained from ESSA analyses with different cutoff distance values and the combined-ESSA representation are shown in Figures 42 and 43, respectively. The display of residues with z-score values above 5 on the structure in the analysis results for different cutoff distance values is also provided in Figure 44.



Residue Name	Residue Count	%	Residue Name	Residue Count	%
A	28	6.75	M	18	4.34
C	13	3.13	N	15	3.61
D	10	2.41	P	17	4.1
E	9	2.17	Q	12	2.89
F	23	5.54	R	14	3.37
G	37	8.92	S	34	8.19
H	2	0.48	T	21	5.06
I	33	7.95	V	31	7.47
K	20	4.82	W	6	1.45
L	54	13.01	Y	18	4.34

Figure 41. Representation of the primary sequence data regarding the Nanobody 87 and NTCP (Antigen) complex. A: Complex structure sequence [PDB ID: 7PQG, (Goutam, et al. 2022)] B: Count and percentage values for each residue in complex structure, C: Pie-chart representation of the residue distribution in complex structure.

As seen in Table 10, the cutoff distance value with the highest number of essential residues is 7.3 Å (29 residues in total). In addition, a total of 17 residues were found in the analysis results using cutoff distance values of 10 Å and 13 Å. In the analysis performed by taking the cutoff distance value as 10 Å, all residues (except Tyr59) in the complex structure Nanobody 87 are in the interaction region given on the PDBsum. When the cutoff distance value is 7.3 Å, essential residues are mostly clustered around the nanobody-antigen interaction site. In addition, Tyr59, one of the residues located on the nanobody side of the complex structure (nanobody 87), is in the CDR-2 region, while the other residues are in the CDR-3 region. In addition, Lys212 and Phe216 (which is an interacting residue), which were found when ESSA analysis was performed on the antigen alone, were not found in any ESSA results using different cutoff distances performed on the nanobody-antigen complex.

Table 10. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the complex structure (Only high z-score values are reported in this table. *: Residues specified as interaction sites in PDBsum).

ESSA Cutoff Distance: 7.3 Å			**ESSA Cutoff Distance: 10 Å**		**ESSA Cutoff Distance: 13 Å**	
Nanobody87	**NTCP**		**Nanobody87**	**NTCP**	**Nanobody87**	**NTCP**
Tyr100*	Phe18		Tyr59	Ile38	Tyr100*	Met112
Tyr101*	Ile30		Tyr100*	Asn103	Tyr101*	Lys113
Trp105*	Leu31		Tyr101*	Met112	Trp105*	Tyr238
	Met34	Ile37	Trp105*	Lys113		Arg252
	Glu47	Lys50	Phe106*	Arg252		Glu256
	Phe77	Asn87*	Leu107*	Gln264		Leu298
	Leu94	Phe108	Ile108*	Leu300		Leu299
	Met112	Leu137		Ile301		Leu300
	Tyr141	Tyr146*		Phe304		Ile301
	Asp149*	Lys153*		Met305		Ala302
	Ile198	Leu205				Ile303
	Phe234	Phe284				Phe304
	Leu287	Met290				Trp305
	Glu296	Ile301				Tyr307
	Trp305					
Total: 3	**Total: 26**		**Total: 7**	**Total: 10**	**Total: 3**	**Total: 14**

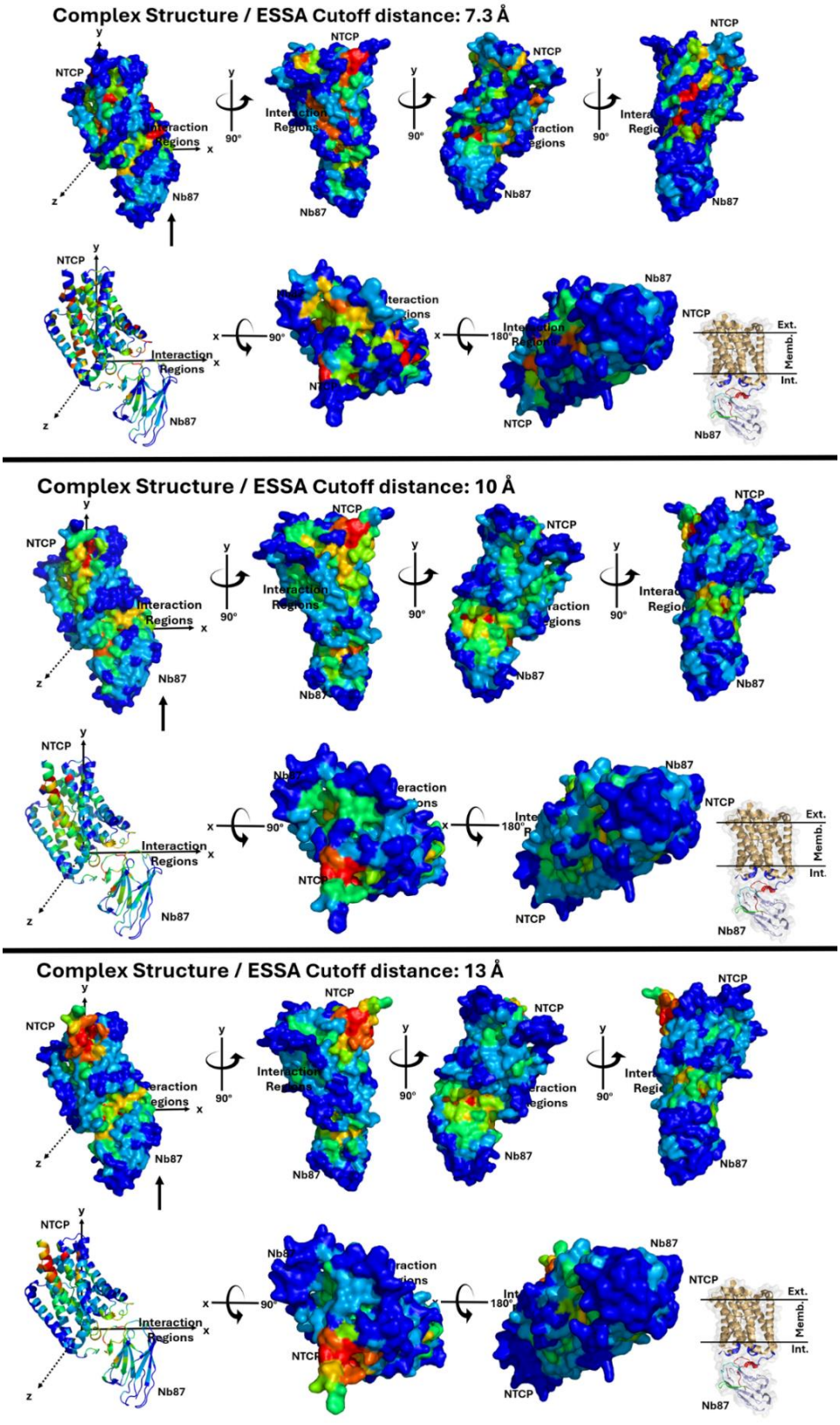


Figure 42. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

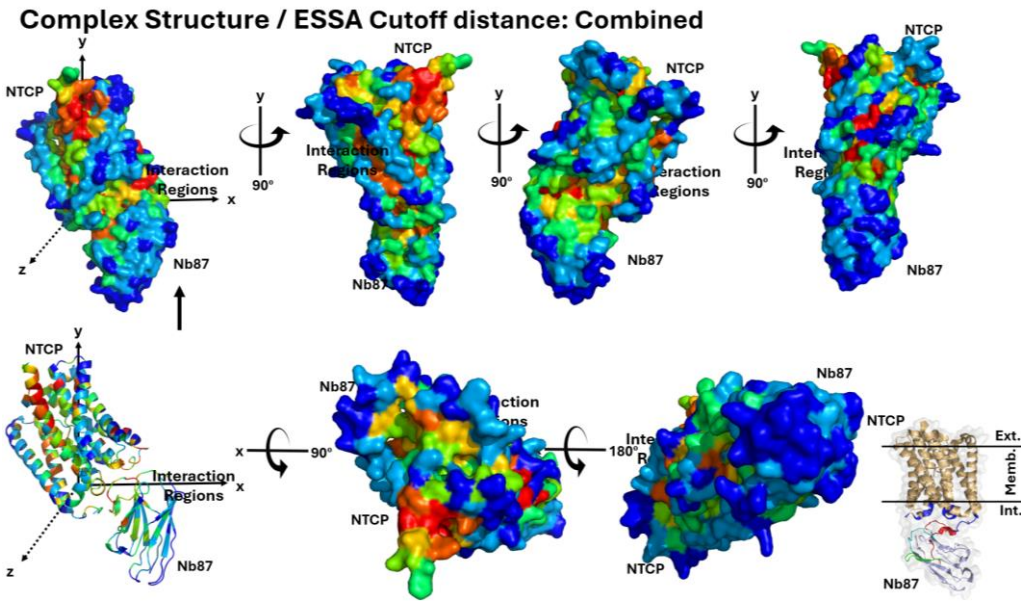


Figure 43. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

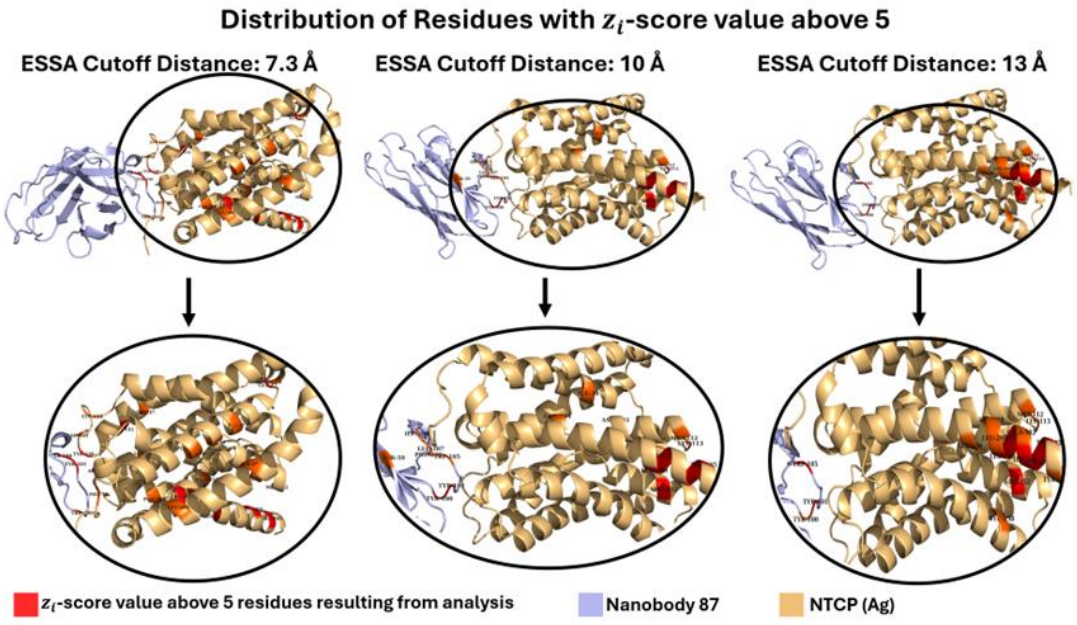


Figure 44. The most essential residues on the complex structure determined using different cutoff distance values in ESSA for the complex structure.

3.3. Nanobody 2-67 – SARS-CoV-2 Spike Protein Complex Structure

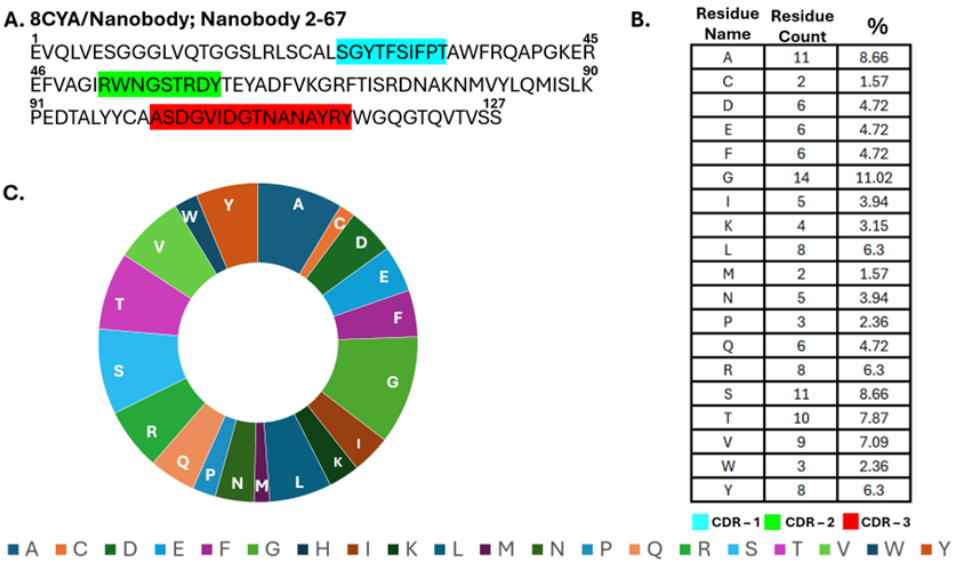
Throughout world history, humankind has faced many different global diseases. The consequences of the SARS-CoV-2 pandemic, which had an impact on all over the world in recent years and was declared by the World Health Organization (WHO), continue today (Xiang, et al. 2022).

This group of viruses, called coronavirus (sarbecovirus) that caused this pandemic, has attracted the attention of researchers due to its high genetic diversity, high recombination abilities, and rapid proliferation. There is a lot of recent research on this family of viruses. Various nanobodies have been used in research targeting the spike protein of the virus. Xiang *et al.* who conducted research on various nanobodies, selected Nanobody 2-67 that showed the strongest binding affinity among the studied nanobodies in the experiments (Xiang, et al. 2022). The results of the ESSA analysis on this nanobody-antigen complex are given in this section.

3.3.1. Nanobody 2-67 Structure

The nanobody 2-67 structure, which contains 127 residues in its crystal structure, forms a complex with the SARS-CoV-2 antigen spike protein. The count and percentage distributions of the remains found in the structure are shown in Figure 45. The residue found in the largest count and percentage in the structure is glycine. The least amount is cysteine. Data regarding the CDR regions of this structure, which contains counts and percentages of residues in the CDR regions, are given in Figure 46.

The ESSA results for three different cutoff distances are presented from different angles (0°, 90°, 180°, and 270°) through 3D structures in Figure 47. The combined result for Nanobody 2-67 from different angles is shown in Figure 48. The essential residues with z-score values above 5 obtained using different cutoff distance values are shown in Figure 49 and provided as a list in Table 11.



Residue Name	Residue Count	%
A	11	8.66
C	2	1.57
D	6	4.72
E	6	4.72
F	6	4.72
G	14	11.02
I	5	3.94
K	4	3.15
L	8	6.3
M	2	1.57
N	5	3.94
P	3	2.36
Q	6	4.72
R	8	6.3
S	11	8.66
T	10	7.87
V	9	7.09
W	3	2.36
Y	8	6.3

Figure 45. Representation of the primary sequence data regarding Nanobody 2-67. A: Nanobody 2-67 sequence [PDB ID: 8CYA, (Xiang, et al. 2022)] B: Count and percentage values for each residue in Nanobody 2-67, C: Pie-chart representation of the residue distribution in Nanobody 2-67.

The most residues were found as a result of ESSA analysis by taking the cutoff distance value as 7.3 Å (14 residues in total). Regarding the total count of residues, the closest result to the ESSA result with 7.3 Å was found in the ESSA analysis with the cutoff distance value being 13 Å (13 residues in total). The analysis with the least count of residues is the analysis performed by taking the cutoff distance value as 10 Å (8 residues in total).

The ESSA analysis with the cutoff distance value of 7.3 Å was able to detect all residues that were detected by the other cutoff distances, except for Gln39 and Lys43 residues, which were found only as a result of the ESSA analysis with the cutoff distance value of 10 Å. In addition, the ESSA analysis with the cutoff distance value of 7.3 Å detected at least one residue on each CDR region, whereas the others yielded residues either on CDR-2 (cutoff 10 Å) or CDR-1 and CDR-2 (cutoff 13 Å only).

Interestingly, these residues in Nanobody 2-67 structure are not located in the interaction region, according to PDBsum data. However, the cutoff distance value with the most residues in the CDR regions is 13 Å. As a result of the ESSA analysis with this parameter, most of the residues in the CDR-2 region appeared as essential residues.

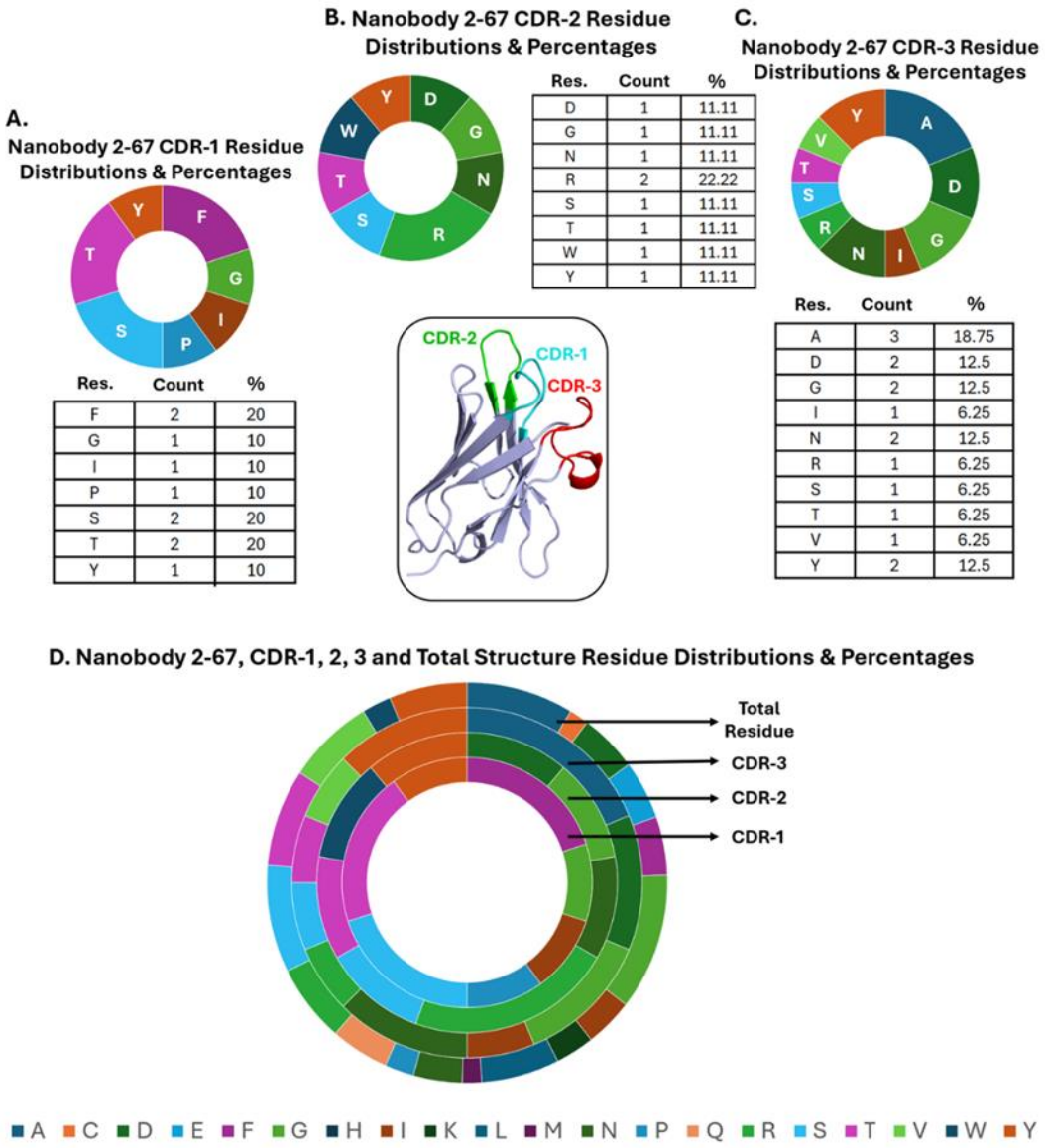


Figure 46. Representation of the primary sequence data of CDR regions located in the Nanobody 2-67 [PDB ID: 8CYA, (Xiang, et al. 2022)] Types and percentage distributions of residues in A: the CDR-1 region, B: the CDR-2 region, and C: the CDR-3 region. D: Pie-chart representation of the residue distribution in Nanobody 2-67 CDR regions.

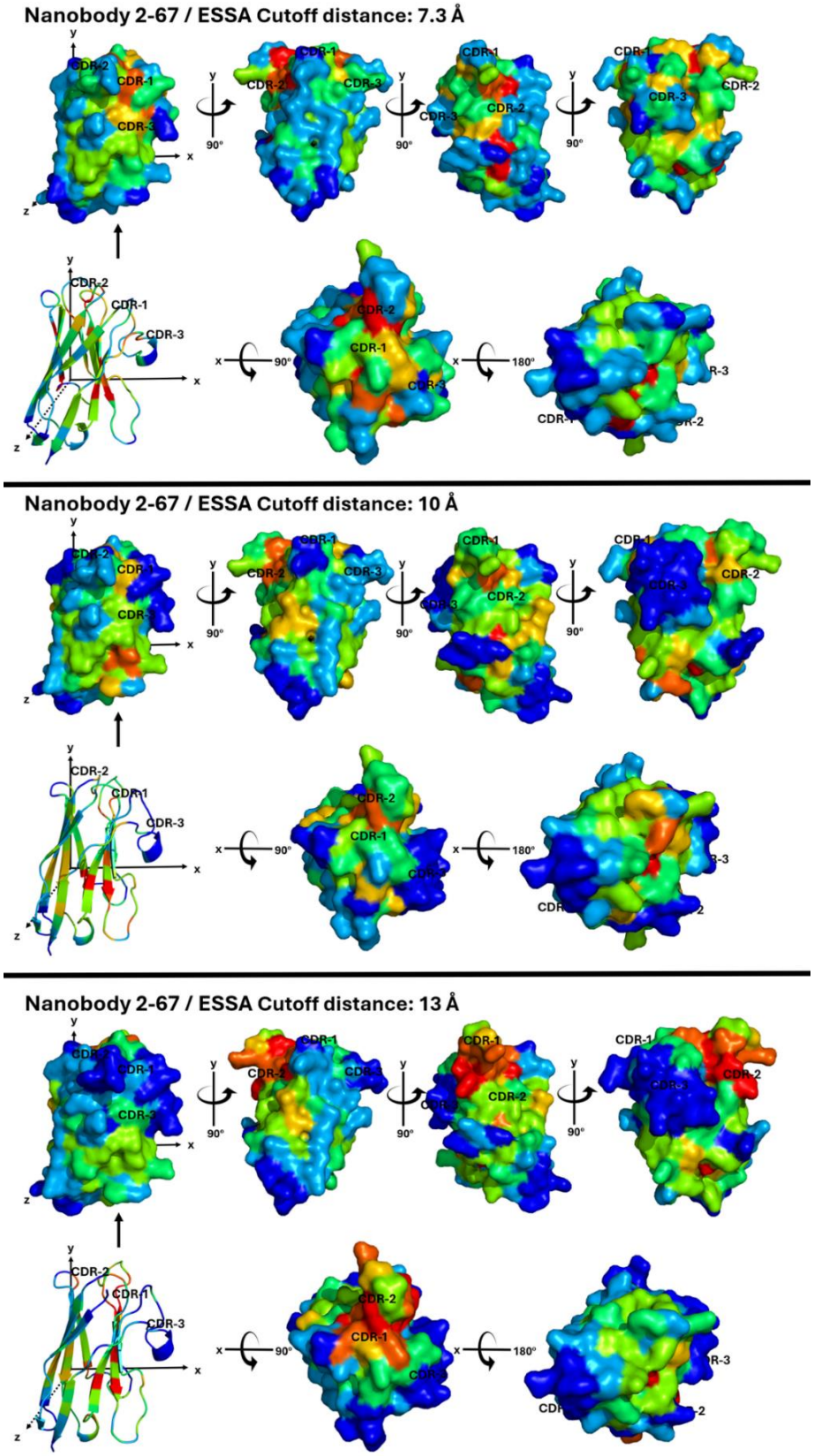


Figure 47. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff distance values for the Nanobody 2-67.

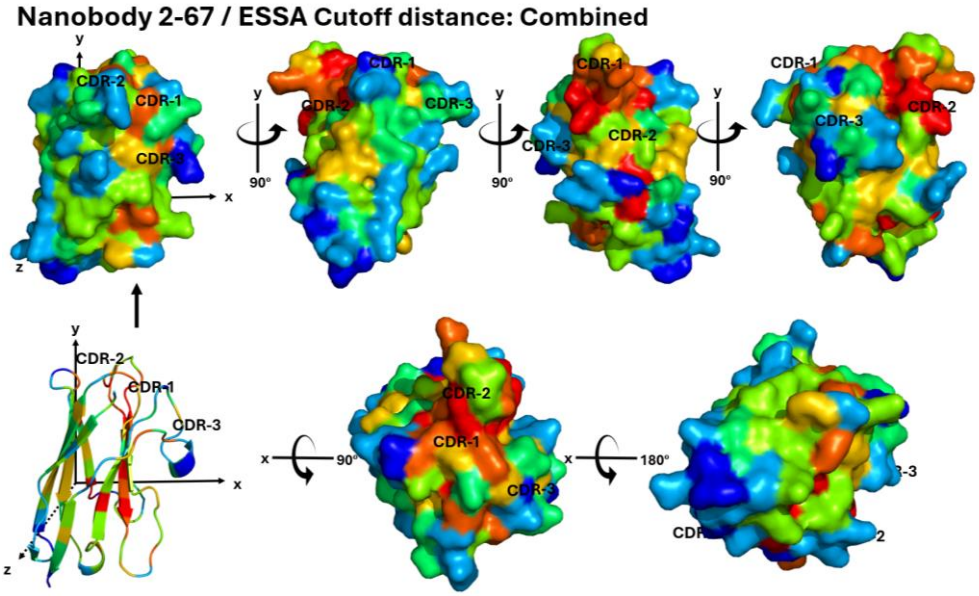


Figure 48. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the Nanobody 2-67.

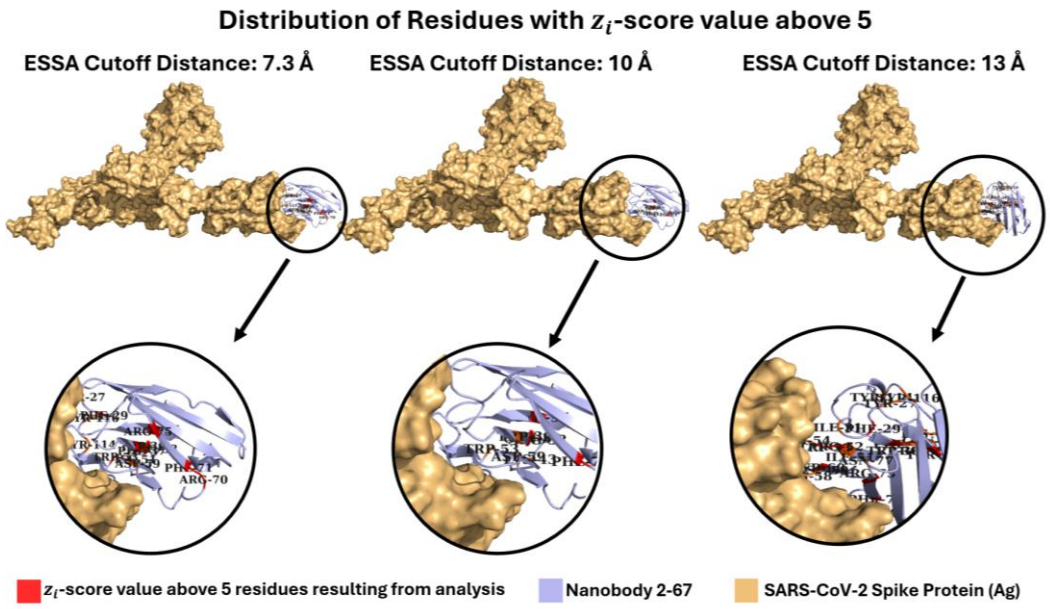


Figure 49. The most essential residues on the Nanobody 2-67 structure determined using different cutoff distance values in ESSA.

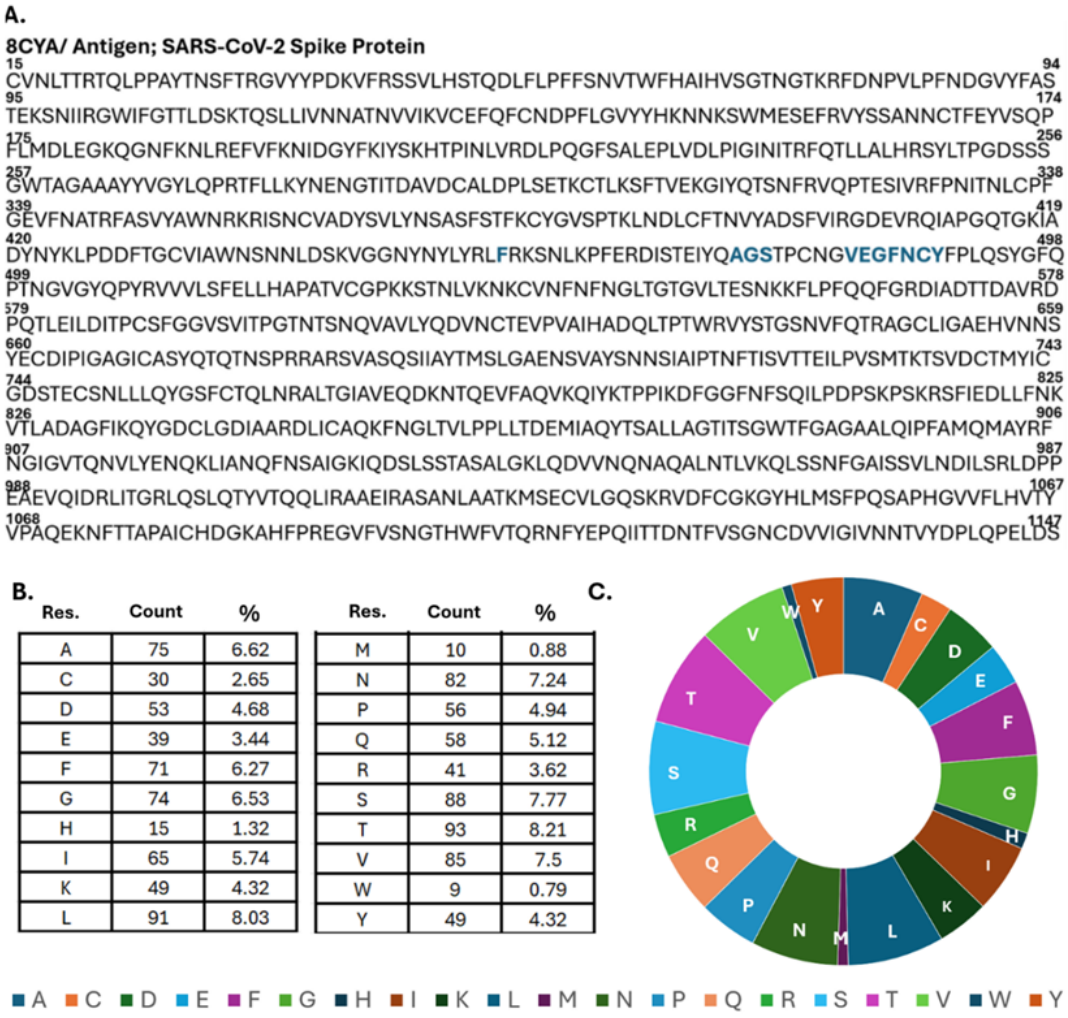
Table 11. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the Nanobody 2-67 (Only high z-score values are reported in the table).

ESSA Cutoff Distance: 7.3 Å	**ESSA Cutoff Distance: 10 Å**	**ESSA Cutoff Distance: 13 Å**
Nanobody 2-67	**Nanobody 2-67**	**Nanobody 2-67**
Tyr27 (CDR-1)	Trp36	Ile31 (CDR-1)
Phe29 (CDR-1)	Arg38	Trp36
Trp36	Gln39	Arg38
Phe37	Lys43	Arg52
Arg38	Trp53 (CDR-2)	Trp53 (CDR-2)
Ile51	Asp59 (CDR-2)	Asn54 (CDR-2)
Trp53 (CDR-2)	Phe71	Arg58 (CDR-2)
Asp59 (CDR-2)	Tyr97	Asp59 (CDR-2)
Arg70		Tyr60 (CDR-2)
Phe71		Phe71
Arg75		Arg75
Tyr97		Asn77
Tyr114 (CDR-3)		Tyr97
Tyr116		
Total: 14	**Total: 8**	**Total: 13**

3.3.2. SARS-CoV-2 Spike Protein (Antigen) Structure

The spike protein of the SARS-CoV-2 antigen, which has 1147 residues in its crystal structure, forms a homo-trimeric complex with Nanobody 2-67. The count and percentage distribution of the ruins found in the structure are given in Figure 50. As seen in this figure, the residue with the highest count is threonine whereas the lowest count is tryptophan in the structure. Additionally, the regions shown in blue color in the figure are the residues located in the interaction regions according to the PDBsum database. The analysis results based on different cutoff distance values are shown in Figure 51 from different angles. The combined ESSA result of the cutoff distance values is given in Figure 52. In addition, residues with z-score values above 5 from ESSA analysis performed for different cutoff distance values (7.3 Å, 10 Å, and 13 Å) are shown in Figure 53. The results shown in Table 12 are the residues with a z-score value above 5 in the

ESSA analysis results done by taking different cutoff distance values (7.3 Å, 10 Å, and 13 Å).



Res.	Count	%	Res.	Count	%
A	75	6.62	M	10	0.88
C	30	2.65	N	82	7.24
D	53	4.68	P	56	4.94
E	39	3.44	Q	58	5.12
F	71	6.27	R	41	3.62
G	74	6.53	S	88	7.77
H	15	1.32	T	93	8.21
I	65	5.74	V	85	7.5
K	49	4.32	W	9	0.79
L	91	8.03	Y	49	4.32

Figure 50. Representation of the primary sequence data regarding the antigen. A: Sequence [PDB ID: 8CYA, (Xiang, et al. 2022)] B: Count and percentage values for each residue in antigen, C: Pie-chart representation of the residue distribution in antigen.

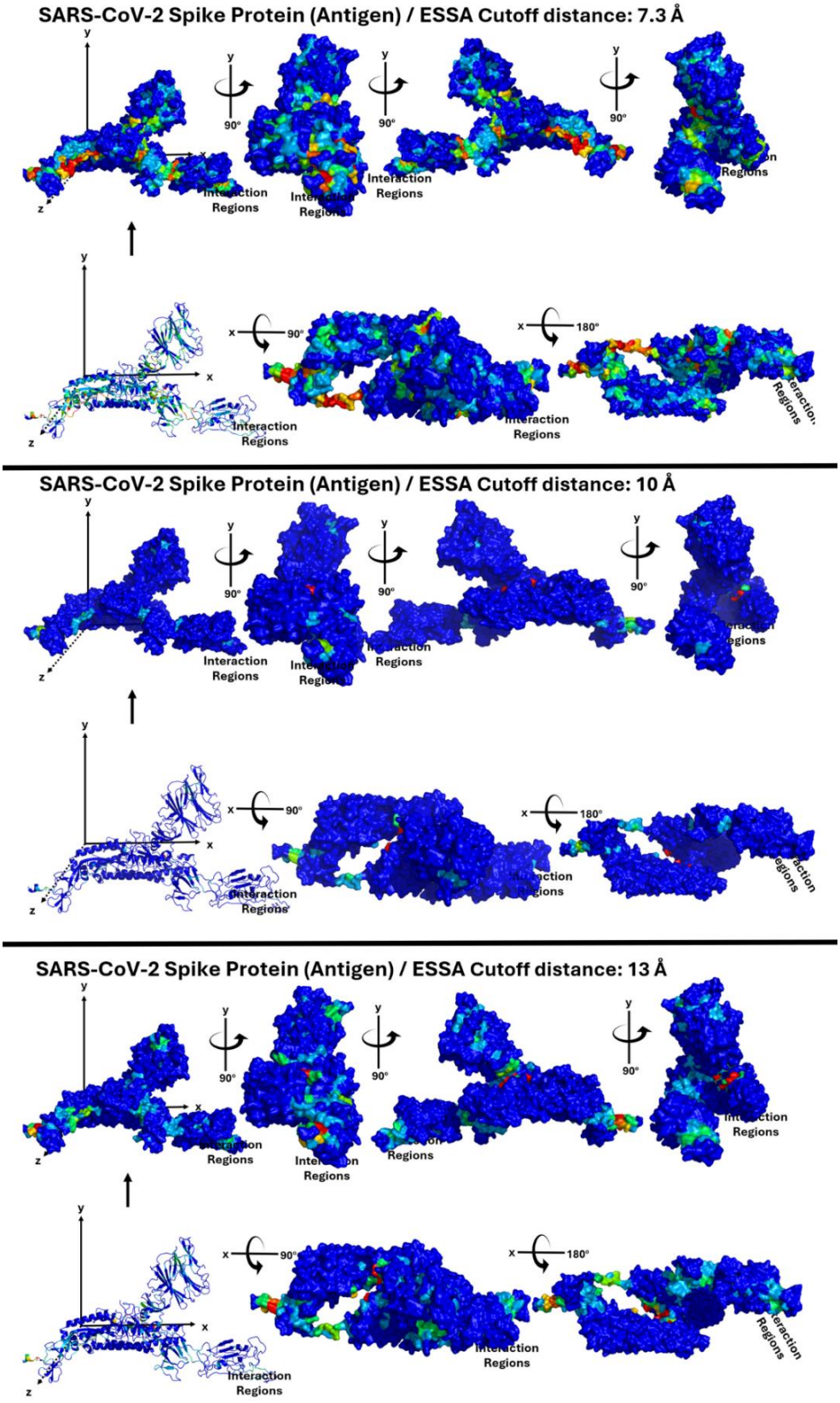


Figure 51. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the antigen.

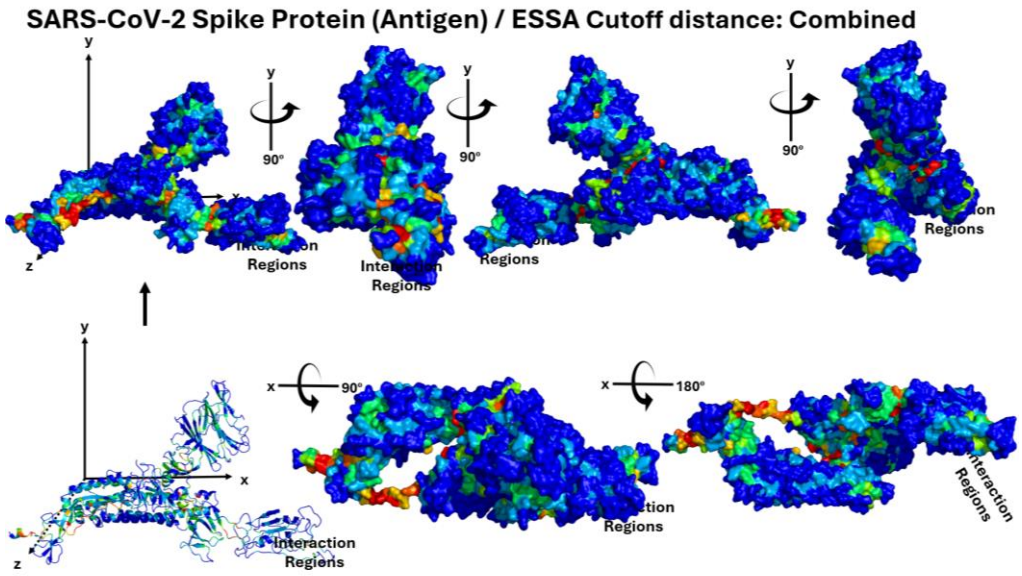


Figure 52. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the antigen.

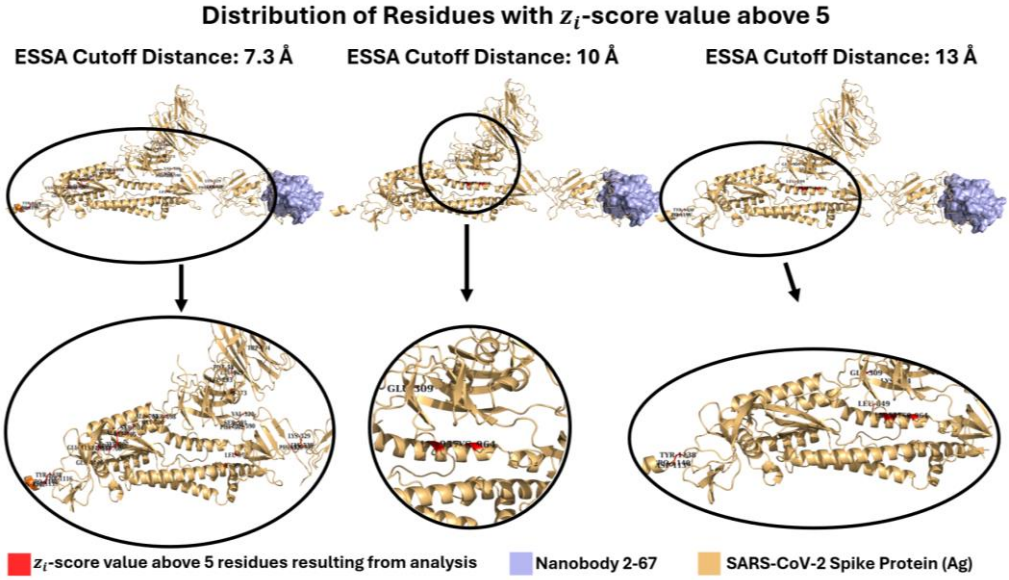


Figure 53. The most essential residues on the antigen structure determined using different cutoff distance values in ESSA.

Table 12. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the spike protein (Only high z-score values are reported in this table Ag: Antigen).

ESSA Cutoff Distance: 7.3 Å		**ESSA Cutoff Distance: 10 Å**	**ESSA Cutoff Distance: 13 Å**
Spike Protein (Ag)		**Spike Protein (Ag)**	**Spike Protein (Ag)**
Phe58	Trp104	Glu309	Lys304
Arg273	Leu293	Gln957	Glu309
Val320	Phe329	Lys1064	Leu849
Asn331	Lys528		Gln957
Lys529	Cys590		Asn960
Ser591	Phe592		Lys964
Leu629	Ser698		Tyr1138
Gly700	Ala701		Asp1139
Ser704	Val705		Pro1140
Ala706	Tyr707		
Ser708	Ser711		
Leu977	Arg1000		
Tyr1067	Gln1106		
Glu1111	Thr1116		
Tyr1138	Asp1139		
Pro1140	Gln1142		
Total: 32		**Total: 3**	**Total: 9**

Although the minimum count of residues (3 residues in total) was found in the ESSA analysis with a cutoff distance value of 10 Å, a considerable number of residues were found as essential as a result of the analysis with a cutoff distance value of 7.3 Å (32 residues in total). A relatively small count of residues was found (9 in total) in ESSA with the cutoff distance value of 13 Å. As in the nanobody 2-67 structure, none of the interaction sites given in PDBsum were found from ESSA analysis.

In addition to all these results, other residues in the B-chain (spike protein) with a z-score value over 5 as a result of ESSA analysis which are not in the interaction region with nanobody 2-67 and the residues in other chains that they interact with are shown in Table 13.

Table 13. Spike protein antigen (B-chain) and list of interaction residues with other monomers spike proteins (A and C chain) found in the structure.

B – Chain	Interaction Residues		Chain Name
Arg1039	Arg1039		C
Asn960	Ala570		
Leu849	Ile569		
Arg1039	Glu1031		A
Gln957	Arg765		
Thr961	Gln762		
Phe592	Met740	Thr859	
	Gly857		
Leu699	Tyr873	Ile788	
Gly700	Ile788		
Ala701	Gln787	Ile788	
Glu702	Ile788	Lys790	
Asn703	Ile788	Tyr789	
	Lys790	Gln787	
Val705	Thr883	Gln895	
	Tyr789	Ser884	
Ala706	Gln895		
Tyr707	Thr883	Ile896	
	Pro792	Pro897	
	Phe898	Asp796	
	Phe797		
Ser708	Pro897		
Asn709	Pro897	Asp796	
Ser711	Pro897		

3.3.3. Nanobody 2-67 – SARS-CoV-2 Spike Protein (Complex Structure)

The count and percentage distributions of residues in nanobody 2-67 and SARS-CoV-2 spike protein, which has 1274 residues in its crystal structure, are shown in Figure 54. The appearance of the complex structure colored according to the z-score values obtained by ESSA analysis with different cutoff distance values and the combined ESSA representation are shown in Figures 55 and 56, respectively. The representation of residues with z-score values above 5 on the structure in the analysis results for different cutoff distance values is also given in Figure 57.

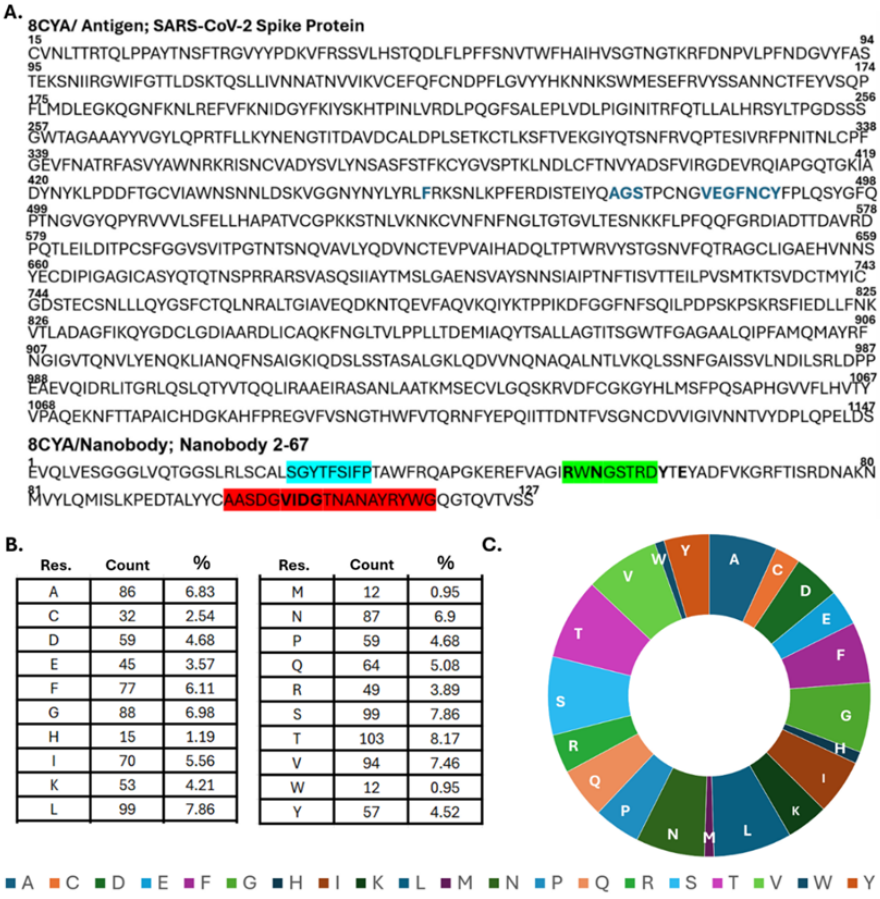


Figure 54. Representation of the primary sequence data regarding the Nanobody 2-67 and spike protein (Antigen) complex. A: Complex structure sequence [PDB ID: 8CYA, (Xiang, et al. 2022)] B: Count and percentage values for each residue in complex structure, C: Pie-chart representation of the residue distribution in complex structure.

The highest number of residues was found as a result of ESSA analysis by taking the cutoff distance value as 7.3 Å (21 residues in total). ESSA analysis performed by taking the cutoff distance value as 10 Å, with a total of three residues, is the value with the fewest residues. In the ESSA analysis with a cutoff distance value of 13 Å, a total of 10 residues were found. According to the PDBsum data, Phe329 interacts with Val 105 (Table 14) in the CDR-3 region of nanobody 2-67. Apart from this, according to PDBsum data, the rest of the essential residues are not in the interaction region. Three residues found in ESSA analysis with the cutoff distance value of 10 Å were also found as a result of ESSA analysis with the cutoff distance value of 13 Å.

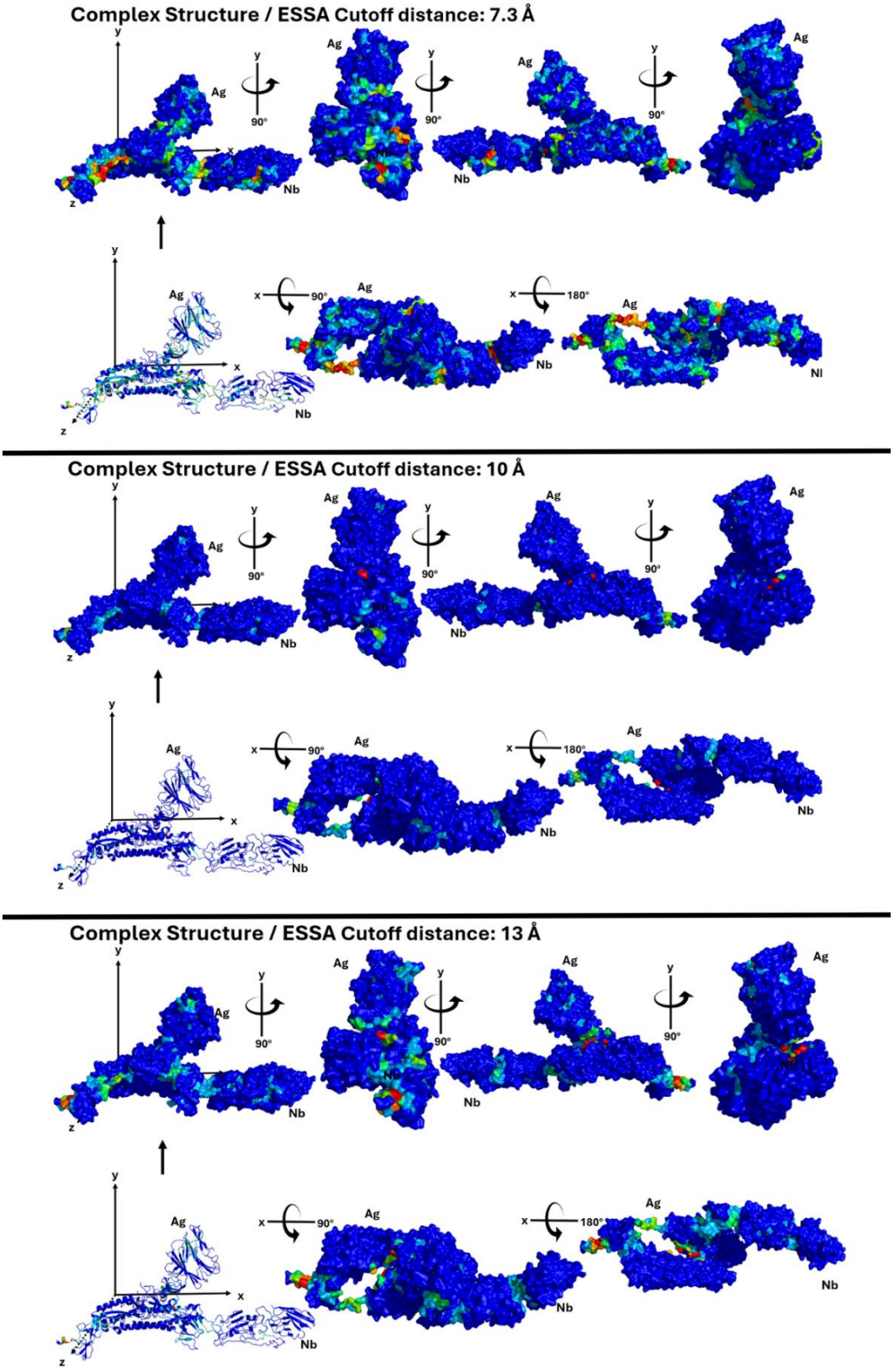


Figure 55. The ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

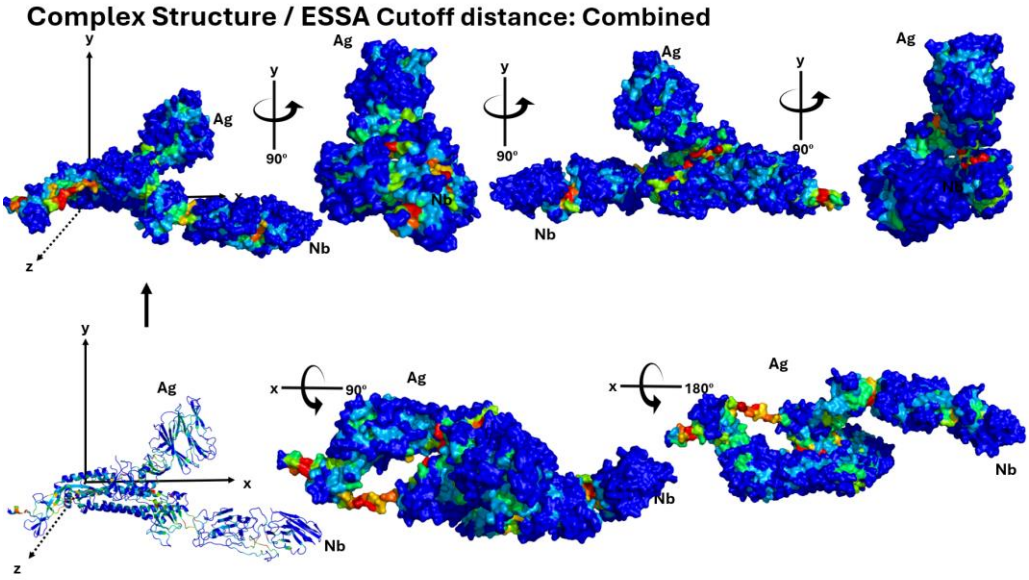


Figure 56. The combined-ESSA results using three different (7.3 Å, 10 Å, and 13 Å) cutoff values for the complex structure.

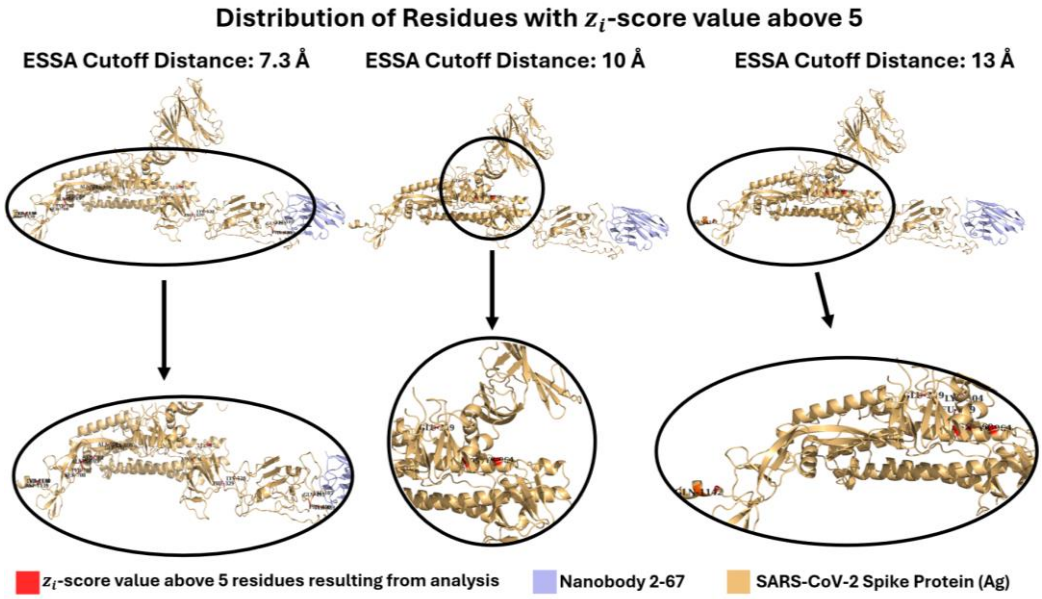


Figure 57. The most essential residues on the complex structure determined using different cutoff distance values in ESSA for the complex structure.

Table 14. The list of the most essential residues obtained from ESSA analysis based on three different cutoff distance values (7.3 Å, 10 Å, and 13 Å) for the complex structure (Only high z-score values are reported in this table. *: Residues specified as interaction sites in PDBsum. Ag: Antigen).

ESSA Cutoff Distance: 7.3 Å		**ESSA Cutoff Distance: 10 Å**	**ESSA Cutoff Distance: 13 Å**
Spike Protein (Ag)		**Spike Protein (Ag)**	**Spike Protein (Ag)**
Phe329*	Leu455	Glu309	Lys304
Phe456	Tyr489	Gln957	Glu309
Phe490	Gln493	Lys964	Leu849
Lys528	Leu629		Gln957
Ser698	Gly700		Asn960
Ala701	Ser704		Lys964
Val705	Ala706		Tyr1138
Tyr707	Ser708		Asp1139
Leu977	Arg1000		Pro1140
Tyr1138	Asp1139		Gln1142
Pro1140			
Total: 21		**Total: 3**	**Total: 10**

In summary, three nanobody-antigen complex structures were studied in this thesis. As a result of the ESSA analysis performed on these examined structures at different cutoff values (7.3 Å, 10 Å, and 13 Å), many potential interaction regions were found in the structures. Interestingly for all nanobodies studied in this thesis, the best results were obtained using the cutoff distance value of 7.3 Å, However, for the antigen and nanobody-antigen complexes a similar interpretation could not be made. It has been determined that these regions are the binding regions of nanobodies and other structures (such as ligands, other proteins) in complex with antigens. ESSA, a binding site prediction application method, was able to predict many interaction sites both in nanobodies and their antigens. Interestingly, it was observed that a potential binding site in an antigen might not disappear when the antigen is in complex with a molecule which is bound to another binding site. This study also showed that there are important regions in nanobodies other than CDR. This is an important finding for future research in understanding the molecular mechanisms of nanobody-antigen complexes.

CHAPTER 4

CONCLUSION AND RECOMMENDATIONS

This thesis focuses on the allosteric changes that occur in the internal dynamics of the structures in nanobody-antigen complexes and the potential importance of residues that appear as important regions even though they are not interaction sites. The Essential Site Scanning Analysis (ESSA) was applied to all structures in the dataset to predict potential binding sites. In order to evaluate the results and make them compatible with a standard, the z-score value for each residue was examined. Remains with a z-score value over 5 were taken into consideration and these residues were considered important.

Some important results of this study, conducted on the nanobodies and antigens in three different nanobody-antigen complexes, are as follows: As well known, nanobodies interacted with antigens through CDR regions. Critical residues on these CDRs were successfully detected using ESSA. Important residues outside the CDR regions in nanobodies were also found. It has been observed that these important residues, which are thought to be the signaling pathway, are mostly in the β-sheets of the nanobody. It is estimated that the presence of important regions other than the CDR regions in nanobody-antigen complex structures might shed light on future research about discovery of potential antigen-binders. Similarly, when these results are examined on an antigen basis, it is observed that some of the important regions seen in the antigen are the interaction regions with the nanobodies in the selected structures. It has been observed that some antigens have important regions outside the interaction region with the nanobody and that they interact with different structures from these regions, as seen in the case of von Willebrand Factor. Another important result is that after the structures form a complex, the interaction regions for other structures continue to emerge as important regions (See Caplacizumab - von Willebrand factor complex structure and Nanobody 87 - NTCP complex structure). In addition, the interaction regions of the spike protein with nanobody 2-67 and other monomers of the spike were successfully predicted. In future studies, if available, complex structures of the studied antigens with other

molecules should be examined using the ESSA method. For example, other molecules, such as ligands, antibodies, or proteins that interact with the von Willebrand Factor antigen should be examined. The PDB IDs of these molecules are 1OAK, 1FNS, 1FE8, 1M10, 1ATZ, 1UEX, 2ADF, and 3ZQG. Likewise, other molecules that bind to NTCP should also be examined. These molecules have PDB IDs such as 7FCI, 7ZYI, and 7VAD. The interactions of the complex structure of the SARS-CoV-2 spike protein antigen with the PDB ID 8CYA examined in this study with other structures should be examined in more detail. All ESSA calculations were performed on a single conformation of the structures studied in this thesis. Conformational sampling could be increased using molecular dynamics simulations or other sampling methods such as ClustENMD, then ESSA could be applied to new conformations.

As a result, all these findings are expected to shed light on the understanding of the working mechanisms of these structures and the therapeutics to be developed in future nanobody-antigen complex research.

REFERENCES

Abbas A. K., Lichtman A. H., and Pillai S., In *Basic Immunology,* Philadelphia: Elsevier, 2020.

Albert B. A. J., and Lewis J., In *Molecular Biology of the Cell. 4th Edition*, New York: Garland Science, 2002.

Arbabi-Ghahroudi M., "Camelid Single-Domain Antibodies: Historical Perspective and Future Outlook." *Frontiers in Immunology,* 2017: 1-8. https://doi.org/10.3389/fimmu.2017.01589

Atilgan A. R., Durell S. R., Jernigan R. L., Demirel M. C., Keskin O., and Bahar I., "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." *Biophysical Journal,* 2001: 505-515. https://doi.org/10.1016/s0006-3495(01)76033-x

Bahar I., Atilgan A. R., and Erman B., "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential." *Folding and Desing,* 1997: 173-181. https://doi.org/10.1016/S1359-0278(97)00024-2

Beisel R. W., In *Military Strategies for Sustainment of Nutrition and Immune Function in the Field,* by Institute of Medicine (US), Washington DC: National Academies Press (US), 1999.

Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., and Bourne P. E., "The Protein Data Bank." *Nucleic Acids Research,* 2000: 235-242. https://doi.org/10.1093/nar/28.1.235

Bo-Kyung J., Odongo S., Radwanska M., and Magez S., "Nanobodies: A review of Diagnostic and Therapeutic Applications." *International Journal of Molecular Sciences,* 2023: 2-22. https://doi.org/10.3390/ijms24065994

Celikel R., Varughese K. I., Yoshioka M. A., Ware J., and Ruggeri Z. M., "Crystal Structure of the von Willebrand Factor A1 Domain in Complex with the Function Blocink NMC-4 Fab." *Nature Structural & Molecular Biology,* 1998: 189-194. https://doi.org/10.1038/nsb0398-189

Celikel R., Ruggeri Z. M., and Varughese K. I., "Von Willebrand Factor Conformation and Adhesive Function is Modulated by an Internalized Water Molecule." *Nature Structural & Molecular Biology,* 2000: 881-884. https://doi.org/10.1038/79639

Chen Y., Fleetwood O., Pérez-Conesa S., and Delemotte L., "Allosteric Effect of Nanobody Binding on Ligand-Specific Active State of the β2 adrenergic Receptor." *J. Chem. Inf. Model.,* 2021: 6024-6037. https://doi.org/10.1021/acs.jcim.1c00826

Cooper A., and Dryden D. T. F., "Allostery Without Conformational Change, A Plausible Model." *European Biophysics Journal,* 1984: 103-109. https://doi.org/10.1007/bf00276625

Dror R. O., Arlow D. H., Borhani D. W., Jensen M. Ø., Piana S., and Shaw D. E., "Identification of Two Distinct Inactive Conformations of the β2-Adrenergic Receptor Reconciles Structural and Biochemical Observations." *PNAS,* 2009: 4689-4694. https://doi.org/10.1073/pnas.0811065106

Franzel A., Schirrmann T., and Hust M., "Phage Display-Derived Human Antibodies in Clinical Development and Therapy." *MABS,* 2016: 1177-1194. https://doi.org/10.1080/19420862.2016.1212149

Goutam K., Lelasi F. S., Pardon E., Steyaert J., and Reyes N., "Structural Basis of Sodium-Dependert Bile Salt Uptake into the Liver." *Nature,* 2022: 1015-1020. https://doi.org/10.1038/s41586-022-04723-z

Harris L. J., Larson S. B., Hasel K. W., and McPherson A., "Refined Structure of an Intact IgG2a Monoclonal Antibody." *Biochemistry,* 1997: 1581-1597. https://doi.org/10.1021/bi962514+

Hassanzadeh-Ghassabeh G., Devoogdt N., Pauw De P., Cécile V., and Muyldermans S., "Nanobodies and Their Potential Applications." *Nanomedicine,* 2013: 1013-1026. https://doi.org/10.2217/nnm.13.86

Janeway C. A., Travers P., Walport M., and Shlomchik M. J., In *Immunobiology, 5th Edition,* New York: Garland Science, 2001.

Jovčevska I., and Muyldermans S., "The Therapeutic Potential of Nanobies." *BioDrug,* 2020: 11-26. https://doi.org/10.1007/s40259-019-00392-z

Kaufmann S. H. E., “Remembering Emil von Behring: from Tetanus Treatment to Antibody Cooperation with Phagocytes.” *American Society for Microbiology,* 2017: 1-6. https://doi.org/10.1128/mbio.00117-17

Kaynak B. T., Bahar I., and Doruker P., “Essential Site Scanning Analysis: A new Approach for Detecting Sites That Module the Dispersion of Protein Global Motions.” *Computational and Structural Biotechnology Journal,* 2020: 1577-1586. https://doi.org/10.1016/j.csbj.2020.06.020

Kelow S. P., Adolf-Bryfogle J., and Dunbrack R. L., “Hiding in Plain Sight: Structure and Sequence Analysis Reveals the Importance of the Antibody DE Loop for Antibody-Antigen Binding.” *mAbs,* 2020: 1-21. https://doi.org/10.1080%2F19420862.2020.1840005

Lee H. T., Park U. B., Jeong T. J., Gu N., Lee S. H., Kim Y., and Heo Y., “High-Resolution Structure of the vWF A1 Domain in Complex with Caplacizumab, the First Nanobody-Based Medicine for Treating Acquired TTP.” *Biochemical and Biophysical Research Communications,* 2021: 49-55. https://doi.org/10.1016/j.bbrc.2021.06.030

Liu H., Irobalieva R. N., Sørensan R. B., Nosol K., Mukherjee S., Agrawal P., Stieger B., Kossiakoff A., and Locher K. P., “Structure of Human NTCP Reveals the Basis of Recognition and Sodium-Driven Transport of Bile Salts into the Liver.” *Nature, Cell Research,* 2022: 773-776. https://doi.org/10.1038%2Fs41422-022-00680-4

Liu J., and Nussinov R., “Allostery: An Overview of Its History, Concepts, Methods, and Applications.” *PLOS Computational Biology,* 2016: 1-5. https://doi.org/10.1371/journal.pcbi.1004966

López-Blanco J. R., and Chacón P., “New Generation of Elastic Network Models.” *Current Opinion in Structural Biology,* 2016: 46-53. https://doi.org/10.1016/j.sbi.2015.11.013

Mitchell L. S., and Colwell L. J., “Comparative Analysis of Nanobody Sequence and Structure Data.” *Proteins Structure Function Bioinformatics,* 2018: 697-706. https://doi.org/10.1002/prot.25497

Nicholson B. L., “The Immune System.” *Essays in Biochemistry,* 2016: 275-301. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5091071/

Oreste U., Ametrano A., and Coscia M. R., “On Origin and Evolution of the Antibody Molecule.” *Biology,* 2021: 2-18. https://doi.org/10.3390/biology10020140

Park J. H., Iwamoto M., Yun H. J., Uchikubo-Kamo T., Son D., Jin Z., Yoshida H., Ohk M., Ishimoto N., Mizutani K., Oshima M., Muramatsu M., Wakita T., Shirouzu M., Liu K., Uemura T., Nomura N., Iwata S., Watashi K., Tame R. H. J., Nishiawa T., Lee W., and Park S. Y., "Structural Insights into the HBV Receptor and Bile Acid Transporter NTCP." *Nature,* 2022: 1027-1031. https://doi.org/10.1038%2Fs41586-022-04857-0

Rao D. S., "Chapter 19-Overview and Compartmentalization of the Immune System." In *Hematology: Basic Principles and Practice, 7th Edition,* by Edward Ronald Hoffman, New York: Elsevier, 2018: 199-209. http://dx.doi.org/10.1016/B978-0-323-35762-3.00019-6

Rapley R., "The Biotechnology and Applications of Antibody Engineering." *Molecular Biotechnology,* 1995: 139-154. https://doi.org/10.1007/bf02789110

Sargentini-Maier M. L., Decker De P., Tersteeg C., Canvin J., Callewaert F., and Winter De H., "Clinical Pharmacology of Caplacizumab for the Treatment of Patients with Acquired Thrombotic Thrombocytopenic Purpura." *Expert Rev. Clin. Pharmacol.,* 2019: 537-545. https://doi.org/10.1080/17512433.2019.1607293

Schroeder Jr. H. W., and Cavacini L., "Structure and Function of Immunoglobulins." *Journal of Allergy Clin Immunol.,* 2010: 41-52. https://doi.org/10.1016/j.jaci.2009.09.046

Schumacher D., Helma J., Schneider A. F. L., Leonhardt H., and Hackenberger C. P. R., "Nanobodies: Chemical Functionalization Strategies and Intracellular Applications." *Angewandte Chemie,* 2018: 2314-2333. http://dx.doi.org/10.1002/anie.201708459

Status D. P., Strachan R. T., Manglik A., Pani B., Kahsai A. W., Kim T. H., Wingler L. M., Ahn S., Chatterjee A., Masoudi A., Kruse A. C., Pardon E., Steyaert J., Weis W. I., Prosser R. S., Kobilka B. K., Costa T., and Lefkowitz R. J., "Allosteric Nanobodies Reveal the Dynamic Range and Diverse Mechanisms of G-Protein-Coupled Receptor Activation." *Nature,* 2016: 448-535. http://www.nature.com/doifinder/10.1038/nature18636

Stoumpos A. I., Kitsios F., and Talias M. A., "Digital Transformation in Healthcare: Technology Acceptance and Its Applications." *Int J Environ Res Public Health,* 2023: 2-44. https://doi.org/10.3390/ijerph20043407

Tsumoto K., and Kuroda D., In *Computer-Aided Antibody Desing,* US: Human Press, Springer Protocols, 2022.

Tulchinsky T. H., and Varavikova A. E., "A History of Public Health," *The New Public Health,* 2014: 1-42. https://doi.org/10.1016/B978-0-12-415766-8.00001-X

Wang Y., Fan Z., Shao L., Kong X., Hou X., Tian D., Sun Y., Xiao Y., and Yu L., "Nanobody-Derived Nanobiotechnology Tool Kits for Diverse Biomedical and Biotechnology Applications." *International Journal of Nanomedicine,* 2016: 3287-3303. https://doi.org/10.2147/IJN.S107194

Weller M. G., "Ten Basic Rules of Antibody Validation." *Analytical Chemistry Insights,* 2018: 1-5. https://doi.org/10.1177/1177390118757462

Wodak S. J., Paci E., Dokholyan N. V., Berezovsky I. N., Horovitz A., Li J., Hilser V. L., et al., "Allostery in Its Many Disguises: From Theory to Applications." *Structure,* 2019: 2-26. https://doi.org/10.1016/j.str.2019.01.003

Xiang Y., Huang W., Liu H., Sang Z., Nambulli S., Tubiana J., Williams Jr., K. L., Duprex W. P., Schneidman-Duhovny D., Wilson I. A., Taylor D. J., and Shi Y., "Superimmunity by Pan-Sarbecovirus Nanobodies." *Cell Reports,* 2022: 1-16. https://doi.org/10.1016/j.celrep.2022.111004

Yang J., Zhang Z., Yang F., Zhang H., Wu H., Zhu F., and Xue W., "Computational Desing and Modeling of Nanobodies Toward SARS-CoV-2 Receptor binding Domain." *Chemical Biology & Drug Desing,* 2021: 1-18. https://doi.org/10.1111/cbdd.13847

Zheng F., Pang Y., Li L., Pang Y., Zhang J., Wang X., and Raes G., "Applications of Nanobodies in Brain Diseases." *Frontiers in Immunology,* 2022: 1-17. https://doi.org/10.3389/fimmu.2022.978513