NOVA SCHOOL OF SCIENCE & TECHNOLOGY

CHEMISTRY DEPARTMENT

SUSANA CATAPIRRA MAGESSI PARREIRAS

Bsc in Biochemistry

RATIONALLY DESIGNED ANTIVIRAL PROTEINS TARGETING SARS-COV-2

MASTER DEGREE IN BIOCHEMISTRY

Universidade NOVA de Lisboa

March, 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

CHEMISTRY DEPARTMENT

RATIONALLY DESIGNED ANTIVIRAL PROTEINS TARGETING SARS-COV-2

SUSANA CATAPIRRA MAGESSI PARREIRAS
Bsc in Biochemistry

Supervisor: Diana Andreia Pereira Lousa, PhD.
Researcher
Instituto de Tecnologia Química e Biológica António Xavier

Co-supervisor: João Filipe Bogalho Vicente, PhD.
Auxiliary Investigator
Instituto de Tecnologia Química e Biológica António Xavier

Júri:

Presidente: Professora Doutora Sofia Rocha Pauleta

Arguentes: Doutor Bruno Lourenço da Silva Victor

Vogais: Doutora Diana Andreia Pereira Lousa

MASTER DEGREE IN BICHEMISTRY

Universidade NOVA de Lisboa
March, 2022

Rationally Designed Antiviral Proteins Targeting SARS-CoV-2

Agradecimentos

Já passou um ano e meio desde que iniciei este percurso e muito se passou. Houve vários percalços pelo caminho, momentos de grande agitação e ansiedade, mas também foram tempos de uma grande aprendizagem, onde adquiri conhecimentos e descobri uma nova área que nunca eu tinha pensado em vir a trabalhar e ainda por cima gostar. E todo este ano e meio e as coisas que aconteceram não teriam sido o mesmo sem as pessoas com que me cruzei, que contribuíram para esta incrível viagem e por isso não podia deixar passar sem deixar um agradecimento muito especial a todas elas.

Primeiramente, queria agradecer aos meus orientadores, Dra. Diana Lousa e Dr. João Vicente, por me terem aceitado para trabalhar com eles neste projeto, por me terem sempre acompanhado ao longo deste tempo, por todos os ensinamentos e conhecimentos que me transmitiram, mas principalmente por terem dado um verdadeiro sentido a palavra orientador. Também gostaria de agradecer ao Professor Doutor Cláudio Soares e a Professora Doutora Maria Arménia Carrondo por me terem recebido nos seus grupos e por terem me permitido ter toda esta experiência e aprendizagem.

A todas as pessoas do Protein Modelling Laboratory o meu muito obrigada, mas em especial a Ana Carolina Buga, a Mariana Valério, ao Carlos Cruz e ao Caio Souza por me terem ajudado em todos os momentos, por terem paciência para todas as dúvidas e ajudas que eu precisei, e sem eles nada deste ano e meio tinha sido possível.

A todas as pessoas da Macromolecular Crystallography Unit mas em especial a Dalila, ao João, a Carlota, ao André, a Andreia e ao Bruno por todas as ajudas, conversas, companhia e por terem tornado os dias no laboratório mais fáceis, leves e divertidos.

Todo o meu percurso académico e pessoal não teria sido o mesmo e às vezes nem possível sem várias pessoas muito importantes que contribuíram para ele e para quem eu sou hoje e também não poderia deixar de lhes agradecer.

"A família é a base de tudo e com ela qualquer coisa se pode conquistar e superar". Um gigante obrigado a minha família, aos meus avós, aos meus tios e ao meu primo, mas em especial a minha mãe e ao meu pai por terem sido sempre os pilares da minha vida, por terem permitido sempre que eu voasse e alcança-se todos os meus objetivos, por estarem sempre presentes nos bons e nos maus momentos, por me apoiarem e amarem incondicionalmente.

Ao meu namorado, Rodrigo Varges, sem ele e sem o seu apoio nada era possível. Muito obrigada por estar sempre ao meu lado, pela paciência, apoio, companheirismo, por acreditar sempre em mim, por me acalmar nos momentos difíceis e obrigada por ter sido o melhor que a faculdade me deu.

Aos meus amigos da licenciatura, João Lopes, Joana Galhana, Susana Santos, mas em especial aos meus hidrolisados, Sara Fialho, Inês Paiva, João Alves, Rodrigo Varges, João Cruz, Margarida Monteiro, Sofia Benedito e Rita Baptista, obrigada por estarem sempre presentes mesmo que longe, por me terem sempre apoiado, por me terem aceitado como eu sou, por todo o apoio, brincadeiras, conversas, todos os cafés e noites no sete, por todos os momentos e memórias partilhadas. Obrigada pela vossa amizade e por se terem tornado tudo mais fácil.

Aos meus amigos do mestrado, Cristina Nunes, Filipa Moura, Catarina Alves e Nelson Pereira, por terem tornado estes dois anos mais fáceis, por todas as ajudas e paciência, pela companhia, que mesmo cada um em sua casa estavam sempre presentes, pelos jogos de UNO e por todos os momentos vividos. Obrigada pela vossa amizade.

Por fim, mas não menos importante, obrigada as minhas duas estrelinhas, que estejam onde estiverem espero que estejam orgulhosos de mim.

Abstract

SARS-CoV-2 is the virus responsible for the current COVID-19 pandemic, which has caused >400 million infections and >5 million deaths, as of February 2022. Despite the vaccination efforts, it remains urgent to develop strategies to control the infection and treat patients. SARS-CoV-2 is a positive-sense RNA virus that is part of the Coronaviridae family. Its outer structure is spherical, it is encapsulated by a viral membrane and, in order to infect the host cell, it needs to fuse its membrane with the host cell membrane.

One of the proteins that is attached to the viral membrane of the virus is the spike (S) protein, which is composed of two subunits: S1, containing a receptor binding domain (RBD) responsible for binding to the host cell receptor, and S2, that facilitates membrane fusion between the viral and host cell membranes. Thus, this protein is primarily responsible for the ability of the virus to enter the host cells, making it one of the most promising therapeutic targets of coronaviruses.

The goal of this work was to design and produce antiviral proteins that might prevent the interaction between the two proteins and therefore block infection. These proteins are engineered to bind to the RBD region and block its interaction with the host receptor, the angiotensin converting enzyme-2 (ACE2).

In a first step, several antiviral proteins were computationally designed with the Rosetta program, based on the interactions between ACE2 and the receptor-binding domain. Next, molecular dynamics (MD) simulations of three candidates were performed, both free in solution as well as in complex with the RBD, in order to test their interaction with the RBD. This was followed by experimental validation that began with the expression and purification of the three candidates. After obtaining pure fractions, the secondary structure and thermal stability of these proteins were tested by far-UV circular dichroism spectropolarimetry and differential scanning fluorimetry, respectively. In order to assess the affinity of each candidate for the RBD, surface plasmon resonance was employed. Finally, neutralization assays were performed to study the neutralization ability of the proteins. The experimental results show that one of the designed proteins is a promising therapeutic lead that will be further improved in the future.

Keywords: SARS-CoV-2, COVID-19, Pandemic, Coronavirus Disease-2019, severe acute respiratory syndrome coronavirus 2, Spike protein (S), protein structure, Receptor binding domain (RBD), Angiotensin-converting enzyme-2 (ACE2), Helix-Turn-Helix (HTH), Molecular dynamics (MD) simulation, Surface plasmon resonance (SPR), Circular dichroism (CD), neutralization assays.

Resumo

O SARS-CoV-2 é o vírus responsável pela atual pandemia COVID-19, que causou >400 milhões de infeções e >5 milhões de mortes, a datar de Fevereiro de 2022. Apesar dos esforços de vacinação, continua a ser urgente desenvolver estratégias para controlar a infeção e tratar as pessoas infetadas que apresentam sintomas. O SARS-CoV-2 é um vírus de RNA de sentido positivo que pertence à família Coronaviridae. A sua estrutura externa é esférica, sendo encapsulado por uma membrana viral e, de forma a infetar a célula hospedeira, precisa de fundir a sua membrana com a membrana desta célula.

Uma das proteínas que está ligada à membrana viral do vírus é a proteína spike (S), que é composta por duas subunidades: S1, que contém um domínio de ligação ao recetor (RBD) responsável pela ligação ao recetor da célula hospedeira, e S2, que facilita a fusão membranar entre as membranas do vírus e da célula do hospedeiro. Assim, esta proteína é a principal responsável pela capacidade do vírus de entrar nas células hospedeiras, tornando-a num dos alvos terapêuticos mais promissores dos coronavírus.

O objetivo deste trabalho era conceber e produzir proteínas antivirais que pudessem impedir a interação entre as duas proteínas e, assim, bloquear a infeção. Estas proteínas são concebidas para se ligarem à região do RBD e bloquear a sua interação com o recetor hospedeiro, a enzima conversora da angiotensina-2 (ACE2).

Numa primeira etapa, várias proteínas antivirais foram computacionalmente concebidas com o programa Rosetta, com base nas interações entre a ACE2 e o domínio de ligação ao recetor da proteína S. Posteriormente, realizaram-se simulações de dinâmica molecular (MD) de três candidatos, tanto livres em solução como em complexo com o RBD, a fim de testar a sua interação com esta proteína. Seguiu-se uma validação experimental que começou com a expressão e purificação dos três candidatos. Após a obtenção de frações puras, a estrutura secundária e a estabilidade térmica destas proteínas foram testadas, respetivamente, por espetropolarimetria de dicroísmo circular no UV distante e fluorimetria de varrimento diferencial. A fim de avaliar a afinidade de cada candidato para com o RBD, foi utilizada a ressonância plasmónica de superfície. Finalmente, foram realizados ensaios de neutralização para estudar a capacidade destas proteínas se ligarem ao RBD. Os resultados experimentais mostram que uma das proteínas concebidas representa uma estratégia terapêutica promissora que será ainda melhorada no futuro.

Palavras-chave: SARS-CoV-2, COVID-19, Pandemia, Síndrome respiratória aguda grave coronavírus 2, Proteína Spike (S), estrutura proteica, Domínio de ligação ao receptor (RBD), enzima conversora da angiotensina-2 (ACE2), Helix-Turn-Helix (HTH), Simulação de Dinâmica Molecular (MD), Ressonância Plasmónica de Superfície (SPR), Dicroísmo Circular (CD), ensaios de neutralização.

Index

Figures Index

Table Index

Equations Index

List of Abbreviations

+ssRNA – Positive-sense Single-stranded RNA genome

α-CoV – Alphacoronavirus

β-CoV – Betacoronavirus

γ-CoV – Gammacoronavirus

δ-CoV – Deltacoronavirus

6-HB – Six-helical Bundle

ACE2 – Angiotensin-converting Enzyme 2

AMBER – Assisted Model Building and Energy Refinement

ARDS – Acute Respiratory Distress Syndrome

ARTIs – Acute Respiratory Tract Infections

BSA – Bovine Serum Albumin

C1 – Flat Carboxymethylated

CD – Central Domain

cDMEM – Dulbecco's Modified Eagle Medium

CFF - Consistent Force Field

CHARMM – Chemistry at Harvard Macromolecular Mechanics

CLD – Collectrin-like Domain

CM5 – Carboxymethylated Dextran

COVID-19 – Coronavirus 2019 Disease

CoVs – Coronavirus

CP – Cytoplasmic Domain

CPD – Computational Protein Design

CR – Connecting Region

CRP – C-reactive Protein

CT – Computerized Tomography

CTD – C-terminal Domain

DC-SIGN – Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin

DPP4 – Dipeptidyl Peptidase 4

DTT – Dithiothreitol

EDTA – Ethylenediamine Tetraacetic Acid

ECEPP – Empirical Conformational Energy Program for Peptides

ECMO – Extracorporeal Membrane Oxygenation (ECMO)

ELISA – Enzyme Linked Immunosorbent Assay

E protein – Envelope Protein

ERGIC – Endoplasmic Reticulum-Golgi Intermediate Compartment

F1 – Short Dextran

far-UV CD – far-UV Circular Dichroism

FD – Fusion Domain

FF – Force Fields

FL – Fusion Loop

FP – Fusion Peptide

HBS-N – HEPES-buffered Saline

HCl – Hydrochloric Acid

hCoVs – Human Coronavirus

HPA – Hydrophobic Mono Layer

HR1/2 – Heptad Repeat 1/2

HTH – Helix-Turn-Helix

IB – Inclusion Bodies

IF – Insoluble Fraction

IFN – Type I Interferons

IMAC – Immobilized Metal Ion Affinity Chromatography

IPTG – Isopropyl β-D-1thiogalactopyranoside

J1 – Unmodified Gold Surface

Kan – Kanamycin

L-SIGN – Liver/lymph node-specific ICAM-3 grabbing non-integrin

L1 – Lipophilic Dextran

LB – Luria-Bertani Medium

LJ – Leonard-Jones

LNPs – Lipid Nanoparticles

MCP – Monocyte Chemoattractant Protein

MD – Molecular Dynamics

MERS – Middle East Respiratory Syndrome

$MgCl_2$ – Magnesium Chloride

MIP – Macrophage Inflammatory Protein

MM – Molecular Mechanics

MMFF – Merck Molecular Force Field

MOE – Molecular Operating Environment

MOF – Multiple Organ Failure

MOI – Multiplicity of Infection

M protein – Membrane Protein

NaCl - Sodium Chloride

nanoDSF – Nano Differential Scanning Fluorimetry

NFkB – Nuclear Factor Kappa B

NHC – Ribonucleoside analog β-D-N4-hydroxycytidine

NK – Natural Killer

N protein – Nucleocapsid Protein

NPs – Nucleoproteins**Nsps –** Non-structural Proteins

NTA – Nickel Chelation

NTD – N-terminal Domain

OPLS – Optimized Potential for Liquid Simulations

ORFs – Open Reading Frames

PBC – Periodic Boundary Conditions

PBS – Phosphate-buffered Saline

PCR – Polymerase Chain Reaction

PD – Protease Catalytic Domain

PME – Particle Mesh Ewald

PPIs – Protein-protein Interactions

PSSM – Position Substitution Score Matrix

PVDF – Polyvinylidene Difluoride

RAAS – Renin Angiotensin Aldosterone System

RBD – Receptor Binding Domain

RBM – Receptor-binding Motif

RdRp – RNA dependent RNA-polymerase

Rg – Radius of Gyration

RMSD – Root-mean-square Deviation

RMSF – Root-mean-square Fluctuations

RNA – Ribonucleic Acid

RSV – Respiratory Syncytial Virus

RTC – Replication-transcription Complex

SA – Streptavidin

sACE2 – Soluble ACE2

SARS – Severe Acute Respiratory Syndrome

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SASA – Solvent Accessible Surface Area

SD1 – Subdomain 1

SD2 – Subdomain 2

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide

SF – Soluble Fraction

SP – Signal Peptide

SpO2 – Oxygen Saturation in blood

SPR – Surface Plasmon Resonance

S protein – Spike Protein

TBS-T – Tris-buffered Saline

TCEP – Tris(2-carboxyethyl)phosphine hydrochloride

TM – Transmembrane Region

TMD – Transmembrane Domain

TMPRSS2 – Transmembrane Serine Protease 2

UH – Upstream Helix

VOC – Variants of Concern

VOI – Variants of Interest

WHO – World Health Organization

Symbols

b – Bond length

CV – Column Volumes

k – Force Constant

K_d – Dissociation Constant

N – Particles

P – Pressure

r – Nuclear Atomic Coordinates

REU – Rosetta Energy Units

RT – Room Temperature

RU – Relative Units

T – Temperature

V – Potential Energy/ Volume

1. Introduction

1.1. COVID-19 Pandemic

The word "Pandemic" comes from the Greek and means "of all the people", which corresponds to an unexpected rise in the number of people presenting a certain group of symptoms, what is called an outbreak, which spreads worldwide. On the other hand, when infections occur locally, they are called epidemics, and are usually due to seasonal strains of viruses, bacteria and other pathogenic agents[1,2]. Pandemics can arise when new strains of viruses jump from another species to infect humans, which will promote the transmission of the virus among mankind before the population develops the immunity necessary to combat these new strains[1,3]. Pandemics have common characteristics: they affect a large fraction of the population and they spread to multiple geographic locations. However, their causes can vary substantially[1].

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the agent causing the ongoing global pandemic, which caused the emergence of coronavirus 2019 disease (COVID-19). The outbreak initially emerged in Wuhan city, within the Hubei province in China, in December 2019, causing pneumonia-like symptoms in a group of people[4–6]. On January 20th, 2020, the World Health Organization (WHO) released the situation report-1 which indicated 282 laboratory-confirmed cases and 6 deaths worldwide[4]. Up to this point, the transmission of the disease cannot be controlled, as the number of confirmed cases and deaths is still increasing[6].

On January 30th, 2020, the WHO designated the infection as a public health emergency of global proportions[7], and on March 11th, 2020, the WHO officially declared the COVID-19 infection a pandemic[4,8]. By March 1st, 2022, there were 435,626,514 confirmed cases and 5,952,215 deaths worldwide (Figure 1.1). In Portugal, there were 3,262,618 confirmed cases and 21,063 deaths.

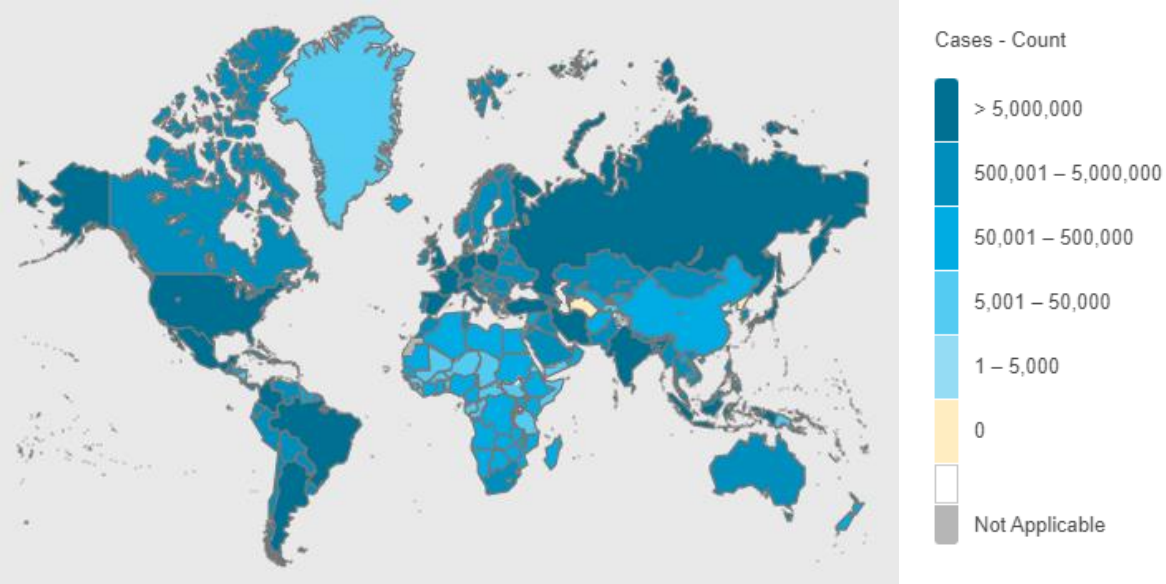

Figure 1.1 – Illustration of the world map regarding the number of confirmed cases per COVID-19 by March 1st, 2022[9].

Current treatment of patients with COVID-19 is based only on the management of their symptoms, and other, more specific therapeutic options are still under investigation. But in order to prevent the increase of infections, different types of vaccines are currently being administered, and in Portugal

these are Pfizer, AstraZeneca, Moderna and Janssen[10]. However, the high infection rates highlight the need for the development of new therapies[4].

1.2. Respiratory virus outbreaks

The COVID-19 pandemic is not the first viral outbreak to have devastating impacts on human health. Among the various infections that exist, acute respiratory tract infections (ARTIs) are the most common diseases that affect all people regardless of age or gender. These diseases are caused by various types of microorganisms including a wide range of bacteria (i.e., Streptococcus pneumoniae) and viruses, such as Influenza A or B ("the flu"), respiratory syncytial virus (RSV), parainfluenza, adenoviruses, coronaviruses, and others. Of all of these, the most relevant infections are associated with coronavirus, Influenza A or B and RSV, which have caused several epidemics and pandemics throughout history (Figure 1.2)[1].

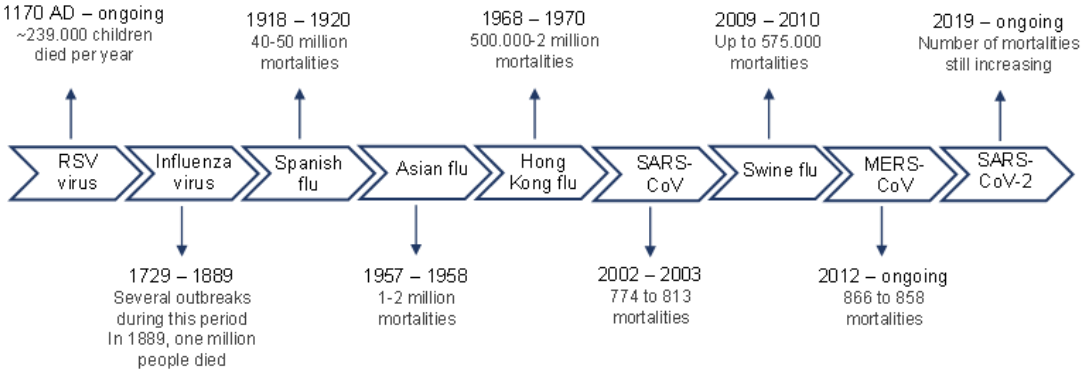

Figure 1.2 – Time diagram of pandemics caused by viruses associated with the respiratory tract and the number of deaths that occurred.[1]

Of all the viruses mentioned, coronaviruses and influenza viruses cause the most severe symptoms and some of the worst and most prolonged outbreaks[1].

Coronaviruses outbreaks

Coronaviruses have been known to cause infections since the 1960s, but their ability to cause deadly epidemics has only become known in the last two decades. COVID-19 is the third major outbreak of coronaviruses in twenty years[1].

Because coronaviruses are zoonotic viruses, they can be transmitted from animals to humans and also from humans to humans by aerosols in the air. To date, several animals have been identified as possible reservoirs for these viruses, such as camels, pigs, rats, turkeys, bats, dogs, cats and others. Among all of them, the best known carrier of human infections is the bat. In 1960, the first cases of infection caused by coronaviruses in humans were reported and presumed to be the reason for the common cold[1].

In 2002, the first lethal disease caused by coronaviruses occurred and was designated severe acute respiratory syndrome (SARS), with the causative agent being named SARS-CoV. Ten years later, in 2012, another outbreak of coronavirus infections occurred in Saudi Arabia, which was designated as Middle East respiratory syndrome (MERS), and the responsible virus was named MERS-CoV[1,2].

In November 2002 in Foshan, China, the first known case of atypical pneumonia caused by SARS-CoV was reported. Since then, the disease has spread worldwide. In China, more than 300 cases were reported in a few months, mostly among health care workers, prompting WHO to declare the disease a "global health threat."[1,2].

The natural reservoir of SARS-CoV was possibly the bat and the intermediate hosts before its spread to humans were possibly palm civets. Among humans, the main transmission routes were droplets and aerosols[2].

SARS-CoV infections caused an illness similar to that caused by the influenza virus but more severe, causing fatigue and high fevers. The most common symptoms were nausea, vomiting and diarrhoea. The disease progressed to atypical pneumonia, in 20 - 30% of infected patients, with poor oxygen exchange in the alveoli and shortness of breath, where patients required mechanical ventilation[2].

In order to deal with this outbreak, in March 2003, the WHO, along with a large network of research centres around the world, launched studies to identify the agent that causes SARS. That same year, they examined several patients affected by the disease and found that a new coronavirus may have caused SARS[1].

In order to control the spread of the disease, the authorities took immediate action such as isolation of suspected patients, quarantines and contact tracing. Later, in July 2003 the SARS pandemic ended, with 8096 to 8437 infected individuals and 774 to 813 deaths in 29 countries. In late 2003 (December - January 2004) a few more cases of SARS emerged again but no further cases of infection have been detected since then[1,2].

Although this disease had a low mortality and morbidity rate, the consequences of the SARS pandemic were not limited to those infected. It caused great anxiety among the population because the virus was new, could spread rapidly, and health care workers and hospitals were vulnerable[1].

In June 2012, in the city of Jeddah in Saudi Arabia, appeared the first case of acute pneumonia caused by a new form of coronavirus, MERS-CoV[1,2].

Individuals infected with MERS-CoV show various clinical features ranging from mild to severe fulminant lung disease. Infection with this virus causes highly lethal acute pneumonia and renal dysfunction with several associated symptoms, including headache, fever, chills, cough, sore throat, myalgia and arthralgia. Other symptoms include nausea, vomiting, diarrhoea and abdominal pain[1,2].

Between 2012 and 2020, 2494 to 2519 cases were confirmed and caused 858 to 866 deaths in 27 countries[1,2].

Some of the laboratory-confirmed cases of MERS-CoV infection have been described as asymptomatic, meaning that these individuals had no clinical symptoms but tested positive for this infection[1].

Once again, the natural reservoir of MERS-CoV was possibly the bat, while this time the intermediate host before its spread to humans changed, having possibly been dromedary camels[1,2].

Due to technological advances, this time around researchers and health professionals were relatively better prepared when this pandemic emerged. With the advancement of molecular diagnostic tools, such as the availability of advanced sequencing tools allowed the complete discovery of the genome[1].

1.3. Biological and structural characteristics of SARS-CoV-2

SARS-CoV-2 is an enveloped, spherical particle approximately 120 nm in diameter with a positive-sense single-stranded RNA genome (+ssRNA) consisting of 29,891 nucleotides coding for 9860 aminoacids[4,8,11]. The RNA of coronavirus (CoVs) is the largest known viral genome with 27-32 kb in length[4].

This virus belongs to the family Coronaviridae, which is included in the order Nidovirales. This family is composed by two subfamilies, Coronavirinae and Torovirinae. The subfamily of SARS-CoV-2, the Coronavirinae, is classified into four main genera: Alphacoronavirus (α-CoV), Betacoronavirus (β-

CoV), Gammacoronavirus (γ-CoV) and Deltacoronavirus (δ-CoV). The first two genera exclusively infect mammalian species, while the other two genera have a wide host range, including avian species[1,4,8,12,13].

Recently, the Betacoronavirus genus has been subdivided into lineages or subgenera A (Embecovirus), B (Sarbecovirus), C (Merbecovirus) and D (Nobecovirus). SARS-CoV-2 is a part of the genus Betacoronavirus and the subgenus Sarbecovirus (lineage B) (Figure 1.3)[1,4,8,12].

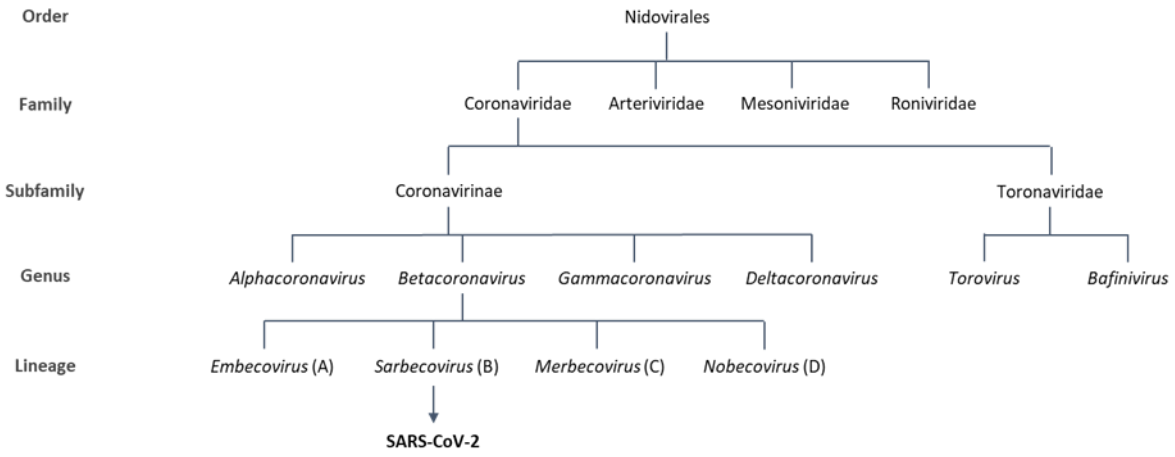

Figure 1.3 – Diagram of taxonomic features of SARS-CoV-2 (Adapted from[1]).

Regarding the structure of SARS-CoV-2, it is constituted by four structural proteins and sixteen non-structural proteins[12]. Two thirds of the viral genome consist of two open reading frames (ORFs), ORF1a and ORF1b. These are translated into polyproteins, pp1a and pp1ab, respectively, which are processed by viral proteases into sixteen non-structural proteins (Nsp 1-16). The remaining one-third of the genome contains overlapping ORFs encoding four major structural proteins, the spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins (Figure 1.4)[12,14].

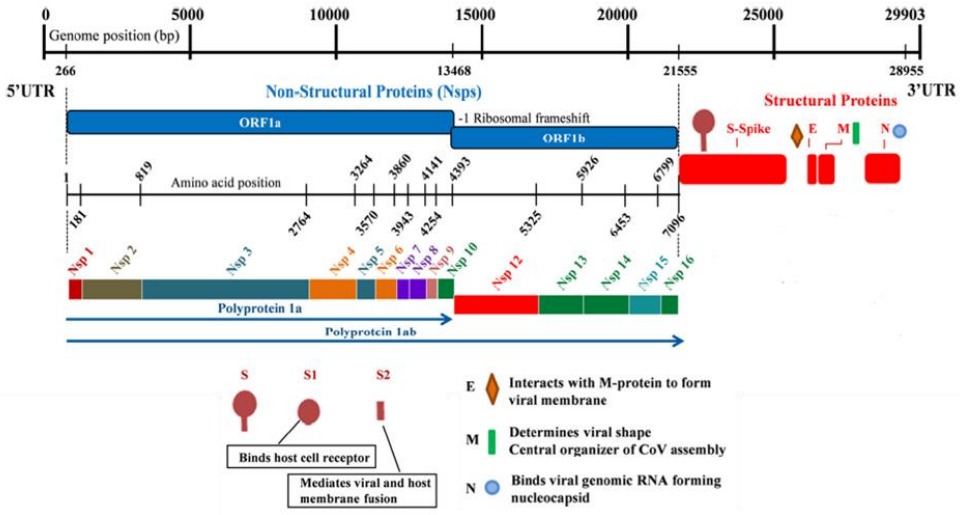

Figure 1.4 – Organization of the SARS-CoV-2 genome. The viral genome consists of structural proteins necessary for the assembly of new virions and sixteen non-structural proteins (Nsps) necessary for replication/transcription (Adapted from[14]).

From the 16 Nsp resulting from pp1a/ab proteolytic cleavage[8], fifteen make up the viral replication-transcription complex (RTC) which includes RNA-processing, RNA-modifying enzymes and an RNA proofreading function needed to keep the integrity of the ~30 kb genome of the virus[13].

The spike (S) protein is a class I viral fusion protein that binds to the host receptor and promotes fusion between the viral and host membranes. It has two subunits: S1 that contains a receptor binding domain (RBD) that binds to the host cell receptor and S2 that promotes fusion between the viral and host cell membranes[15]. The membrane (M) protein is a small protein (~25-30 kDa) responsible for giving the virion its shape and is the most abundant structural protein of the viral particle. This protein, which is a dimer, can adopt two different conformations, which favours membrane curvature and its binding to the nucleocapsid[11]. The envelope (E) protein (~8-12 kDa) contains an N-terminal ectodomain and a C-terminal endodomain and functions as an ion channel. It is present in small quantities in the virion but facilitates virus assembly and release. While the ion channel activity is not required for the virus to replicate, it is important for viral pathogenicity[11]. Finally, the nucleocapsid (N) protein consists of two separate domains and is the only protein present in the nucleocapsid. This protein is responsible for binding the viral genome in the form of beads-on-a-string (Figure 1.5)[11].

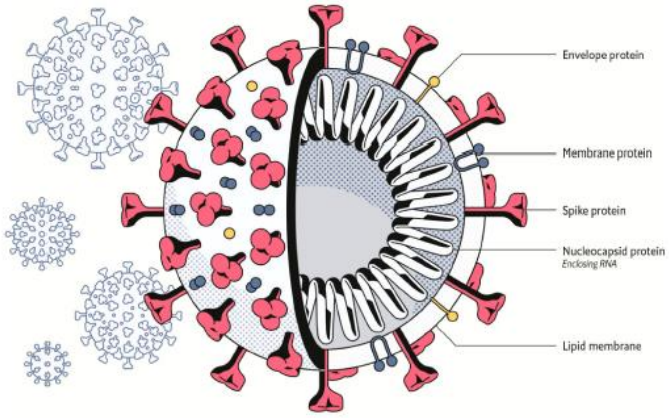

Figure 1.5 – Schematic illustration of the structural proteins of SARS-CoV-2. The envelope (E) protein enables assembly and release of the virus; membrane (M) protein gives the virion its shape; spike (S) protein mediates binding to the host receptor; nucleocapsid (N) protein binds the viral genome into a beads-on-a-string conformation[16].

1.4. Similarities between SARS-CoV-2, SARS-CoV and MERS-CoV

In two decades, three highly deadly and pathogenic coronaviruses have emerged, specifically SARS-CoV, MERS-CoV and finally SARS-CoV-2. In morphology and genome organization, MERS-CoV has the largest genome with about 30.11 kb, followed by SARS-CoV-2 with about 29.9 kb and SARS-CoV with 29.75 kb. In genomic homology, SARS-CoV-2 is about 80% identical to SARS-CoV, while it is 50% identical to MERS-CoV, the latter being from the C lineage of betacoronaviruses[17,18]. The S protein of SARS-CoV-2 also shows high homology with the other two viruses. With SARS-CoV this protein shares ~77% identity[14], whereas with MERS-CoV the identity of the S protein is lower: 31.9%[4].

All of these outbreaks by human coronavirus (hCoVs) are related to human-animal interactions (Figure 1.6). Most notably, SARS-CoV and SARS-CoV-2, both emerged from wet markets in China. SARS-CoV was thought to have emerged from wild animals that were sold in such markets. A CoV strain isolated from a palm civet from wildlife markets shared 99.8% identity with SARS-CoV. Therefore, this animal was considered to be the intermediate host for this strain[18], although this matter remains unclear.

MERS-CoV also originated from bats. An RNA fragment isolated from bat faeces showed 100% nucleotide identity with MERS-CoV from an infected person in the same area, indicating that bats were likely the source of this virus. Subsequently, the ability to replicate in bats without causing symptoms was demonstrated, suggesting that bats were an ideal reservoir for MERS-CoV. The dromedary camel was considered an intermediate reservoir when it was found that the sequence identity between the

virus in this animal and the virus in humans was high (99.2-99.5%)[18]. The origin of SARS-CoV-2 has also been linked to wet markets, like it happened in SARS-CoV. SARS-CoV-2 and BatCoV were found to share a genomic similarity, thus it was assumed that this animal was the natural reservoir of this virus. Genomic similarity was also found between SARS-CoV-2 and pangolin-CoV (91.02%), although lower than that with BatCoV (96.2%). With these findings, the doubt of which is the natural reservoir of this virus remains[18].

Regarding how these viruses enter host cells (Figure 1.6), in all of them do so by binding the receptor-binding domain to functional receptors on the surface of the host cell. Angiotensin-converting enzyme 2 (ACE2) is the dominant receptor in both SARS-CoV and SARS-CoV-2 hosts, although SARS-CoV has other co-receptors (DC-SIGN and L-SIGN). The MERS-CoV receptor is the dipeptidyl peptidase-4 (DPP4)[18].

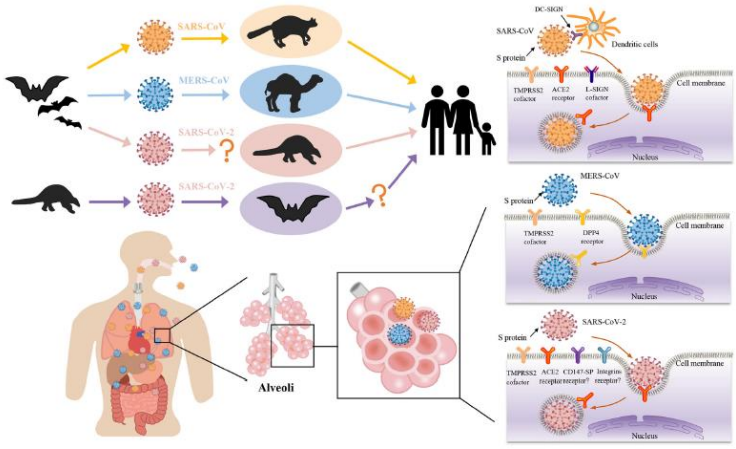

Figure 1.6 – Illustration of the host cell receptors, potential animal hosts and biodistribution of SARS-CoV, MERS-CoV and SARS-CoV-2[18].

As for the mode of transmission of these viruses, both SARS-CoV and SARS-CoV-2 have identical modes of transmission, through close person-to-person contact by inhalation airborne droplets or by contact with contaminated surfaces. In the case of MERS-CoV, humans can become infected when they come into contact with dromedary camels that are infected. Of the three viruses, SARS-CoV-2 is the one with the highest transmissibility, followed by SARS-CoV and MERS-CoV[18].

Both SARS-CoV and SARS-CoV-2 have more or less the same incubation time, which on average is 4 to 6 days. The incubation time of MERS-CoV is similar to both SARS (5.2 days)[5,18]. After the incubation period, when clinical manifestations begin to occur, in all three cases these are similar, including fever (≥38.0°C), sore throat, cough, dyspnoea, fatigue, diarrhoea and headache[18]. In order to diagnose these three viruses, polymerase chain reaction (PCR) with viral RNA from clinical samples is mainly used because of its sensitivity, specificity, and simplicity[18].

1.5. Transmission, symptoms and diagnostics of SARS-CoV-2

An episode of COVID-19 can be divided into three phases, depending on the degree of disease progression. In the initial phase, viral replication and mild symptoms occur[6]. In the initial phase of infection, SARS-CoV-2 infiltrates the lung parenchyma and begins to replicate. The S1 subunit of the spike protein binds to the ACE2 receptor on the host cell surface, while a furin protease cleaves the protein at the S1/S2 site and the transmembrane serine protease 2 (TMPRSS2) at the S'2 site to allow fusion of the viral and host cell membranes. Replication and translation of the viral genome then begins in the cytoplasm of the cell. This phase is characterised by the development of mild symptoms and an initial response of the innate immune system. The spread of the virus is prevented by the presence of macrophages and natural killer (NK) cells. At this stage the synthesis of type I interferons (IFN) is activated.

Concerning adaptive immunity in this phase, the Th1/Th17 type immune response is crucial for virus clearance. Helper T cells activate T-dependent B cells in order to promote antibody production and will also promote the NFkB signalling pathway in order to introduce pro-inflammatory cytokines. Cytotoxic T cells destroy infected cells (Figure 1.7)[4,6,19].

The second stage is the pulmonary phase which involves the stimulation of adaptive immunity (inflammatory response) and respiratory failure and tissue damage begin to occur. This respiratory failure caused by SARS-CoV-2 presents different characteristics from that of typical acute respiratory distress syndrome (ARDS). ARDS is a complicated clinical syndrome of acute respiratory failure that develops from non-cardiogenic pulmonary oedema. With the development of ARDS, the most common diseases associated with this development are bacterial and viral pneumonia. Extreme levels of cytokines are present at this stage. Alveolar cell damage and necrosis may occur due to the presence of excessive and deregulated inflammation[6].

The third stage is the hyperinflammatory phase where hyperinflammatory conditions such as ARDS occur. Damage to distant organs and systemic inflammation are present due to increased host inflammatory response, which will result in multiple organ failure (MOF). At this stage, a high number of leukocytes with lymphopenia and increased levels of pro-inflammatory cytokines can be observed, some of these being IL-2, IL-6, IL-7, IL-10, C-reactive protein (CRP), monocyte chemoattractant protein (MCP) 1 and macrophage inflammatory protein (MIP) 1α (Figure 1.7). This "cytokine storm" causes lung injury with life-threatening complications (septic shock, ARDS, MOF, haemorrhage/coagulopathy, secondary bacterial infections and acute heart/liver/kidney injury)[4,6,19].

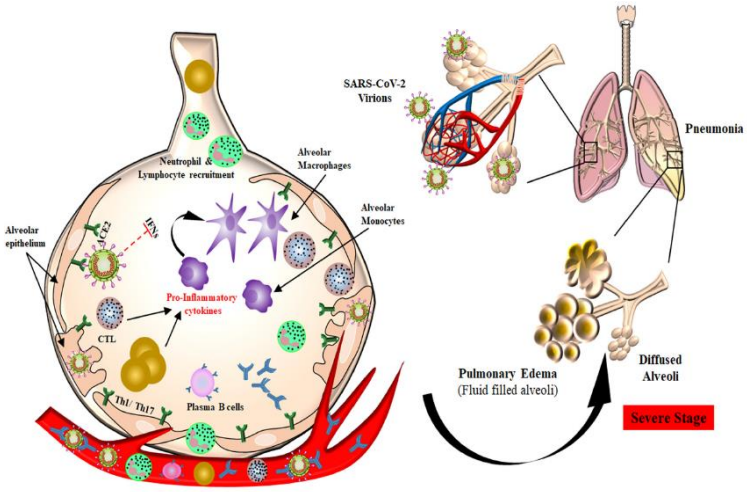

Figure 1.7 – The various immune responses of the host during infection with SARS-CoV-2. SARS-CoV-2 virus infects the host through the naso-oral route. Subsequently, the infection spreads to the cells expressing the ACE2 receptors in the lungs, which are the alveolar type 2 cells. This virus evades the innate immune cells due to uncontrolled virus replication, which leads to attenuation of the antiviral IFN responses. The rise in pro-inflammatory cytokines is due to the influx of neutrophils, monocytes/macrophages, and other adaptive immune cells. When Th1/Th17 cells are stimulated by viral epitopes, this can lead to an increased immune response. This inflammatory response leads to 'cytokine storms' that cause immunopathies such as pulmonary edema and pneumonia. In the innate immune response, cytotoxic T cells are recruited to the site of infection to try to kill the infected cells. B cells/plasma cells recognise viral proteins and are stimulated to produce antibodies specific for SARS-CoV-2 to provide systemic immunity[20].

SARS-CoV-2 Transmission

The spread of this infection depends on several routes of human-to-human transmission (Figure 1.8), which include direct contact with airborne droplets released when infected individuals speak, cough, and sneeze, which can lead to the inhalation of infected droplets by those nearby. Thus, respiratory droplet transmission is the principal mode of transmission of this disease[4,6]. Another way through

which this virus spreads is via environmental contamination. For example, the disease can be spread indirectly when a person touches a contaminated object and then touches their eyes, mouth, or nose. This is because this virus can remain stable and infectious in aerosols, surfaces made of plastic or stainless steel for hours or even days[6]. Recent research has suggested that another possible way of spreading is by airborne transmission through infected people[4].

The average incubation period for this virus is between 5 and 6 days, while it can go up to 14 days or more. In a study by Lauer et al, 97.5% of people developed symptoms within 11.5 days of infection, whereas approximately 1% of those infected did not develop symptoms until 14 days of quarantine[6,21]. The average time between the onset of symptoms in a patient with a primary case (infector) and a patient with a secondary case (infected person) was 3.96 days[6]. Given this information, presymptomatic transmission is likely. Symptomatic people are a source of infection, but asymptomatic can also be an important source of transmission[6].

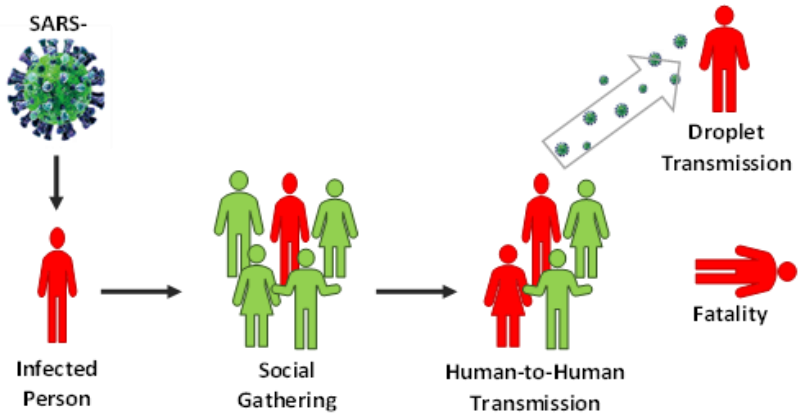

Figure 1.8 – Mode of transmission of SARS-CoV-2 in humans. Transmission can occur by human-to-human spread and by droplets. Red colour corresponds to infected persons and green colour to healthy persons (Adapted from[22]).

Symptoms caused by COVID-19

Complications from COVID-19 are associated with several risk factors such as age (>65 years), chronic respiratory disease, cardiovascular disease, hypertension, diabetes, and obesity[23]. The most frequent complication is ARDS, and there are other fatal complications that include pneumonia, type I respiratory failure, sepsis, metabolic acidosis, arrhythmia, acute cardiac and renal damage, heart failure and hypoxic encephalopathy[23]. The clinical manifestations of this disease can be divided into mild, moderate, severe, or critical, depending on the strength of the infected person's immune system. Affected individuals may be symptomatic or asymptomatic, with most patients having only mild symptoms and recovering from the disease[4,12,23–25].

The most common COVID-19 symptoms are fever (70%-90%), dry cough (60%-86%), shortness of breath (53%-80%), fatigue (38%), myalgia (15%-44%), nausea/vomiting or diarrhoea (15%-39%), headache and fatigue (25%)[24,26,27]. Loss of smell (anosmia) and taste (ageusia) are present in 3% of patients with COVID-19[6,27]. When the disease is considered mild, the typical symptoms that the patient may present are fever, dry cough, sore throat, runny nose, sneezing, fatigue, myalgia, tiredness, muscle pain, headache, anosmia and ageusia[6]. There are also atypical symptoms, these being nausea, vomiting, diarrhoea and abdominal pain[12,23]. When the disease is considered moderate, symptoms of pneumonia, fever (usually persistent) higher than 37.8 °C with a dry cough start to appear[12,23]. When the disease is considered severe, dyspnoea, hypoxia (blood oxygen saturation (SpO_2) ≤ 93%), diarrhoea and nausea begin to occur[12,23]. Finally, when the disease is already in a critical state, patients begin to have severe difficulty breathing and shortness of breath, chest pain and there is a continuous worsening of the patients' clinical prognosis. In this condition the most prevalent complications are ARDS, myocardial damage, arrhythmias, heart failure, acute kidney and liver damage, encephalopathy, septic

shock, and multiple organ abnormalities[12,23]. In most cases, approximately 90% of patients present more than one of these symptoms. Most cases are between the ages of 30-79 (86.6%) and the highest number of fatalities are in the group of patients over 80 years[23].

COVID-19 diagnostics

In order to detect whether a person is infected with the SARS-CoV-2 virus, there are two types of tests[28]. One of these tests is used to detect SARS-CoV-2 in the body by checking for the presence of this virus in samples taken from the throat, nose, nasal secretions (mucus) or sputum (saliva). These tests can be divided into nucleic acid-based tests (PCR), that detect the genetic material of the virus and antigen tests, which detect specific viral proteins. Associated with these tests, computerized tomography (CT) scans and chest X-rays are usually performed in order to complement the diagnosis[28,29]. Real-time reverse transcription-polymerase chain reaction (real time RT-PCR) is the main diagnostic test for the detection of SARS-CoV-2 in the human organism[25,27]. Swab samples, usually taken from the upper respiratory tract, are required to perform RT-PCR, depending on the ability to amplify a low concentration of nucleic acids present in these samples[27]. Regarding the type of sample, samples from the lower respiratory tract, such as bronchoalveolar lavage fluid, are more sensitive than samples from the upper respiratory tract, as shown in a study conducted in China[26,27,30]. Another alternative sample type is saliva, being a faster and less painful method than the other ways of acquiring samples[27,31]. The most important advantage of this method is that it can be designed for two different target systems, i.e., it can serve as a primer that can detect different coronavirus types including SARS-CoV-2 or as a primer that can identify SARS-CoV-2 with high specificity[27,29,31,32].

Rapid diagnostic tests to detect the presence of viral antigens have been developed, which are expressed by SARS-CoV-2, in respiratory tract samples from infected persons. These rapid tests are available for use on a large scale and the results are obtained shortly after the test, usually within minutes. In addition, these tests do not need to be performed in a laboratory and can be performed at the same site where the sample is collected. As these tests do not require laboratory processing and due to the type of technology, they become much cheaper than PCR tests. Regarding the type of sample used, these are very similar to those used in PCR tests, which are nasopharyngeal exudates obtained with a nasal and/or oral swab[33].

Another type of test is serological testing, which are mostly used to find out if a person has ever been infected[29]. Normally, antibodies against SARS-CoV-2 are found in up to 50% of infected patients by day 7 and in all infected patients by day 14[29,34]. There are two ways to perform these serological tests, by enzyme linked immunosorbent assay (ELISA) or colloidal gold immunochromatographic kits[35]. These kits are sensitive, specific, rapid (less than ten minutes to obtain results) and require only a few microlitres of blood via a finger-prick to perform the assay[34]. Both techniques have been shown to have equal sensitivity with 100% specificity for detection of SARS-CoV-2[35].

1.6. Entry mechanism in the host organism and mode of infection

The first step in infection with SARS-CoV-2 is the recognition of its receptors on the surface of host cells. The virus first infects ciliated bronchial epithelial cells and type II pneumocytes[36]. This step occurs due to the viral S protein interaction with the human receptor ACE2[13,19,36,37].

The S protein consists of two functional domains, the S1 domain containing the receptor binding domain (RBD) that binds to ACE2 and the S2 domain which mediates the fusion of the viral membrane to that of the host cell[4,19]. For SARS-CoV-2 to enter the host cell, two cleavage events are required in the S protein, one at the junction site of the S1 and S2 subunits and the other at the S2' site, which is located within the S2 subunit (Figure 1.9). The junction site of the two subunits is cleaved by a furin protease only during maturation of the virus in an infected cell, but the S2' site will only be cleaved after binding to ACE2, the latter being dependent on the former, and both are required for membrane fusion to take place[38].

Binding of the virus to ACE2 will induce conformational changes in the S1 subunit that will allow exposure of the S2' cleavage site in the S2 subunit. This cleavage action at the S2' site can be carried out by different proteases, transmembrane serine protease 2 (TMPRSS2) or cathepsin L, depending on the entry pathway chosen by SARS-CoV-2. If the target cell does not express sufficient TMPRSS2 or a virus-ACE2 complex does not encounter this protease, this complex will be internalised via endocytosis, mediated by the clathrin protein in the endosomes, where cleavage at S2' is carried out by cathepsins, but these require an acidic environment for their activity to take place. On the other hand, in the presence of TMPRSS2 the cleavage of S2' occurs on the surface. In both entry pathways, cleavage of the S2' site triggers the dissociation of the S1 and S2 subunits and the exposure of the fusion peptide, which leads to conformational changes in the S2 subunit, particularly in heptad repeat 1. The S2 subunit becomes extended, which enables the insertion of the FP into the host membrane, thus initiating fusion between the viral membrane and the host cell membranes. A fusion pore is formed when the fusion between these membranes is completed, through which the viral RNA is released into the host cell cytoplasm to be uncoated and replicated[19,38].

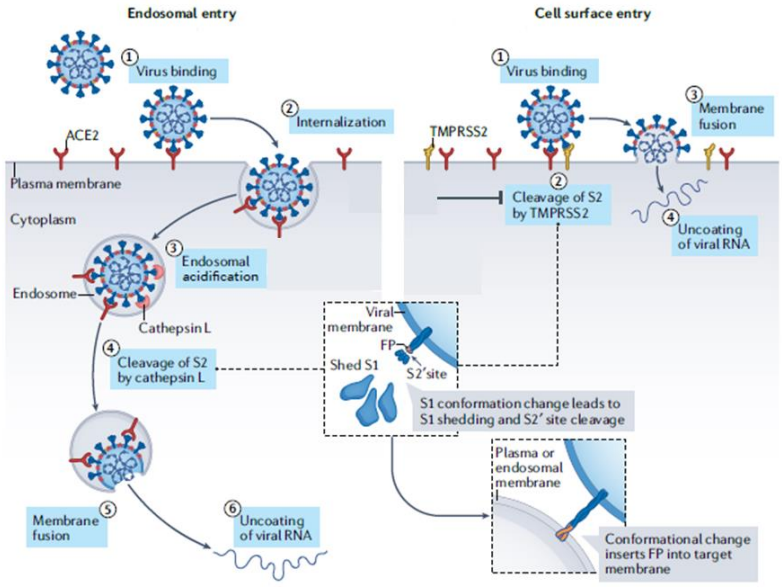

Figure 1.9 – Two different SARS-CoV-2 entry mechanisms. After binding to ACE2 (step 1), the virus can enter the host cell by two different routes. In endosomal entry (left), internalisation of the virus-ACE2 complex is via endocytosis and for cleavage at S2' to occur, carried out by the cathepsins, these require acidification of the endosome to be able to carry out their activity (steps 3 and 4 left). In cell surface entry (right), cleavage at S2', carried out by TMPRSS2, occurs at the cell surface (step 2 right). Then, in both forms of entry, membrane fusion occurs between the viral membrane and the host cell membrane (step 5 left and step 3 right) and the subsequent release of the viral RNA into the host cell cytoplasm for uncoating and replication (step 6 left and step 4 right). (Adapted from[38])

Viral replication begins with the translation of the replicase-polymerase gene, i.e., the translation of genomic RNA into viral replicase polyproteins (pp1a and 1ab), and the assembly of the replication-transcription complex (RTC). Then, in the process of replication and transcription mediated by the RTC, intermediates negative-sense RNA are synthesized, which serve as templates for the synthesis of positive-sense genomic RNA and of subgenomic RNA, which will encode structural proteins[21,39]. Then, after translation, the M, S and E proteins are inserted into the endoplasmic reticulum membrane and transported from here to the assembly site, the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). The remaining structural protein, the N protein, which is replicated, transcribed, and synthesised in the cytoplasm, will complex with the genomic RNA synthesised to form nucleoproteins (NPs)[40,41]. This nucleoprotein, along with the S, M and E proteins, and other viral proteins, are further assembled and formed into mature virions by budding in ERGIC. These are eventually released from the host cell again[40].

The whole process described above, referring to what happens after viral entry, is represented in Figure 1.10.

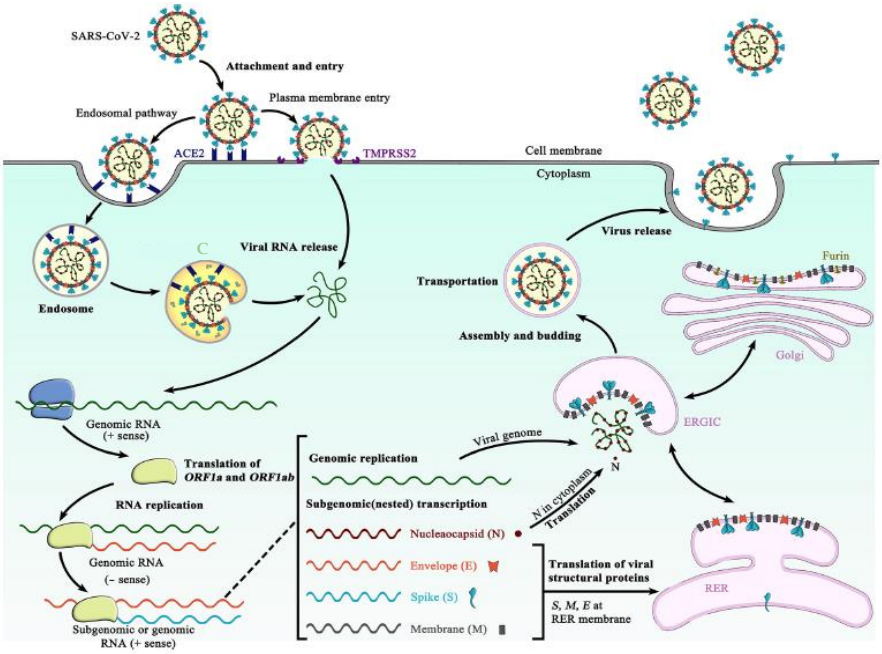

Figure 1.10 – Description of the SARS-CoV-2 infection process, showing what happens after viral entry. After membrane fusion, viral genetic material is released into the cytoplasm and viral replication begins with translation of the replicase-polymerase gene and assembly of the replication-transcription complex (RTC). Intermediates of negative-sense RNA (- sense) are then synthesized in order to serve as templates for the synthesis of positive-sense genomic RNA (+ sense) and of the subgenomic RNA, which will encode the structural proteins. After translation, the M, S and E proteins are inserted into the endoplasmic reticulum membrane and transported from there to the assembly site, the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). The remaining structural protein, the N protein, which is replicated, transcribed, and synthesised in the cytoplasm, forms a complex of nucleoproteins (NPs) together with the genomic RNA, which are further assembled along with the S, M, and E proteins and other viral proteins and subsequently develop into mature virions in the ERGIC. These are eventually released from the host cell[40].

1.7. Spike (S) Protein

In order to bind to the host cell receptor, the ACE2 protein, and to fuse its membrane with the cell membrane, SARS-CoV-2 uses a trimeric protein, the spike (S) protein, which is attached to the viral membrane by a small transmembrane domain and contains a large soluble region that protrudes from the viral surfaces. For this protein to be able to bind to ACE2 and promote fusion, several steps, involving different cleavage events in the S protein are required. All this involves different domains of this protein, the two functional domains S1 and S2, interacting with the host cell, as well as with extracellular and intracellular components, such as furin protease, TMPRSS2, cathepsin L and clathrin protein[14,38,42].

It is the efficiency of each of these steps that will dictate the virulence and infectivity of the virus, so disruption of any one of these steps may lead to possible treatment[42].

The S protein (Figure 1.11), with 1273 residues, consists of two main subunits, S1 and S2[14,43]. Several three-dimensional structures of this protein are available, such as the full-length S trimer where the RBD can be found in the close or open conformation[43], certain components such as the RBD and the NTD of this protein, the S protein of some variants, and this protein complexed with the ACE2 receptor. The first solved structure of the S protein of the SARS-CoV-2 virus dates from early 2020.[44]

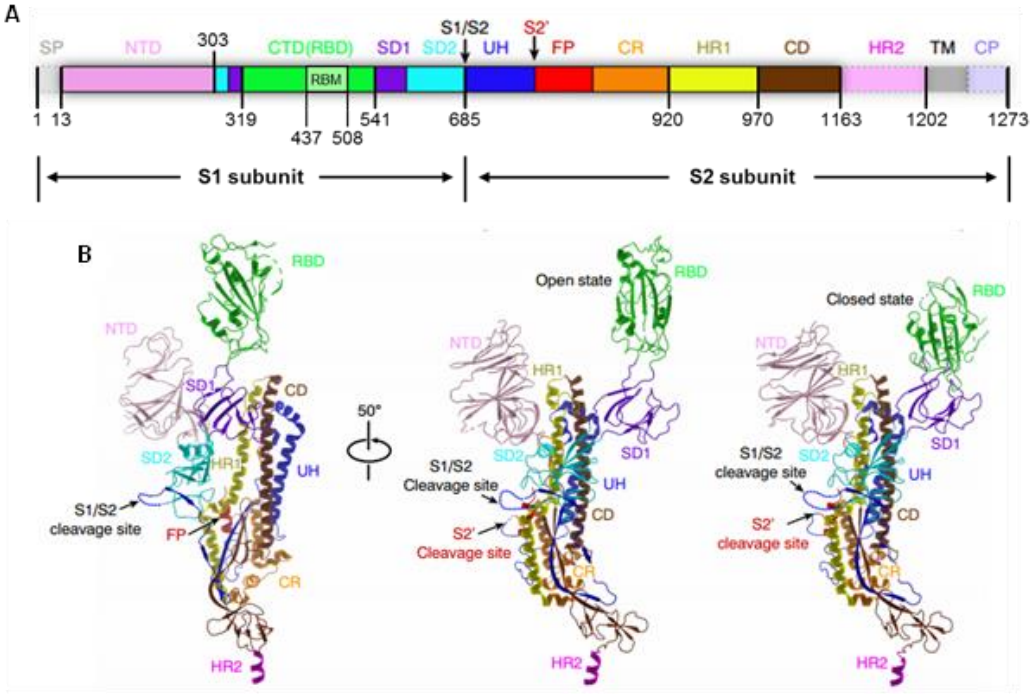

Figure 1.11 – SARS-CoV-2 virus spike protein structure. (A) Schematic representation of the S protein domain organization, where each constituent is represented by a colour. (B) Structural representation of an S protomer (PDB: 6VXX and 6VYB). Abbreviations: Signal peptide (SP), N-terminal domain (NTD), C-terminal domain (CTD), receptor-binding domain (RBD), receptor-binding motif (RBM), subdomain 1 (SD1), subdomain 2 (SD2), upstream helix (UH), fusion peptide (FP), connecting region (CR), heptad repeat 1/2 (HR1/HR2), central domain (CD), transmembrane region (TM), and cytoplasmic domain (CP) and two distinct protease sites (S1/S2 and S2/S2') are indicated. (Adapted from[14,45])

The S1 subunit has the function of recognising and binding to host cell receptors and consists of the signal peptide (SP), the N-terminal domain (NTD) and the receptor binding domain (RBD)[12,14,42]. The SP corresponds to the initial 13 amino acids and its function is to guide the transport of the protein to its target membrane[42]. The RBD is composed by a central subdomain and the receptor-binding motif (RBM) that will interact directly with host ACE2. The sequence of this domain varies among different coronaviruses, which means that vaccines and RBD-based antibodies developed against SARS-CoV-2 do not offer protection against other coronaviruses and vice-versa[12,42]. It also has the property that in the pre-fusion state it can adopt different conformations: one in which the binding interface of the receptor is covered by the adjacent protomer, corresponding to the closed conformation, and another one in which the binding interface is open for access by ACE2, corresponding to the open conformation (Figure 1.12). Only the latter conformation is suitable for binding to the ACE2 receptor[45].

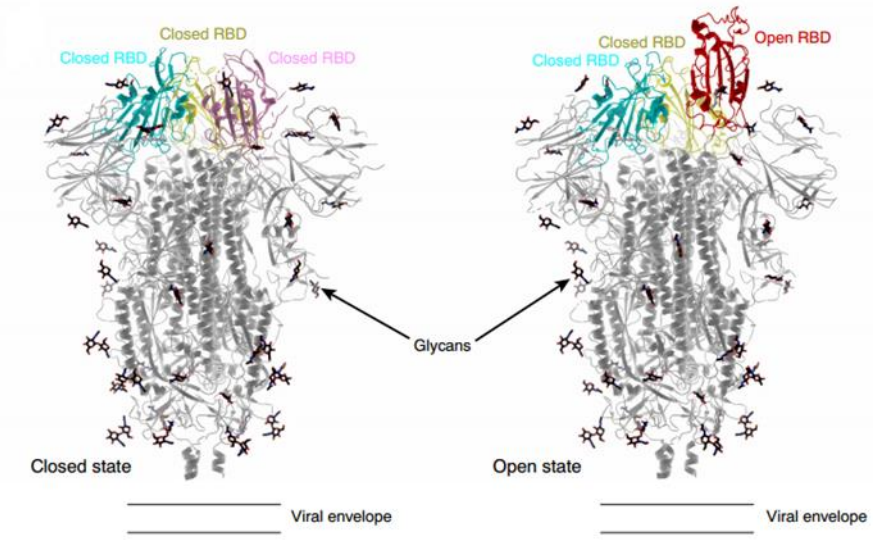

Figure 1.12 – SARS-CoV-2 S trimer structure in different conformations, closed on the left and open on the right[45].

In Figure 1.12 it can be observed that each trimeric spike has three non-synchronized RBDs, which implies asymmetric interactions with the receptor. However, binding of ACE2 to one open RBD may promote the conformational change of the other closed RBDs, allowing them to interact with the receptor[45]. In addition to this, it can also be seen in the figure that S protein has a large amount of glycans in both its subunits. Besides altering the antigenicity of the S protein, this shell of glycans may also alter the conformation of specific domains[45]. More specifically, the S1 subunit has 14 N-glycosylsation sites (N-glycosites; N17, N61, N74, N122, N149, N165, N234, N282, N331, N334, N343, N603, N616, and N657) and 2 O-glycosylation sites (O-glycosites; T323 and S325), five of which are located in the RBD (T323, S325, N331, N334, and N34), while the S2 subunit has 9 N-glycosites (N709, N717, N801, N1074, N1098, N1134, N1158, N1173, and N1194)[46].

The function of the S2 subunit is to promote membrane fusion to facilitate the insertion of the viral genome into the cell. The S2 subunit consists of the fusion peptide (FP), two heptad-repeat domains (HR1 HR2), the transmembrane domain and the C-terminal domain (or cytoplasmic domain, CT)[12,14,42]. The region that corresponds to the SARS-CoV-2 FP is still under debate, with two main candidate regions having been proposed. Fusion peptides (FP) constitute a small segment in the S protein found in the fusion domain (FD) of about 15 to 25 residues[47,48]. The FDs are highly conserved regions, especially in the more pathogenic members of the virus family and are maintained even in the absence of substitutions in the most common variants of SARS-CoV-2 (alpha, beta, gamma, delta, omicron) through evolution[49]. This region has a high hydrophobicity, being located at the N-terminus of the S2 subunit, near or right next to the proteolytic site of the protein[50,51]. These interact with host cell membranes and considerably alter its properties of fluidity, orderliness, curvature, and hydration, ultimately leading to membrane fusion, and is thus a critical region for the fusion process[51]. The FD of SARS-CoV-2 (S816 - F855) consists of two distinct regions, the fusion peptide (FP, S816 - G838) and the fusion loop (FL, D839 - F855)[49]. These being the two main proposed regions, FP1 from residue 816 to 835 and FP2 from residue 836 to 854[48,52]. However, other proposals exist in the literature and this matter remains unclear.

Two anchors are required for fusion of the viral and host membranes, one on the virion side and another one on the host cell side. A region at the C-terminal end of the S2 subunit, known as the transmembrane domain (TMD), attaches this protein to the host membrane, whereas the FP, located on or near the N-terminal end of S2, enters the target cell membrane to initiate fusion. To insert the FP into the host cell membrane, the S2 subunit needs to undergo a large conformational change and become extended, which is triggered by an external stimulus when the virus is close to the target cell. In

the case of SARS-CoV-2, the stimulus is the cleavage by the protease TMPRSS2, expressed on the surface of the host cell[12,14,42]. After inserting the FP into the host membrane, S2 undergoes another conformational change, which brings HR1 and HR2 close to each other. These two domains will interact with each other and form a six-helical bundle (6-HB), bringing the viral and host membranes into close proximity, facilitating the fusion process[53].

Regarding the characteristics of the RBD interaction with ACE2, its core is formed by a twisted five-stranded antiparallel β-sheet (β1, β2, β3, β4, and β7) together with short helices and connecting loops. Beyond this, between the β4- and β7-strands of the core, there is an extended insertion that contains the short β5- and β6-strands, α4- and α5-helices and loops. This extended insertion contains most of the SARS-CoV-2 contact residues that bind to ACE2 (Figure 1.13)[54].

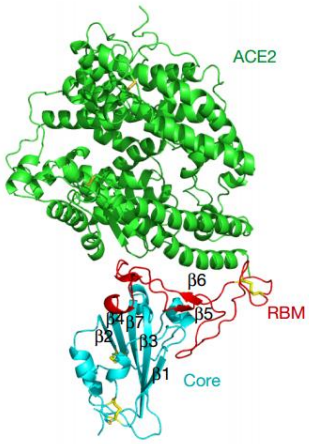

Figure 1.13 – SARS-CoV-2 RBD bound to ACE2 general structure, where the region of the extended insertion that has most of the receptor contact residues can be seen. ACE2 is in green. The RBD of SARS-CoV-2 is shown in cyan, with the RBM in red. And in yellow are the disulfide bonds of the RBD[54].

Specifically, the RBD residues that exhibit interaction with those of ACE2 are K417, Y449, Y453, L455, F456, Q474, F486, N487, Q493, Y495, G496, Q498, T500, N501, G502, and Y505 (Figure 1.14 A and B), six of which (K417, F456, F486, Q493, Q498 and N501) being identified as the key residues responsible for the strong binding of SARS-CoV-2 to ACE2. In addition to this, Jawad, B. et al saw that residues Y449, N487, Q493, G496, T500 and G502 in RBD are linked to residues Q24, E35, D38, Y41, Q42 and K353 of ACE2 via specific hydrogen bonds. Many of these hydrophilic residues located along the interface form a solid hydrogen bond network, as mentioned above, but also form salt bonds interactions (Figure 1.14 C)[54–56].

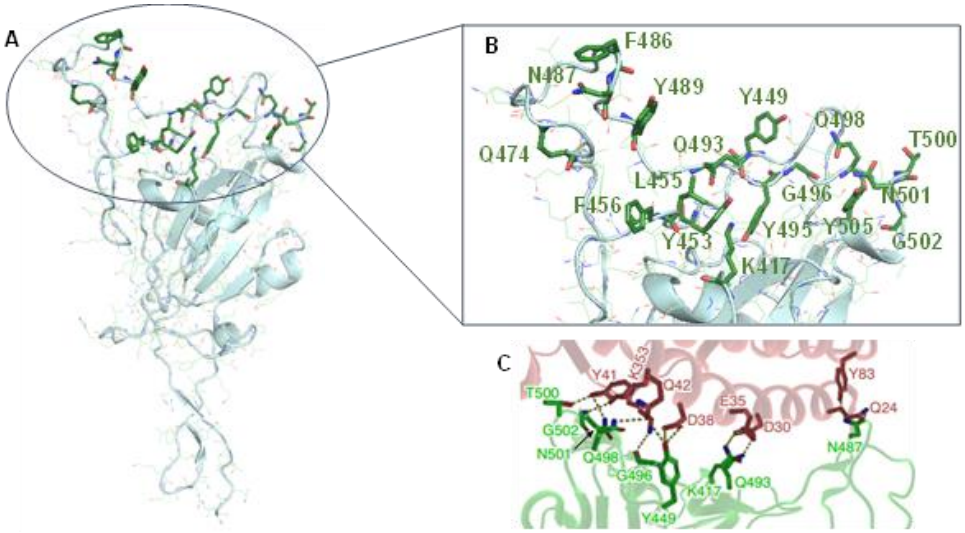

Figure 1.14 – Representation of the RBD residues that are interacting with those of ACE2. (A) Open RBD with the residues that interact with those of ACE2 highlighted in green color. (B) Close-up view of the open RBD region that has the interacting residues with their respective identification. (C) Representation of some interactions, where salt bonds and hydrogen bonds are represented by dashed lines. (Adapted from[45]).

In addition to the interaction with ACE2, this protein also interacts with antibodies that neutralize SARS-CoV-2, and there are several antibodies structures in complex with this protein. These antibodies generally target both the RBD and the N-terminal domain of the S protein, these being the regions with the highest number of epitopes. Neutralizing antibodies directed to the RBD recognize multiple non-overlapping epitopes, whereas neutralizing antibodies directed to the NTD seem to target a single supersite[57–59].

Since the pandemic began, several SARS-CoV-2 variants have appeared, some being of particular importance because of their potential to increase transmissibility, virulence or reduce vaccine efficacy, thus contributing to the continuation of the COVID-19 pandemic. Some of these variants, those that have been deemed the most relevant and of concern are categorized as variants of interest (VOI) or variants of concern (VOC) by the WHO. The VOI contain mutations that can modify the phenotypic properties of the virus and may alter the rates of transmission or spread of the disease, while the VOC have been shown to have adverse effects on human health, because they have substitutions that can increase transmission rates causing more severe effects, and may also reduce the effectiveness of public health interventions such as vaccination[60]. Those categorized as VOI (Table 1.1) are Epsilon, Zeta, Eta, Theta, Iota, Kappa, and Lambda, while the VOC (Table 1.2) are Alpha, Beta, Gamma, Delta, and Omicron[60–62]. All of these variants have mutations in several genes that code for different proteins of the virus, all of them bearing mutations causing amino acid substitutions in the Spike protein. This is of extreme concern because, as we have seen, this is the protein that most participates in the infection and when substitutions affect it, they can lead to a reduction in the effectiveness of vaccines.

Table 1.1 – Variants of the SARS-CoV-2 virus classified as VOI and their characteristics, such as date of appearance, geographical origin, number of substitutions in the S protein, most relevant substitutions, and phenotypic differences.

Variant	Date of appearance	Geographical Origin	Number of substitutions in the S protein	Substitutions with pathogenic effects	Phenotypical differences
Epsilon	September 2020	California	4	L452R	significant resistance to antibodies, decrease in neutralization by vaccines, 18-22% increase in transmissibility
Zeta	October 2020	Brazil	3	-	no public health impact
Eta	December 2020	Many countries	5	E484K	reduction in the ability to be neutralized
Theta	January 2021	Philippines	7	E484K	transmissibility increased, reduction in neutralization
Iota	November 2020	New York	7	E484K, D253G	reduction in neutralization
Kappa	October 2021	India	8	E484Q, L425R	transmissibility increased, reduction in neutralization
Lambda	December 2020	Peru	6	-	high prevalence, higher infectivity and transmissibility

In more detail, the Epsilon variant appeared in California in September 2020. This variant has four substitutions in the S- protein, one of them being L452R which affords significant resistance to antibodies, causing a decrease in neutralization by vaccines. In addition, it has an 18-22% increase in transmissibility compared to the original strain. The Zeta variant appeared in Brazil in October 2020. This variant has only three substitutions affecting the S protein. It has been reported that this variant has no public health impact. The Eta variant was identified in many countries in December 2020. It has five substitutions affecting the S protein, the E484K substitution raising concerns for immune evasion, thus causing a potential reduction in the ability to be neutralized by some monoclonal antibody treatments. The Theta variant appeared in the Philippines in January 2021. This variant results from seven substitutions affecting the S protein, which caused the transmissibility to increase and there could be a potential neutralization reduction by some monoclonal antibody treatments, again displaying the E484K substitution. The Iota variant first appeared in New York in November 2020, and has seven substitutions affecting the S protein, including two substitutions, E484K and D253G, which participate in immune escape, and there may be a potential neutralization reduction ability by some monoclonal antibody treatments. The Kappa variant appeared in India in October 2021. This variant has eight substitutions affecting the S protein and, for the first time, two of its constituent substitutions appeared together, resulting in the E484Q and L425R substitutions. Once again, it caused transmissibility to increase and there could be a potential neutralization reduction by some monoclonal antibody treatments. The Lambda variant appeared in Peru in December 2020, having only six substitutions affecting the S protein. It spread rapidly with high prevalence, showing it to have a higher infectivity and transmissibility[60,62,63].

Table 1.2 – Variants of the SARS-CoV-2 virus classified as VOC and their characteristics, such as date of appearance, geographical origin, number of substitutions in the S protein, most relevant substitutions, and phenotypic differences.

Variant	Date of appearance	Geographical Origin	Number of substitutions in the S protein	Substitutions with pathogenic effects	Phenotypical differences
Alpha	September 2020	United Kingdom	7	N501Y	transmissibility increased by 50 to 100%, 39 to 72% more lethal, no adverse effects on vaccine effectiveness, rapid spread
Beta	August 2020	South Africa	8	K417N, E484K, N501Y	transmissibility increased by 20 to 113%, the potential risk of death in hospitalized persons increased by 20%
Gamma	July 2020	Brazil	12	K417N, E484K, N501Y	70 to 140% more transmissible, evades immunity 21 to 46% more, mortality increased by 20 to 90%.
Delta	December 2020	India	8	T478K, P681R, L452R	transmissibility increased 97%
Omicron	December 2021	South Africa	30	-	most transmissible of all, most cases have mild symptoms

Regarding VOCs, all the variants not only have substitutions affecting the S protein, but also have many other substitutions affecting other components of the virus. The Alpha variant first appeared in the UK in September 2020. This variant has seven mutations affecting the S protein, which have increased transmissibility by 50 to 100%, and has become 39 to 72% more lethal but has no adverse effects on vaccine effectiveness. Its rapid spread is attributed to the N501Y substitution, which increases its affinity for binding to the receptor. The Beta variant appeared in South Africa in August 2020. It has eight mutations affecting the S protein, which has caused transmissibility to increase by 20 to 113%, the potential risk of death in hospitalized persons increased by 20% and has impacted immune escape after infection and vaccination. This variant has three amino acid substitutions of concern: the K417N and E484K, related to immune escape, and N501Y that increases receptor binding affinity. The Gamma variant appeared in Brazil in July 2020. It has twelve substitutions affecting the S protein, again exhibiting the previously mentioned three substitutions of concern, K417N, E484K and N501Y. This variant has become 70 to 140% more transmissible, evades immunity 21 to 46% more and its mortality has increased by 20 to 90%. The Delta variant appeared in India in December 2020. It has eight substitutions affecting the S protein, which have caused an increase in transmissibility of 97%. This significant increase is due to the presence of the T478K, P681R and L452R substitutions that are known to affect transmissibility. This variant was dominant in new cases worldwide from its emergence until the Omicron variant appeared. The Omicron variant appeared in South Africa in December 2021. Of all the variants, it has the most substitutions affecting the S protein, with a total of thirty substitutions. This variant has the greatest transmissibility and is currently the dominant variant in new cases worldwide, but most cases have mild symptoms[60,62,63].

1.8. Properties of the ACE2 protein used as a receptor by SARS-CoV-2

Angiotensin-converting enzyme 2 (ACE2) is a membrane-bound protein that has a wide variety of (patho)physiological functions. It is a negative regulator of the renin angiotensin aldosterone system (RAAS), facilitates transport of amino acids, and is the receptor for SARS-CoV-2, which allows virus entry into human cells[6,64–66]. RAAS is a complex system (Figure 1.15) that has an important role in maintaining blood pressure, electrolyte and fluid homeostasis and affects the function of many organs and constituents of the human body, such as the heart, kidneys, and blood vessels[64,66–68].

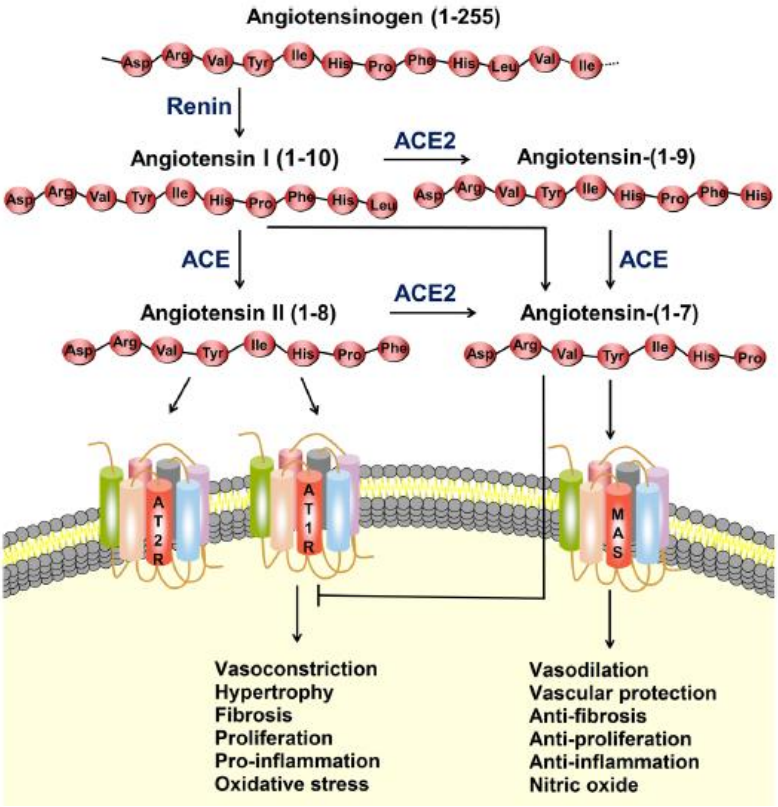

Figure 1.15 – Illustration of the renin angiotensin aldosterone system (RAAS). Renin protease converts angiotensinogen to Ang-I and this will be converted to Ang-II by the angiotensin-converting enzyme (ACE). Ang-II binds to the angiotensin type 1 receptor (AT1R), which will cause actions such as vasoconstriction, hypertrophy, fibrosis, proliferation, inflammation, and oxidative stress. ACE2 can convert Ang-I and Ang-II into angiotensin-(1-7). The latter will bind to the MAS receptor and causing effects such as vasodilation, vascular protection, anti-fibrosis, anti-proliferation, and anti-inflammation. Ang-II may also bind to the angiotensin type 2 receptor (AT2R) in order to counteract the effects mediated by AT1R. The pathway most likely to occur is from Ang-II to angiotensin-(1-7), due to the high affinity between ACE and Ang-I[68].

This protein (Figure 1.16) will interact with the RBD located in the S domain of the S protein of this virus. Thus, increased expression of ACE2 may allow greater susceptibility to viral cell entry.

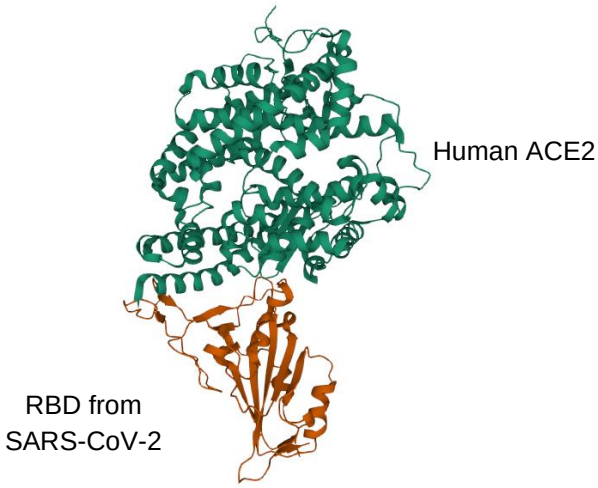

Figure 1.16 – SARS-CoV-2 RBD structure (orange) linked to human ACE2 protein (green). (Taken from PDB. Code: 6M0J).

The first crystallographic structures of this protein were solved in 2004 in response to the SARS-CoV epidemic in the apo- and inhibitor-linked forms (PDB code: 1R42 and 1R4L)[69]. With the emergence of SARS-CoV-2, several structures were solved, such as the full-length, peptidase domain in complex with the coronavirus spike or with the amino acid transporter B^0AT1[70]. This protein is an integral type I membrane protein consisting of 805 amino acids, also acting as a carboxypeptidase. It is a functional homodimer comprising an N-terminal peptidase domain (Protease catalytic domain, PD) where the spike protein binding site is located, and a collectrin-like domain (CLD), containing a neck domain, a single transmembrane sequence (TM) and a cytoplasmic 43 amino acid tail. The neck domain is tethered to the TM by a long linker (Figure 1.17). In addition, this protein at the interface region has two helices called α1 and α2, which are sites that interact with the RBD[69–73].

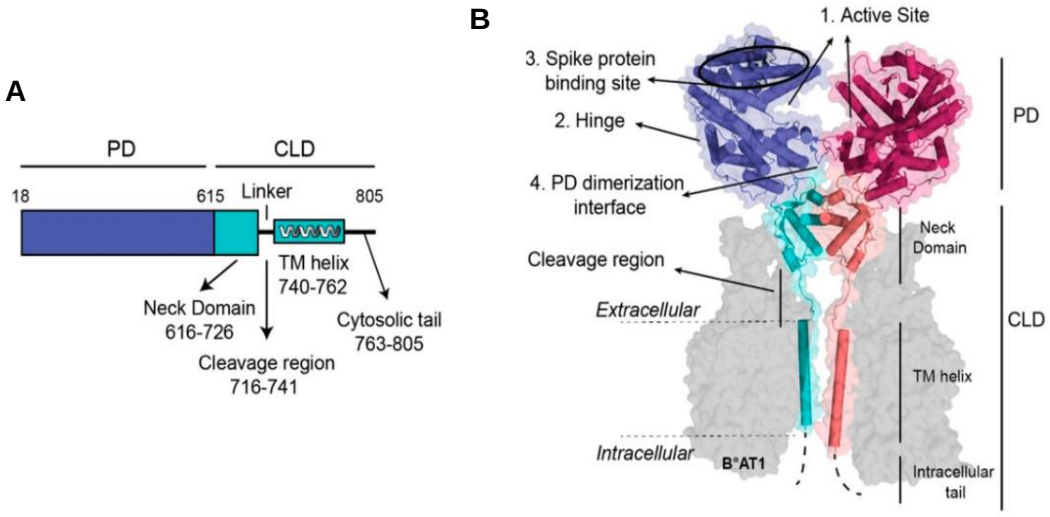

Figure 1.17 – Structure of angiotensin-converting enzyme 2 (ACE2). (A) Schematic representation of ACE2 domains, namely an N-terminal catalytic protease domain, PD (blue) and a C-terminal collectrin-like domain, CLD (cyan). The signal peptide, cleaved during protein maturation, corresponds to the first 17 amino acids of the signal peptide. The extracellular neck domain, the linker, the single transmembrane (TM) helix and the intracellular tail of 43 amino acids are domains of CLD. (B) Structure of full-length ACE2 in complex with B°AT1 (PDB: 6M17). ACE2 is shown as cylindrical helices and loops with surface, whereas the surface of B°AT1 is shown in grey. The different ACE2 regions and the 4 major regulatory sites (1. active site; 2. hinge; 3. claw-like or Spike (RBD)-binding site; 4. PD dimerization interface) are indicated[72].

It is possible for two S protein trimers to bind to one dimeric ACE2, in other words, each S protein trimer binds via one of the ACE2 protein monomers (Figure 1.18)[74,75].

In addition, this protein has seven potential N-linked glycosylation sites at its N-terminal peptidase end, namely asparagines 53, 90, 103, 322, 432, 546 and 690 (Figure 1.18). But it also has several O-glycosylation sites, such as serine 155 and threonine 730. These glycosylation sites, especially those at the N-terminus, contribute substantially to virus binding. On the other hand, the abundance of O-glycosylation sites is low and, since they are far from the interface of interaction with the virus, they probably do not contribute to binding[70,76–78].

Of these seven potential N-linked glycosylation sites, there are three that are more prominent, namely N90, N322 and N546. N90 may offer resistance against infection, and it is believed that its removal will increase the likelihood of spike protein binding to ACE2. On the other hand, the glycan N322 binds to a conserved region of the spike protein, strategically interacting with the RBD of this protein, thus enhancing binding. Finally, through MD simulations, it was observed that the N546 glycan is also involved in the interaction with the RBD, interacting with the N74 and N165 glycans of the S protein (Figure 1.18)[74,76–78].

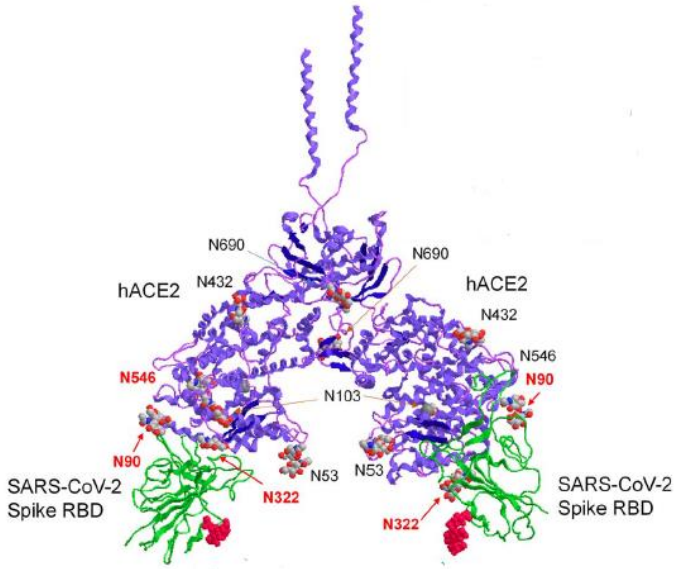

Figure 1.18 – Representation of human ACE2, where the binding of individual monomers of the S protein to the peptidase domain (PD) of each hACE2 monomer is shown. Also depicted are the seven glycosylations present in ACE2, with the three most prominent, N90, N322 and N546, in red[74].

The interface between ACE2 and SARS-CoV-2 is mediated primarily by hydrophobic interactions, and there are several sites in this region of ACE2 that interact with the RBD of the virus (Figure 1.19 A). An extended loop region of the RBD spans the α1-helix of the ACE2 PD as a bond. The α2 helix and the loop that is connecting the β3- and β4-antiparallel strands (loop 3-4) of the PD also contribute to RBD coordination, but in a more limited way. As such, the contact can be divided into three clusters, and there are interactions of the RBD with the N- and C-terminal of α1-helix, but also with small areas of α2-helix and loop 3-4[75].

Regarding the residues that participate directly in the interaction with the spike protein, they are mostly located in the α1-helix of ACE2. At the N-terminus of this helix, residues Q498, T500 and N501 of RBD form a network of hydrogen bonds with Y41, Q42, K353 and R357 of ACE2 (Figure 1.19 B). In the middle α1, residues K417 and Y453 of RBD interact with D30 and H34 of ACE2, respectively, thus reinforcing the interaction due to the involvement of two polar residues (Figure 1.19 C). Finally, at the C-terminal of α1, Q24 of RBD interacts with Q24 of ACE2 by a hydrogen bond, while F486 of RBD interacts with M82 of ACE2 through van der Waals forces (Figure 1.19 D)[72,75].

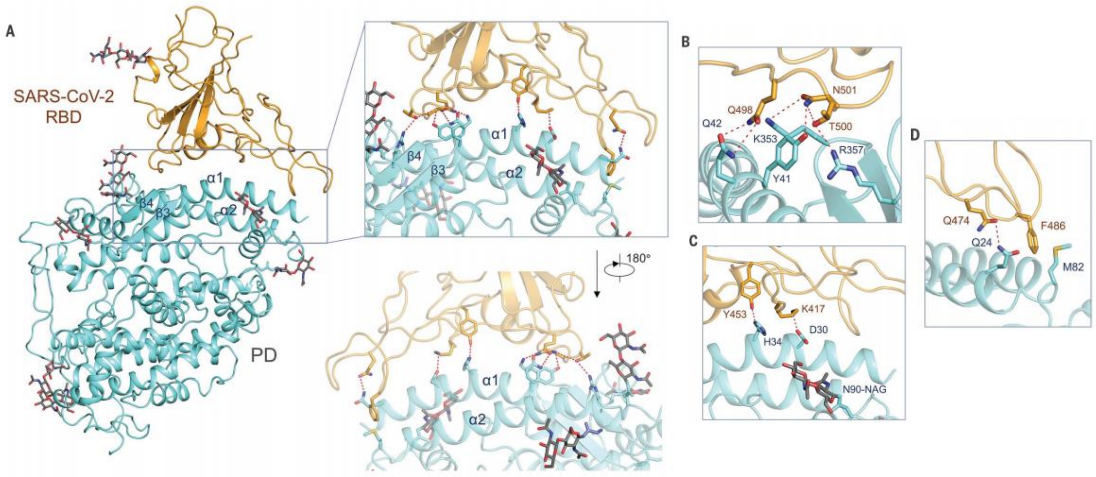

Figure 1.19 – Interactions between ACE2 and SARS-CoV-2 RBD. (A) For RBD recognition, the PD of ACE2 mainly involves α1-helix. However, α2-helix and the linker between β3 and β4 also contribute to the interaction. (B to D) More detailed representation of the interface between RBD and ACE2, regarding the residues of both proteins that are interacting. Polar interactions are indicated by dashed lines in red[75].

The interaction between ACE2 and the RBD of SARS-CoV-2 can be affected by changing the binding affinity through naturally occurring mutations. For the ACE2 gene, several missense mutations are known to exist in the population[79]. In order to try to understand how these mutations influence binding, deep mutagenesis was used. A library was created containing all possible single amino acid substitutions at 117 sites spanning this entire interface, where it was observed that several ACE2 variants present in this library did not bind to RBD, while a smaller number of variants showed higher binding signals. In this study, it was reported that residues buried in the interface tended to be conserved, while residues on the periphery of the interface were tolerant to mutations. It was also observed that the region around the C-terminal of the α1-helix and the β3-β4 strands showed low tolerance to polar residues, while the amino acids at the N-terminal of this helix and at the C-terminal of α2 have a preference for hydrophobic residues[80].

In addition, ACE2 residues N90 and T92, which together form a consensus N-glycosylation motif, were identified as notable hotspots for enrichment substitutions. All the substitutions performed at these two sites were highly favourable for RBD binding, with the exception of T92S which retains the N-glycan, showing once again that this N90-glycan partially hinders the interaction. An example of a favourable substituting amino acid is proline, which was enriched at five positions in this library (S19, L91, T92, T324 and Q325) where it may entropically stabilize the first turns of helices[80]. Thus, all this may explain the greater susceptibility to this virus that a part of the population presents.

Lastly, the ACE2 protein is widely expressed in a wide variety of cells in different human organs. This protein is highly expressed in lung alveolar epithelial cells, where 83% of the cells expressing this protein are located, suggesting that the lung is the primary target of this virus[68,81]. This expression is also found in other tissues such as heart, kidneys, endothelium, intestine, eyes, bladder, pancreas, prostate, testicles, placenta and central nervous system[41,81]. It is also highly expressed in the luminal surface of intestinal epithelial cells, where it functions as a co-receptor for nutrients, namely for the uptake of amino acids from food[64,65,81].

1.9. Different types of treatments and vaccines

At the moment, there are already vaccines against COVID-19, but specific antiviral treatments for this disease are still under development.

In the meanwhile, in order to help patients, supportive treatment has been used to relieve the symptoms and discomfort associated with this disease[25,82]. Some treatments can be provided based

on early symptoms, such as mild analgesics, cough syrup, rest and a high amount of fluid intake[22]. The administration of various antibiotics in order to prevent the appearance of secondary infections is also being used[20]. Other supportive care that can be provided is oxygenation and ventilation in more severe cases[4]. Ventilators help infected patients to breathe and support lung function. Extracorporeal membrane oxygenation (ECMO) is another useful technique to support the body during infection[22,25].

Although there are no specific antiviral treatments in use, there are several strategies being developed (in different stages), using possible targets that are involved in the entry mechanism and replication of this virus. These include small molecules with the ability to inhibit the cleavage of the S protein and, consequently, the ACE2-RBD interaction. An example of this approach are small molecules that inhibit host proteases required for viral entry, such as camostat mesylate which is an inhibitor of the serine protease TMPRSS2[4,12,53]. Other approaches include the inhibition of enzymes involved in viral replication, the development of peptides inhibiting the HR1-HR2 interaction, monoclonal antibody-based therapy and convalescent plasma therapy[53], which consists of using plasma from people who have already recovered from SARS-CoV-2 infection so that the antibodies against this virus present therein specifically target this pathogen and aid the patient's recovery[4,53].

Another example is the development of peptides or small proteins capable of inhibiting viral binding of SARS-CoV-2 to the ACE2 receptor. Disruption of the SARS-CoV-2 RBD/ACE2 complex is the most promising strategy for blocking viral entry at the very first stage of infection. As such, there are several examples of possible approaches to this disruption. The development of ACE2 antagonist peptides/proteins that will interact with the protein-protein interface in order to block the interaction is a promising therapy due to the high specificity and affinity these peptides may possess against the required target, the low interference with other biological processes, their rapid discovery, and their low toxicity due to low possibility of accumulation in the body[53,83,84].

De novo design is another approach that was used, through two ways: (i) computer-generated scaffolds that were designed around a helix of ACE2 that interacts with RBD or (ii) docked against RBD in order to identify new binding types. These have several advantages such as the fact that they cover a range of binding modes and together, escape by viral mutation would be quite unlikely. In addition to this, the high solubility, stability and small size mean that little or no immune response has been observed[85,86].

Another approach is to use decoy ACE2 molecules in order to trap the virus. For this, soluble ACE2 (sACE2) was used, which is able to bind to the spike protein but since it is unable to reform membrane-bound ACE2, it will block the entry mechanism of the virus and consequently its replication. But there are several hurdles to overcome such as an ectodomain of ACE2 having 740 amino acids, so a truncated form of sACE2 is needed to reduce immunogenicity. Furthermore, sACE2 is unstable and can degrade rapidly, which will decrease its ability to trap the virus efficiently. To try to solve these obstacles, a chimeric sACE2 with a human IgG Fc fragment at the C-terminus was generated to stabilize the protein and the resulting construct had a very prolonged plasma half-life. Another attempt at solving the obstacles was the use of novel LNPs (lipid nanoparticles) to package a soluble form of human ACE2, where it was reported that introducing sACE2 into the lungs of a mouse caused SARS-CoV-2 infection to be strongly inhibited and that it can be detected after 48 hours of administration[87].

The use of pseudoligands to dominate the ACE2 binding site for SARS-CoV-2 has also been thought of as an approach. Here a pseudoligand that has a high affinity for the receptor was generated, which could be, for example, a truncated form of viral S protein that only contains the RBD or an engineered artificial receptor-binding motif. But this approach has the risk that the pseudoligand upon binding to the receptor may trigger other intracellular signalling pathways or subcellular responses, such as a cytokine storm[87].

Another proposed approach is to use agents that promote or inhibit ACE2 internalization, either by designing small molecules that can bind to ACE2 and trigger its internalization or by using clathrin

inhibitors to block virus-ACE2 complex endocytosis, where several clinical trials of these types of inhibitors are underway. A limitation in promoting internalization is the reduction of enzymatically active ACE2 on the cell surface, which can lead to increased inflammatory responses and lung damage[87].

Another way of approaching this disease that has been used is the study of the use of existing drugs in the treatment against COVID-19, which have shown to have some positive effect against this virus in clinical trials conducted. Some examples of these drugs are oseltamivir, arbidol, nucleoside analogues and the monulpiravir[6,25,36,53,88–93].

Oseltamivir is a drug approved by clinical trials to treat influenza viruses. Its way of fighting infection is to prevent the secretion of neuraminidase from viral cells, an enzyme that initiates the release of new viruses into host cells. In SARS-CoV-2, this drug causes the exocytosis of new viruses from the host cell[53].

Arbidol, also known as umifenovir, is another drug used to treat influenza virus infections. It inhibits the fusion of the viral membrane with the host cell membrane. It may therefore prevent fusion between the spike protein and ACE2[53].

Nucleoside analogs are inhibitors of the RNA dependent RNA-polymerase (RdRp) enzyme and when entering the viral RNA chain, will cause premature termination. These drugs are shown to have a broad spectrum of antiviral activity against RNA viruses such as Ebola and hepatitis C. Clinically approved (ribavirin and favipiravir) and experimental (remdesivir and galidesivir) nucleoside analogs may have potential to treat COVID-19[36,53,88].

Molnupiravir is a prodrug of the ribonucleoside analog β-D-N4-hydroxycytidine (NHC). In the plasma, it is transformed into NHC and then into the active 5'-triphosphate form by kinases present in the host cell. The latter compound acts as a competitor substrate for virally encoded RdRp and, by incorporating into the viral nascent RNA, will induce an antiviral effect by leading to the accumulation of errors that increase with each cycle of viral replication. With the accumulation of damaging errors throughout the viral genome they cause the virus to become non-infectious and unable to replicate[89,90,92]. This drug was developed by Emory University, USA, originally for the treatment of influenza, and was in pre-clinical testing at the time of the start of the pandemic. After the spread of COVID-19 and as it was seen to have activity against a diverse amount of viruses, including SARS-CoV-2, the development program for this drug shifted to the treatment of COVID-19. It is now being developed jointly by Emory University, Ridgeback Biotherapeutics, and Merck[93]. This drug has already undergone several trials and is currently in phase-III trials. The results of these trials have demonstrated the tolerability, safety, and antiviral efficiency of molnupiravir in reducing virus replication and accelerating the elimination of the infectious virus, thereby preventing COVID-19 progression and eliminating SARS-CoV-2 transmission. This drug could be available to the public in the second quarter of 2022, with the patent having been granted in Australia, Brazil, Canada, China, Europe, England, Japan, Korea, Philippines, and Singapore, and its simple structure will allow for easy synthesis[89,93].

All these drugs are still in clinical trials regarding their positive effects in treating COVID-19.

Regarding existing vaccines in Portugal, those approved and administered to the population are the BioNTech Pfizer, Johnson & Johnson, Moderna and AstraZeneca. Of these four vaccines, two are mRNA vaccines and the other two are viral vector vaccines[10,94–97].

The mechanism of action of mRNA vaccines is based on human cells using DNA as a start material for the production of proteins via an RNA intermediate. Here the mRNA will be used as the template material, and these specially designed mRNA strands will encode a specific antigen of a particular disease. From the moment these RNAs enter the host cells, cellular translational machinery initiates the production of a fully functional antigenic protein from this mRNA. This antigenic protein is then taken up by antigen-presenting cells to activate the innate immune response[53]. Vaccines based on this mechanism of action are the BioNTech Pfizer and Moderna vaccines[94,95]. The synthetic mRNA strand encodes the viral spike protein which will stimulate an antiviral response against this protein[12].

The mechanism of action of viral vector vaccines is based on the use of a modified version of a virus (a vector), in order to deliver genetic instructions to the body's cells. Thus, cells will produce virus-specific antigens that will trigger an innate immune response. Vaccines based on this mechanism of action are Johnson & Johnson and AstraZeneca[98].

1.10. Objectives

Since the S protein is essential for SARS-CoV-2 virus entry into the human body, and the RBD of this protein is the region of interest in this work, the aim of this thesis was the design and production of antiviral proteins that bind to this region and block the interaction between the S protein and the host receptor, the ACE2 protein.

For this purpose, several antiviral proteins were computationally designed and then molecular dynamics simulations of three candidates were performed. After this, experimental methods were applied to produce and test these antiviral proteins.

2. Materials and Methods

2.1. Molecular Modelling

Molecular Modelling is defined as the scientific approach that studies molecular structures using a simplified description of a process or system in order to facilitate predictions and calculations, i.e., a "model"[99,100]. Hence, this technique is concerned with having ways to mimic the behaviour of molecules and molecular systems[100], and can be applied to simple solids and fluids, complex systems, and even biological matter[101].

The appearance and constant advances of computational techniques have revolutionised molecular modelling to the point that, today, most calculations cannot be performed without the use of a computer. All this is due to the great increase in the range of models and systems to which molecular modelling can be applied[100].

Most molecular modelling studies comprise three steps. The first step consists of selecting an approach to describe the intra- and intermolecular interactions that the system has. The two most common approaches used in molecular modelling are quantum mechanics and molecular mechanics. These will calculate the energy of any arrangement of molecules and atoms in systems, allowing to determine how the energy varies depending on changes in the positions of the molecules and atoms. The second stage is where the actual calculation is performed, such as energy minimization and molecular dynamics or Monte Carlo simulation. Finally, the third step is the analysis of the calculations performed and verification that they were performed correctly[100].

Computational studies based on molecular models have started to play an important role in biological chemistry, biophysics, and biology. This is due to the fact that not all molecular properties of biological systems can be analysed by experimental means. Thus, computer simulations can complement the information obtained experimentally, such as conformational distributions or certain interactions between parts of the system[102].

When choosing the type of molecular models that will be used, the type of question that we want to answer must be taken into account, because these models can vary in the level of detail that is used to represent the system (Table 2.1)[102].

Table 2.1 – Examples of molecular modelling levels (Adapted from[102]).

Methods	Degrees of freedom	Time scale
Quantum Dynamics	Atoms, nuclei, electrons	Picosecond
Quantum Mechanics (*ab initio*)	Atoms, nuclei, electrons	No time scale
Classical Statistical Mechanics (MD, MC, force field)	Atoms, solvent	Nanoseconds
Statistical Methods (database analysis)	Groups of atoms, amino acid residues, bases	No time scale
Continuum Methods (hydrodynamics and electrostatics)	Electrical continuum, velocity continuum, etc.	Supramolecular

Molecular Mechanics

Molecular mechanics (MM) is the method selected for performing protein simulations using potential energy functions (force fields)[100,103].

Many of the systems that are intended to be studied by molecular modelling are too large to consider quantum mechanics. This method deals with the degree of freedom of the electrons in a system, so even if some electrons are ignored, a large number of particles still needs to be considered which makes the calculations time-consuming or even impractical in most cases[100].

Molecular mechanics calculates the energy of the system only as a function of velocities and nuclear positions, ignoring electronic motions. This is only possible due to the assumption that the electronic and nuclear motions are independent, called the Born-Oppenheimer approximation. In addition to this approximation, this method considers the atomic nuclei as classical particles, which will imply that their trajectories can be calculated by Newton's laws of motion[100,104]. As such, this method is normally used to perform calculations in systems containing a significant number of atoms[100].

Force Fields

In this method, force fields (FF) or potential energy functions are used to describe the intra- and intermolecular interactions between the various atoms of a system[100]. In the literature there is a wide range of force fields, and the choice depends on the type of study to be performed[105].

The potential energy (V) corresponds to a function of the nuclear atomic coordinates (r) of the particles that constitute the system, being defined as the sum of the terms that represent the covalent and non-covalent interactions between the atoms[100]. A typical force field is represented by the expression in Equation 2.1:

$$V(r) = V_{bonds} + V_{angles} + V_{improper\ dih.} + V_{proper\ dih.} + V_{vdw} + V_{eletrostatic}$$

Equation 2.1 – Equation representing a typical force field. The contributions consist of bond-stretching (V_{bonds}), bond angle bending (V_{angles}), bond torsion ($V_{proper\ dih.}$, proper dihedral torsions) and plane bending ($V_{improper\ dih.}$, improper dihedral torsions). The nonbonded interactions are the van der Waals interactions (V_{vdw}) term and the electrostatic Coulomb interactions ($V_{electrostatic}$)[100].

Bonded contributions consist of bond-stretching, bond angle bending, bond torsion (proper dihedral torsions) and plane bending (improper dihedral torsions), while the nonbonded interactions are the van der Waals interactions presented in the form of a Lennard-Jones term and the electrostatic Coulomb interactions (Figure 2.1)[102].

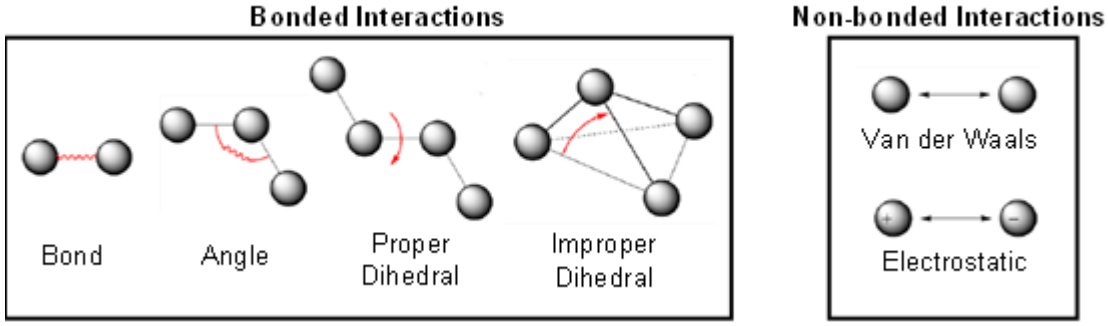

Figure 2.1 – Types of interactions between particles during a molecular dynamics simulation. In bonded interactions there are four contributions, the covalent bonds, bending of the angle, proper dihedrals, and improper dihedrals. In the non-bonded interactions, there are only two contributions, the van der Waals and electrostatic interactions[106].

There are several types of force fields and the choice of which one to use always depends on the user's preference for their target application. An example of force fields is given by Chemistry at Harvard Macromolecular Mechanics (CHARMM), which consists of a set of force fields that can be used for simulations of saturated and unsaturated lipids, protein-nucleic acid complexes, protein-lipids, and DNA-lipids[107,108]. GROMOS can be used to study biomolecular systems and this force field also covers proteins, nucleotides, and sugars[108]. Optimized Potential for Liquid Simulations (OPLS) is a range of force fields for condensed-phase simulations and was originally developed for liquid simulations on a variety of organic compounds. MM3 is one of the most widely used force fields for simulations of organic molecules. The consistent force field (CFF) has also been developed to handle a wide selection of small molecules, including parameters for peptides. Another force field is the Merck Molecular Force Field (MMFF) which has been shown to be successful in condensed phase proteins simulations, although the results have been inferior to those obtained with the AMBER and CHARMM force fields. Two other types of force fields are the Empirical Conformational Energy Program for Peptides (ECEPP) and Coarse-grained force fields, based on a different approach from the previously mentioned ones. Whereas the previously mentioned force fields are based on potential energy functions, ECEPP is the best one known that was optimised to directly reproduce free energies instead of potential energies[107,108]. Coarse-grained force fields provide a systematic way to reduce the number of degrees of freedom representing the system under study and to achieve this, entire groups of atoms are represented by single beads and these force fields describe their effective interactions. An example of this force field is MARTINI which is a coarse-grain parameter set capable of the modelling many systems, including proteins and membranes[109].

Finally, the force field that was used in the simulations performed in this work was the assisted model building and energy refinement (AMBER), more specifically the Amber14sb. This is a force field more commonly used in simulations of nucleic acids and proteins[107]. There are some studies that demonstrate a better performance of this force field in comparison with the previously existing ones, like the ff99SB. Mean errors in the relative energies of conformation pairs were lower and modifications of the backbone and side chains performed better in reproducing their reference parameters and also improved the parameters of the secondary structure content in small peptides. Therefore, the use of the ff14sb force field, in protein simulation began to be recommended[110].

Currently, most biomolecular simulations are realized with the AMBER, CHARMM and GROMOS force fields[107].

The energy terms of a classical force field used in MD simulations are present in Equation 2.2:

A $$V_{bonds} = \sum_{bonds} k_b (b - b_o)^2$$

B $$V_{angles} = \sum_{angles} k_\theta (\theta - \theta_o)^2$$

C $$V_{imp.\ dihedrals} = \sum_{imp.\ dihedrals} k_\omega (\omega - \omega_o)^2$$

D $$V_{proper\ dihedrals} = \sum_{proper\ dihedrals} k_\phi (1 + \cos[n\phi - \gamma)$$

E $$V_{van\ der\ Waals} = \sum_{pairs\ i,j} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right]$$

F $$V_{electrostatic} = \sum_{pairs\ i,j} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}}$$

Equation 2.2 – Equations representing the energy terms of a classical force filed used in molecular dynamics simulations. Each equation corresponds to the calculation of the potential energy of each type of contribution[106].

Bonded interactions

The bonded interactions (Figure 2.2) are based on a fixed number of atoms and consist of different contributions[108]. Each contribution is represented by one of the previous equations.

Equation 2.2A represents the stretching of covalent bonds which will keep the bond length (b) close to the respective reference value (b_o) and is modelled by a harmonic potential, treating bonds as springs. The rigidity of this bond is regulated by a force constant (k_b)[103,106,111].

Equation 2.2B represents the bending of the angle between two covalent bonds that will be modelled by a harmonic potential function, which will increase as the angle (θ) moves away from its reference value (θ_o). It is through the value of the force constant (k) that the energy penalization due to deviations from the reference angle is controlled[103,106,111].

In equation 2.2C is represented the improper dihedral, a term which is used to maintain the chirality of the tetrahedral centres, that is, in order to prevent the molecules from turning towards their mirror images. It is also used to maintain the geometry of planar groups. This is a harmonic potential, which will penalise deviations from the improper dihedral-angle reference (ω_o), according to the force constant (k)[106,111].

Finally, in equation 2.2D are represented the proper dihedrals, which model the energy associated with the rotation over a single chemical bond. In this equation, the proper dihedrals are modelled by a periodic function ($k\phi$) that will determine the barrier height to rotation, the phase shift (γ) corresponds to the location of the maxima in the dihedral energy surface and n represents the multiplicity, that is, the number of cycles per 360° rotation about the dihedral[106,111].

Non-bonded interactions

The non-bonded interactions (Figure 2.2) consist of two groups, the van der Waals and the Coulombic (electrostatic) interactions that occur between all pairs of unbonded atoms[111]. The terms that constitute them account for the potential energy of non-bonded interactions between atoms in the same molecule or atoms in different molecules, the latter separated by more than two or three covalent bonds[100,103].

In computational studies of biological systems, these terms are very important, due to the large number of these types of interactions occurring in biological molecules themselves and the strong influence that the surrounding environment has on the properties of macromolecules[111].

In equation 2.2E are represented the van der Waals interactions, classified as Leonard-Jones (LJ) 12-6 type. These are represented by two terms, a repulsive term that varies with distance raised to the 12^{th} and an attractive term that varies with distance raised to the 6^{th}, rm is the distance at which the potential reaches its minimum and r_{ij} is the distance between the atom i and j[106,111].

The electrostatic interactions are represented in equation 2.2F, and these are described by a Coulomb's law between pairs of atomic partial charges q_i and q_j that are divided by the distance r_{ij}, between atoms i and j, always taking into account the dielectric constant (ε_o). These interactions depend on the charge of the atoms and the distance between them[106,111].

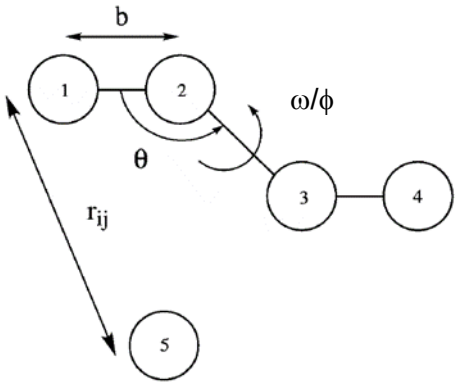

Figure 2.2 – Illustration of two molecules in order to represent the energy terms. Molecule A consists of atoms 1 to 4 and molecule B consists of atom 5. The covalent terms occur between atoms 1 to 4: the bonds (b) between atoms 1-2, 2-3 and 3-4, the angles (θ) between atoms 1-2-3 and 2-3-4 and the improper dihedrals (ω)/proper dihedrals (ϕ) between the 4 atoms. The non-bonded terms occur between all pairs of atoms, which are separated by more than two bonds, over an interatomic distance r_{ij}, which include electrostatic and van der Waals interactions (Adapted from[111]).

Energy Minimization

The changes that occur in the potential energy of the system along the coordinates can be considered as movements on a multidimensional "surface", also called potential energy surface[100]. In MM there are regions of this surface that are more relevant corresponding to stable states of the system, that is, energy minimum points[100].

In order to identify the minimum points on the energy surface, a minimization algorithm is used. When choosing the most appropriate algorithm for the problem under study, several factors must be taken into account, and the energy minimization algorithm is one that provides a quick answer[100].

In the various studies carried out in this Thesis, the steepest descent algorithm was used to perform the system energy minimization. This is a first order derivative minimization algorithm that descends in a straight line on the energy surface of the system. This process stops after a fixed number of interaction steps occur or when the energy gradient is less than a certain predefined value[100].

2.1.1. Protein Designs

Protein design is a very useful tool for creating proteins that possess certain desired properties, but also for investigating their sequences and structural and functional relationships[112]. This is based on joining energy functions to evaluate candidates with the use of search algorithms in order to analyse large sets of candidates[113]. Despite great advances in recent decades, a reliable design of a stable, well-folded and soluble protein with the desired structure usually requires several trials, improvements, and significant resources[114].

The emergence of computational protein design (CPD) enabled exploring complete sequences of new topologies and, more recently, to design new topologies[114]. In most CPD methods, a discrete set of amino acid states and sidechain orientations relative to a single backbone structure is optimized according to a scoring function[115]. Furthermore, this technique is also used for problem solving in protein engineering, such as the *de novo* design problem, i.e., the creation of novel protein folds, binding interfaces, or enzymatic activities[113].

De novo protein design is often related to the specific fold of a certain structure relative to its amino acid sequence, but it will be the conformational changes that often mediates the protein function[113].

There are several computational programs that can be used for protein design, some of them being Schröndinger software[116], Molecular operating environment (MOE)[117] and the Rosetta software[118].

The software used for this work was Rosetta. This consists of a suite of programs, which include various algorithms for the computational modelling, design, prediction, and analysis of protein structures. This program started as a structure prediction tool but has evolved to offer a wide variety of effective algorithms to explore certain features of proteins, such as the backbone and sidechain. It also has scoring (energy) functions, containing a wide list of applications from folding, to docking and design[118].

RosettaScripts is an XML-like language belonging to the Rosetta program that specifies modelling tasks, such as protocols specified for certain experiments like sequence redesign or rigid-body docking[119].

Regarding the designed proteins employed in this Thesis (Table 2.2), to design the first binder candidate, the initial step consisted of selecting the ACE2 RBD-binding motif with 84 residues, between residue 19 and 103. This protein was designated as the Helix-Turn-Helix 1 (HTH1). The structure of HTH is composed of two long, parallel helices involved in ACE2 binding to Spike RBD (helix 1 and 2) and a short helix that links the other two in order to stabilise them. The first candidate was named HTH1 and although it is isolated from the context of ACE2, it alone is expected to be able to maintain its native structure, preserve its protein integrity and the binding mode of ACE2 to the RBD.

Since amino acid residues that were previously in contact with other residues in the ACE2 protein become exposed in HTH, which may destabilize it. In order to reduce the impact of removing HTH from its context and at the same time improve the binding affinity to RBD, the design of the HTH binding interface in the presence of RBD was proposed using the Position Substitution Score Matrix (PSSM), obtained from 181 non-redundant homologous sequences. This led to the design of a second candidate (HTH3). The solubility and structural stability of HTH3 was subsequently improved by 3 human-guided amino acid substitutions distal from the binding site. Another version of the same construct with two disulfide bonds (HTH2) was strategically created to maintain the native fold and avoid disruption of the tertiary structure under critical dynamics stresses. The Rosetta disulfidize tool[118] was used for this purpose. The algorithm searches for pairwise beta-carbons close enough to build disulfide bonds with a mutagenic energy cut-off of 1.0 rosetta energy unit.

Table 2.2 – Sequences of the three proteins, the respective 10 substitutions performed in order to try to improve the interaction between the HTH3 protein and RBD and the four cysteines that are in the sequence of HTH2.

Protein	Sequence
HTH1	STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQSTLAQM YPLQEIQNLTVKLQLQALQQN
HTH2	SDIEEQAKTFCDKFNVRAEDCYYQSSLASKNYNTNITEENVQIMNNAGDKCSAFLKEQSTEAQKY PLQEIQNDTVKRCLQALQQN
HTH3	SDIEEQAKTFLDKFNVRAEDLYYQSSLASKNYNTNITEENVQIMNNAGDKWSAFLKEQSTEAQKY PLQEIQNDTVKRQLQALQQN

After this, protein structure prediction methods are mandatory to determine the native folding propensity. Thus, the HTH1, HTH2 and HTH3 folding was obtained by an Ab Initio prediction protocol that is available in the Rosetta suite program, designated AbinitioRelax[120]. This consists of the modeling of tertiary structures by means of 3 to 9 residue fragment libraries of known structure homologous sequences and Monte Carlo simulated annealing algorithms driven by an energy score function[121]. These libraries form local and non-local interactions that are subsequently grouped into folded proteins (decoys). They are then compared to the native structure in terms of RMSD and total energy, until the energy versus RMSD plot is shaped like a protein folding funnel. In this study, the fragment libraries

were created by Robetta web server[122]. All Ab Initio structure designs and predictions were performed using RosettaScript and the AbinitioRelax tool, respectively, both using the ref2015 score function[118].

Although I have followed the procedures described above, the actual designs were essentially obtained by Dr. Carlos Cruz.

2.1.2. Molecular Dynamics Simulation

Molecular Dynamics (MD) simulations is a method that generates successive configurations of the system with the correct distribution. Through Newton's laws of motion integration, the trajectory (positions over time) will specify how the velocities and positions of the particles present in the system will vary over time[100].

The trajectory is calculated with Newton's second law of motion integration, which states (Equation 2.3):

$$F_i = m_i a_i$$

Equation 2.3 – Equation of Newton's second law of motion. F_i is the force that is exerted on a particle i, m_i is the mass of this particle and a_i is the acceleration[111].

The acceleration of the particle is the first derivative of its velocity (v_i) and the second derivative of its position (r_i), thus Newton's second law can be represented as it is in Equation 2.4:

$$F_i = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r_i}{dt^2}$$

Equation 2.4 - Equation of Newton's second law of motion. The acceleration of the particle can be represented as the first derivative of its velocity (v_i) and the second derivative of its position (r_i)[100,111].

This method is a tool widely used in various areas of science, as it provides the methodology for detailed microscopical modelling at the atomic scale. Thus, allowing the prediction of the static and dynamic properties of substances, directly from the analysis of the interactions between molecules over time[111].

Integration Algorithms

In order to integrate Newton's laws of motion and update the velocities and positions of particles over time, several algorithms can be used. The leap-frog algorithm can also be called half-step scheme, i.e., the velocities are evaluated in the middle of the evaluation point of the position and vice versa[111]. This algorithm can be written in two ways as it is in Equation 2.5:

$$r_{n+1} = r_n + v_{n+1/2}\Delta t$$

$$v_{n+1/2} = v_{n-1/2} + \frac{F_n}{m}\Delta t$$

Equation 2.5 – Equations representing the leap-frog algorithm. $v_{n\pm1/2}$ represents the velocity at the mid-step ($t \pm \frac{1}{2}\Delta t$)[111].

This algorithm consists of three steps: in the first step the current position r_n is used in order to calculate the current force F_n. In the second step, this force and the previous mid-step velocity $v_{n-1/2}$ are used to calculate the next mid-step velocity $v_{n+1/2}$. Finally, in the third step, the current position r_n and the next mid-step velocity $v_{n+1/2}$ from the previous step are used to calculate the position in the next step, r_{n+1}[111].

The name leap-frog comes from the fact that this algorithm allows velocities to "leap over" positions and then positions to "leap over" velocities[111].

The choice of the time interval to be used, t, must be made carefully, always taking into account the properties of the molecular system. Besides this, it should also be considerably smaller than the time characteristic of the fastest motion studied. But, on the other hand, the time step should be as large as possible, in order to speed up the simulation[111].

A strategy that is widely used to increase the integration time step is to limit the length of the bonds to their reference values by using constraint algorithms. In order to allow the use of larger time steps, the LINCS constraint algorithm was used, which will remove fast vibrations from the system, which are usually not coupled with relevant slow motions[123].

MD simulations with periodic boundary conditions

In molecular dynamics simulations, one must think about how to deal with system boundaries when working with a system that is finite. Because of this, periodic boundary conditions (PBC) were used in the studies performed in order to minimize the effects of boundaries.

When you are performing a simulation in a box of a certain substance in vacuum, the forces that will act on the atoms near the edges of the box will be different from those on the atoms near the centre. Also, if there were no walls, the density would decrease because the atoms would move out of the box over time[100,107].

In order to solve this problem, periodic boundary conditions are used. This allows the simulation to be performed using a relatively small number of particles, in a way that the particles feel the forces as if they were in bulk fluid. When using PBC, it is considered that the simulation of the system is being replicated in all directions with similar boxes, with the molecules being able to move between these adjacent replicated boxes (Figure 2.3). Thus, each atom in the central cell is under the influence of all the other atoms in the box, both those in the central cell and in the other cells, so there are no effects due to walls. In reality, this periodicity is repeated due to mathematical operations, and duplicate boxes are not explicitly simulated[100,107].

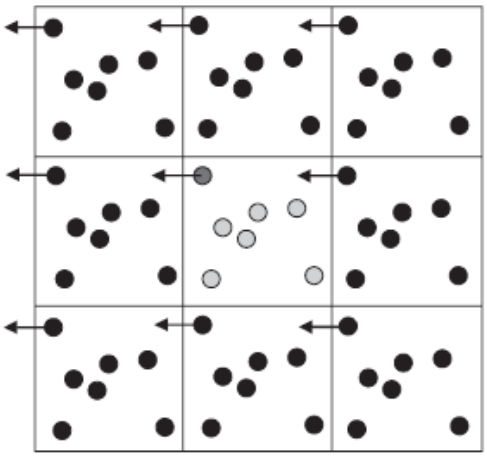

Figure 2.3 – Illustrating a two-dimensional slice through a small portion of the system (the central box where the atoms are represented in gray) and the copies (the atoms are represented in black). Each copy is identical at the atomic level and each atom is subjected to the same time development as its image in all the other copies. When the dark gray atom leaves the central cell, its image will enter from an adjacent copy, which is represented by the vector displacements in the figure, which will cause the density to remain constant[107].

When PBCs are used, the boxes for the simulations should be large enough to prevent unwanted interactions, i.e., a molecule should only interact with the nearest neighbour of each particle, which is called the minimum image convention[100,105].

MD simulations at constant temperature and/or pressure

The MD simulation methods aim at simulating the bulk properties of the system, such as the number of particles (N), the temperature (T) and the volume (V) of the box in which the simulation was performed. In order to obtain thermodynamic averages of the system, different conformations of the system are tested and thus an average calculation of this vast number of conformations can be obtained. This is called ensemble averaging because it encompasses all possible states of a macroscopic system[111].

MD is commonly performed in the constant NVE or NVEP ensemble, also called microcanonical ensemble, where the energy (E), particle number (N) and volume (V) are kept constant in each cell, the system being isolated from the surrounding environment[100,111]. This is a very simple ensemble because energy cannot flow from one cell to the other[107]. Another existing ensemble is called canonical ensemble, where the temperature is kept constant, by coupling to an external bath, so N, V and T do not change during the simulation[111].

On the other hand, most real systems are not thermally isolated from the surrounding environment, so simulations performed with N, T and pressure (P) kept constant, which is called isothermal-isobaric ensemble, or with V, T and the chemical potential kept constant, which is called grand canonical ensemble, will be more realistic[107,111].

In the canonical ensemble or NVT, it is through the Boltzmann distribution, represented in Equation 2.6, that one obtains distribution of the microscopic states of the system[124]:

$$P_{\mathrm{i}}(N,V,T) = \frac{e^{\,-\mathrm{Ei}\,(\mathrm{N,V})/\mathrm{kT}}}{Q(N,V,T)}$$

Equation 2.6 – Equation of the Boltzmann distribution. E_i is the energy of state i, k is the Boltzmann constant and T is the absolute temperature[124].

Currently, the vast majority of MD simulations are performed at constant T, enabling the comparison of MD results with those obtained in *in vivo* or *in vitro* experiments that performed at a fixed temperature. On the other hand, when constant temperature simulations are performed at different temperature values, it makes it possible to analyse how the system's behaviour is influenced by this parameter.

Regarding temperature, this is a thermodynamic property that is given by the average kinetic energy of the system over the time that the simulation is running. In an MD simulation, the most common way to control the temperature is to couple the system to an external heat bath, fixed at a desired temperature and will serve as a thermal energy source that provides or removes energy from the system[100]. An example of one type of temperature coupling algorithm is the Berendsen bath[123], which ensures that the rate of temperature change is proportional to the difference between the bath temperature (T_{bath}) and the system temperature (T(t)). Here the deviation of the system temperature from the bath temperature is slowly corrected (Equation 2.7)[100].

$$\frac{dT(t)}{dt} = \frac{T_{\mathrm{bath}} - T(t)}{\tau}$$

Equation 2.7 – Equation of the deviation of the system temperature from the bath temperature. τ is a coupling parameter that determines how much the bath temperature, and the system temperature are coupled. The higher the τ, the weaker the coupling will be[100].

In practice, the velocities will be multiplied by a time-dependent scaling factor (λ) at each step (Equation 2.8), which is given by:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_{\text{bath}}}{T(t)} - 1\right)\right]$$

Equation 2.8 – Equation of a time-dependent scaling factor (λ) at each step[123].

In some cases, it is necessary to perform the simulation at constant temperature and pressure, i.e., in an NPT ensemble. In these cases, the pressure can also be controlled through the coupling of the system to a pressure bath, as with temperature. An example will be the Berendsen pressure algorithm (Equation 2.9), which will scale the box coordinates and vectors at each time step so that the pressure can relax to a given reference pressure (P_{bath})[100,123]:

$$\frac{dP(t)}{dt} = \frac{P_{\text{bath}} - P(t)}{\tau_P}$$

Equation 2.9 – Equation of the deviation of the system pressure from the bath pressure[123].

In order to rescale the vectors and the box coordinates, this algorithm uses a scaling factor (μ) (Equation 2.10).

$$\mu = 1 - \frac{\beta \Delta t}{3\,\tau_P}(P_{\text{bath}} - P)$$

Equation 2.10 – Equation of a pressure-dependent scaling factor (λ) at each step. β is the isothermal compressibility of the system[123].

2.1.3. Simulation Setup for HTH proteins

The methodology used in the MD simulations performed in this work for the three proteins in water, it started with the choice of the initial coordinates, which it was from the PDB code 6M0J. The force field was Amber14sb, as previously mentioned. The proteins were solvated with water, using the TIP3 water model, in a truncated octahedron box, setting the minimum distance between the peptide and the box walls to 1.5 nm. Cl^- and Na^+ ions were added as required to maintain the neutrality of the system.

An energy minimization procedure was performed using a steepest descent algorithm with 50,000 steps.

Then, the system was initialized through four steps of 100 ps. In the first step, the initial velocities were generated, and the positions of all atoms were restrained at their initial positions by a harmonic function with 100 kj/Å^2 force constant. The temperature was kept at 300 K with the Berendsen coupling algorithm[123], using 0.01 ps as the temperature coupling constant. Temperature coupling was applied to the protein and ions together and to the solvent atoms separated from the others. No pressure coupling was applied in this step, which was performed at NVT.

In the second step, the same parameters were used as in the previous step, except that the temperature coupling constant was increased to 0.1 ps. In this step, pressure coupling was still not applied. Once more, the positions of all atoms were restrained at their initial positions.

In the third step, the temperature coupling parameters remained the same as in the second step, but now isotropic pressure coupling was activated, maintained at 1 bar, using the Berendsen pressure algorithm with 5.0 ps pressure coupling constant. Thus, the system was now simulated in the NPT ensemble. In this step, only the C-α atoms were position-restrained.

In the last step, the thermostat was switched to the V-rescale algorithm[125] and the barostat to the Parrinello-Rahman algorithm[126], with a lower pressure coupling constant of 2.0 ps, with the pressure remaining at 1 bar. Once more, only the C-α atoms were position-restrained.

We then proceeded to perform the production of the MD simulation, for each of the three proteins, but also for the three proteins in complex with the Spike protein RBD. Six different systems were simulated (Table 2.2) using the program GROMACS version 2020.3[127], each with 1000 ns. For each simulation, 5 replicates were run. During the simulation, the equations of motion were integrated using a time step of 2 fs. Periodic boundary conditions were used for the simulations. The temperature was maintained at 300 K, using the V-rescale algorithm with a time constant of 0.1 ps, and the pressure was maintained at 1 bar using the Parrinello−Rahman pressure coupling algorithm, with a time constant of 2.0 ps, with an isotropically applied pressure coupling.

The long-range interactions were treated with the Particle Mesh Ewald (PME)[129], using a grid spacing (fourier spacing) of 0.12 nm, with a cubic interpolation (pme-order =4). The neighbor list was updated every twenty steps, with a Verlet cutoff scheme with a 0.8 nm radius[128]. All bonds were constrained using the LINCS algorithm.

Table 2.3 – The systems simulated for each of the HTH, free in water and in complex with the RBD.

Proteins	System simulated
HTH1	Protein free in water
HTH2	
HTH3	
HTH1-RBD	Protein in complex with the RBD
HTH2-RBD	
HTH3-RBD	

2.2. *Escherichia coli* expression strain transformation of the Helix-Turn-Helix Proteins

To express the designed proteins in a bacterial expression system, the corresponding vectors have been designed and ordered from GeneCust®. The synthetic genes, containing a sequence for an N-terminal hexa-histidyl tag followed by a sequence for the proteolysis site of HRV-3C protease, were subcloned into pET28a(+) (Figure 2.4). All constructs were sequenced to confirm the correct insertion of the synthetic genes.

Transformation is a process that involves inserting foreign DNA into the host cell so that it begins to express the protein of interest.

The plasmid pET_HTH1 was transformed into different strains of *Escherichia coli* for heterologous expression: BL21 (DE3) pLysS, BL21 (DE3) and BL21 (DE3) Star. The plasmids pET_HTH2 and pET_HTH3 were only transformed into *E. coli* BL21 (DE3). When necessary, plasmid DNA amplification was performed using the *E. coli* DH5α strain.

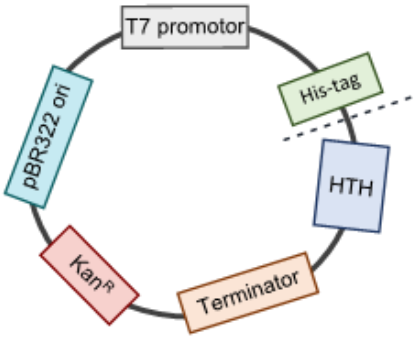

Figure 2.4 – Illustration of plasmid pET_HTH1 based on pET28a(+). This plasmid is constituted by several critical elements, a selection marker that corresponds to the antibiotic kanamycin (Kan^R), the origin site, the T7 promoter, the zone corresponding to the affinity tag (His-tag), the proteases cutting site (----), the coding sequence for the protein of interest (HTH) and the termination site.

Bacterial cells were mixed with 150 ng plasmid and incubated on ice for 30 minutes. The mixture was then heat shocked for 45 seconds at 42 °C in a water bath and then immediately placed on ice to rest for 5 minutes. After the heat shock, 600 µL of Luria-Bertani (LB) medium (made in the lab - Annex 1) without antibiotic was added and the cells were incubated at 37 °C in an Eppendorf® Thermomixer compact for 1 hour at 500 rpm. The cell suspension was centrifuged (Eppendorf® Centrifuge 5415 D) at 3600 rpm for 2.5 minutes and 450 µL of the supernatant was removed. The pellet was resuspended in the remaining medium and plated on a LB agar plate (made in the laboratory - Annex 1) supplemented with kanamycin (50 µg/mL), in order to select the *E. coli* cells that contain the construct, these constructs having kanamycin as selection marker. For colony formation to occur, the plated cells were incubated at 37 °C overnight and stored at 4 °C. No colony was used with a storage time longer than 2 weeks.

2.3. Expression and purification of the Helix-Turn-Helix Proteins

2.3.1. Small-scale expression

In order to determine the optimal expression conditions, several different conditions were tested.

Three different *E. coli* strains (BL21 (DE3), BL21 (DE3) pLysS, and BL21 (DE3) Star) and two different types of culture media, LB and M9 minimal medium (made in the lab - Annex 1) were used.

Freshly grown colonies were used to inoculate 10 mL overnight cultures in LB medium. All these cultures were grown in 50 mL Falcon tubes, in a shaker incubator (ARALAB™ Agitorb 200 incubator) at 37 °C overnight and 140 rpm, supplemented with kanamycin (50 µg/mL). The next day, these overnight cultures were used to prepare small-scale cultures at an initial OD of 0.05 in 30 mL LB and M9 medium supplemented with kanamycin (30 µg/mL). These small-scale cultures were grown in a shaking incubator at 37 °C with a cooling system (New Brunswick™ Innova® 44 Incubator Shakers) at 120 rpm. When the OD of the cultures reached values between 0.6 and 0.8, they were induced with 0.4 mM Isopropyl β-D-1thiogalactopyranoside (IPTG) and the temperature was lowered to 18 °C. Aliquots of 1 mL were collected before induction, 4 hours after induction and overnight.

The collected aliquots were centrifuged (Eppendorf® Centrifuge 5415 R) at 12000 rpm, for 10 minutes at 4°C. The sample pellets were resuspended in 150 µL of lysis buffer (150 µl BugBuster (made in the laboratory - Annex 1), 0.5 µL Deoxyribonuclease I (DNAse I, Sigma-Aldrich®) at 1 mg/mL and 1.5 µL Lysozyme (Sigma-Aldrich®) at 2 mg/mL, adjusting the volume of each compound to the number of samples), incubated for 20 minutes on ice and centrifuged at 1300 g for 30 minutes at 4°C. After centrifugation, the supernatant was transferred to a new eppendorf and 50 µL 4× Loading buffer (made in

the laboratory - Annex 1) (soluble fractions) was added. The pellets (insoluble fractions) were resuspended in 200 µL 1× Loading buffer, boiled at 95 °C for 10 minutes and mixed by vortexing. Finally, all samples (soluble and insoluble) were boiled at 95 °C for 5 minutes.

All samples were analysed by SDS-PAGE (NuPAGE™ 10% Bis-Tris Gel at 200 V for 35 minutes with Bolt™ MES SDS Running Buffer (Invitrogen™)) and western blotting.

2.3.2. Large-scale expression

According to the small-scale expression trials, freshly grown colonies were used to inoculate 50 ml overnight cultures in LB medium. These cultures were grown in 250 ml Ultra Yield™ Flasks, in a shaker incubator at 37 °C overnight and at 140 rpm, supplemented with kanamycin (50 µg/ml). The next day, these overnight cultures were used to prepare larger scale cultures. Several batches of 4L and 8L of these were grown for the three proteins, where each 1L of LB supplemented with kanamycin (30 µg/ml) was placed in 2.5L Tunair™ Shake Flasks with an initial OD of 0.05. When the OD of the cultures reached values between 0.6-0.8, they were induced with 0.4 mM IPTG and incubated for 4 hours at 18 °C. Finally, and cells were collected by centrifugation (Avanti® J-26 XPI Centrifuge, Beckman Coulter®, JA-10 rotor) at 7500 rpm for 10 minutes at 4 °C. Then, the obtained pellets were placed in 50 ml falcons and in order to remove the remaining medium, they were centrifuged (Eppendorf® Centrifuge 5804 R) again at 7500 rpm, for 10 minutes at 4 °C and finally stored at -20 °C.

2.3.3. Soluble Fraction Purification

Frozen HTH1- (~ 19.2 g from an 8L growth) and HTH3-expressing (~ 21.3 g from a 4L growth) cells were resuspended, with mild stirring for about 30 minutes, in 60 mL of buffer A (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 100 µM TCEP and 10 mM imidazole) supplemented with 1 mg/mL Lysozyme (Sigma-Aldrich®), 20 mU of Benzonase® Nuclease (ChemCruz™), 1:500 protease inhibitor cocktail (500 µL PBS 1× and 1 Protease inhibitor cocktail tablet, Sigma-Aldrich) and 2 mM $MgCl_2$. HTH2-expressing cells (~ 14.1 g from a 4L growth) were resuspended in 45 mL of buffer A supplemented with the same compounds as above. Cells were further disrupted by sonication (Hielsher UP200S sonicator) with 50% amplitude and 0.6 of cycle, 30 seconds pulse on followed by 30 seconds off, for 8 minutes. In order to isolate the soluble proteins, the lysed cells were centrifuged (Avanti® J-26 XPI Centrifuge, Beckman Coulter®, JA-25.50 rotor) at 18000 rpm, for 30 minutes at 4 °C, and finally the supernatant was filtered (membrane with a 0.45 µm pore size).

The clarified lysate was injected into an affinity chromatography 5-mL Nickel HisTrap™ HP column (GE Healthcare) at room temperature (RT), previously equilibrated with buffer A using a ÄKTA FPLC system (GE Healthcare). Bound proteins were eluted with buffer A supplemented with 500 mM imidazole (buffer B) through a linear gradient to 50% buffer B in 20 column volumes (CV) and finally a 100% step in order to remove tightly bound proteins.

The fractions containing the proteins were all pooled and then the sample was concentrated in concentrators with a 3 kDa cut-off (Amicon® Ultra-15 Centrifugal Filter Units, Merck Millipore®) through several centrifugations (Eppendorf® Centrifuge 5804 R) at 5000 rpm each for 15 minutes at 4°C. The sample was concentrated to a sample volume of 2.5 mL.

The sample was then centrifuged (Eppendorf® Centrifuge 5415 R) at 12000 rpm for 10 minutes at 4 °C and then injected into a HiLoad® 16/600 Superdex® 75 size-exclusion chromatography (GE Healthcare), previously equilibrated with buffer C (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol and 100 µM TCEP). In the case of HTH1 and HTH3, only the peak of interest was collected, which corresponded to protein. For HTH2, two peaks were collected as they corresponded to different oligomeric forms of the protein. The fractions containing the protein were concentrated and quantified by Nanodrop (Alfagene® NanoDrop® Spectrophotometer ND-100).

Finally, in order to remove the tag from the proteins, the obtained samples were incubated with 1:100 of 1 mg/mL of His-3C Protease and this incubation was left overnight at 4°C.

The following day, the sample was injected into a 1 mL Nickel HisTrap™ HP column (GE Healthcare), previously equilibrated with buffer D (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 100 µM TCEP). As the protein of interest was no longer bound, the tag was eluted with this buffer and then the remaining proteins, such as the 3C protease, the tag and any impurities still present in the sample, were eluted with a step of 100% buffer E (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 100 µM TCEP and 500 mM imidazole). The fractions of interest were pooled and concentrated to be injected in a Superdex® 75 10/300 GL size exclusion chromatography (GE Healthcare), equilibrated with buffer D. The fractions of interest peaks were pooled and proceeded to the last concentration step, after which they were quantified by the Bradford method, frozen and stored at -80 °C.

Between each purification step a 15% sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gel electrophoresis (done in the laboratory - Annex 1) was performed, run at 150 Volt for 1 hour, in order to monitor the purity of the sample under study. At the end of the whole protocol a western blot was performed.

This protocol was performed for the three proteins separately.

2.3.4. Insoluble Fraction Purification

Two more 8L growths were performed, one to obtain HTH1 and the other HTH2, in order to try to recover the proteins from the inclusion bodies (IB).

The frozen cells from HTH1 (~ 33.6 g) and HTH2 (~ 20.8 g) were resuspended in 60 mL and 40 mL of buffer A from the purification of the soluble fraction, respectively. The remaining protocol for the treatment of the lysed cells was the same as that performed previously.

The difference is that after centrifugation at 18000 rpm for 30 minutes at 4°C, the pellets were removed from the centrifuge tubes and stored in falcons for later weighing and initiation of the inclusion bodies solubilization protocol.

On weighting the pellets of both proteins, the sample of the HTH1 inclusion bodies weighed 16.5 g while that of HTH2 weighed 11.5 g. The pellets were then resuspended in a volume of wash buffer (10 mM sodium phosphate pH 6.5, 500 mM NaCl, 2% Triton X-100 and 2 M Urea) equal to 3× the weight of the inclusion bodies. Then, a centrifugation (Avanti® J-26 XPI Centrifuge, Beckman Coulte®) at 10000 g for 10 minutes at 9°C was performed and then the pellet from this last centrifugation was resuspended in a volume of buffer A (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 500 µM TCEP, 20 mM imidazole and 6M Urea) equal to 10× the initial weight of the inclusion bodies for 30 minutes on ice. Finally, an ultracentrifugation (BECKMAN COULTER® Ultraoptima™ LE-80K C0L02K09) was performed at 118000 g for 1 hour at 4°C and the supernatant after this centrifugation was stored at 4°C.

The next day, the sample was injected into a 5 mL Nickel HisTrap™ HP column (GE Healthcare), pre-equilibrated with buffer A. The sample was injected into the column using a P-1 peristaltic pump (GE Healthcare). Proteins that did not bind to the column were eluted with buffer A and the protein was eluted with buffer B (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 500 µM TCEP, 500 mM imidazole and 6M Urea) with 4 steps of 10%, 20%, 40% and 100% buffer B.

The fractions of interest were then pooled and diluted with buffer C (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 500 µM TCEP and 6M Urea) to a final imidazole concentration of 40 mM. The sample was then injected into a 5 mL Nickel HisTrap™ HP column (GE Healthcare), previously equilibrated with buffer C, with the aid of a peristaltic pump P-1 (GE Healthcare). The protein was renatured by passing buffer D (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 500 µM TCEP) through a 20-column volume gradient to 100% of this buffer at 1 mL/min and a step of

100% of buffer D for 20 minutes. After this step, the protein was eluted with buffer E (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol, 500 µM TCEP and 500 mM imidazole) in a 15-column volume gradient to 100% of this buffer.

The fractions of interest were pooled and concentrated and then injected into a Superdex® 75 10/300 GL size exclusion chromatography (GE Healthcare), equilibrated with buffer D.

Again, the fractions of interest were pooled and concentrated to be finally quantified by Bradford method, frozen and stored at -80°C.

Between each purification step a 15% SDS-PAGE gel electrophoresis (done in the laboratory - Annex 1), run at 150V for 1 hour and a western blot was performed in order to monitor the purity of the sample under study and if the renaturation of the proteins had gone as expected.

2.4. Bradford Method

The Bradford method was performed using FLUOstar® OPTIMA (BMG LABTECH) microplate reader using Coomassie Plus™ Protein Assay reagent (Thermo Fisher Scientific) in order to rapidly measure the total concentration of proteins by comparison with a standard protein with a known concentration (Bovine serum albumin (BSA) at 2 mg/ml, CALBIOCHEM®). Initially dilutions were prepared with the BSA, in which the sample buffer under study was used to perform these dilutions, in order to determine the calibration curve. Two concentration ranges were employed: 'high' (up to 1 mg/mL) and 'low' (up to 350 µg/mL). The standard solutions and the test sample were pipetted (5 µL in the case of the 'high' range and 20 µL for the 'low' range) into a Nunc™ MicroWell™ 96-Well microplate (Thermo Fisher Scientific), then the Coomassie Plus™ Protein Assay reagent was pipetted (150 µL for the 'high' range and 200 µL for the 'low' range) and incubated at RT for 30 minutes before reading the absorbance at 595 nm. In all assays triplicates of each solution and sample were performed.

2.5. Western Blotting

Following electrophoresis, the SDS-PAGE proteins were transferred to a Polyvinylidene difluoride (PVDF) membrane using two systems. One of the systems used was the iBlot™ 2 Gel Transfer device (Thermo Fisher Scientific) at 25 V, for 7 minutes, where the gel was placed into an iBlot™ 2 PVDF Regular Stacks transfer system. The other system used was Trans-Blot® SD Cell Semi-Dry Transfer Cell (BIO-RAD) at 90 mA, for 30 minutes, where the transfer system was assembled using an Amersham™ Hybond™ P 0.2 m PVDF western blotting membrane.

After the transfer occurred, the membrane was blocked with 5% (w/v) Skim Milk Powder (Sigma-Aldrich®) in Tris Buffered Saline (made in the lab - Appendix 1) with 0.05% Tween 20 (TBS-T 1×) for 20 minutes at 60 rpm and 37°C. After this time, the membrane was incubated with an antibody mixture consisting of 1.6 µL Monoclonal Anti-polyHistidine antibody produced in mouse (Sigma-Aldrich®) and 2.5 µL Anti-Mouse IgG-Alkaline Phosphatase produced in goat (Sigma-Aldrich®) in 10 mL of TBS-T containing 5% (w/v) Skim Milk Powder, for 1 hour at 60 rpm and 37°C. The membrane was then washed three times for 7 minutes with TBS-T. To reveal the result, 1 mL of a ready-to-use 1-StepTM NBT/BCIP solution (Thermo Fisher Scientific) was used and added to the membrane until the bands were visible to the naked eye. Finally, to stop the reaction, the membrane was washed with water.

2.6. Far-UV Circular Dichroism Spectropolarimetry

Circular dichroism (CD) is an excellent spectroscopy to quickly evaluate the secondary structure of proteins and their folding[130]. This method can be used to follow the unfolding and folding of proteins as a function of temperature. This technique is routinely used to analyse if an expressed and purified protein is folded[131].

An advantage of this technique in the study of proteins is that certain structural information can be obtained from various regions of the spectrum[130].

These types of spectra will give us the composition of the secondary structure, such as α-helix, β-sheet, turn, etc., of the peptide bond region. Different types of secondary structure have different spectral signatures in the far-UV region, i.e., between 180 and 240 nm. Thus, it is possible to obtain characteristic CD spectra in the ultraviolet (Figure 2.5)[132].

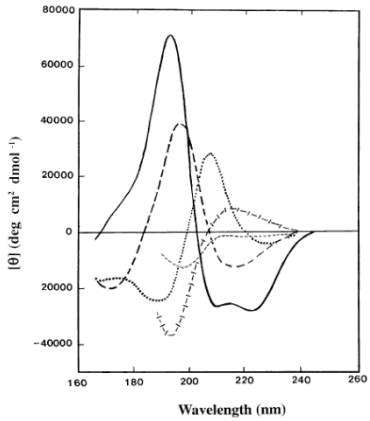

Figure 2.5 - CD spectrum associated with various types of secondary structure. Solid line – α-helix, long dashed line - anti-parallel β-sheet, dotted line - type I β-turn, crossed dashed line - extended 3-helix and short dashed line - irregular structure[132].

It is difficult to make measurements below 180 nm in conventional CD instruments because the intensity of the radiation drops in this region, but also because the nitrogen (N_2) used to purge the optics and the sample site, and the solvent H_2O absorb very significantly[132].

Assays were performed for all three HTHs. For the HTH1 and HTH3 samples obtained from the soluble fraction, dilutions were performed with the buffer 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol and 100 µM TCEP to obtain a final concentration of 0.15 mg/mL. For the samples obtained from the insoluble fraction, a dilution with 10 mM sodium phosphate buffer pH 6.5, 200 mM NaCl, 10% (v/v) glycerol and 500 µM TCEP was performed in order to obtain a final concentration of 0.15 mg/mL for HTH1. While the HTH2 assay was performed with a sample that was at 0.16 mg/mL.

CD spectra were obtained between 200-260 nm on a Jasco J-815 CD Spectropolarimeter with the following acquisition parameters: bandwidth 2.0 nm, data pitch 0.1 nm, scanning speed 50 $nm.min^{-1}$, data integration time 8 sec.

These spectra were acquired at a temperature of 20 °C, followed by a thermal denaturation, with the Jasco CDF-426S/15 Peltier device, increasing 1 °C/min until 80 °C and then a re-cooling step to 20 °C. And again, a new spectrum was obtained after the re-cooling.

2.7. Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a powerful technique used to non-invasively monitor non-covalent molecular interactions in real time. As it is a label-free technique, it does not require the use of tags, dyes or specialised reagents to obtain results[133].

This technique has become important in several areas, such as in biochemistry, biology, and medical sciences due to its characteristics, such as its ability to characterize and quantify low molecular weight molecules[134]. But also, for addressing issues such as the specificity of an interaction, affinity constants, kinetics, equilibrium constants and concentration of certain molecules present in a sample[133,135,136].

SPR is based on a phenomenon where electrons in the surface layer of the metal are excited by incident light photons with a specific angle of incidence, which then propagate parallel to the metal surface[137]. A certain angle that triggers SPR depends on the refractive index of the material near the

metal surface. A small change in this reflective index in the detection medium will prevent SPR from occurring, which makes detection possible[137].

Regarding the SPR instrument (Figure 2.6), there are three main components. One major component is the optical light source which is usually near-infrared high-efficiency light-emitting diode (LED). The second component is a sensor chip with a thin gold layer (~ 40-50 nm thick) attached to a glass layer, this layer is coupled to a prism and on the other side of the sensor is the solution side where the sample flow and interaction occur. The third component is a detection system[133].

There are several sensor chips available for different types of biological application. The most common is carboxymethylated dextran attached to different functional groups in a way that makes it suitable for immobilization of any type of ligand. Currently, CM5 (carboxymethylated dextran), J1 (unmodified gold surface), SA (streptavidin), NTA (nickel chelation), F1 (short dextran), HPA (hydrophobic mono layer), B1 (low charge carboxymethylated dextran), L1 (lipophilic dextran) and C1 (flat carboxymethylated) are available. For the experiments performed in this work, the CM5 chip that is suitable for routine analysis was used[135].

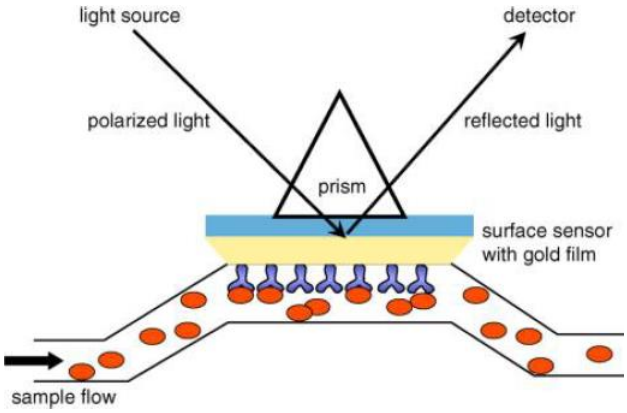

Figure 2.6 – The most common illustration of the setup of an SPR experiment, called the Kretschman configuration[133].

One of the interacting partners has to be immobilised on the surface of the gold layer and is called the "ligand". The second interacting partner is called the "analyte" and flows over the surface of the chip. The analyte is introduced by injection into a continuous flow of a running buffer and this flow will drive the molecules of the analyte close to the gold surface of the sensor and the ligand molecules. When non-covalent interactions begin to occur between the analyte and the immobilised ligand molecule, changes in molecular weight will lead to changes in signal[133].

A typical diagram (Figure 2.7) of an SPR experiment, which is called a sensorgram or a binding progress curve shows the detector response on the y-axis that corresponds to the change in SPR angle expressed in resonance units. Time is represented in seconds on the x-axis[133].

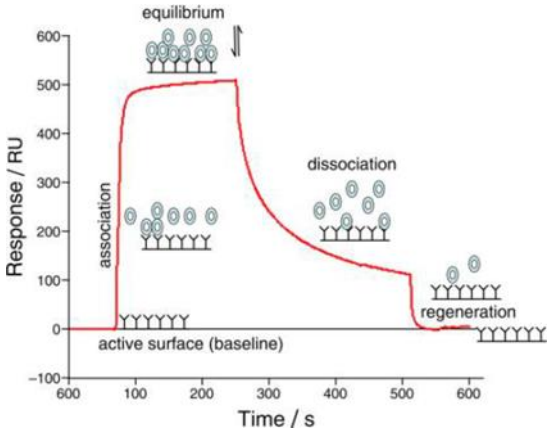

Figure 2.7 – A generic sensorgram with the main phases of an SPR experiment. It includes the establishment of the baseline, the injection of the sample (represented by the circles) into the running buffer, the formation phase of non-covalent complexes between the ligand (represented by the letter Y) and the sample molecules, the equilibrium phase, the dissociation phase and finally regeneration. The latter is defined as the process where the sample molecules that are still bound to the sensor chip through changes in the pH of the running buffer[133].

The rates of change of an SPR signal can be analysed in order to obtain association and dissociation constants of the reaction[135].

All the SPR assays were performed, in collaboration, by Dr. Pedro M. F. Sousa from Merck Healthcare KGaA Satellite Laboratory at iBET, but regarding the methodology used, the Biacore 4000 instrument (GE Healthcare) was used. The assays were performed at 25 °C. RBD was diluted to 2 $\mu g.mL^{-1}$ in 10 mM Bis-Tris buffer pH 6.5 and immobilized on the sensor chip with a carboxymethylated dextran matrix (CM5) using the standard amine coupling procedure. HBS-N buffer, composed of 10 mM HEPES pH 7.4 and 150 mM NaCl, was used as background buffer. Prior to immobilization, the carboxymethylated chip surface was activated with a 1:1 ratio of 400 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and 100 mM N-hydroxysuccinimide for 10 min. RBD was coupled to the surface with an injection time of 11 and 16 min at a flow rate of 10 µL/min in order to reach 50 to 200 RU. The remaining activated carboxymethylated groups were blocked with a 7 min injection of 1 M ethanolamine at pH 8.5.

HTH1, HTH3 and ACE2 were directly diluted in running buffer (10 mM Tris-HCl, 300 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.05% (v/v) Tween-20, 1 mM DTT, pH 7.5) and injected at ten different concentrations using double dilution series, from 0.019 to 10 µM for the HTHs and from 0.98 nM to 0.5 µM for ACE2.

All sensorgrams were treated by subtracting the binding response recorded across the control surface (reference point), followed by subtracting an average of the blank buffer injections from the reaction point. Interactions were evaluated from plots obtained from the steady-state SPR response levels against HTH (or ACE2) concentration using the supplied Biacore 4000 evaluation software.

2.8. Differential Scanning Fluorimetry (nanoDSF)

NanoDSF is an advanced differential scanning fluorimetry methodology that will detect small changes in the intrinsic tryptophan, when folding or unfolding as a function of temperature or other solution conditions. By tracking changes in fluorescence, the chemical and thermal stability of proteins can be assessed. These analyses can be performed with a much wider range of concentrations, consumes considerably fewer materials and provides significantly higher throughput[138–140].

There are some advantages in using this technology for thermal stability analysis such as the analysis of samples with concentrations between 0.05 and 0.2 mg/mL, considerably less volume is used (10 µL/per sample), and finally it is capable to run 48 samples in about 1 hour[138].

NanoDSF assays were performed in a Prometheus NT.48 instrument (Nano-Temper Technologies GmbH), with the HTH2 sample from the purification of the soluble fraction. This sample was centrifuged for 10 minutes prior to assay preparation. The final reaction mixtures were each at a protein concentration of 0.05 mg/mL in different types of buffers. Twenty two buffer compositions were tested, with the pH ranging between 6.5 and 8.5 and with the absence and presence of 0.5M salt (NaCl) (Annex 2).

Then, high sensitivity capillaries (NanoTemper Technologies) were filled with 10 µl of sample and placed in the sample holder. A temperature gradient of 1 °C/min was applied between the temperatures of 20 °C to 90 °C and the fluorescence emission at 330 and 350 nm (excitation at 275 nm) was recorded. The data were analysed using the 350/330 ratio values and the corresponding first derivatives.

2.9. Neutralization assays

Pseudoviruses are very useful in neutralization assays because of their safety and versatility, making them safe and effective alternatives to the use of wild-type virus[141,142].

Pseudoviruses are recombinant viral particles in which their core/backbone and surface proteins are derived from different viruses. Usually, the genes of this virus are modified or even altered in a way that suppresses the expression of the native surface proteins. Therefore, an additional plasmid must be used for the expression of the alternative surface proteins, so that this pseudovirus can infect susceptible host cells but can only replicate for one round in these cells[143,144].

Studies conducted with the active SARS-CoV-2 virus must be carried out in biosafety level 3 laboratories. But because certain genetic sequences have been removed from the virulent virus, the pseudovirus can be handled in biosafety level 2 laboratories[143,144].

As the structure of the surface proteins of the pseudovirus have a high similarity to those of the native virus, and this type of protein is very important for viral entry into host cells, studies on viral entry and assessments of potential neutralization are thus feasible[143,144].

In this type of system there are some limitations. Although the surface protein largely mediates viral entry in a manner similar to that of the native virus, these viruses can only replicate for 1 round and do not always induce pathogenesis as the native virus do. Furthermore, the conformation and distribution pattern of the protein in the pseudovirus may not reflect the reality with respect to the native virus protein[143,144].

Neutralization assays were performed in collaboration with the cell biology laboratory of viral infection at the Instituto Gulbenkian da Ciência by Dr. Marta Alenquer.

These were performed for all three HTH. The samples used were HTH1 at 0.95 mg/mL and HTH2 at 1.40 mg/mL in buffer 10 mM sodium phosphate pH 6.5, 200 mM NaCl, 10% (v/v) glycerol and 500 µM TCEP, and HTH3 at 2.12 mg/mL in buffer 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol and 100 µM TCEP.

On the first day of the neutralization assay protocol 293T-ACE2 seeding was performed. This consisted of treating the 96-well plates (black walls, transparent bottom, corning #3904) with poly-d-lysine three times. This procedure began by diluting 400 µL stock in 20 mL PBS to have a final poly-lysine concentration of 0.1 mg/mL, then adding 50 µL/well of a 96-well plate and incubating for 5 minutes at room temperature. After this time, wash 1× with PBS. Finally, add 1×10^4 cells/well.

On the second and third day neutralization and infection was performed. The first step was to prepare peptide dilutions in a 96-well plate (U bottom, not-treated) using complete DMEM. Serial dilutions were made starting with a 500 µg/mL solution (Annex 3). To perform these dilutions, the indicated amount of DMEM was added to the first dilutions (Table 2.4), 90 µL to the remaining dilutions and virus only, and 180 µL to the cell only wells.

Table 2.4 – Values needed to add peptide and cDMEM (µL) for the first dilution to have a concentration of 500 µg/mL.

Peptide	Concentration (mg/mL)	1st dilution		
		Final concentration (µg/mL)	µL DMEM	µL peptide
HTH1	0.95	500	61.6	68.4
HTH2	1.40	500	83.6	46.4
HTH3	2.12	500	99.3	30.7
Buffer HTH1/2	-	-	61.6	68.4
Buffer HTH3	-	-	99.3	30.7

Then, the corresponding amount of peptide is added to the first dilution, pipetting up and down several times in order to homogenize the solution well. Finally, 40 µL of the first dilution is transferred to the second dilution, repeating this step for all dilutions.

For preparation of the pseudovirus suspension, the multiplicity of infection (MOI) was 0.2 per well, assuming the cells doubled, there were 4×10^3 ($0.2\times2\times10^4$) transduction units/25 µL.

Peptide and pseudovirus dilutions were mixed by adding 90 µL of the pseudovirus suspension to each well of the dilution plate and leaving the plate to incubate at 37°C for 1 hour.

After this incubation, 50 µL of the pseudovirus dilutions with the peptide were added to the cells, after aspiration of the supernatant of the cell plate (Cell plates layout - Annex 3). The plate was left at 37°C for 5 to 6h. When this time was up, 100 µL cDMEM was added to each well and incubated again, for 48h.

On the fourth day fluorescence was measured in the GloMax apparatus but first the supernatant was aspirated, and 100 µL PBS was added.

3. Results and Discussion

3.1. Molecular Modelling

3.1.1. Protein Designs

For this study, three protein designs were obtained, in which we are starting with HTH1, which consists of 85 residues of the binding motif of the ACE2 protein (Figure 3.1), more specifically from residue 19 to 103.

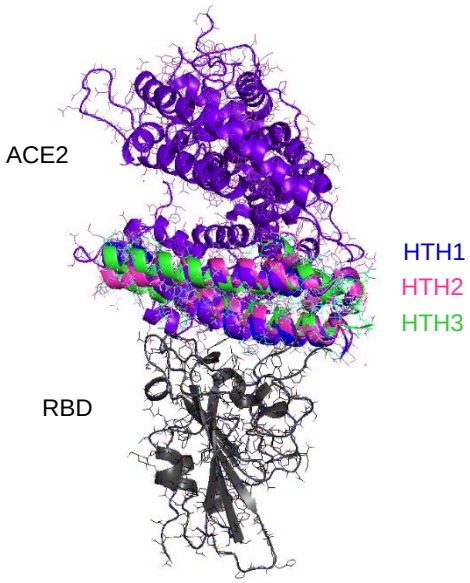

Figure 3.1 – Representation of the overlap site of these three proteins with ACE2 and the interaction site between them and the RBD. The ACE2 protein is shown in purple, the RBD in gray, HTH1 in blue, HTH2 in pink, and HTH3 in green. The figure was built with PyMOL[146] (PDB code: 6M0J) using a cartoon representation for all the proteins.

To obtain the other two designs, amino acid changes were made to try to improve their interaction with the spike protein RBD. HTH2 and HTH3 share 10 substitutions with respect to HTH1 (Figure 3.2). HTH2 has four additional substitutions for cysteine residues, yielding two disulfide bonds in order to improve the stability of the structure (Figure 3.3). While the bond between cysteines 21 and 51 reinforces the structure between the helix 1 and helix 2, the bond between cysteines 11 and 78 reinforces the structure between the first helix and the smaller helix. These two bonds can thus be expected to rigidify the structure by establishing a bond between helix 1 and 2 and then between 1 and 3, thus making it more structurally stable.

```
HTH1      STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQST  60
HTH2      SDIEEQAKTFCDKFNVRAEDCYYQSSLASKNYNTNITEENVQIMNNAGDKCSAFLKEQST  60
HTH3      SDIEEQAKTFLDKFNVRAEDLYYQSSLASKNYNTNITEENVQIMNNAGDKWSAFLKEQST  60
          * ******** **** .*** :******* ************ ******* *********

HTH1      LAQMYPLQEIQNLTVKLQLQALQQN     85
HTH2      EAQKYPLQEIQNDTVKRCLQALQQN     85
HTH3      EAQKYPLQEIQNDTVKRQLQALQQN     85
           ** ******** ***  *******
```

Figure 3.2 – Clustal Omega multiple sequence alignment for the three HTHs. The residues where the substitutions were made do not have the symbol *.

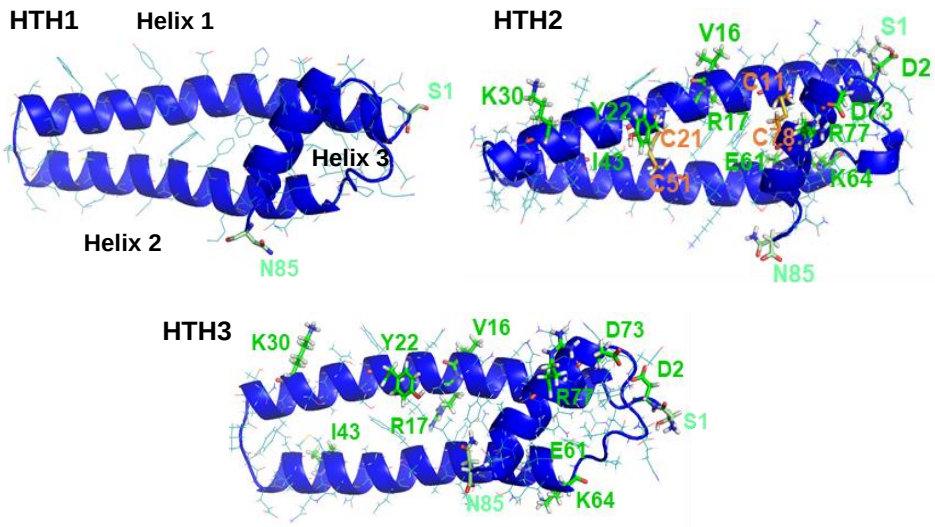

Figure 3.3 – Structural representation of three proteins under study, with the respective N and C-termini, the substitutions and the cysteines of the disulfide bonds marked.

Some of the substitutions were made using the Position Substitution Score Matrix (PSSM), which determines sites of higher probability of certain residue changes occurring that will favor the solubility and structural stability of the proteins. In addition to these substitutions, three human-guided changes were performed in order to further improve these two features of the proteins. These three substitutions, from hydrophobic residues to charged hydrophilic residues, were located at positions 30 (tryptophan to lysine), 73 (leucine to aspartate), and 77 (leucine to arginine). Having a protein with hydrophobic residues on the outside of its structure surrounded by water may destabilize its structure and make it highly insoluble in water.

In order to determine if the designed proteins have a stable fold, the following Protein Folding plots of each HTH were acquired (Figure 3.4).

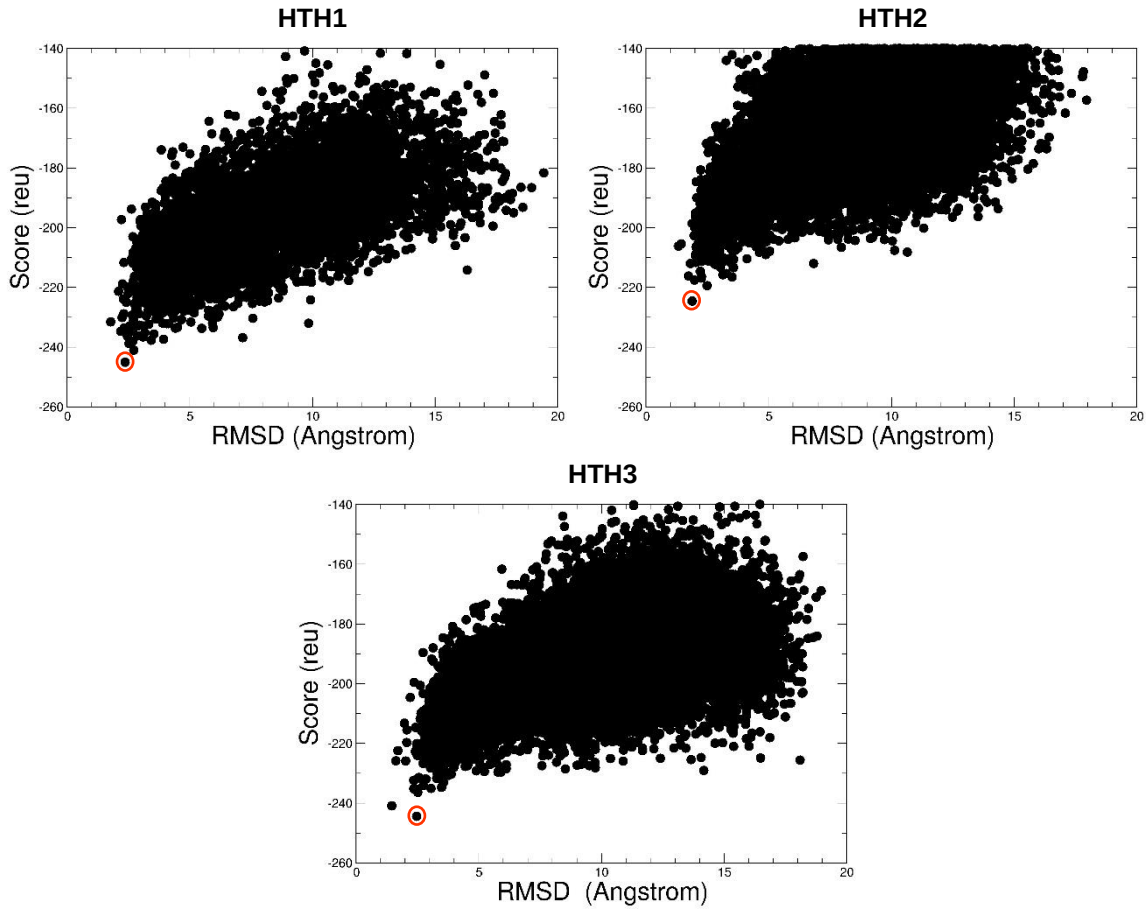

Figure 3.4 – Rosetta ab initio protein folding plots. These graphics are plotted by the energy Score in Rosetta energy units (REU) versus the root mean square deviation (RMSD) in Angström. Each point represents the computer calculation of the Rosetta-predicted protein folding pattern for each of the HTH. When the graph shows a folding funnel, it indicates that a well-defined structure has been obtained.

These plots were acquired through the Rosetta program with the AbinitioRelax tool, where each point is a structure prediction. The lower the value of the root mean square deviation (RMSD) the closer the structure is to the native structure, whereas low energy Score values mean that the conformation is stable. Thus, low values of RMSD and score are desired. The points circled in orange show the best protein folding with a low calculated energy for each of the proteins, while most of the calculations (remaining points) produced predicted structures that deviated from the native structure and have a higher energy. With this, 3D structure of each of the HTHs was predicted, which corresponds to the circled point. Of the three designs, HTH1 and HTH3 are those with the best energy score vs RMSD values.

In addition to calculating the folding stability, we also calculated the binding energy of each HTH to the RBD, using the Rosetta REF15 scoring function. The binding energy values obtained were: -55.52 Rosetta energy units (REU) for HTH1, -52.34 REU for HTH2 and -63.10 REU for HTH3.

3.1.2. Molecular Dynamics Simulations

In this work simulations of these three proteins were performed free in solution (Figure 3.3) but also in complex with the RBD (Figure 3.5), obtaining in total six simulations each with five replicates (Materials and Methods, Table 2.3).

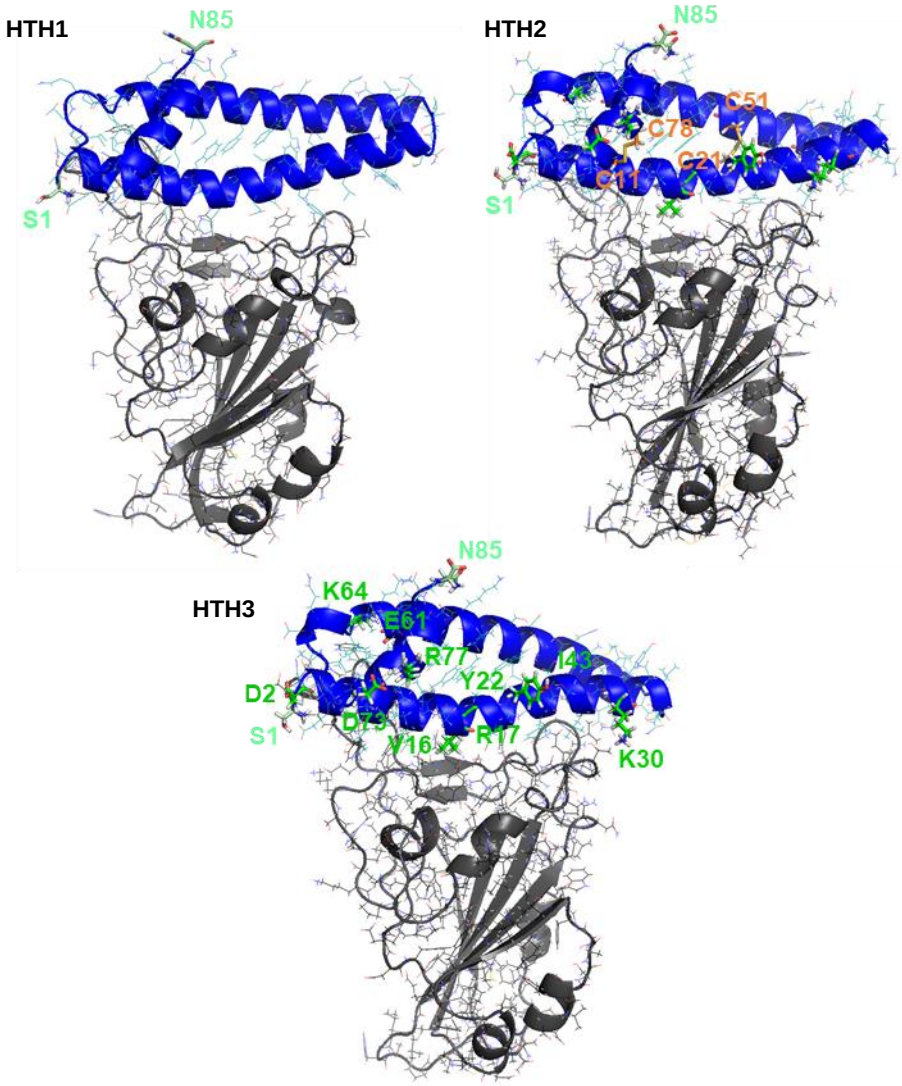

Figure 3.5 – Representation of the secondary structure of the three proteins in complex with the RBD, with the respective N and C-termini, the substitutions and the disulfide bonds highlighted.

Root-Mean-Square Deviation (RMSD)

The RMSD (Root-mean-square deviation of atomic positions) was calculated for the three simulated HTH proteins free in solution and in complex with the RBD. To perform this analysis the GROMACS tool gmx rms was used, through which each structure in the trajectory file is compared to a reference structure. Lower RMSD values and few oscillations in the values will correspond to more stable structures, since they are less distant from the original conformation.

The RMSD was calculated for all performed simulations, both in water (Figure 3.6) and in complex.

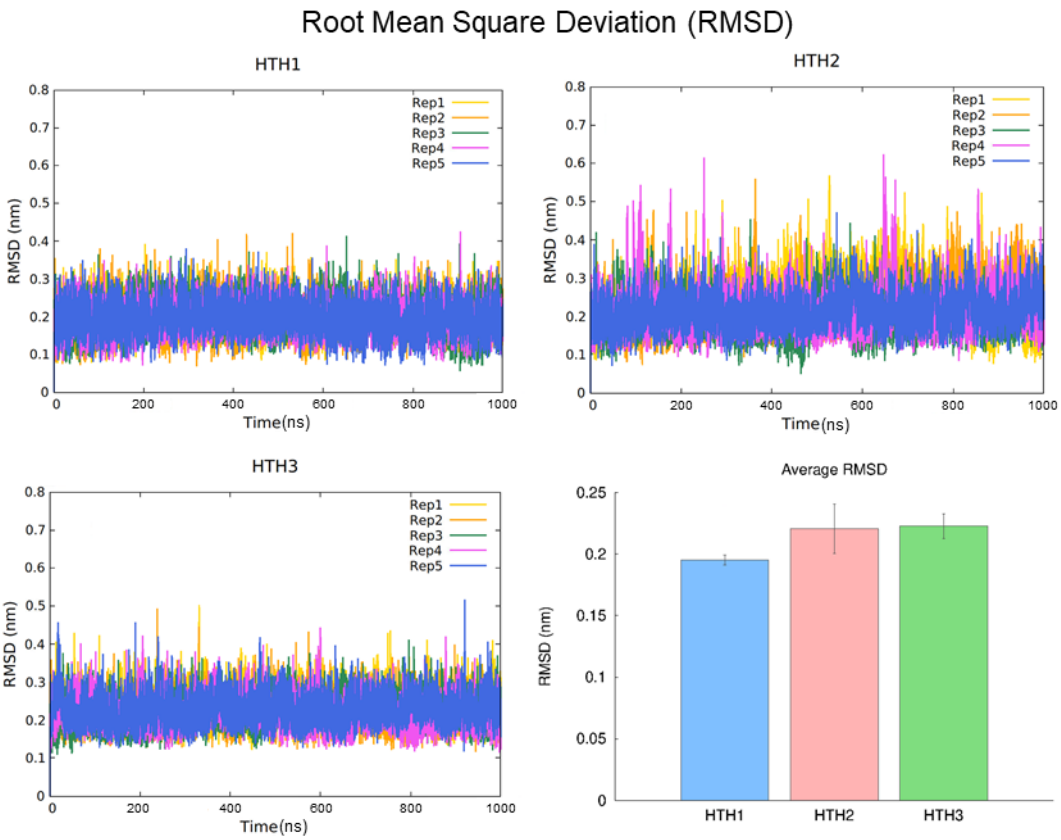

Figure 3.6 – RMSD of the three HTH, simulated in water for 1000 ns. In the plots named HTH1, HTH2 and HTH3 the temporal evolution of the RMSD is shown and each line corresponds to an independent replicate. In the lower right corner the average values obtained from the five replicates for each HTH are represented. Average RMSD values are 0.20 ± 0.01 nm for HTH1, 0.22 ± 0.02 nm for HTH2, and 0.22 ± 0.01 nm for HTH3.

When analyzing the results in Figure 3.6, it can be observed that the three designs present quite stable RMSD values over time. When looking at the average values a slight increase is noted in HTH2 and HTH3 when compared with the HTH1, although when considering the error bars it can be said that the difference is only significant between HTH1 and HTH3. The substitutions made in this two HTH2 and HTH3 caused the RMSD values to increase slightly, probably because as the objective of these changes was to increase binding to ACE2 and not to increase stability, they may have slightly altered the structural stability compared to HTH1.

Regarding the plot of RMSD for HTH2 it can be observed that this design shows larger fluctuations of values compared to the other two. This is somewhat unexpected since this protein has two disulfide bonds. Although these two bonds were designed stabilizing the structure, they may be imposing some strain, thus generating these oscillations and deviations from the initial RMSD values of the simulation. In the HTH3 RMSD plot it is also possible to observe some oscillations of the values over the simulation time. Overall, from these plots it can be conclude that all three designs are structurally quite stable proteins when simulated in water, with HTH2 having larger fluctuations and HTH3 having slightly lower stability.

Regarding the simulations in complex, in order to compare the simulated system, the RMSD was calculated for the complex (Figure 3.7) but also calculated separately for the region comprising the HTH (chain A - Figure 3.8) and the region comprising the RBD (chain B - Figure 3.9).

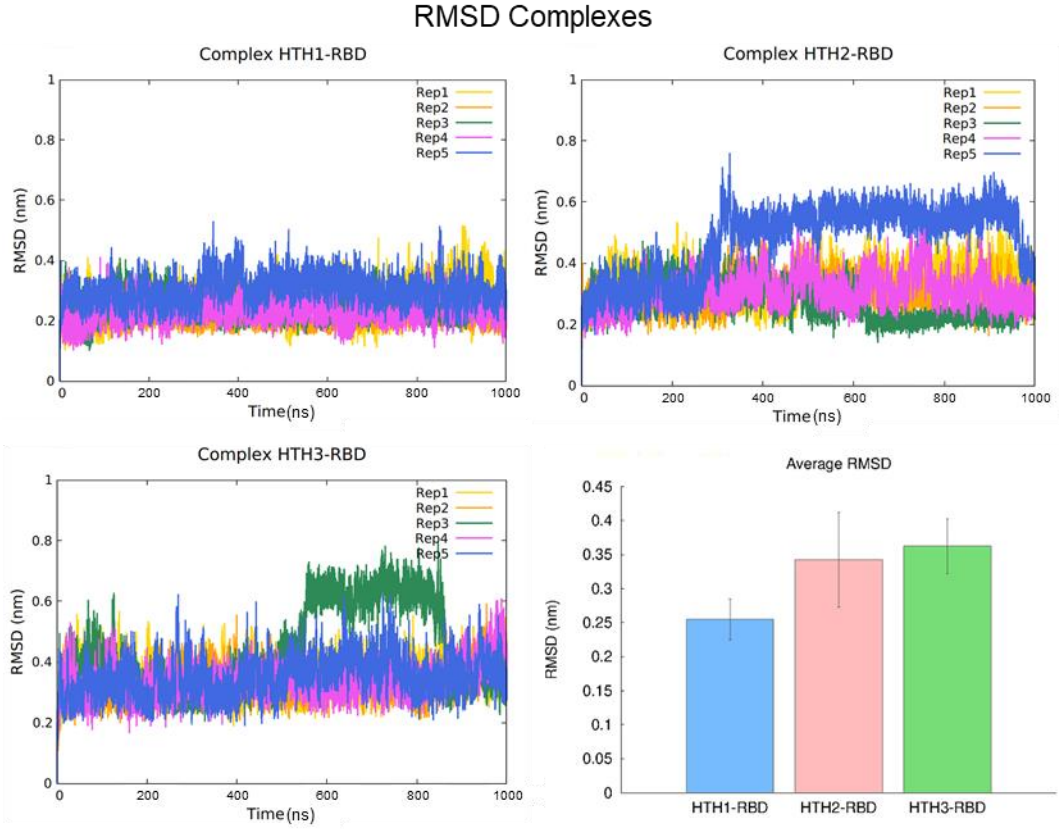

Figure 3.7 – RMSD of the three complexes between the HTHs and RBB, simulated during 1000 ns. In the plots named Complex HTH1-RBD, Complex HTH2-RBD and Complex HTH3-RBD the temporal evolution of the RMSD is shown and each line corresponds to an independent replicate. In the lower right corner the average values obtained from the five replicates for each complex are represented. Average RMSD values are 0.25 ± 0.03 nm for HTH1-RBD, 0.34 ± 0.07 nm for HTH2-RBD and 0.36 ± 0.04 nm for HTH3-RBD.

The results obtained in Figure 3.7 provide information on the conformational stability of both proteins in relation to each other and how much they need to deviate from their initial structure to obtain a more stable conformation. By looking at the plots present in the previous figure, it is possible to see that the complex between HTH1 and RBD is the one that remains more stable over time because there are little fluctuations of the values, indicating that these two proteins show conformational stability in relation to each other. The same cannot be said for the other two complexes, especially the complex between HTH2 and RBD.

In the plot concerning this complex, during the vast majority of the simulation time, mostly in replicate 5, this structure deviates from the initial one. This indicates that both proteins converge to a conformation different from the original one. However, at the end of the simulation there is a decrease in the values, hinting for the need of an extension of this simulation in order to better understand what is happening at longer times. In the other replicates we do not observe such sharp oscillations, but it is still possible to see that the conformation with greater structural stability for this complex is not the same as the initial structure.

In the plot concerning the complex of HTH3 with RBD, it is also possible to observe, especially in replica 3, that there is an accentuated change in the values, with a deviation during 400 ns of the conformation with respect to the initial structure. Although in the remaining replicates this deviation is not observed, there are also several fluctuations of the values and they have larger RMSD than the HTH1-RBD complex. This again indicates that both proteins need to have a different conformation from the initial one for the interaction between them to occur in a more favorable way.

These changes can also be observed in the average values of the three complexes, where there was a significant increase in the values of the complexes with HTH2 or HTH3 in comparison to the one with HTH1.

Overall, it can be concluded that HTH1 appears to possess greater stability when in the presence of RBD, compared to the other two complexes.

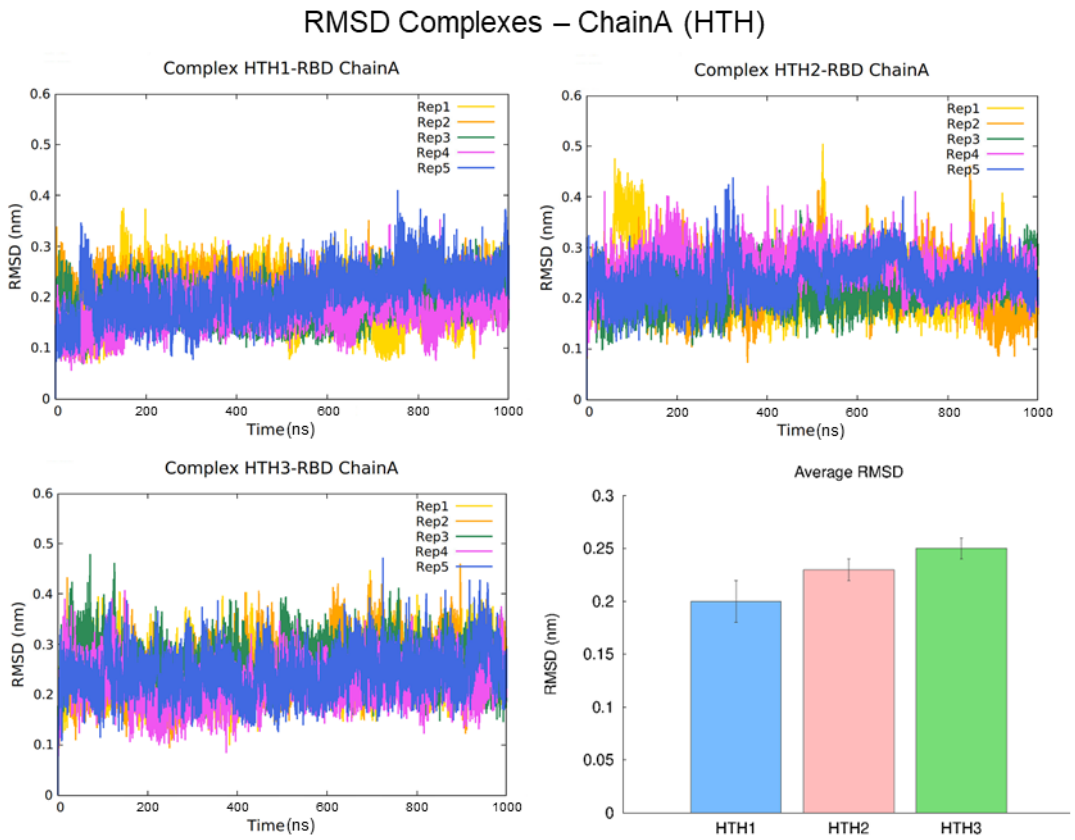

Figure 3.8 – RMSD of the chain A, which corresponds to the HTHs proteins, of the three HTH-RBD complexes, simulated during 1000 ns. In the plots named Complex HTH1-RBD, Complex HTH2-RBD and Complex HTH3-RBD the temporal evolution of the RMSD is shown and each line corresponds to an independent replicate. In these plots the RMSD values correspond only to each HTH protein, but in the context of their complex with RBD. In the lower right corner is represented the average values obtained from the five replicates for each HTH in the complex. Average RMSD values are 0.20 ± 0.02 nm for HTH1, 0.23 ± 0.01 nm for HTH2, and 0.25 ± 0.01 nm for HTH3.

In Figure 3.8, it can be observed that the three HTH proteins within the complex exhibit lower stability than when isolated (Figure 3.6). There are some fluctuations of the values over time indicating that the most favorable conformation for the three proteins is different from the one with the initial structure and that again it would be necessary to extend these simulations in order to try to understand when the three HTH would reach stability with respect to RBD.

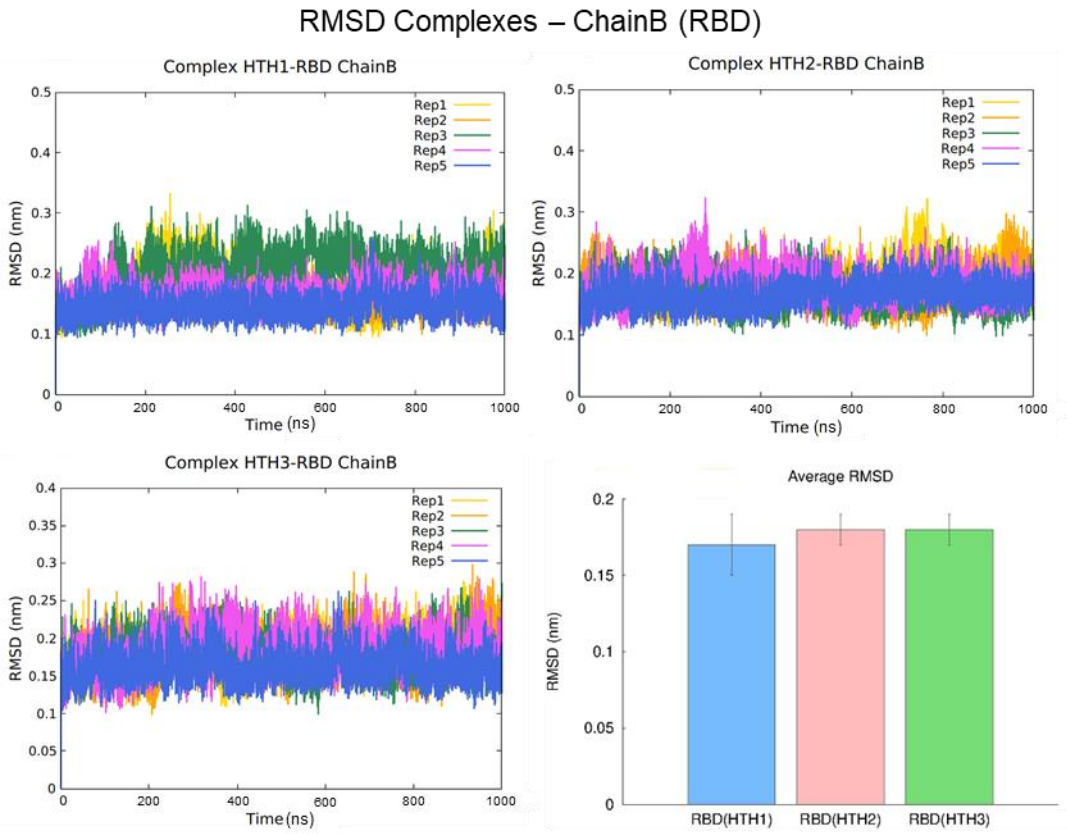

Figure 3.9 – RMSD of the chain B, which corresponds to the RBD, of the three HTH-RBD complexes, simulated for 1000 ns. In the plots named Complex HTH1-RBD, Complex HTH2-RBD and Complex HTH3-RBD the temporal evolution of the RMSD is shown and each line corresponds to an independent replicate. In these the RMSD values correspond only to the RBD, but in the context of their complex with each HTH. In the lower right corner is represented the average values obtained from the five replicates for the RBD in each complex. Average RMSD values are 0.17 ± 0.02 nm for RBD(HTH1), 0.18 ± 0.01 nm for RBD(HTH2) and 0.18 ± 0.01 nm for RBD(HTH3).

The scenario observed for the RBD in the three complexes is different from the one observed in Figure 3.7. Here the RBD is shown to possess greater stability right from the beginning of the simulations, with small fluctuations of the values.

Finally, despite all the oscillations of values observed throughout the previous figures, taking into account that in the three complexes the RMSD values are quite low (value ranges between 0.1 - 0.8 nm), it can be stated that they are structures with a great stability *per se*, but also in the context of the complex. This is curious, since simulations of the isolated RBD performed by our lab[147] show that it is quite dynamic. However, this may be due to the stabilization provided by binding to the HTHs, but longer simulation times would be required to clarify this matter.

Radius of gyration (Rg)

The radius of gyration (Rg) was calculated for the three HTH, free in solution. To perform this analysis, the GROMACS tool gmx gyrate was used, which calculates the radius of gyration of a molecule over time. This parameter is defined as the weighted distance of a set of atoms from their common center of mass. In addition to this, it provides information about the compactness of protein molecules and gives hints on the conformational changes of a protein.

The results obtained for this analysis are shown in Figure 3.10.

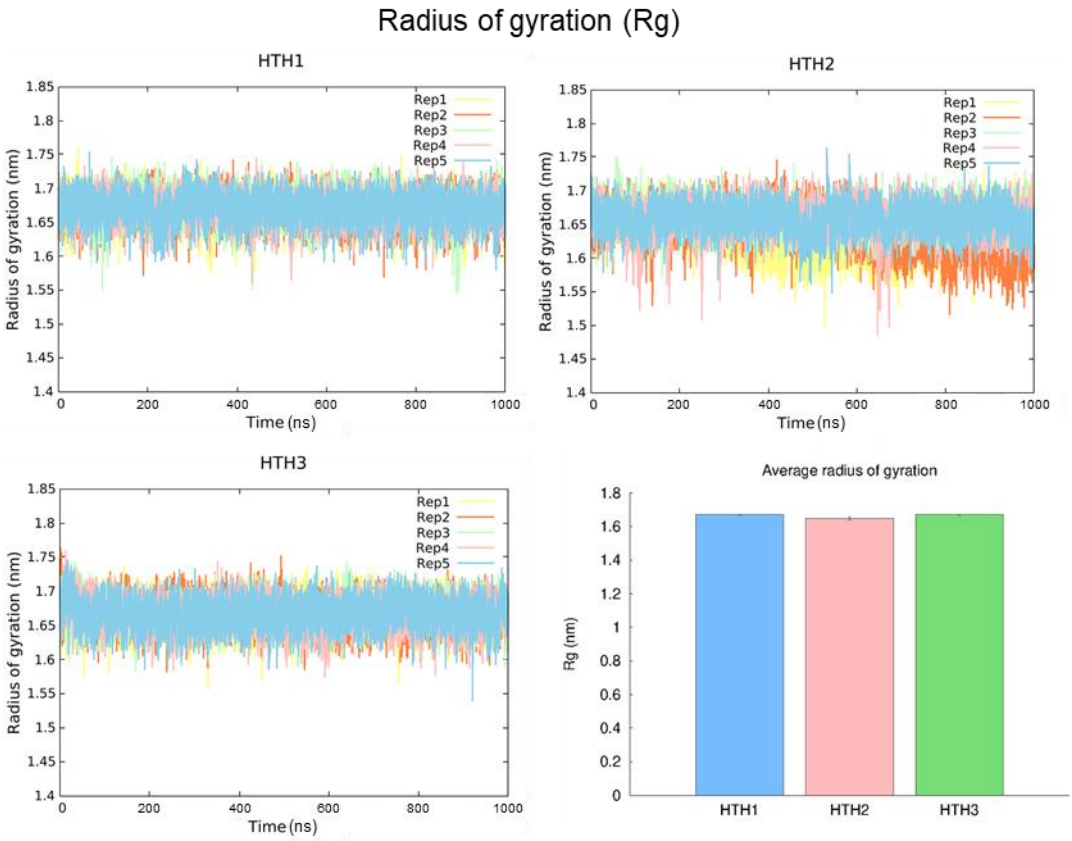

Figure 3.10 – Radius of gyration of the three HTH, simulated in water for 1000 ns. In the plots named HTH1, HTH2 and HTH3 the temporal evolution of the radius of gyration is shown and each line corresponds to an independent replicate. In the lower right corner is represented the average values obtained from the five replicates for each HTH. HTH1 has an average Rg value of 1.67 ± 0.002 nm, HTH2 has an average value of 1.65 ± 0.01 nm, and HTH3 has an average value of 1.67 ± 0.001 nm.

In the Rg plot for the HTH1 it can be observed that the values remain quite stable throughout the simulation time, indicating that there are practically no conformational changes.

The same cannot be said for the HTH2 and HTH3. In the plot of HTH2 there are several fluctuations of Rg values mainly around 100 ns in replicates 4 and 5, from 400 to 700 ns in replicates 1 and 5, in the last 300 ns of the simulation in replicate 2 and also in the last 100 ns of the simulation in replicate 5. These observations allow us to state that there are several conformational changes in this HTH throughout the simulation time in the several replicates performed. In the plot of HTH3 the scenario is different, having observed a slight decrease in Rg values in the first 100 ns, the values keeping quite stable in the remaining simulation time, indicating that a slight conformational change occurred at the beginning.

All these observations correlate with what was concluded from the RMSD results.

Root-Mean-Square Fluctuations (RMSF)

The RMSF (Root-mean-square fluctuations) were calculated for the three free HTH in solution. To perform this analysis the GROMACS tool gmx rmsf was used, which calculates the RMSF values of the atomic positions in the trajectory after fitting to a reference frame. This analysis is a measure of the fluctuations between the position of a certain particle relative to its average position and can reveal which regions of the structure are more mobile, corresponding to an area of the structure with higher RMSF values.

The results obtained for this analysis are shown in Figure 3.11.

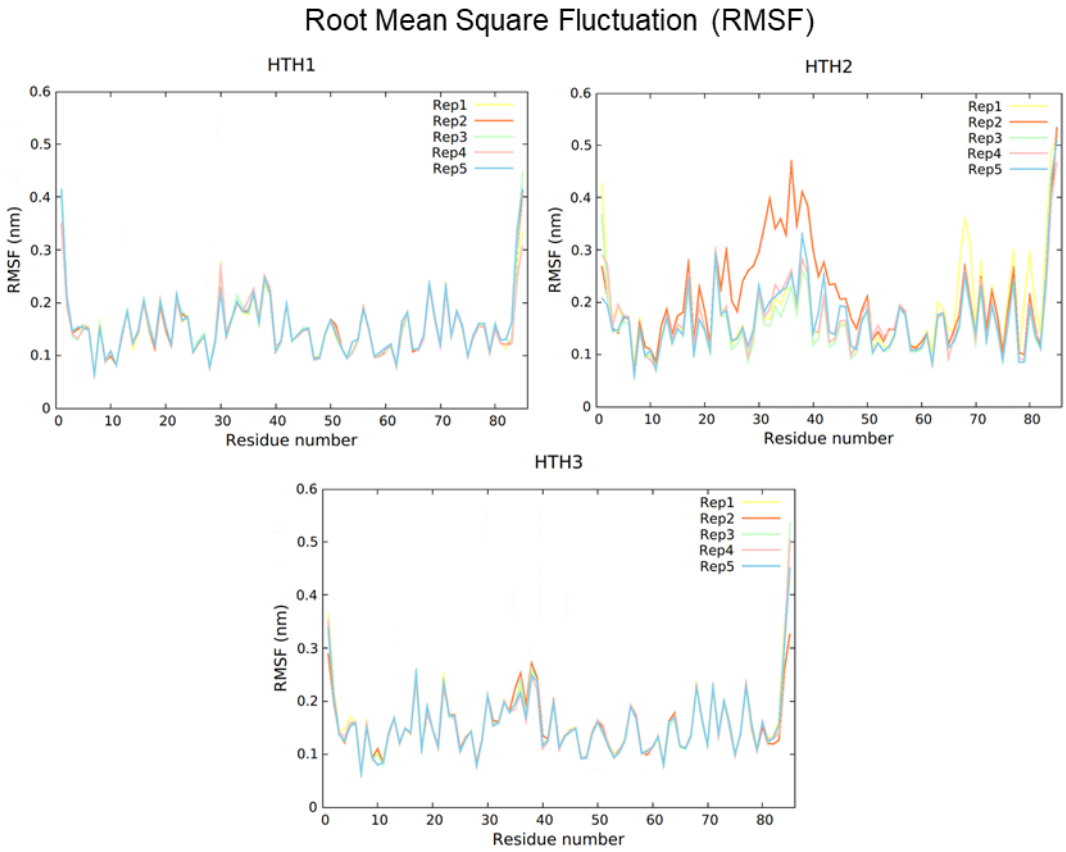

Figure 3.11 – RMSF of the three HTHs, simulated in water for 1000 ns. In the plots named HTH1, HTH2 and HTH3 the temporal evolution of the RMSF is shown and each line corresponds to an independent replicate.

As observed in Figure 3.11, some regions of the HTH structure, especially in HTH2, exhibit increases in RMSF values indicating that they acquire greater mobility.

In the RMSF plot for HTH2 it can be seen that a significant increase in mobility occurred in the residues between positions 15 and 50 and between positions 75 and 80. This may be related to the presence of the two disulfide bonds, which by stabilizing the region and the residues between the regions they are formed, might cause this increase in mobility in the flanking regions. Although counterintuitive this may be due to the fact that the stiffening of the structure in the region of the bonds may cause other regions to become more unstable.

In the RMSF plot for HTH3 it is possible to see that there are two regions of the structure that present some increases in mobility, these being residues 15 to 25 and 70 to 80. In these regions some of the substitutions that were made with respect to HTH1 can be found, and these are the residues that exhibit a higher mobility. Residue 17 changed from an acidic hydrophilic residue (glutamate) to a basic hydrophilic one (arginine). Although both are polar residues, they have opposite charges and the chemical environment around the substitution may be affecting this new amino acid in a different way than the previous one, causing the mobility to increase. At residues 22, 73, and 77, the substitutions were from hydrophobic to hydrophilic residues, and since the simulations were performed in water these residues became more mobile because they face outwards from the protein surface and experience the contact with the water.

In addition to this, the regions of the residues that have the highest RMSF values in the three HTHs are those that are present at the RBD interface interaction.

Secondary structure

To analyze whether the secondary structure of the three HTH was maintained during the 1000 ns of the various simulations performed, the GROMACS tool gmx do_dssp was used which reads the trajectory file and calculates the secondary structure for each time frame.

The secondary structure was determined for the three simulated HTH free in water (Figure 3.12, 3.13 and 3.14), but also for the three HTH in complex with RBD, with only the secondary structure of replicate 1 being presented. The remaining secondary structure analyses are shown in Annex 4. Also shown in these figures are the initial and final structure of the replicate 1 simulation, while the remaining final structures are in Annex 5.

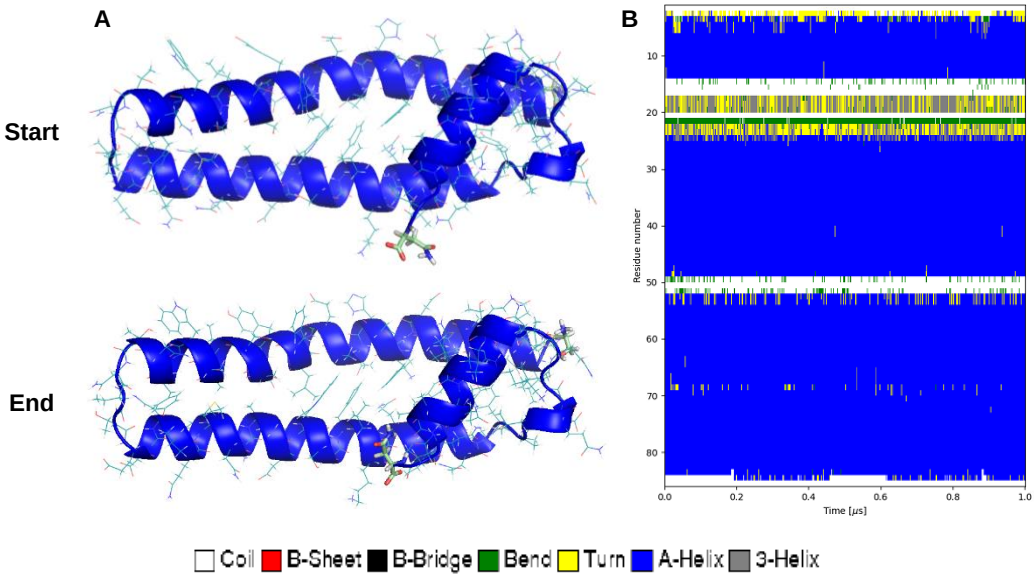

Figure 3.12 – HTH1 secondary structure, free in solution. (A) Structure of the protein at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of the protein during the simulation.

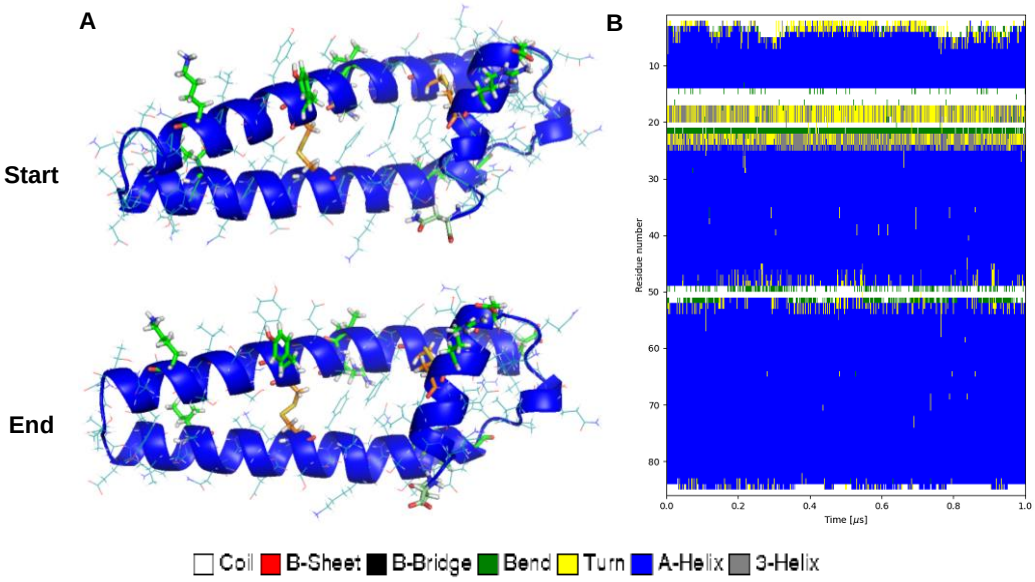

Figure 3.13 – HTH2 secondary structure, free in solution. (A) Structure of the protein at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of the protein during the simulation.

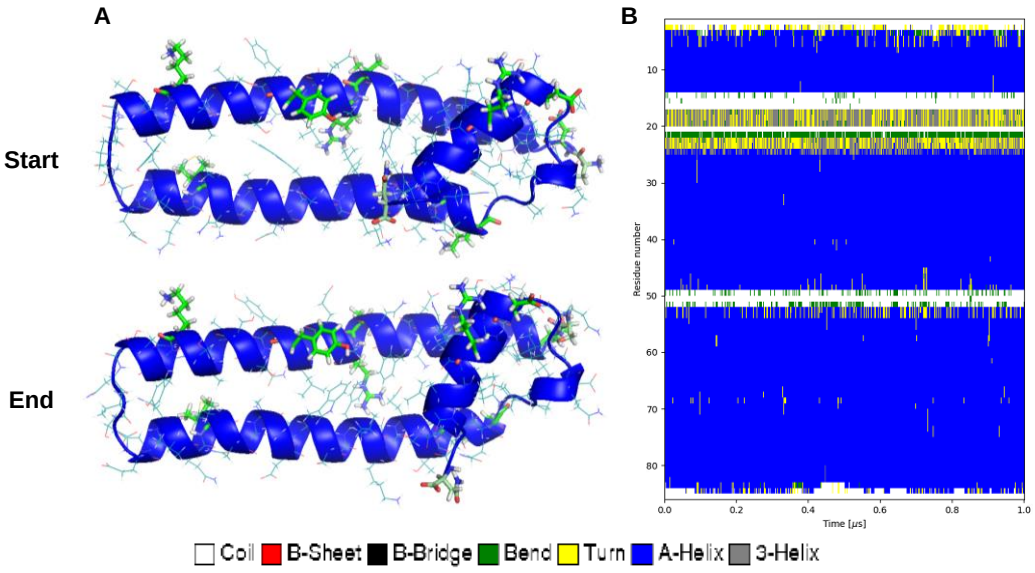

Figure 3.14 – HTH3 secondary structure, free in solution. (A) Structure of the protein at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of the protein during the simulation.

As observed in Figures 3.12, 3.13, 3.14, HTHs alone in solution, maintain their secondary structure over the simulation time, as no structure losses are observed. In addition to this, it is also possible to state that the three designs consist mostly of α-helices, making up around 80% of the structure.

By observing the images of the beginning and end of the simulations of each HTH it is possible to confirm that these proteins present mobility, since in the three designs, in the end of the simulation, helices 1 and 2 are in a more stretched form than in the beginning, losing the curvature that they had, and both helices seem to move away from each other.

Regarding the secondary structure of HTHs in complex, these are shown in the following figures (Figure 3.15, 3.16, 3.17, 3.18).

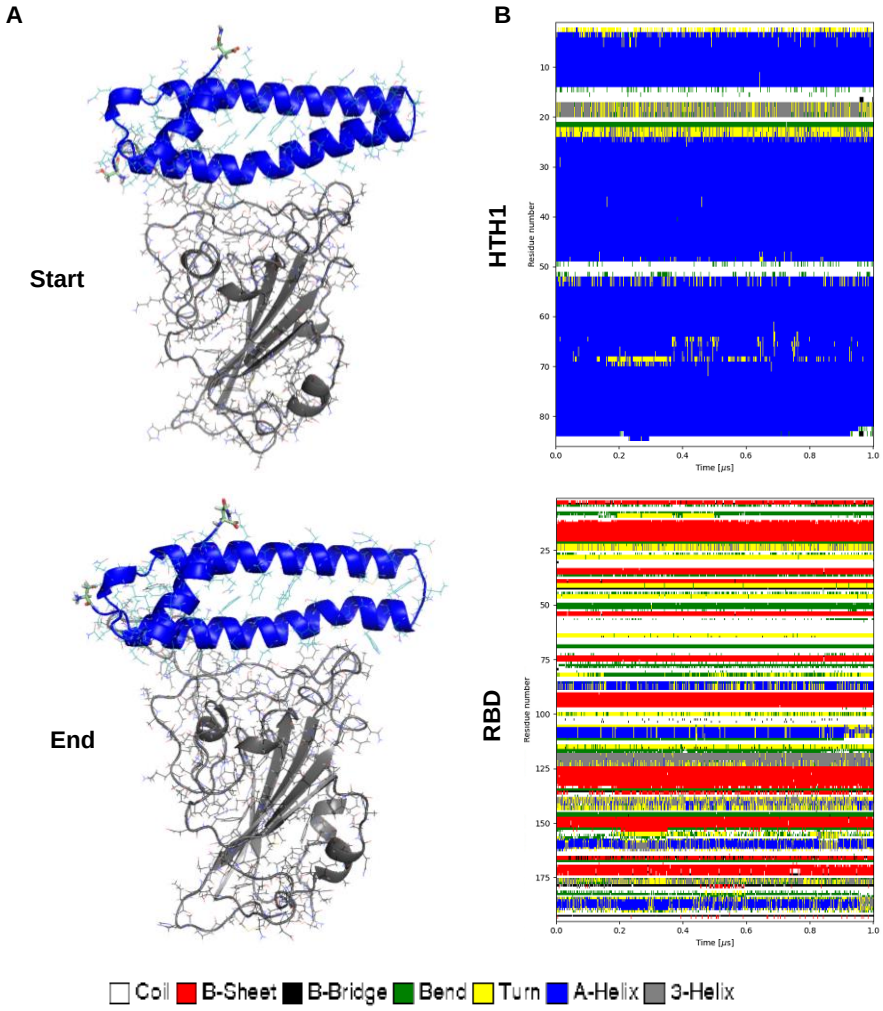

Figure 3.15 – Secondary structure of the Complex HTH1-RBD. (A) Structure of both proteins at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of HTH1 and RBD during the simulation.

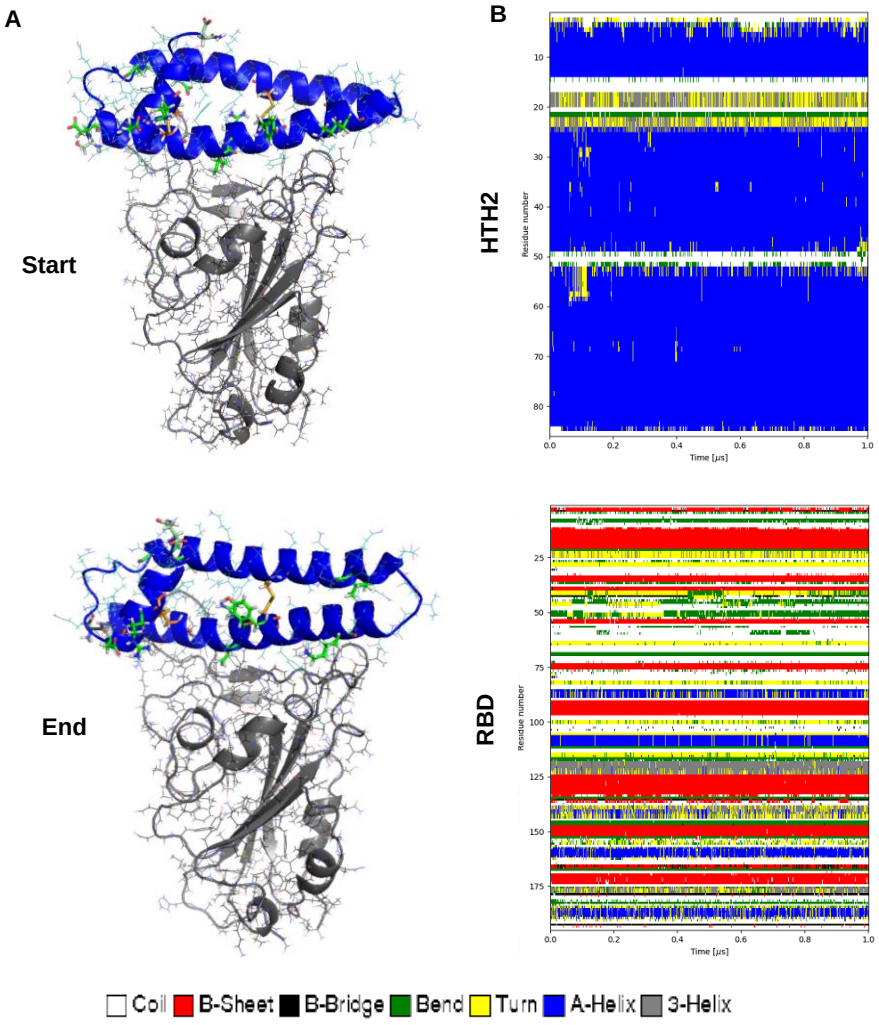

Figure 3.16 – Secondary structure of the Complex HTH2-RBD. (A) Structure of both proteins at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of HTH2 and RBD during the simulation.

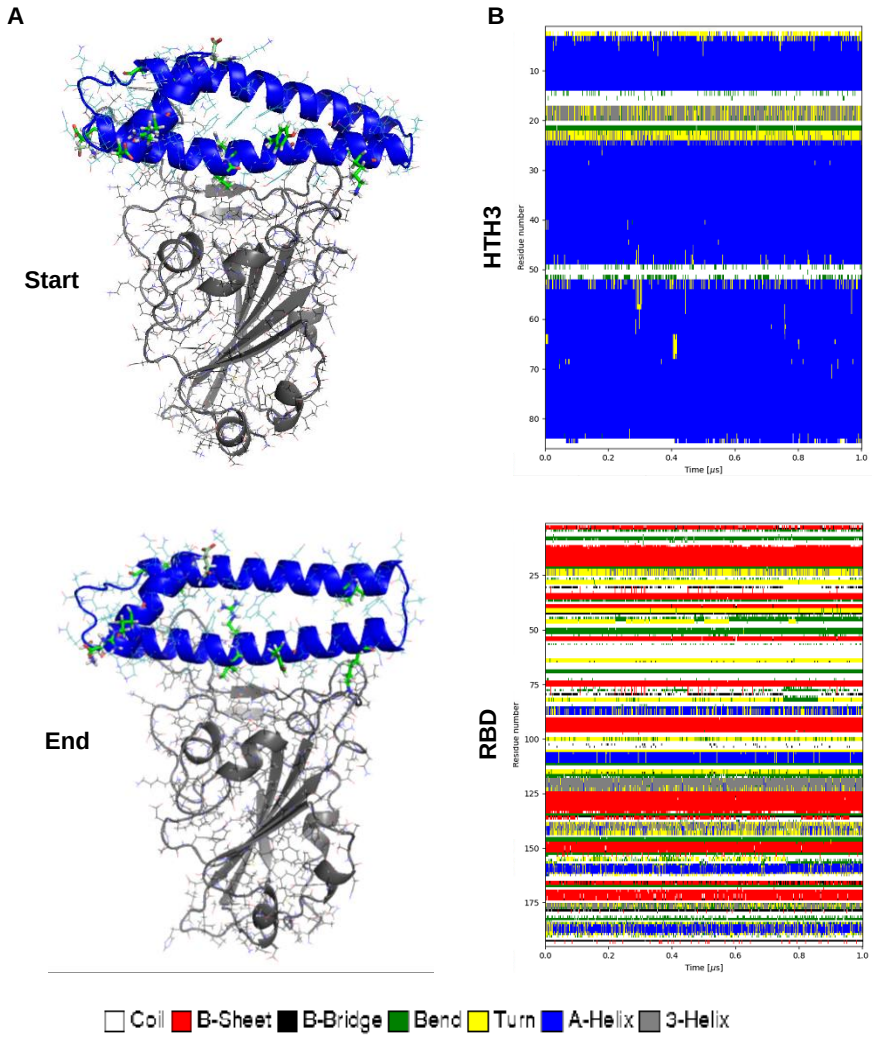

Figure 3.17 – Secondary structure of the Complex HTH3-RBD. (A) Structure of both proteins at the beginning and end of the simulation. (B) Temporal evolution of the secondary structure of HTH3 and RBD during the simulation.

As observed, these proteins maintain their secondary structure over time. However, in the context of the complex there are more turn formations than when these proteins were alone. It is also possible to observe that these losses of α-helical structure to turn occur most markedly in the final regions of the three HTHs (highlighted in Figure 3.18 by the black rectangle), mainly and to a greater extent in HTH2. Thus, there is a small perturbation of the structure when in complex with RBD.

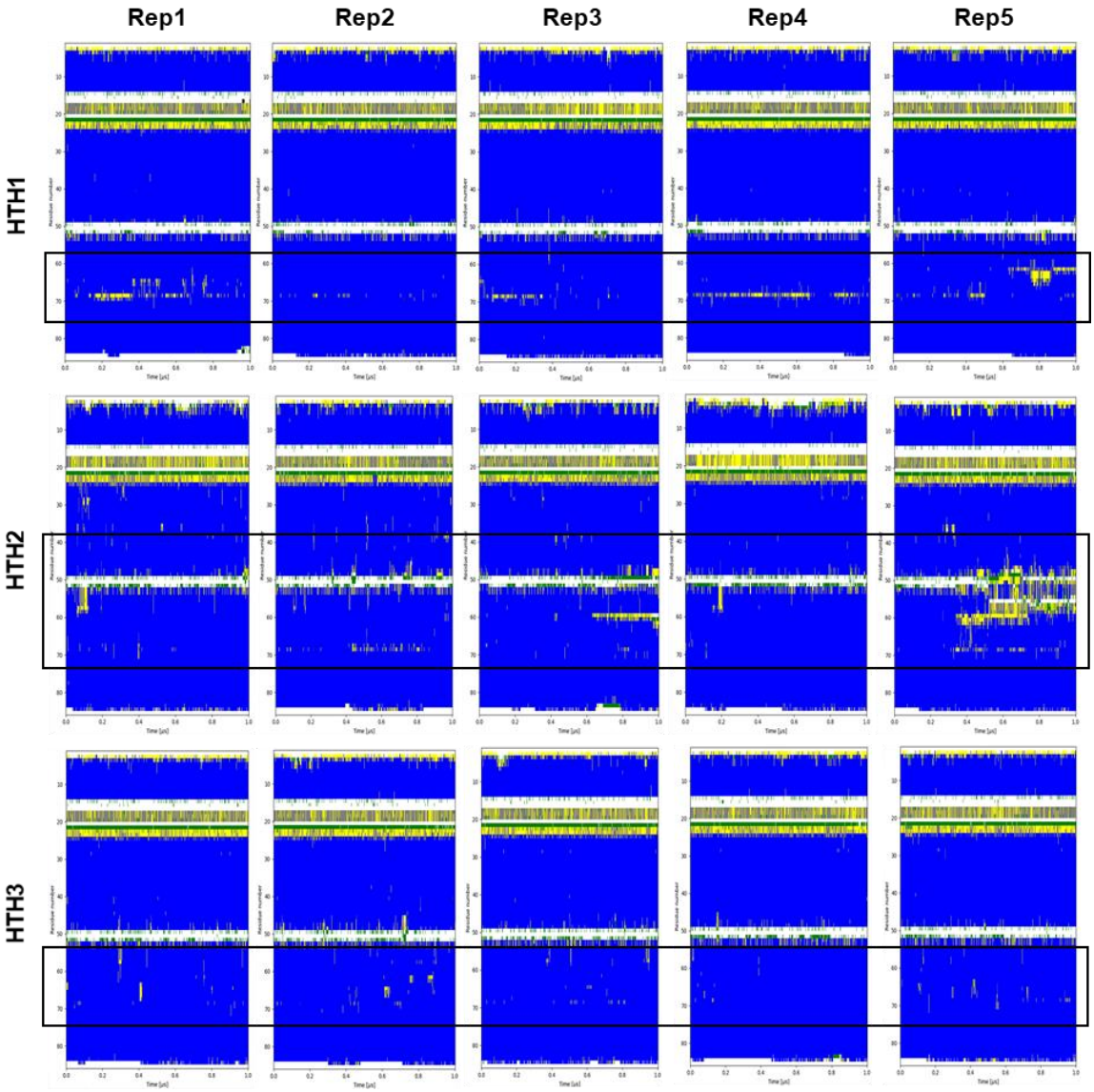

Figure 3.18 – Secondary structure of five replicates of each HTH in complex with the RBD. In each black rectangle are shown the regions, with the most pronounced α-helix to turn structure changes.

Regarding the secondary structure of RBD, it is constituted mostly by β-sheets, and this is the major structure present in these plots. Thus, it is possible to state that RBD also maintains its structure throughout the simulation time.

In addition, it is also possible to observe in these images of the beginning and the end of the simulations the previously referred mobility of the helices relative to one another.

In order to better understand the interaction between HTH and RBD four more analyses were performed, concerning protein-protein interactions (PPIs) and solvent accessible surface area (SASA). PPIs are highly specific physical contacts that are established between two or more proteins and include interactions such as hydrogen bonds and hydrophobic interactions.

Number of contacts between HTHs and RBD

The residue-residue contacts present in protein structures are pairs of spatially close residues. To perform this analysis the GROMACS tool gmx mindist calculates the number of contacts within a given distance, which for this case was a distance smaller than 0.6 nm (6 Å).

The results obtained from this analysis are present in Figure 3.19 where it can be observed that HTH1 has a higher number of contacts with the RBD when compared with HTH2 and HTH3.

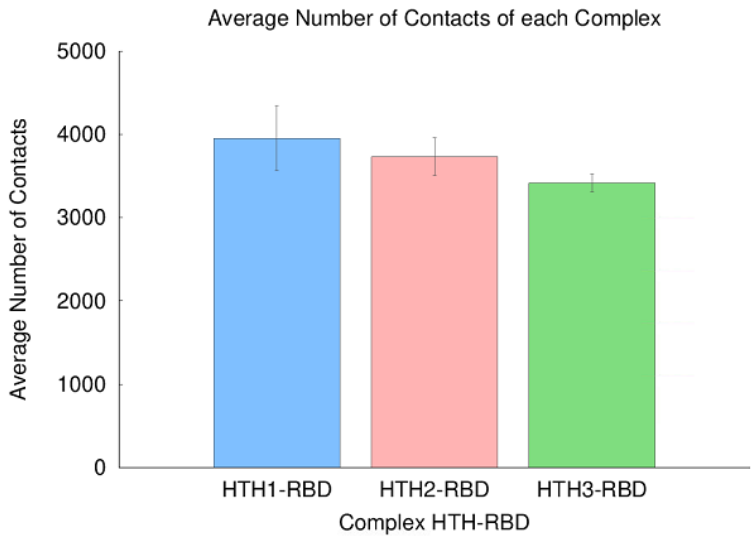

Figure 3.19 – Average number of contacts formed between each of the HTHs and the RBD: 3951 ± 388 for HTH1, 3729 ± 226 for HTH2 and 34134 ± 107 for HTH3.

However, taking into account the error bars, the number of contacts of HTH2 with RBD is similar to those of HTH1. Of the three, the one with significantly lower number of contacts is HTH3. This may be due to the substitutions that were made in HTH2 and HTH3, despite the fact that they were made to improve the interaction of the designs with the RBD of S protein.

Although both HTH2 and HTH3 have the same substitutions, HTH2 also has the two disulfide bonds, and it may be due to the presence of these two bonds that the number of contacts is higher compared to the HTH3. However, it is not easy to understand how the presence of these SS bonds leads to a larger number of contacts.

To understand which residues from both HTH and RBD most contribute to the interaction between the two proteins, an analysis of the average number of contacts for each of the residues was performed (Figures 3.20 - 3.22).

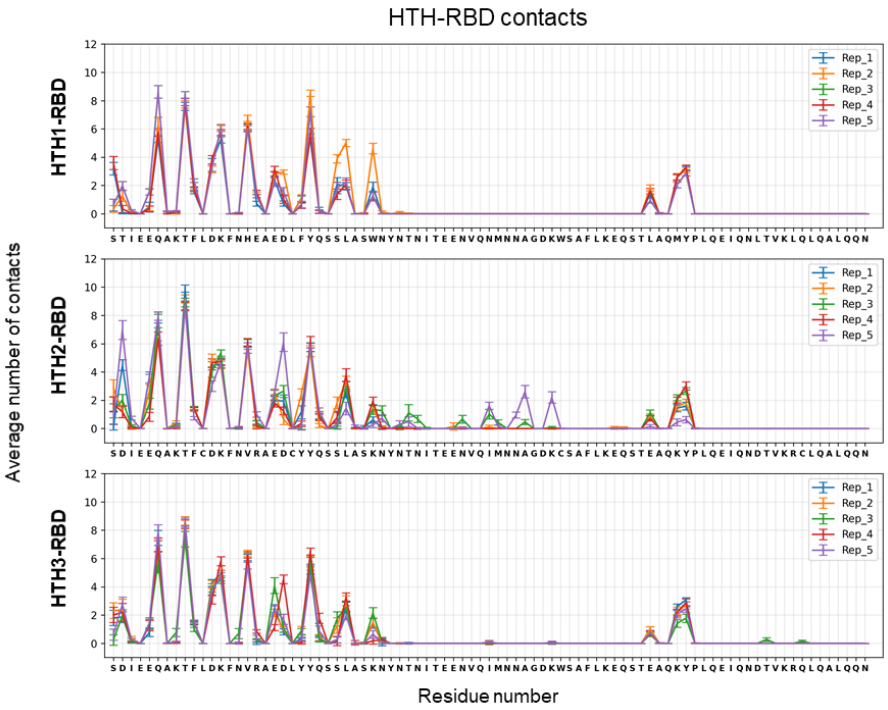

Figure 3.20 – Average number of contacts between HTH and RBD relative to each of the HTH residues, for each of the replicates.

In Figure 3.20, it is possible to estimate the number of residues from each HTH that interact. HTH1 has 29 interacting residues, HTH2 has 40 and HTH3 has 33. Although the number of interacting residues increases in HTH2 and HTH3 in relation to HTH1, it is possible to see that in the first 30 residues the average contact values in HTH2 and HTH3 have dropped when compared to the HTH1.

Furthermore, although both HTH2 and HTH3 show areas in the graph of residues that establish contacts but do not appear in the HTH1 graph, they have such low values that they may not be contributing to a better interaction.

Therefore, the fact that the HTH1 shows a stronger interaction with the RBD can be due to a loss of interaction in the N-terminal region which is not compensated by new interactions that appear but have a low frequency.

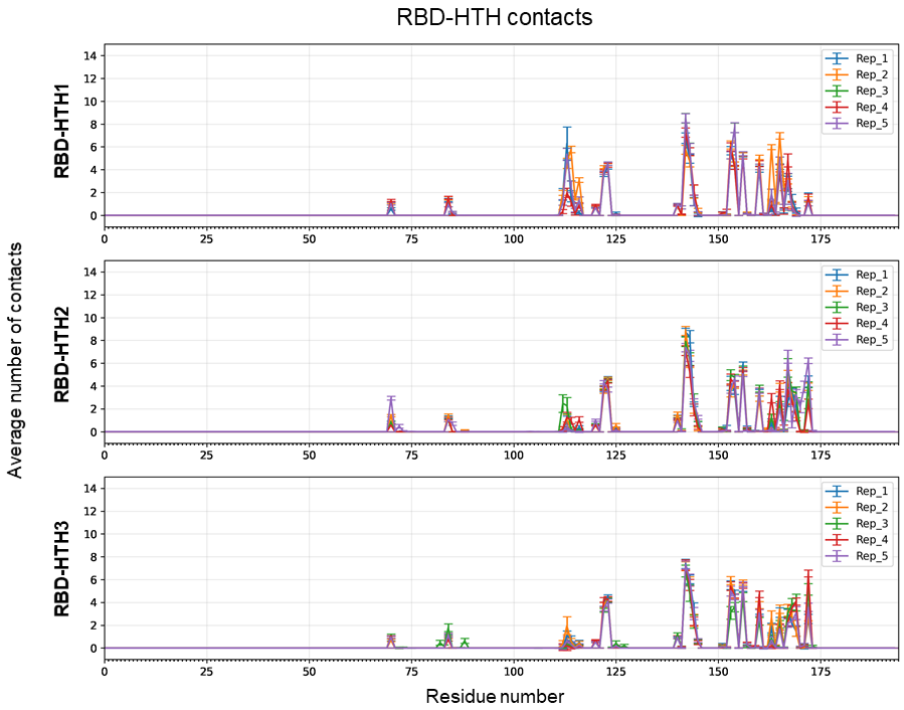

Figure 3.21 – Average number of contacts between HTH and RBD but concerning each of the RBD residues, for each of the replicates.

In Figure 3.21, it is also possible to estimate the number of RBD residues that interact in the presence of each HTH. The RBD has 44, 50 and 54 interacting residues in the presence, respectively of HTH1, HTH2 and HTH3. Although the number of interacting residues has increased in the RBD-HTH2 and RBD-HTH3, the average values of contacts in these two scenarios have decreased, when compared to RBD-HTH1. Again, it may be due to this higher number of contacts that the interaction of HTH1 with RBD appears to be better.

Furthermore, it can also be seen that the regions of RBD that are interacting are identical in the presence of all three HTH.

In Figure 3.22, the interacting residues of each of the proteins are highlighted in order to better observe the areas of the HTH and RBD where these are located.

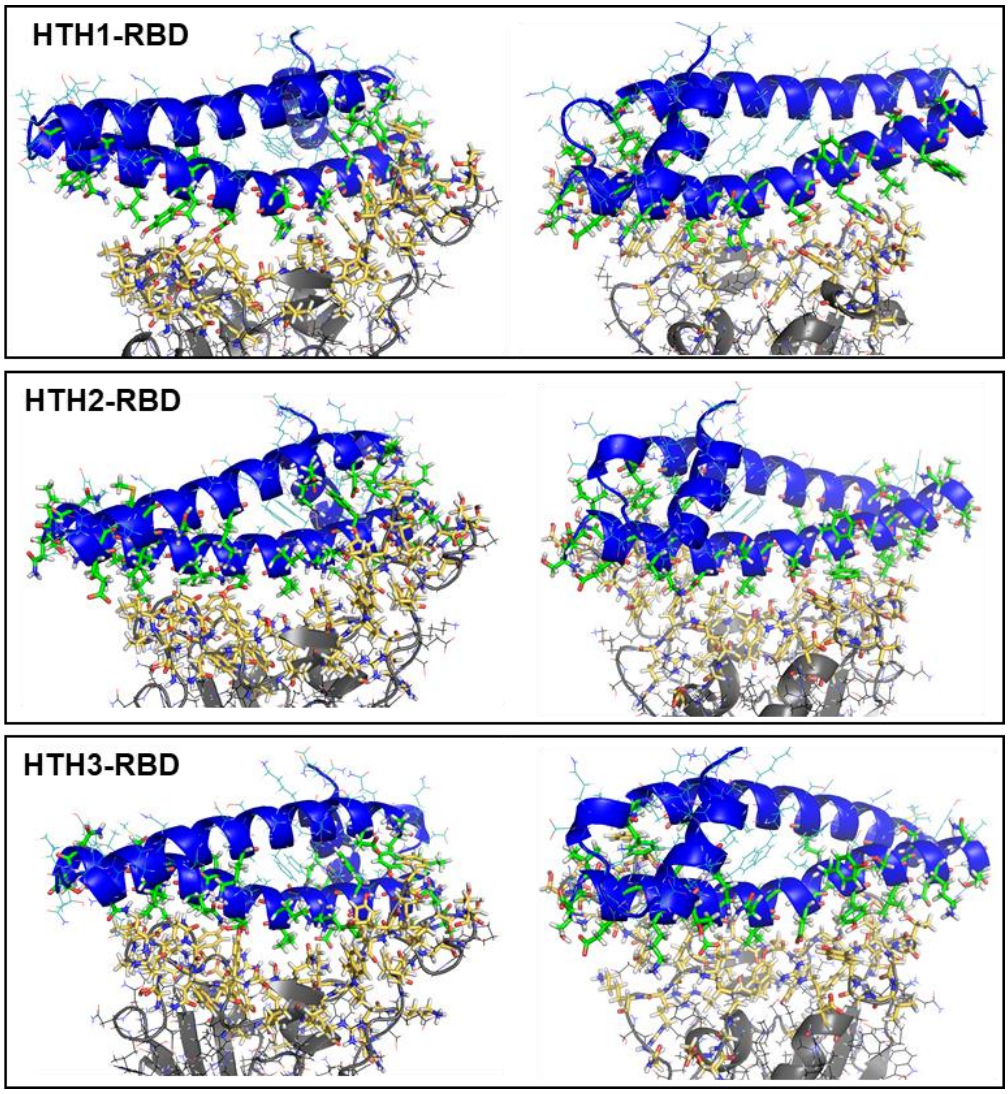

Figure 3.22 – Representation with the residues that are interacting from each protein highlighted. In blue are present the HTH with their residues in green and in gray is the RBD with their residues in yellow. In each rectangle two displays of the same complex are shown, with a 90º rotation, so that a better observation is possible.

Number of hydrophobic contacts between HTH and RBD

The same GROMACS tool was used to perform this analysis as in the previous one, but the residues selected were only the hydrophobic ones. The distance for calculating the hydrophobic contact number was the same as previously used (less than 0.6 nm (6 Å)). Hydrophobic interactions are one of the most important effects for stabilizing the conformation of proteins in aqueous solutions. These interactions describe the tendency of non-polar groups to associate.

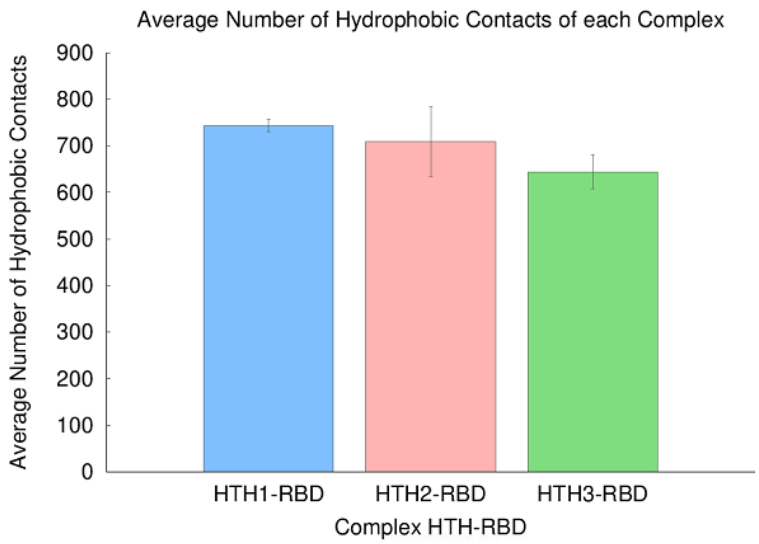

Figure 3.23 – Average number of hydrophobic contacts formed between each of the HTHs and the RBD. Average number of hydrophobic contacts were 743 ± 12 for HTH1, 709 ± 74 for HTH2 and 644 ± 36 for HTH3.

The results obtained from this analysis are shown in Figure 3.23, where it can be observed that once again HTH1 and RBD show a higher average number of hydrophobic contacts between them when compared to HTH2 and HTH3, and this also results from the amino acid substitutions. However, taking into account the error bars, the difference between HTH1-RBD and HTH2-RBD is not significant. Again, of the three, the one with the lowest number of hydrophobic contacts is HTH3.

The substitutions employed to generate HTH2 and HTH3 from HTH1 decreased the number of hydrophobic residues (respectively 32, 25 and 28), which partially explains the decrease in hydrophobic contacts with RBD.

Although HTH2 has a lower value of hydrophobic residues than HTH3, it has a higher number of contacts and again this could be due to the bonds, although it is unclear how this happens.

Number of hydrogen bonds between HTH and RBD

To perform this analysis, the GROMACS tool gmx hbond which analyzes all hydrogen bonds between two groups of atoms was used. A hydrogen bond confers rigidity to protein structures and specificity to intramolecular interactions. It is formed by the interaction between a hydrogen atom that is covalently bonded to an electronegative atom (donor D) with another electronegative atom (acceptor A). For the hydrogen bond geometry to be accepted they must be at a distance ≤ 3.5 Å between hydrogen D and A, which was the distance used in this analysis.

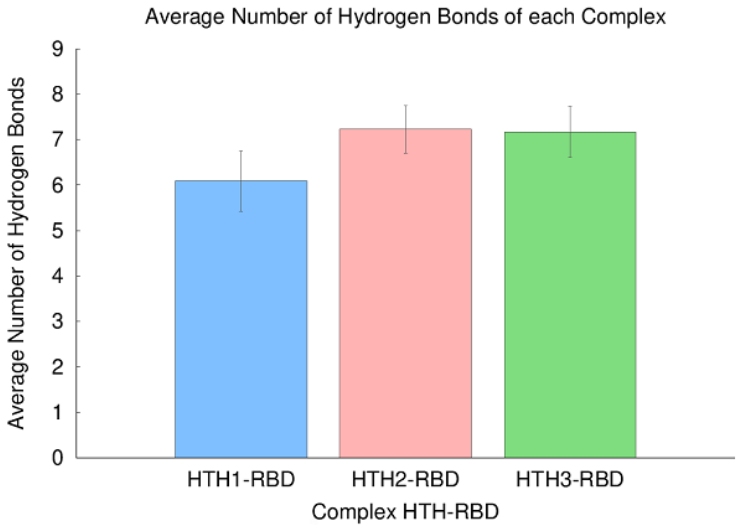

Figure 3.24 – Average number of hydrogen bonds formed between each of the HTH and the RBD. Average value of hydrogen bonds is 6.1 ± 0.6 for HTH1, 7.2 ± 0.5 for HTH2 and 7.2 ± 0.5 for HTH3.

As observed in Figure 3.24, the interaction of HTH1 with RBD is the one with the least average number of hydrogen bonds, while the HTH2 and HTH3 exhibit an equal average number. This may be due to the fact that the substitutions present in both HTH2 and HTH3 have added hydrophilic residues potentiating the formation of a greater amount of hydrogen bond interactions, compared to HTH1 which has a greater number of hydrophobic residues.

The mean value of the HTH2 and HTH3 is the same because they both have the same substitutions in terms of adding hydrophilic residues and removing hydrophobic residues. The disulfide bonds created by the addition of four cysteines neither contribute to nor prevent the formation of these hydrogen bonds.

Solvent-Accessible Surface Area (SASA)

The interface or contact area between HTH and RBD was calculated using the solvent-accessible surface area (SASA).

For this analysis it was necessary to calculate the SASA of HTH, the SASA of RBD and the SASA of HTH complexed to RBD. The sum of the areas obtained for HTH and RBD correspond to the total area of the complex plus two times the interface area between HTH and RBD. To obtain only the interface area it is necessary to exclude the area of the complex and then divide this value by two. The area of each individual residue was calculated in order to obtain the total, the hydrophobic and the hydrophilic interfacial areas.

The average values of the interface areas (nm^2) of each HTH with the RBD are shown in Figure 3.25.

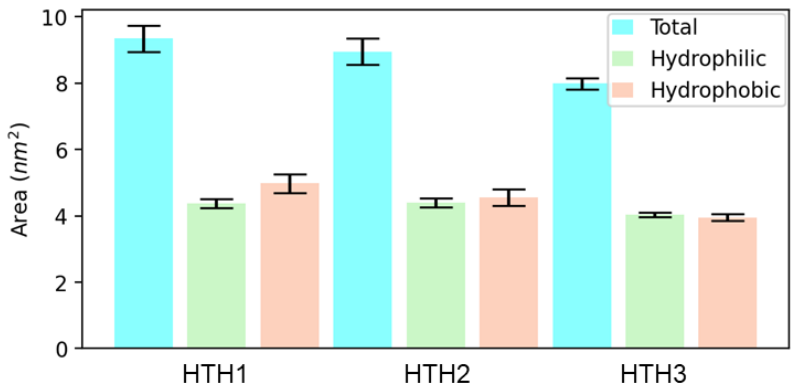

Figure 3.25 – Graphical representation of the average values of the interface areas (nm^2) of each HTH with RBD. Blue bars, total area; green bars, hydrophilic area; pink bars, hydrophobic area. The specific values are: HTH1: Total area - 9.3 ± 0.4 nm^2, Hydrophilic area - 4.4 ± 0.1 nm^2, Hydrophobic area - 4.9 ± 0.3 nm^2; HTH2: Total area - 8.9 ± 0.4 nm^2, Hydrophilic area - 4.4 ± 0.2 nm^2, Hydrophobic area - 4.6 ± 0.3 nm^2; HTH3: Total area - 7.9 ± 0.2 nm^2, Hydrophilic area - 4 ± 0.1 nm^2, Hydrophobic area - 3.9 ± 0.1 nm^2.

This analysis revealed that HTH1-RBD has the highest value of the total interface area, although the difference with the value of HTH2-RBD is not significant when considering the error bars.

Regarding the hydrophobic areas of the interface, HTH1-RBD also has a higher value, something expected considering that the protein has more hydrophobic residues than HTH2 and HTH3. And of the three hydrophobic areas, the HTH3-RBD has the lowest value, which is also expected, taking into account the previous analyses. Through figure 3.23 it was possible to ascertain that of the total residues that interact with RBD, HTH3 is the one that has the lowest value of hydrophobic residues contributing to the interaction.

Regarding the hydrophilic areas, HTH1-RBD and HTH2-RBD have similar values, although the numbers of hydrophilic residues that is contributing to the interaction is lower in HTH1 (20 and 31, respectively). In the case of HTH3, although it has the lowest average hydrophilic area value, it has more hydrophilic residues than HTH1 (27).

Again, correlating these results with those in Figure 3.20, although the HTH1 has fewer residues contributing to interaction, as they have higher values of contacts it makes the interface area value higher than HTH2 and HTH3.

With all the results obtained, both from the simulations of the free HTH in solution and of these in complex, it is possible to conclude that the three proteins are stable, maintaining their secondary structure throughout the simulation time.

According to our results, HTH1 is predicted to be the most promising design. Of the three proteins, this is the one that has a higher stability, particularly we see that the HTH1-RBD complex has a very low and stable RMSD indicating that the two proteins maintain the configuration observed in the X-ray structure during the simulations. Moreover, HTH1 also presents greater number of total and hydrophobic contacts, and its interface area with the RBD also presented greater values, which indicates that HTH1 presents a stronger interaction with the RBD.

3.2. Expression and purification of the Helix-Turn-Helix Proteins

The target proteins' sequences have 98 residues, in which 6 of them correspond to the affinity tag, His-tag, and the 8 other residues correspond to the 3C cut site (Table 3.1).

Table 3.1 – Sequence of the affinity tag, His-tag, that each protein contains and the respective molecular weights with and without the tag.

Protein	His-tag	Molecular weight (kDa) (with tag)	Molecular weight (kDa) (without tag)
HTH1	HHHHHHLEVLFQGP	11.6	9.9
HTH2		11.5	9.8
HTH3		11.6	9.9

3.2.1. Small-scale expression

In order to determine the optimal expression conditions of HTH1 (11.6 kDa), small-scale tests were performed with different culture media, various *E. coli* strains and with two incubation times after induction, as explained in section 2.3.1.

The results obtained in the small-scale tests were analysed and compared by western blotting and SDS-PAGE.

As observed in Figure 3.26, no significant HTH1 expression in the soluble fraction was detected in cells grown in M9 minimal medium, which ruled out the possibility of using this medium for large-scale expression.

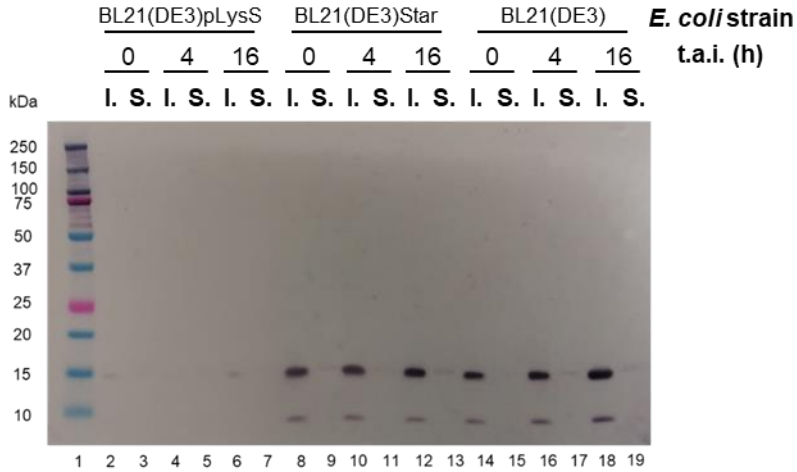

Figure 3.26 – Small-scale expression test with M9 minimal medium analysed by anti-His western blotting (details in Materials and Methods). Lane 1, molecular weight marker (Bio-Rad dual-colour). Lanes 2-19, soluble (S.) and insoluble (I.) fractions obtained from cultures of different *E. coli* strains incubated at different times after induction (t.a.i.).

Conversely, by changing the growth medium LB, HTH1 expression could be detected in the soluble fractions in several conditions (Figure 3.27). Therefore, in order to decide the best condition, an SDS-PAGE of these samples was also performed in order to compare the results (Figure 3.28). In these two gels a stronger band can be observed, which stands out from the rest, between the bands of molecular weights of 15 kDa and 10 kDa.

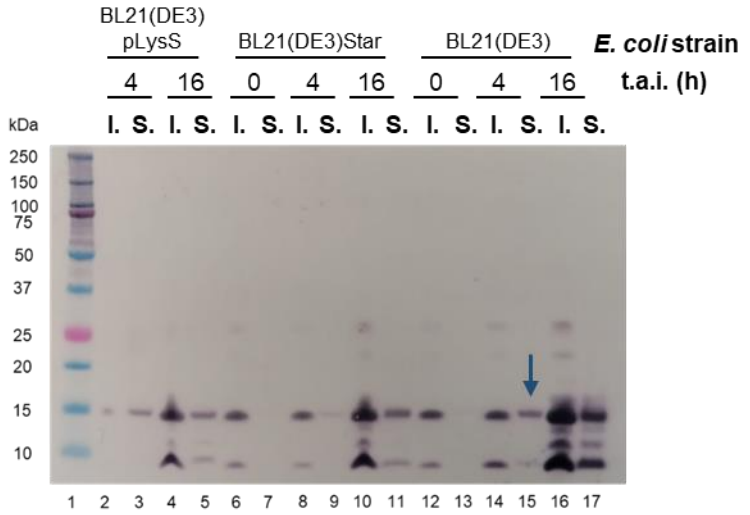

Figure 3.27 – Small-scale expression test with LB medium analysed by anti-His western blotting (details in Materials and Methods). Lane 1, molecular weight marker (Bio-Rad dual-colour). Lanes 2-17, soluble (S.) and insoluble (I.) fractions obtained from cultures of different *E. coli* strains incubated at different times after induction (t.a.i.).

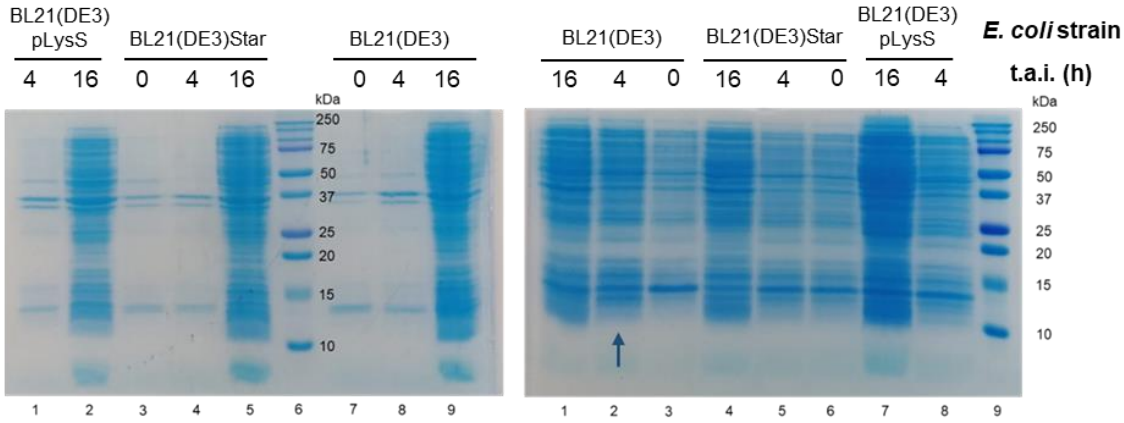

Figure 3.28 – SDS-PAGE gels performed for the LB medium samples. In the left gel was analysed the samples of the insoluble fractions (Lanes 1-5 and 7-9) and on the right gel was analysed the samples of the soluble fractions (Lanes 1-8) from cultures of different *E. coli* strains incubated at different times after induction (t.a.i.). Lane 6 (left); lane 9 (right), molecular weight marker (Bio-Rad dual-colour).

Taking into account the results depicted in Figures 3.27 and 3.28, the chosen conditions were expression in *E. coli* BL21 (DE3) strain in LB medium, 4 hours of induction at 18°C, since it exhibited considerable expression without evidence of protein degradation as in the over-night samples.

Since the structures of the three HTH are relatively similar, it was decided to use these conditions for the expression of HTH2 and HTH3.

3.2.2. Soluble Fraction Purification

From the three proteins present in this work, HTH1 was tested first. Since these proteins have a 6-histidine affinity tag (Figure 2.4 of Materials and Methods), the first purification step consisted of a 5 mL Nickel HisTrap™ HP column.

The sample was injected into this column and after the non-specifically bound proteins were removed by a wash step with 10 mM imidazole, a gradient of 20 CVs was started until 50% buffer B

was reached, i.e. 250 mM imidazole, where the proteins present in peaks I, II and III were eluted. Then, the imidazole concentration was kept constant and after nothing was being eluted, the imidazole concentration was changed to 100% (500 mM) where peak IV was eluted (Figure 3.29).

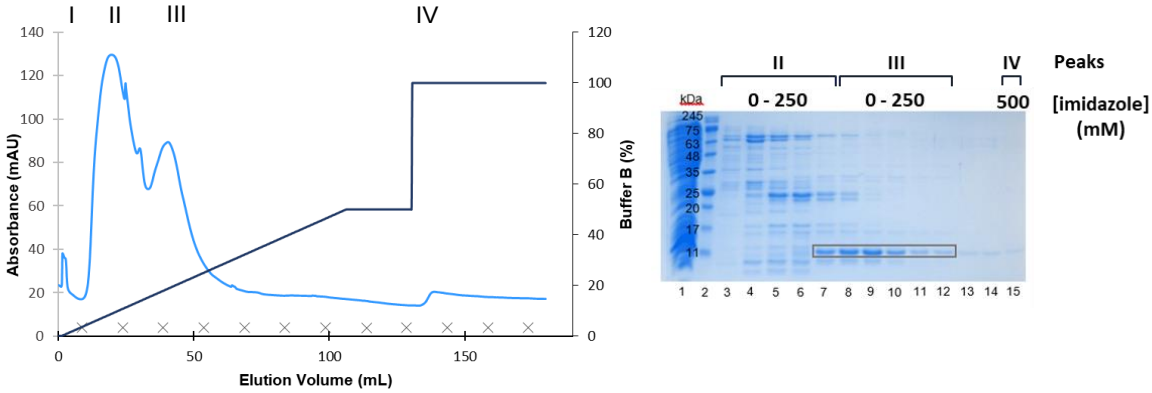

Figure 3.29 – Immobilized metal ion affinity chromatography (IMAC) of HTH1 (Nickel HisTrap™ HP column) (left panel) and SDS-PAGE analysis of eluted fractions (right panel). Peaks I, II and III were eluted with a gradient of 20 CVs up to 50% buffer B and peak IV with 100% buffer B. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 15 to 70, with intervals of 5 fractions. Lane 1, injected sample; lane 2, molecular weight marker (NZYColour Protein Marker II); lanes 3-15, fractions 16, 18, 21, 22, 23, 24, 26, 28, 30, 31, 42, 44 and 59.

The fractions containing HTH1 (black rectangle in Figure 3.29), eluted with 75 mM to 160 mM imidazole, were pooled and concentrated. The sample was then injected in a HiLoad® 16/600 Superdex® 75 pg size exclusion chromatography (Figure 3.30).

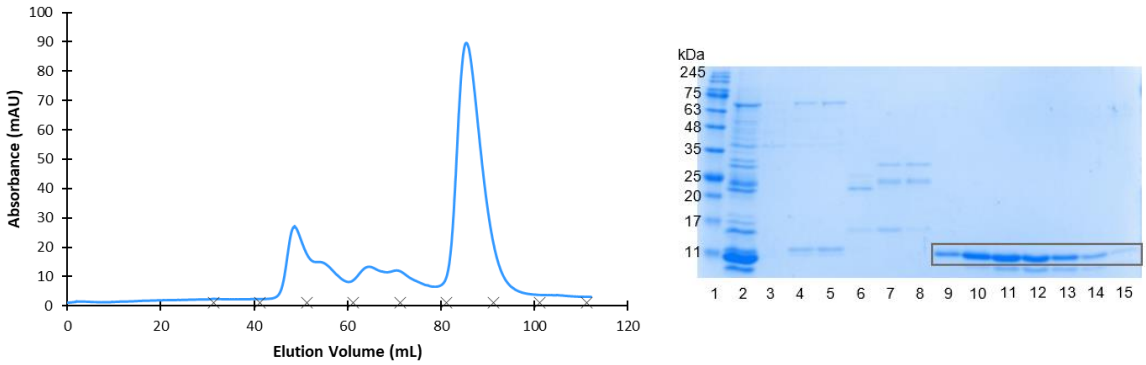

Figure 3.30 – Size exclusion chromatography (HiLoad® 16/600 Superdex® 75 pg) of HTH1 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The peak corresponding to HTH1 was eluted between 80 mL and 100 mL. Absorbance —. Crosses (×) represent the fractions from 5 to 45, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, injected sample; lanes 3-15, fractions 13, 16, 17, 21, 24, 25, 30, 31, 32, 33, 34, 35 and 36.

Fractions containing HTH1 (black rectangle in Figure 3.30) were pooled and concentrated to a final volume of 600 µL (1.02 mg/mL). This sample was incubated overnight at 4°C with 1:100 (w/w) of 1 mg/mL His-3C Protease to remove the His-tag. To separate HTH1 from the His-3C protease and its tag, the sample was injected into a 1 mL Nickel HisTrap™ HP column (Figure 3.31).

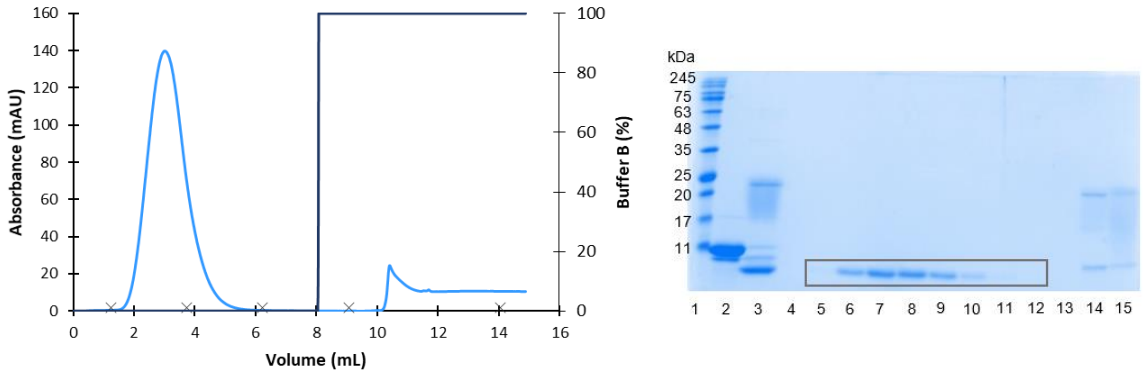

Figure 3.31 – Immobilized metal ion affinity chromatography (Nickel HisTrap™ HP column) of HTH1 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The major peak eluted in the flow-through corresponds to HTH1 without the tag. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 25, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, fraction before reaction with protease; lane 3, injected sample; lanes 4-15, fractions 5, 6, 7, 8, 9, 10, 11, 13, 14, 21 and 22.

Fractions containing HTH1 (black rectangle in Figure 3.31), now with a molecular weight of 9.9 kDa, were pooled and concentrated to a final volume of 270 µL at 1.07 mg/mL. A western blot was performed to confirm the absence of the His-tag (Figure 3.32).

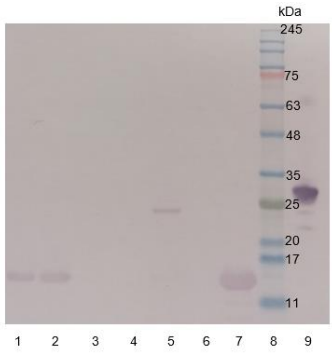

Figure 3.32 – Western blot performed for samples from the various steps of the HTH1 purification protocol. Lanes 1-7, fractions 33 and 32 after the second purification step, untagged fractions 9 and 8, injected in the HisTrap after cutting with His-3C, untagged fraction after being concentrated, tagged fraction before the cut; lane 8, molecular weight marker (NZYColour Protein Marker II); lane 9, positive His-tagged control.

For the HTH2 protein, the first purification step consisted of a 5 mL Nickel HisTrap™ HP column. The rest of the protocol was carried out in the same way as described above.

Accordingly, the sample was injected into this column and the proteins present in peaks I and II were eluted during the 20 CVs gradient. Peak III was eluted with 100% of imidazole (Figure 3.33).

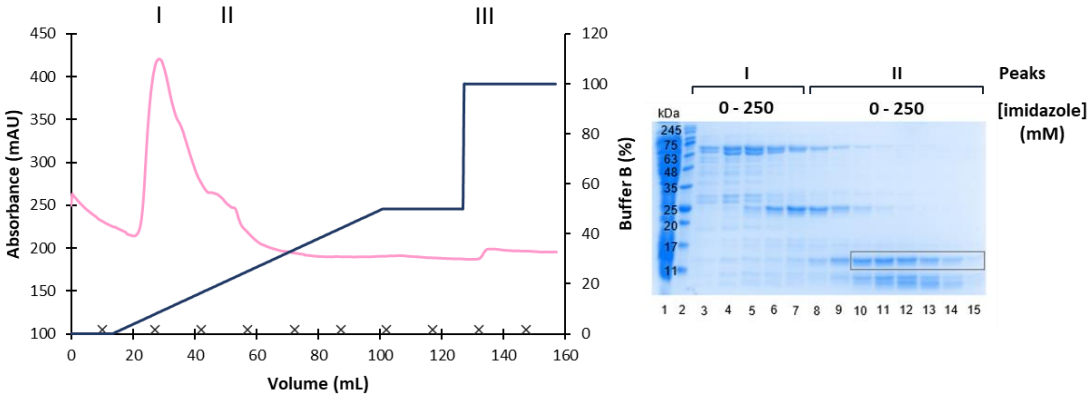

Figure 3.33 – Immobilized metal ion affinity chromatography (IMAC) of HTH2 (Nickel HisTrap™ HP column) (left panel) and SDS-PAGE analysis of eluted fractions (right panel). Peaks I and II were eluted with a gradient of 20 CVs up to 50% buffer B and peak III with 100% buffer B. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 50, with intervals of 5 fractions. Lane 1, injected sample; lane 2, molecular weight marker (NZYColour Protein Marker II); lanes 3-15, fractions 9 to 20 and 22.

The fractions containing HTH2 (black rectangle in Figure 3.33), eluted with 85 mM to 110 mM imidazole, were pooled and concentrated. The sample was injected into a HiLoad® 16/600 Superdex® 75 pg size exclusion chromatography (Figure 3.34).

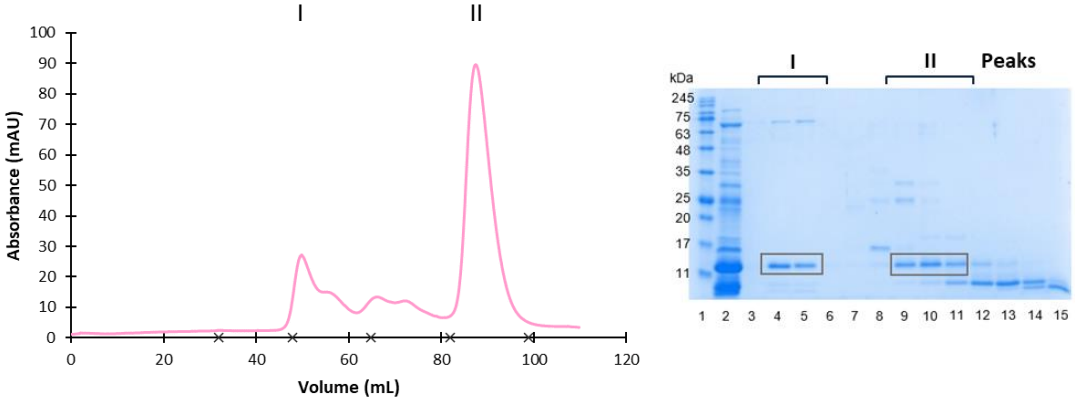

Figure 3.34 – Size exclusion chromatography (HiLoad® 16/600 Superdex® 75 pg) of HTH2 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). Peak I and II correspond to HTH2 in different oligomeric states. Absorbance —. Crosses (×) represent the fractions from 5 to 25, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, injected sample; lanes 3-15, fractions 8, 10, 12, 14, 15, 17, 19, 20, 21, 22, 23, 24 and 25.

The HTH2 protein is present in two peaks but in different oligomeric states, peak I corresponds to a dimer while peak II corresponds to the normal monomeric state of the protein. Fractions containing HTH2 (black rectangles in Figure 3.34) were pooled in two samples and sample 2 was concentrated to a final volume of 600 µL (0.22 mg/mL). This sample was incubated overnight at 4°C with 1:100 (w/w) of 1 mg/mL His-3C Protease to remove the His-tag. To separate untagged HTH2 from the His-3C protease and its tag, the sample was injected into a 1 mL Nickel HisTrap™ HP column (Figure 3.35).

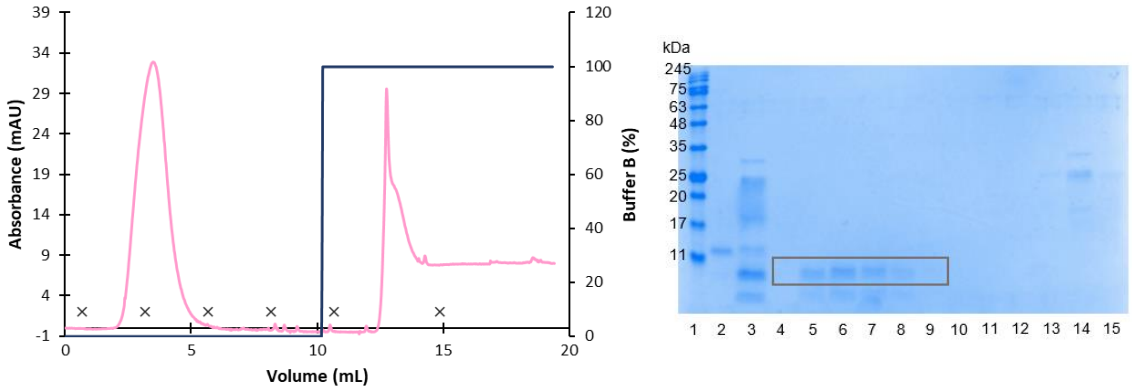

Figure 3.35 – Immobilized metal ion affinity chromatography (Nickel HisTrap™ HP column) of HTH2 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The major peak eluted in the flow-through corresponds to HTH2 without the tag. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 30, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, fraction before proteolytic cleavage; lane 3, injected sample; lanes 4-15, fractions 8 to 13, 20, 22, 27, 28 and 29.

Fractions containing HTH2 (black rectangle in Figure 3.35), now with a molecular weight of 9.8 kDa, were pooled and concentrated to a final volume of 900 µL in order to inject into a Superdex® 75 10/300 GL size exclusion chromatography to verify the oligomeric state of the protein (Figure 3.36).

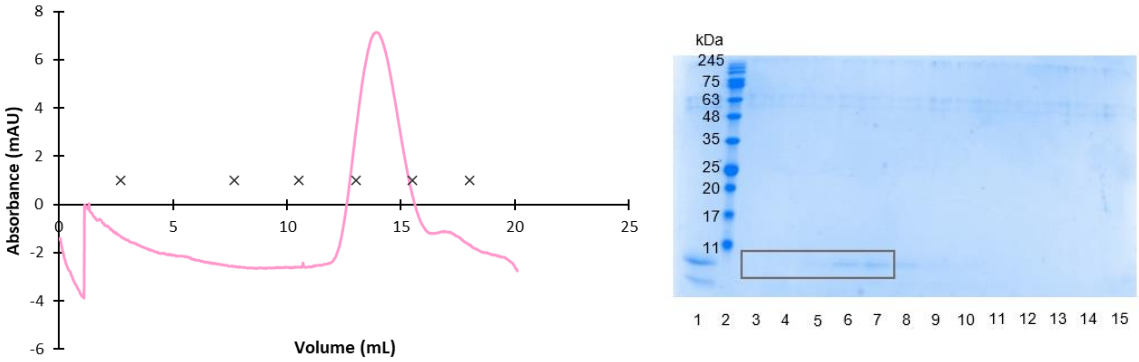

Figure 3.36 – Size exclusion chromatography Superdex® 75 10/300 GL of HTH2 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). Absorbance —. Crosses (×) represent the fractions from 5 to 30, with intervals of 5 fractions. Lane 1, injected sample; lane 2, molecular weight marker (NZYColour Protein Marker II); lanes 3-15, fractions 17 to 29.

Through this last purification step, it was found that the protein was the monomeric state. Fractions containing HTH2 (black rectangle in Figure 3.36) were pooled and concentrated to a final volume of 750 µL at 0.19 mg/mL. To verify that what was being collected was the protein of interest, a western blot was performed where (Figure 3.37) a signal that corresponds to the protein before the tag was cut can been seen.

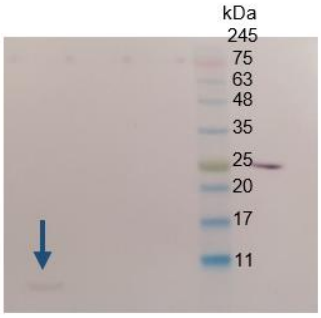

Figure 3.37 – Western blot performed for samples from the various steps of the HTH2 purification protocol. The signal marked with a blue arrow corresponds to the one injected into the column in the second purification step.

Finally, this protocol was also used for the HTH3 protein and the first purification step consisted of a 5 mL Nickel HisTrap™ HP column.

Therefore, the sample was injected into this column and the proteins present in the majority peak were eluted during the 20 CVs gradient (Figure 3.38).

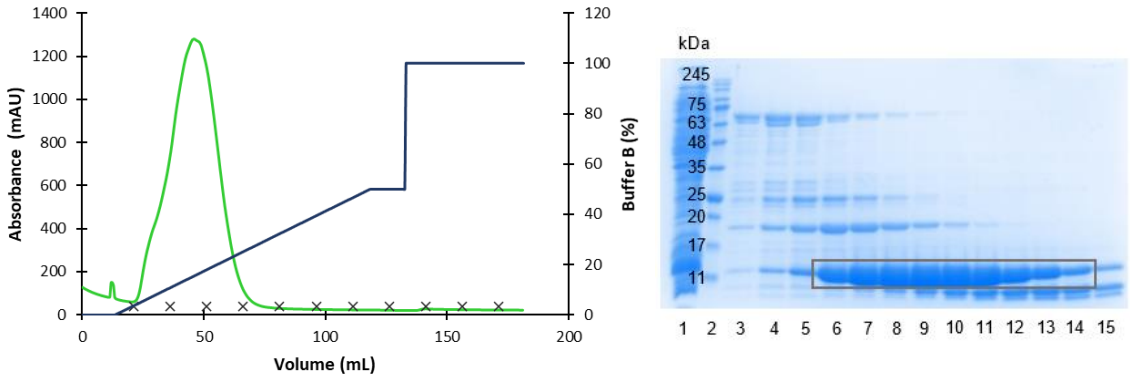

Figure 3.38 – Immobilized metal ion affinity chromatography (IMAC) of HTH3 (Nickel HisTrap™ HP column) (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The major peak was eluted with a gradient of 20 CVs up to 50% buffer B. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 55, at intervals of 5 fractions. Lane 1, injected sample; lane 2, molecular weight marker (NZYColour Protein Marker II); lanes 3-15, fractions 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19.

The fractions containing HTH3 (black rectangle in Figure 3.38), eluted with 45 mM to 125 mM imidazole, were pooled in two samples and sample 2 was concentrated. The sample was injected into the size exclusion chromatography HiLoad® 16/600 Superdex® 75 pg (Figure 3.39).

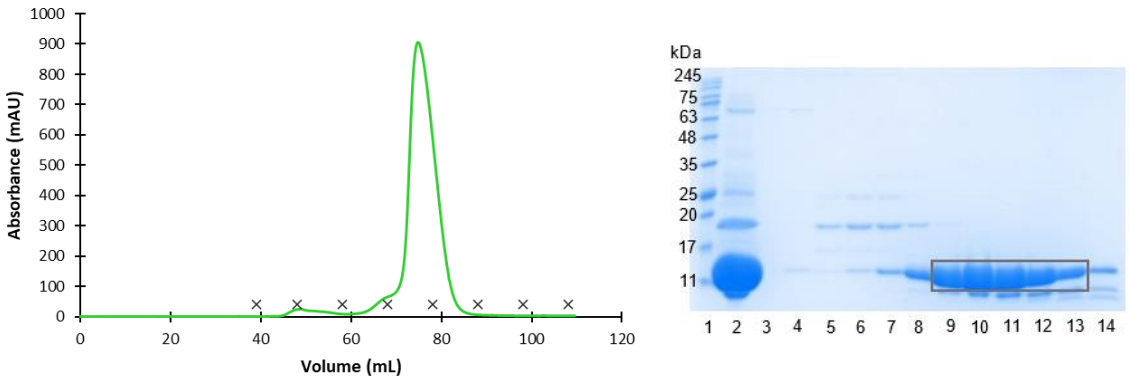

Figure 3.39 – Size exclusion chromatography (HiLoad® 16/600 Superdex® 75 pg) of HTH3 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The major peak corresponds to HTH3. Absorbance —. Crosses (×) represent the fractions from 5 to 40, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, injected sample; lanes 3-14, fractions 10, 12, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27.

The HTH3 protein is present in the major peak that corresponds to the monomeric state of the protein. Fractions containing HTH3 (black rectangle in Figure 3.39) were pooled and the final volume of 8 mL (2.14 mg/mL) was incubated overnight at 4 °C with 1:100 (w/w) of 1 mg/mL His-3C Protease. The sample was injected into a 1 mL Nickel HisTrap™ HP column (Figure 3.40).

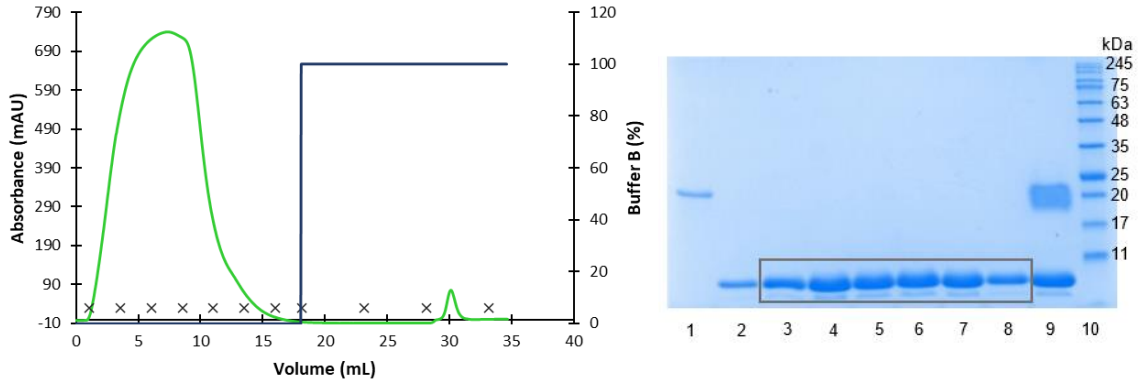

Figure 3.40 – Immobilized metal ion affinity chromatography (Nickel HisTrap™ HP column) of HTH3 (left panel) and SDS-PAGE analysis of eluted fractions (right panel). The major peak eluted in the flow-through corresponds to HTH3 without the tag. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 55, with intervals of 5 fractions. Lanes 1-8, fractions 51, 28, 24, 20, 18, 16, 12 and 8; lane 9, injected sample; lane 10, molecular weight marker (NZYColour Protein Marker II).

Fractions containing HTH2 (black rectangle in Figure 3.40), now with a molecular weight of 9.9 kDa, were pooled, concentrated and injected into a Superdex® 75 10/300 GL column to verify the oligomeric state of the protein (Figure 3.41).

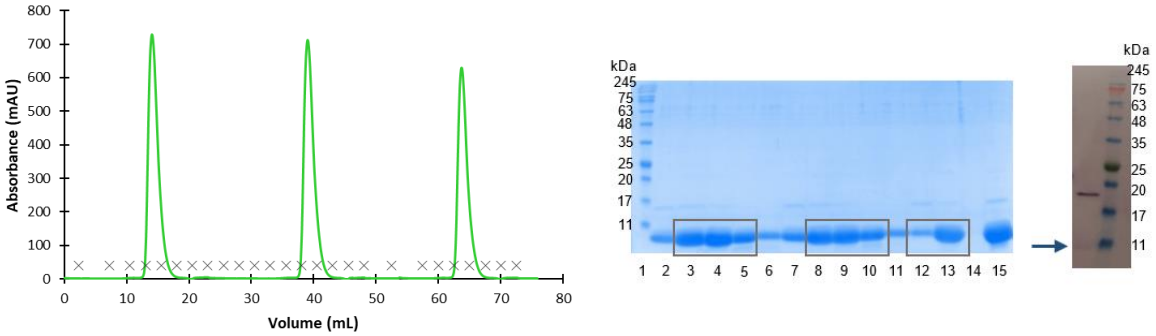

Figure 3.41 – Size exclusion chromatography Superdex® 75 10/300 GL of HTH3 (left panel), SDS-PAGE analysis of eluted fractions and final western blot (right panel). Absorbance —. Crosses (×) represent the fractions from 5 to 130, with intervals of 5 fractions. Lane 1, molecular weight marker (NZYColour Protein Marker II); lanes 2-15, fractions 18, 20, 21, 22, 24, 26, 70, 71, 72, 76, 110, 112; lane 15, injected sample.

The protein was eluted as a monomer. Fractions containing HTH3 (black rectangles in Figure 3.41) were pooled and concentrated to a final volume of 2 mL at 2.12 mg/mL.

3.2.3. Insoluble Fraction Purification

As described above, in the purification of the HTH1 and HTH2 proteins a low yield was obtained. Because of this and because through the expression tests we could see that there was a greater amount of protein in the insoluble fraction than in the soluble fraction, we decided to try to recover these two proteins from the inclusion bodies (section 2.3.4.). This strategy was also based on the fact that we observed that thermal unfolding of HTH1 monitored by Far-UV CD spectropolarimetry was apparently reversible (detailed in section 3.3.1).

First, this protocol was performed for HTH1, which after obtaining a soluble fraction from the inclusion bodies, was injected into a 5 mL Nickel HisTrap™ HP column and after the non-specifically bound proteins were removed by a wash step with 20 mM imidazole (peak I), 4 steps of 10% (50 mM imidazole), 20% (100 mM imidazole), 40% (200 mM imidazole) and 100% (500 mM imidazole) buffer B were performed where the proteins present in peaks II, III, IV and V were eluted (Figure 3.42).

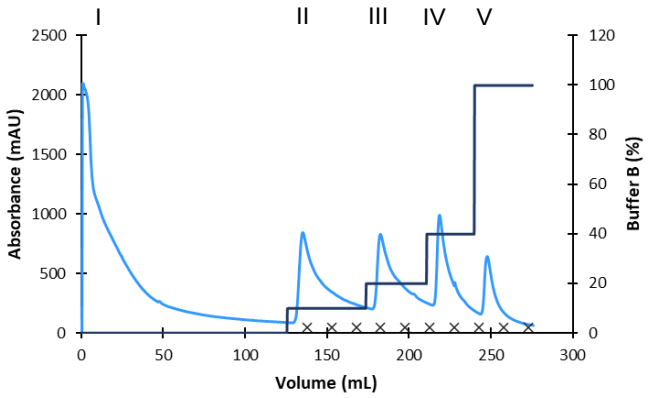

Figure 3.42 – Immobilized metal ion affinity chromatography (IMAC) of HTH1 (Nickel HisTrap™ HP column). Peak I, flow-through; peaks II, III, IV and V, eluted in different steps of buffer B, 10%, 20%, 40% and 100%. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 50, with intervals of 5 fractions.

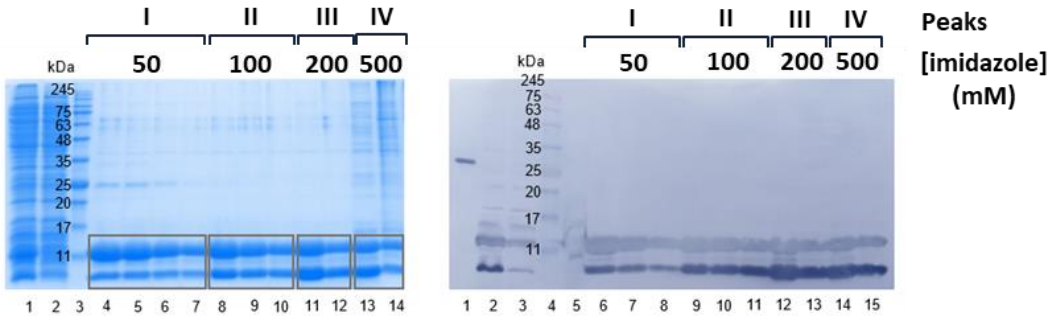

Figure 3.43 – SDS-PAGE and western blot of fractions from the chromatogram in Figure 3.42. In the gel, lane 1, injected sample; lane 2, flow-through; lane 3, molecular weight marker (NZYColour Protein Marker II); lanes 4-14, fractions 4, 5, 8, 13, 20, 22, 26, 32, 35, 41 and 43. In the western blot, lane 1, positive His-tagged control; lane 2, injected sample; lane 3, flow-through; lane 4, molecular weight marker; lane 5, initial soluble fraction; lanes 6-15, fractions 4, 8, 13, 20, 22, 26, 32, 35, 41 and 43.

The fractions containing HTH1 (black rectangles in Figure 3.43) were pooled in four samples, each one corresponding to each imidazole step. By western blot it was confirm the presence of His-tagged protein that, as expected, was not yet fully renatured, due to the presence of several signals at different molecular weights. Through the gel, it is possible to see that the fractions from the 10% and 100% imidazole steps (first and fourth rectangle in Figure 3.43) still contained a lot of impurities, so samples 2 and 3 were pooled, added and proceeded to the subsequent purification steps.

These samples were re-injected into a 5 mL Nickel HisTrap™ HP column for protein renaturation on column. To do this, a gradient of 20 CVs to 100% buffer D, which is a buffer without the denaturing agent, was performed. Then, to elute the protein, a gradient of 15 CVs up to 100% of buffer E was performed (Figure 3.44).

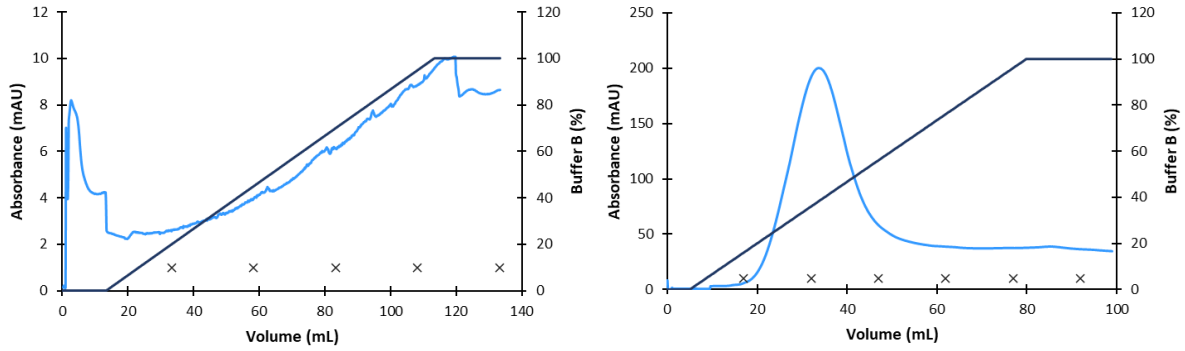

Figure 3.44 – Chromatograms of the elution gradient and on-column renaturation of HTH1. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 25 on the renaturation gradient and from 5 to 30 on the elution gradient, with intervals of 5 fractions.

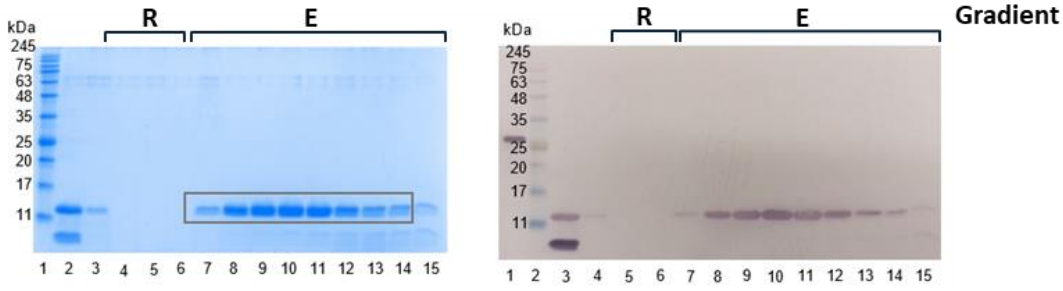

Figure 3.45 – SDS-PAGE and western blot of fractions from the chromatograms in Figure 3.44. In the gel, lane 1, molecular weight marker (NZYColour Protein Marker II); lane 2, injected sample; lane 3, flow-through; lanes 4-15, fractions 19, 21, 23, 6, 8, 9, 10, 11, 13, 15, 17 and 27. In the western blot, lane 1, positive His-tagged control; lane 2, molecular weight marker (NZYColour Protein Marker II); lane 3, injected sample; lane 4, flow-through; lanes 5-15, fractions (renaturation gradient) 21, 23, (elution gradient) 6, 8, 9, 10, 11, 13, 15, 17 and 27. R, renaturation gradient; E, elution gradient.

Fractions containing HTH1 (black rectangle in Figure 3.45) were pooled, concentrated and injected into a Superdex® 75 10/300 GL column (Figure 3.46). By western blot a single His-tagged band was observed around 11 kDa.

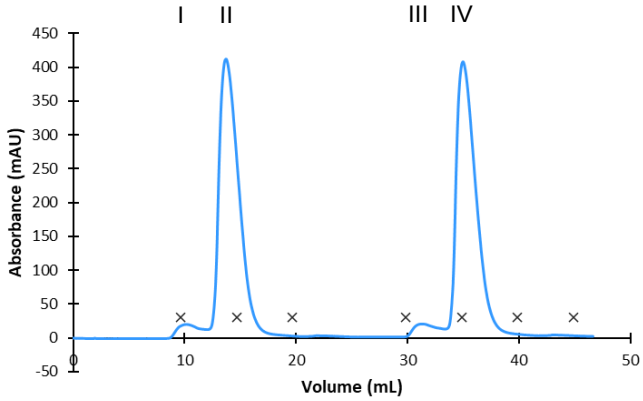

Figure 3.46 – Chromatogram form Superdex® 75 10/300 GL of HTH1 (two runs). Absorbance —. Crosses (×) represent the fractions from 5 to 35, with intervals of 5 fractions.

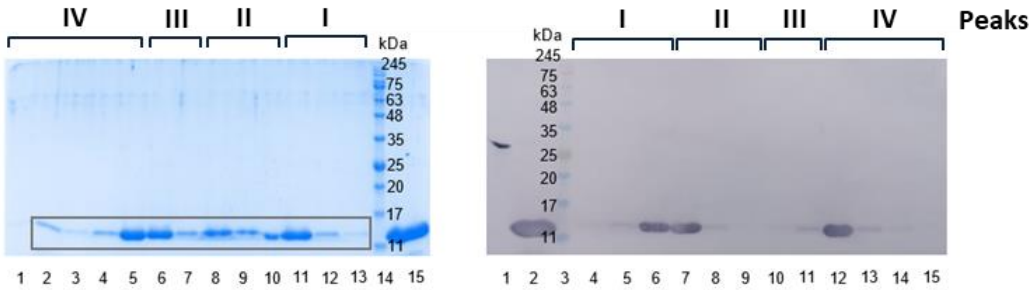

Figure 3.47 – SDS-PAGE and western blot analysis of fractions from the chromatogram in Figure 3.46. In the gel, lanes 1-13, fractions 35, 33, 31, 28, 26, 24, 20, 13, 12, 10, 8, 6 and 4; lane 14, molecular weight marker (NZYColour Protein Marker II); lane 15, injected sample. In the western blot, lane 1, positive His-tagged control; lane 2, injected sample; lane 3, molecular weight marker (NZYColour Protein Marker II); lanes 4-15, fractions 4, 6, 8, 10, 12, 13, 20, 24, 26, 28, 29 and 33.

The four peaks observed in Figure 3.46 correspond to HTH1, but since peaks I and III were eluted before the two major peaks, which could indicate that the protein was aggregating in some way that gave it a higher molecular weight. The fractions containing HTH1 were pooled in two samples and concentrated to a final volume of 2.5 mL of sample 1 (0.98 mg/mL) and 1.5 mL of sample 2 (0.03 mg/mL).

For HTH2 the purification protocol for the soluble fraction obtained from the inclusion bodies was the same as previously described (Figure 3.48).

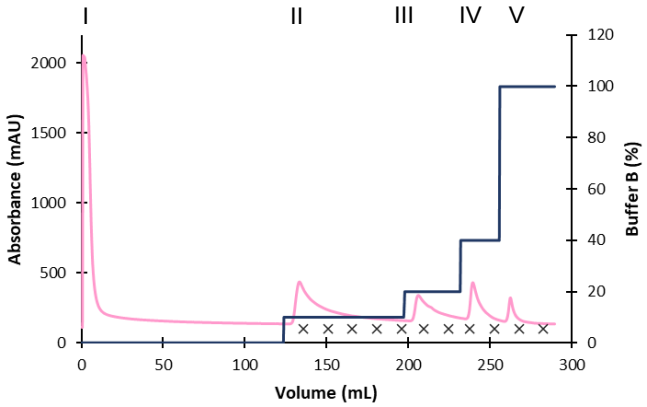

Figure 3.48 – Immobilized metal ion affinity chromatography (IMAC) of HTH2 (Nickel HisTrap™ HP column). Peak I, flow-through; peaks II, III, IV and V, eluted in different steps of buffer B, 10%, 20%, 40% and 100%. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 55, at intervals of 5 fractions.

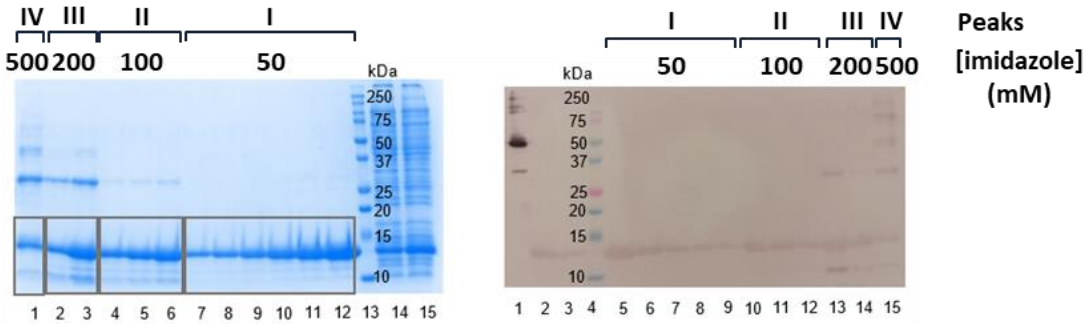

Figure 3.49 – SDS-PAGE and western blot analysis of chromatogram in Figure 3.48. In the gel, lanes 1-12, fractions 48, 43, 40, 34, 31, 29, 24, 18, 12, 8, 6 and 4; lane 13, molecular weight marker (Bio-Rad dual-colour); lane 14, flow-through; lane 15, injected sample. In the western blot, lane 1, positive His-tagged control; lane 2, injected sample; lane 3, flow-through; lane 4, molecular weight marker (Bio-Rad dual-colour); lanes 5-15, fractions 4, 8,12, 18, 24, 29, 31, 34, 40, 43 and 48.

The fractions containing HTH2 (black rectangles in Figure 3.49) were pooled in four samples, each one corresponding to the imidazole steps. By western blot it was possible to check the presence of His-tagged HTH2 and some contaminants in the fractions eluted at higher imidazole concentrations. Therefore, samples 1, 2 and 3 were pooled and the protocol continued.

These samples were re-injected into a 5 mL Nickel HisTrap™ HP column for protein renaturation on column, followed by elution with an imidazole gradient (Figure 3.50).

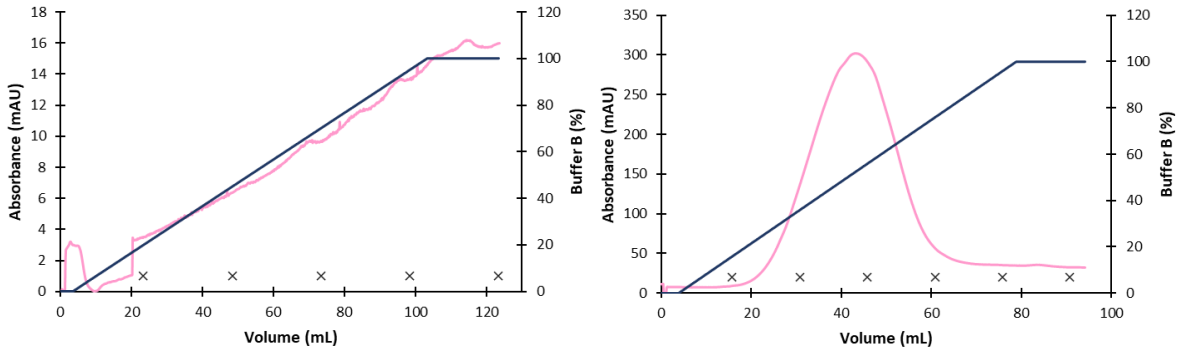

Figure 3.50 – Chromatograms of the elution gradient and on-column renaturation of HTH2. Absorbance —, Buffer B (%) —. Crosses (×) represent the fractions from 5 to 25 on the renaturation gradient and from 5 to 30 on the elution gradient, with intervals of 5 fractions.

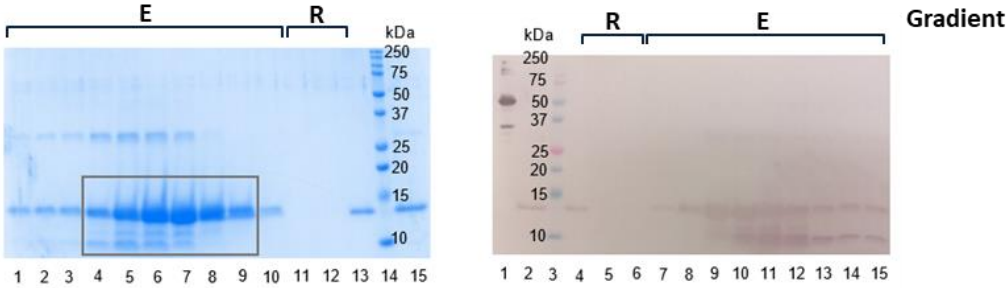

Figure 3.51 – SDS-PAGE and western blot analysis of chromatograms in Figure 3.50. In the gel, lanes 1-12, fractions (elution gradient) 27, 24, 22, 19, 17, 15, 13, 10, 8, 6, (renaturation gradient) 23, 19; lane 13, flow-through; lane 14, injected sample; lane 15, molecular weight marker (Bio-Rad dual-colour). In the western blot, lane 1, positive His-tagged control, lane 2, injected sample; lane 3, molecular weight marker (Bio-Rad dual-colour); lane 4, flow-through; lane 5-15, fractions (renaturation gradient) 19, 23, (elution gradient) 6, 8, 13, 15, 17, 19, 22, 24 and 27.

Fractions containing HTH2 (black rectangle in Figure 3.51) were pooled and concentrated to a final volume of 4 mL and injected into a Superdex® 75 10/300 GL column (Figure 3.52). By western blot in Figure 3.51, more than one signal is observed in bands of different molecular weights. Despite this, we proceeded with the protocol (Figure 3.52).

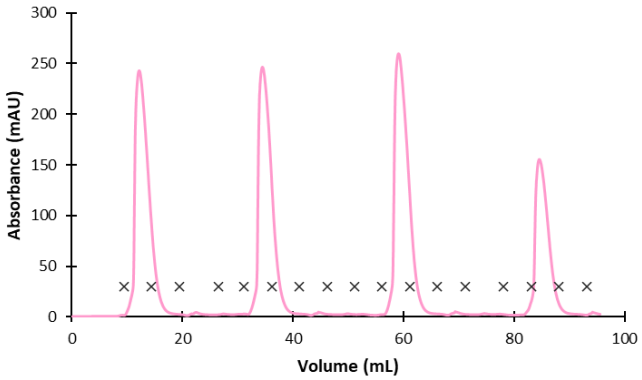

Figure 3.52 – Chromatogram form Superdex® 75 10/300 GL of HTH2 (four runs). Absorbance —. The crosses (×) represent the fractions from 5 to 85, at intervals of 5 fractions.

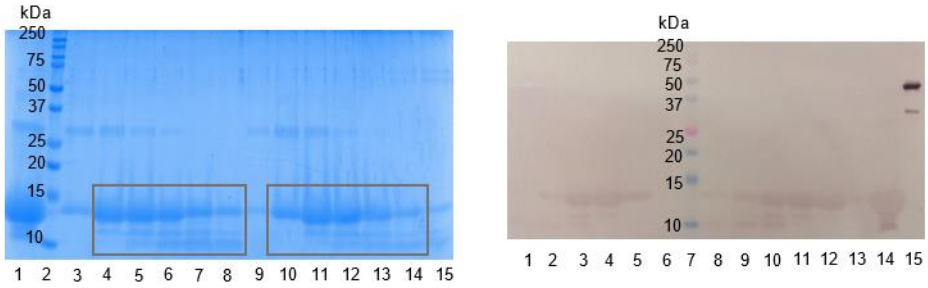

Figure 3.53 – SDS-PAGE and western blot analysis of chromatogram in Figure 3.52. In the gel, lane 1, injected sample; lane 2, molecular weight marker (Bio-Rad dual-colour); lanes 3-15, fractions 6 to 11 and 26 to 32. In the western blot, lanes 1-6, fractions 32, 31, 29, 28, 27, 26; lane 7, molecular weight marker (Bio-Rad dual-colour); lanes 8-13, fractions 11 to 6; lane 14, injected sample; lane 15, positive His-tagged control.

The four peaks observed in Figure 3.52 correspond to HTH2. The fractions containing HTH2 (black rectangles in Figure 3.53) were pooled in four samples and concentrated to a final volume of 1.7 mL of sample 1 (1.31 mg/mL), 3 mL of sample 2 (0.13 mg/mL), 2 mL of sample 3 (1.40 mg/mL), and 2.3 mL of sample 4 (0.81 mg/mL). In the western blot once again, it is possible to observe the presence of the protein of interest.

3.3. Structural characterization and conformational stability of the Helix-Turn-Helix Proteins

3.3.1. Far-UV Circular Dichroism Spectropolarimetry

Through circular dichroism (CD) spectropolarimetry we tried to understand: i) if the proteins obtained from the purification protocol of the soluble fraction have the secondary structure that was predicted by molecular dynamics simulations (Figure 3.12 – 3.14) and ii) if the proteins obtained from the purification protocol of the insoluble fraction had recovered their structure after renaturation. In addition to this, an analysis of the folding and unfolding of the proteins was also performed by varying the temperature, where there was an increase up to 80 °C and then a cooling down to 20 °C.

Through the CD studies, performed as described in section 2.6 of Materials and Methods, the following spectra were obtained (Figure 3.54 and 3.55).

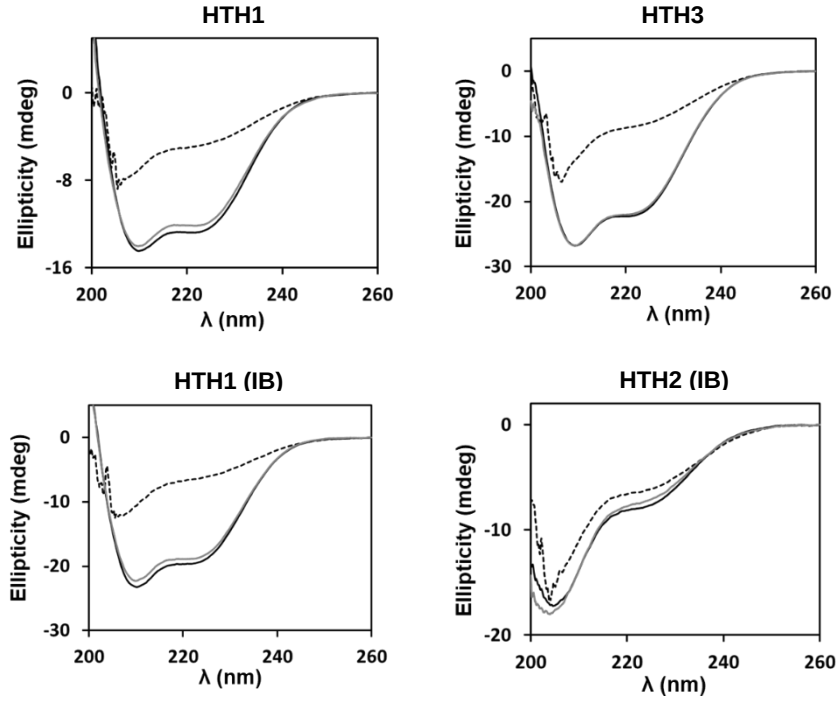

Figure 3.54 – CD spectra of HTH1, HTH2 and HTH3 proteins. The top spectra were obtained from samples coming from the purification of soluble fractions, while the bottom spectra were obtained from samples coming from the purification of insoluble fractions. Black line (—), spectra obtained at 20 °C; dashed line (---), spectra obtained at 80 °C; grey line (—), spectra obtained at 20 °C after re-cooling.

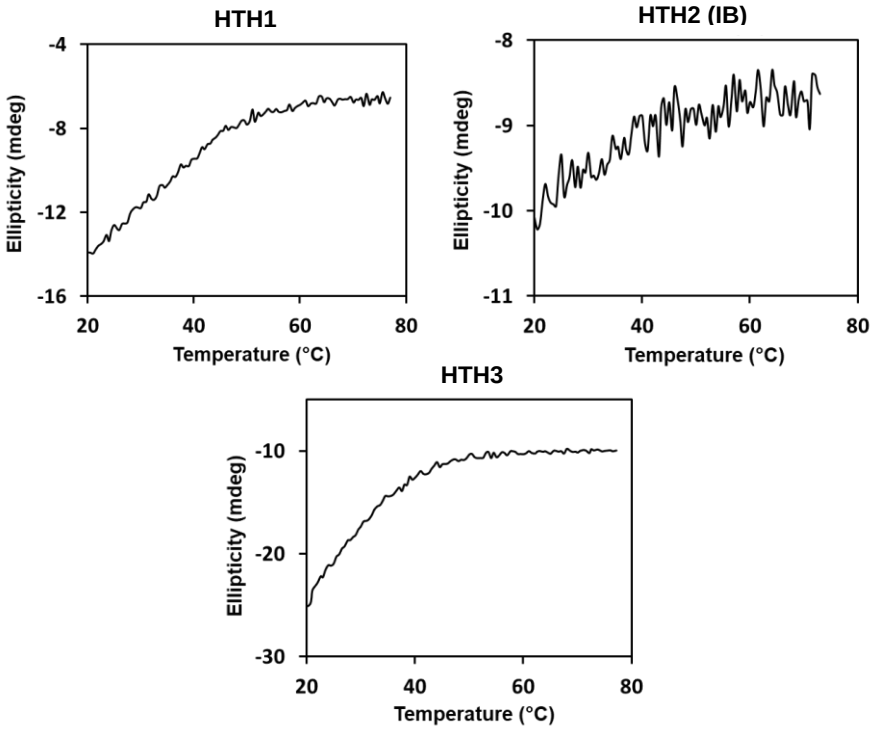

Figure 3.55 – Thermal denaturation curves of the three proteins by increasing temperature.

The CD spectra (Figure 3.54) confirm that the secondary structure of the three proteins consists largely of α-helices, according to the presence of peaks centred approximately at 210 nm and 225 nm. The CD spectra of HTH2 slightly differs from those of HTH1 and HTH3 in terms of relative intensity of the two major bands, possibly due to an effect of the disulfide bonds on the overall secondary structure, i.e., as in Montoliu-Gaya et al.[145].This information is in agreement with the predictions of secondary structure that were made in molecular dynamics, and thus it is possible to correlate the information obtained.

When the denaturation of the proteins was performed employing a linear temperature gradient and monitoring the spectral feature typical of α-helical content (222 nm), we can observe that thermal unfolding starts immediately with no lag phase as observed in typical sigmoidal thermal denaturation curves. Indeed, for HTH1 and HTH3, most secondary structure appeared to be completely lost by ~50 °C (Figure 3.55). This indicates that HTH1 and HTH3 probably exhibit an extremely high conformational flexibility with the resulting decrease in stability. Conversely, HTH2 exhibited a small decrease in secondary structure content up to 80 °C, as seen both in the spectra and thermal denaturation curve. Again, this is likely due to the presence of the two disulfide bonds rendering the structure of the protein more stable and resistant.

Upon cooling the temperature back down to 20 °C, all the proteins regained their secondary structure, indicating that they possess the ability to regain their native structure after undergoing denaturation.

In the spectra obtained from the purification of the insoluble fractions, it is also possible to see that the proteins recovered their secondary structure after renaturation, indicating that the protocol went as expected.

3.3.2. Differential Scanning Fluorimetry (nanoDSF)

As in the HTH1 and HTH2 purification protocol from the soluble fractions a small amount of protein was obtained, nanoDSF was used to conduct a buffer screening, aiming at protein stabilization of HTH2 thermal denaturation curves monitoring intrinsic tryptophan fluorescence as a function of temperature are shown in Figure 3.56, along with the corresponding first derivatives. With the protein denaturation there are changes in the native structure of the protein, which leads to a decrease in fluorescence by inhibition of the excitation of the amino acids responsible for it, due to a possible shielding effect of these.

The T_{onset} corresponds to the temperature at which unfolding starts, and it is possible to get this value at the start of the downward curve in Figure 3.57 and 3.58, and T_m corresponds to the point where in theory half of the protein present in the sample is unfolded. T_m can be estimated from the minimum of the first derivative curves (Figure 3.57), where it can be observed that the T_m did not change significantly in all the buffers tested (43.8-45.4 ºC).

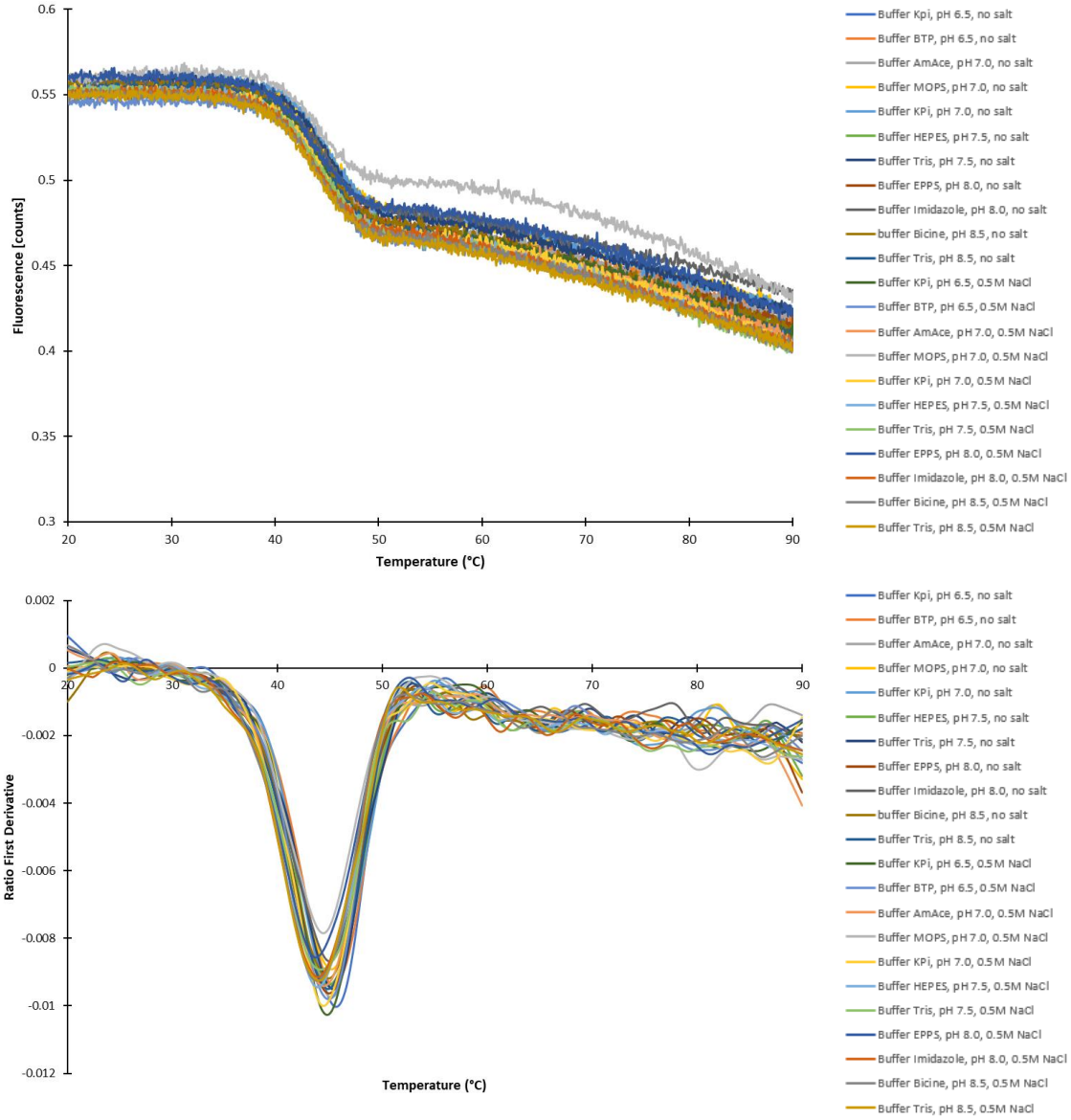

Figure 3.56 – Denaturation curves obtained with buffer screening, through nanoDSF, performed for HTH2, which relate the native fluorescence of the protein with increasing temperature. Bottom panel, first derivative of the curves in the top panel.

Despite these small variations, some valuable information could be extracted. Since purified HTH2 protein was in Tris-HCl buffer, pH 8.0, 0.3 M NaCl, the closest buffer compositions herein tested were Tris-HCl buffer, pH 7.5, 0.5 M NaCl (T_{onset} 36.6°C and T_m 44.3°C), and Tris-HCl buffer, pH 8.5, 0.5 M NaCl (T_{onset} 35.8°C and T_m 43.8 °C). As observed in Figure 3.58, phosphate buffer (KPi) at pH 6.5, 0.5 M NaCl, afforded a shift of both the T_{onset} and the T_m to slightly higher temperature values (respectively, 37.6 °C and 44.7 °C). Therefore, KPi buffer at pH 6.5 was selected as a more favourable condition for this protein.

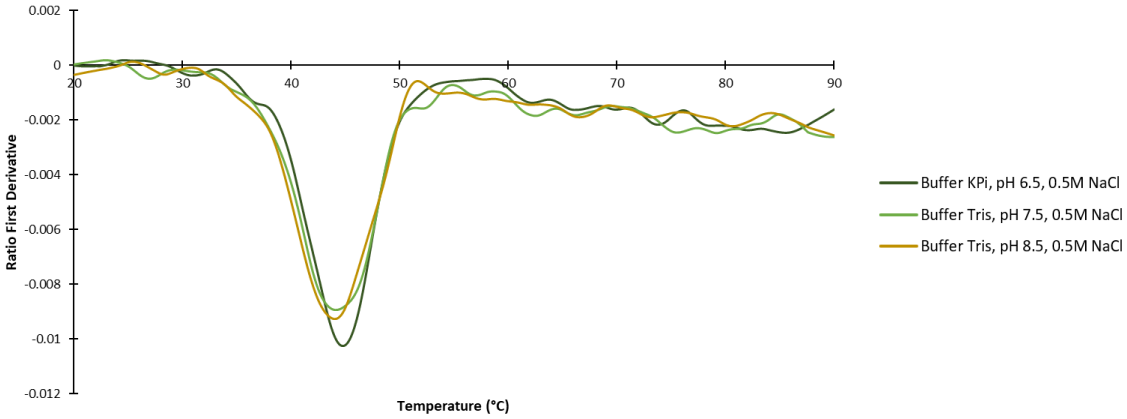

Figure 3.57 – First derivative relative to the potassium phosphate buffer (KPi) at pH 6.5 with a salt concentration (NaCl) of 0.5M, which corresponds to the condition that appears to be most favourable for the protein, and the buffers (Tris-HCl buffer, pH 7.5/8.5, 0.5M NaCl) with the composition most similar to the buffer (Tris-HCl buffer, pH 8.0, 0.3M NaCl) where the protein was.

3.4. Evaluation of the antiviral properties and interaction with the S protein of the Helix-Turn-Helix Proteins

3.4.1. Surface Plasmon Resonance

To evaluate the interaction of these proteins with the RBD from the S protein, Surface Plasmon Resonance (SPR) assays were performed with the HTH1 and HTH3 proteins, since HTH2 purification had not yet been attained.

Through the SPR results obtained for HTH1 (Figure 3.58) it can be seen that this protein interacts with the RBD with a relatively high dissociation constant (K_D) of few µM, between 5 and 7 µM. The fact that the affinity is relatively low may be due to the fact that this assay was done with RBD alone without being in the whole spike protein context. Since in the whole spike context RBD changes conformation between an open and a closed position, in this assay RBD conformation may not be the most favourable for the interaction to occur. In addition to this, other interaction sites may be missing in the context of the entire S protein that are also important. Another reason that may also explain what happened is that the HTH1 protein was taken out of the context of the ACE2 protein and that in the construct that was designed two glycans were missing, at sites N90 and N322, which in previous studies have been seen to be important for the interaction of the ACE2 protein with the RBD of SARS-CoV-2[76].

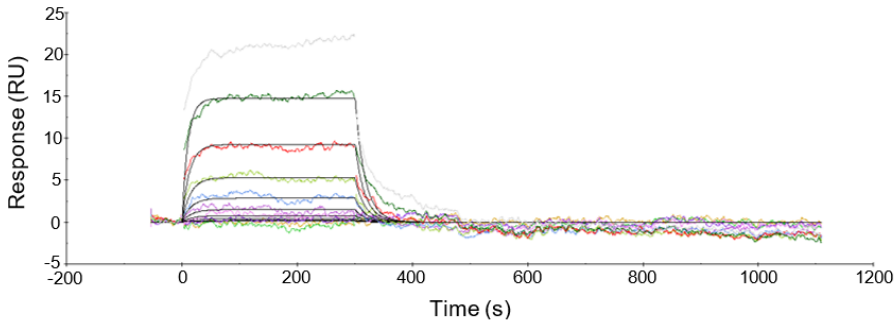

Figure 3.58 – Sensorgram of the SPR experiment performed with the HTH1 protein. For this experiment RBD was immobilized on a CM5 sensor chip and different dilutions of the protein were tested (0.019 µM - 10 µM).

Regarding the HTH3 protein (Figure 3.59), through the SPR data it can be seen that this protein did not interact with the RBD of protein S. Despite the presence of the substitutions made in order to

improve the interaction, this may have been, again, due to the fact that the protein was taken out of the context of the ACE2 protein, and thus HTH3 consisted only of the two helices that appeared to have more interactions with the RBD. But, on the other hand, there are other regions in ACE2 necessary, such as the presence of the two glycans mentioned above, for the interaction to occur, and this may be the explanation why the control assay with ACE2 (Figure 3.60) gave much more positive values.

Furthermore, through molecular dynamics simulations it could be observed that HTH1, which did not have the substitutions performed on HTH3, showed to have a greater interaction with RBD. Indicating, in this manner, that the HTH1 protein would have a higher affinity for the RBD. Therefore, we can correlate the results obtained in the simulations with those obtained here.

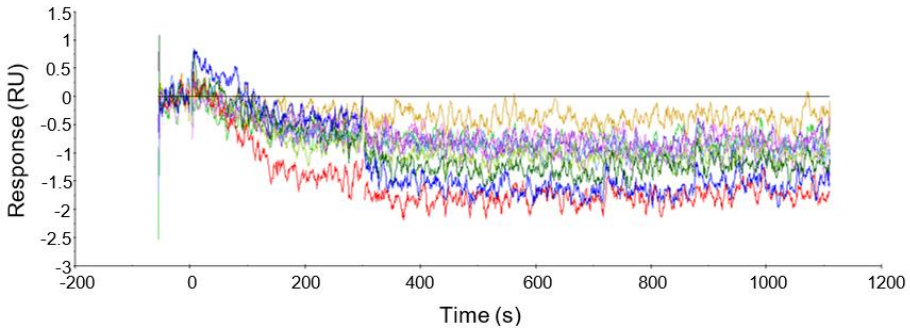

Figure 3.59 – Sensorgram of the SPR experiment performed with the HTH3 protein. For this experiment different dilutions of the protein were used (0.019 µM - 10 µM).

ACE2 protein was used as a positive control for this experiment, which showed a high interaction with RBD, with a dissociation constant of ≈ 143 nM.

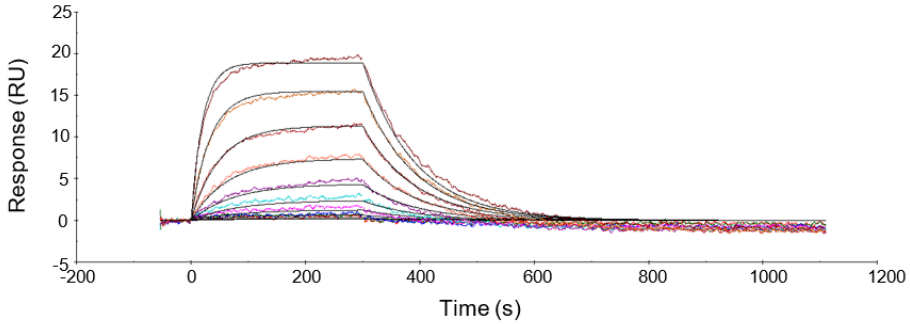

Figure 3.60 – Sensorgram of the SPR experiment performed with the ACE2 protein. RBD was immobilized on a CM5 sensor chip and different ACE2 concentrations (0.00098 µM - 0.5 µM) were analysed.

3.4.2. Neutralization assays

In order to evaluate the antiviral properties of these three proteins, neutralization assays were performed using pseudoviruses.

In Figures 3.61 and 3.62, comparing the values obtained from the neutralization with the respective buffer vehicle control, it can be observed that no neutralization was obtained with any of the constructs. This may have been due to the use of low protein concentrations available, and considering the low affinity observed in the SPR assays.

Another possible reason is related to a possible limitation of this system, in that the conformation and distribution pattern of the protein in the pseudovirus may not reflect the reality of the spike protein.

Because of this possibility, there is interest in performing assays with the SARS-CoV-2 virus in order to obtain further verification of the possible neutralizing capacity of this protein.

Regarding HTH2 and HTH3, with the SPR results for the HTH3 protein, these were results that were sort of expected. Although HTH2 is different from HTH3, the substitutions made to improve their interaction with the RBD were the same in both proteins.

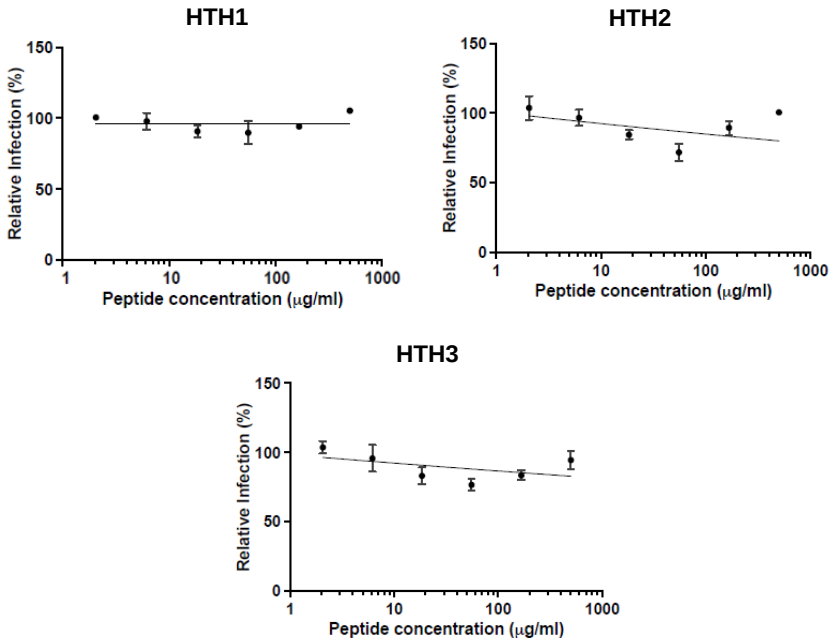

Figure 3.61 – Graphics of the values obtained in the neutralization assays for the three proteins, HTH1, HTH2, HTH3, where the relative infection in percentage of various peptide concentrations tested (μg/mL) is being analysed.

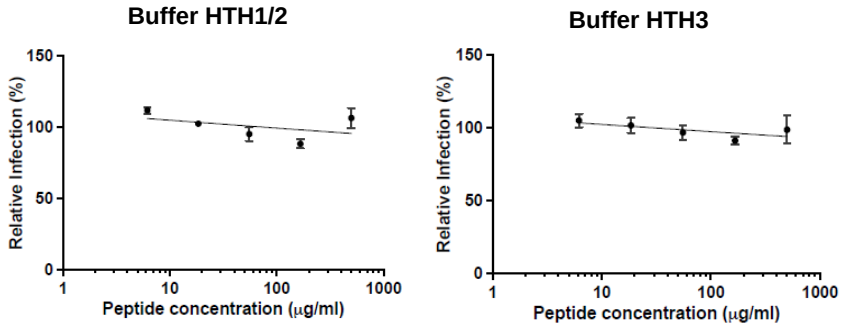

Figure 3.62 – Graphics of the values obtained in the neutralization tests for the buffers where the proteins were, in order to have a control of the effects that the compounds present in the buffers may have.

4. Conclusions and Future Perspectives

The goal of this work was the design and production of antiviral proteins herein called helix-turn-helix, or HTH, that would interact with the RBD of S protein in order to prevent its binding to the human receptor, ACE2, and thus block infection. To this end, several antiviral proteins were computationally designed with the Rosetta program, based on the interactions between ACE2 and the RBD. For their stability and solubility to be improved, some substitutions were made (HTH3) and two disulfide bonds (HTH2) were added. Then, with the three best candidates (HTH1, HTH2 and HTH3) molecular dynamics simulations were performed, both free in solution and in complex with RBD, in order to analyse their stability and behaviour in both environments but also to ascertain, in a preliminary stage, which of the three proteins would have a better interaction with RBD.

Through the results obtained in the simulations of the free proteins in solution it was possible to conclude that all three designs are quite stable, maintaining their secondary structure throughout the simulation time. Thus, these proteins, despite being removed from the context of the native ACE2 protein, are still stable and maintain the structure they were predicted to have.

In the results obtained in the complex simulations it was possible to verify that HTH1, despite having a smaller number of residues interacting with the RBD, have higher contact values than HTH2 and HTH3. Between these two, HTH2 is the one that presents a higher number of contacts, which may be due to the two disulfide bonds that, by stabilizing the structure in the places where they are located, may cause an increase in the mobility of the residues in the remaining structure so that they can perform their function and thus cause this difference with HTH3 in the number of contacts. Furthermore, since HTH1 has a greater number of hydrophobic residues, it showed a higher value of hydrophobic contacts and a lower value of hydrogen bonds made with the RBD residues. This was due to the substitutions performed in the other two designs, where some hydrophobic residues were changed to hydrophilic ones, thus explaining these results. Finally, through the interface area between the HTHs and the RBD it was also possible to observe the effect that these substitutions had, since the hydrophobic interface area of HTH1 is larger than those of the other two HTHs. In the total interface area, the one that presented a higher value was HTH1, and thus it was possible to correlate these results with that of the number of contacts. Since this design shows a higher value of contacts with RBD, the interface area between the two proteins will also be higher.

With this, it was possible to conclude that of the three designs the one that shows having a better interaction with RBD is HTH1, although the differences between the three are not very pronounced.

Regarding the secondary structure of these proteins in the context of the complex, they are maintained, again, throughout the simulation despite there being more turn formations and loss of α-helix, which may be due to the presence of RBD and interactions with it.

In the experimental part of this work, the three HTH proteins were successfully expressed and purified, showing to have the secondary structure that they had been predicted to have in the simulation results. This leads us to conclude that during the entire process of expression and purification these proteins did not lose their secondary structure. In addition to this it was also possible to observe that, despite not being thermodynamically stable proteins, they recover their secondary structure almost entirely when the temperature cools down again after a heating step. Although in HTH2 this loss of structure due to heating was not very pronounced, which may be due to the presence of the bonds.

Through SPR assays it was concluded that HTH1 was the one that exhibited the best interaction with the RBD, although with a high dissociation constant (K_D 5-7 μM), and very different from that of ACE2 (K_D 143 nM). There could be several explanations why the affinity is low: for example, the assays were done with RBD out of the context of the S protein, and since RBD has conformational changes between the open and close position when it is in the context of the S protein, the situation may have

occurred where the conformation adopted in this assay is not the most favourable for the interaction to occur. In addition, other interaction sites may be missing in the context of the whole ACE2 protein that are also important, such as the two glycans that are missing in this region, at sites N90 and N322, which in previous studies have been found to be important for the interaction of the ACE2 protein with the RBD of SARS-CoV-2

Regarding HTH3 not showing any interaction, this may have been due to the fact that the substitutions performed caused some important contacts to be lost in conjunction with the reasons explained above.

Finally, through neutralization assays, it was possible to conclude that the three HTHs did not show neutralization of the virus, which may have been due to the use of low concentrations of the proteins to perform the assay, taking into account the low affinity observed in the SPR assays (or undetected for HTH3). Therefore, it can be concluded that HTH1 is a promising design because although it had a weak interaction with RBD, with larger amounts of protein neutralization may have been achieved.

In the future, further work should be done with this design in order to improve the interaction with the RBD, either through possible different substitutions that might actually improve the interaction.

In addition, testing could be done with the SARS-CoV-2 native virus, instead of the pseudovirus, to further verify the possible neutralizing ability of this protein.

Bibliography

(1) Khan, M.; Adil, S. F.; Alkhathlan, H. Z.; Tahir, M. N.; Saif, S.; Khan, M.; Khan, S. T. COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far. *Molecules* **2021**, *26 (1)*, 1–25.

(2) Piret, J.; Boivin, G. Pandemics Throughout History. *Frontiers in Microbiology* **2021**, *11* (January). https://doi.org/10.3389/fmicb.2020.631736.

(3) Saunders-Hastings, P. R.; Krewski, D. Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. *Pathogens* **2016**, *5* (4). https://doi.org/10.3390/pathogens5040066.

(4) Muralidar, S.; Visaga, S.; Sekaran, S.; Krishnan, U. The Emergence of COVID-19 as a Global Pandemic: Understanding the Epidemiology, Immune Response and Potential Therapeutic Targets of SARS-CoV-2. *Biochimie* **2020**, No. January.

(5) Akande, O.; Akande, T. COVID-19 Pandemic: A Global Health Burden. *Nigerian Postgraduate Medical Journal* **2020**. https://doi.org/10.4103/npmj.npmj.

(6) Tsang, H. F.; Chan, L. W. C.; Cho, W. C. S.; Yu, A. C. S.; Yim, A. K. Y.; Chan, A. K. C.; Ng, L. P. W.; Wong, Y. K. E.; Pei, X. M.; Li, M. J. W.; Wong, S. C. C. An Update on COVID-19 Pandemic: The Epidemiology, Pathogenesis, Prevention and Treatment Strategies. *Expert Review of Anti-Infective Therapy* **2020**, 1–12. https://doi.org/10.1080/14787210.2021.1863146.

(7) Hiscott, J.; Alexandridi, M.; Muscolini, M.; Tassone, E.; Palermo, E.; Soultsioti, M.; Zevini, A. The Global Impact of the Coronavirus Pandemic John. *Cytokine & Growth Factor Reviews* **2020**, No. January.

(8) Liu, Y. C.; Kuo, R. L.; Shih, S. R. COVID-19: The First Documented Coronavirus Pandemic in History. *Biomedical Journal* **2020**, *43* (4), 328–333. https://doi.org/10.1016/j.bj.2020.04.007.

(9) WHO Coronavirus (COVID-19) Dashboard https://covid19.who.int/ (accessed 2021 -09 -13).

(10) Vacinas aprovadas (quadro-resumo) https://www.infarmed.pt/web/infarmed/vacinas-aprovadas (accessed 2021 -10 -29).

(11) Ross, J.; Maier, H. J.; Walker, J. M. *Coronaviruses Methods and Protocols*; 2020.

(12) Wang, M. Y.; Zhao, R.; Gao, L. J.; Gao, X. F.; Wang, D. P.; Cao, J. M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology* **2020**, *10* (November), 1–17. https://doi.org/10.3389/fcimb.2020.587269.

(13) V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nature Reviews Microbiology* **2021**, *19* (3), 155–170. https://doi.org/10.1038/s41579-020-00468-6.

(14) Arya, R.; Kumari, S.; Pandey, B.; Mistry, H.; Bihani, S. C.; Das, A.; Prashar, V.; Gupta, G. D.; Panicker, L.; Kumar, M. Structural Insights into SARS-CoV-2 Proteins. *Journal of Molecular Biology* **2020**, No. January.

(15) Herrera, N. G.; Morano, N. C.; Celikgil, A.; Georgiev, G. I.; Malonis, R. J.; Lee, J. H.; Tong, K.; Vergnolle, O.; Massimi, A. B.; Yen, L. Y.; Noble, A. J.; Kopylov, M.; Bonanno, J. B.; Garrett-Thomson, S. C.; Hayes, D. B.; Bortz, R. H.; Wirchnianski, A. S.; Florez, C.; Laudermilch, E.; Haslwanter, D.; Fels, J. M.; Dieterle, M. E.; Jangra, R. K.; Barnhill, J.; Mengotto, A.; Kimmel, D.; Daily, J. P.; Pirofski, L. A.; Chandran, K.; Brenowitz, M.; Garforth, S. J.; Eng, E. T.; Lai, J. R.;

Almo, S. C. Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis. *ACS Omega* **2021**, *6* (1), 85–102. https://doi.org/10.1021/acsomega.0c03512.

(16) Knowlton, K. U. Pathogenesis of SARS-CoV-2 Induced Cardiac Injury from the Perspective of the Virus. *Journal of Molecular and Cellular Cardiology* **2019**, *147* (July), 12–17. https://doi.org/10.1016/j.yjmcc.2020.08.002.

(17) Hu, T.; Liu, Y.; Zhao, M.; Zhuang, Q.; Xu, L.; He, Q. A Comparison of COVID-19, SARS and MERS. *PeerJ* **2020**, *8*, 1–30. https://doi.org/10.7717/peerj.9725.

(18) Zhu, Z.; Lian, X.; Su, X.; Wu, W.; Marraro, G. A.; Zeng, Y. From SARS and MERS to COVID-19: A Brief Summary and Comparison of Severe Acute Respiratory Infections Caused by Three Highly Pathogenic Human Coronaviruses. *Respiratory Research* **2020**, *21* (1), 1–14. https://doi.org/10.1186/s12931-020-01479-w.

(19) Guadarrama-Ortiz, P.; Choreño-Parra, J. A.; Sánchez-Martínez, C. M.; Pacheco-Sánchez, F. J.; Rodríguez-Nava, A. I.; García-Quintero, G. Neurological Aspects of SARS-CoV-2 Infection: Mechanisms and Manifestations. *Frontiers in Neurology* **2020**, *11* (September), 1–14. https://doi.org/10.3389/fneur.2020.01039.

(20) Shah, V. K.; Firmal, P.; Alam, A.; Ganguly, D.; Chattopadhyay, S. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Frontiers in Immunology* **2020**, *11* (August), 1–17. https://doi.org/10.3389/fimmu.2020.01949.

(21) Lauer, S. A.; Grantz, K. H.; Bi, Q.; Jones, F. K.; Zheng, Q.; Meredith, H. R.; Azman, A. S.; Reich, N. G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (CoVID-19) from Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine* **2020**, *172* (9), 577–582. https://doi.org/10.7326/M20-0504.

(22) Baloch, S.; Baloch, M. A.; Zheng, T.; Pei, X. The Coronavirus Disease 2019 (COVID-19) Pandemic. *Tohoku Journal of Experimental Medicine* **2020**, *250* (4), 271–278. https://doi.org/10.1620/tjem.250.271.

(23) Baj, J.; Karakuła-Juchnowicz, H.; Teresiński, G.; Buszewicz, G.; Ciesielka, M.; Sitarz, E.; Forma, A.; Karakuła, K.; Flieger, W.; Portincasa, P.; Maciejewski, R. COVID-19: Specific and Non-Specific Clinical Manifestations and Symptoms: The Current State of Knowledge. *Journal of Clinical Medicine* **2020**, *9* (6), 1–22. https://doi.org/10.3390/jcm9061753.

(24) Alimohamadi, Y.; Sepandi, M.; Taghdir, M.; Hosamirudsari, H. Determine the Most Common Clinical Symptoms in COVID-19 Patients: A Systematic Review and Meta-Analysis. *Journal of Preventive Medicine and Hygiene* **2020**, *61* (3), E304–E312. https://doi.org/10.15167/2421-4248/jpmh2020.61.3.1530.

(25) Amawi, H.; Abu Deiab, G. I.; A Aljabali, A. A.; Dua, K.; Tambuwala, M. M. COVID-19 Pandemic: An Overview of Epidemiology, Pathogenesis, Diagnostics and Potential Vaccines and Therapeutics. *Therapeutic Delivery* **2020**, *11* (4), 245–268. https://doi.org/10.4155/tde-2020-0035.

(26) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **2021**, *19* (3), 141–154. https://doi.org/10.1038/s41579-020-00459-7.

(27) Wiersinga, W. J.; Rhodes, A.; Cheng, A. C.; Peacock, S. J.; Prescott, H. C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA - Journal of the American Medical Association* **2020**, *324* (8), 782–793. https://doi.org/10.1001/jama.2020.12839.

(28) How testing works for COVID-19 https://www.tga.gov.au/how-testing-works-covid-19 (accessed 2021 -10 -14).

(29) Kevadiya, B. D.; Machhi, J.; Herskovitz, J.; Oleynikov, M. D.; Blomberg, W. R.; Bajwa, N.; Soni, D.; Das, S.; Hasan, M.; Patel, M.; Senan, A. M.; Gorantla, S.; McMillan, J. E.; Edagwa, B.; Eisenberg, R.; Gurumurthy, C. B.; Reid, S. P. M.; Punyadeera, C.; Chang, L.; Gendelman, H. E. Diagnostics for SARS-CoV-2 Infections. *Nature Materials* **2021**, *20* (5), 593–605. https://doi.org/10.1038/s41563-020-00906-z.

(30) Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; Zhao, Y.; Li, Y.; Wang, X.; Peng, Z. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA - Journal of the American Medical Association* **2020**, *323* (11), 1061–1069. https://doi.org/10.1001/jama.2020.1585.

(31) Ghodake, G. S.; Shinde, S. K.; Kadam, A. A.; Saratale, R. G.; Saratale, G. D.; Syed, A.; Elgorban, A. M.; Marraiki, N.; Kim, D. Y. Biological Characteristics and Biomarkers of Novel SARS-CoV-2 Facilitated Rapid Development and Implementation of Diagnostic Tools and Surveillance Measures. *Biosensors and Bioelectronics* **2021**, *177* (December 2020), 112969. https://doi.org/10.1016/j.bios.2021.112969.

(32) Udugama, B.; Kadhiresan, P.; Kozlowski, H. N.; Malekjahani, A.; Osborne, M.; Li, V. Y. C.; Chen, H.; Mubareka, S.; Gubbay, J. B.; Chan, W. C. W. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS nano* **2020**, *14* (4), 3822–3835. https://doi.org/10.1021/acsnano.0c02624.

(33) Rapid tests for covid-19: what are the differences? https://www.hospitaldaluz.pt/en/health-guide/health-and-wellness/rapid-tests-covid-19-what-are-differences (accessed 2021 -10 -18).

(34) Burmer, G.; Burmer, M.; Pabuwal, V. SARS-CoV-2 and COVID-19 Pathogenesis : A Review Biology and Life Cycle of Coronaviruses. **2020**, *2019*, 1–34.

(35) Pokhrel, P.; Hu, C.; Mao, H. Detecting the Coronavirus (CoVID-19). *ACS Sensors* **2020**, *5* (8), 2283–2297. https://doi.org/10.1021/ACSSENSORS.0C01153.

(36) Alanagreh, L.; Alzoughool, F.; Atoum, M. The Human Coronavirus Disease Covid-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens* **2020**, *9* (5). https://doi.org/10.3390/pathogens9050331.

(37) Astuti, I. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An Overview of Viral Structure and Host Response Indwiani. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **2020**, *14* (4).

(38) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 Entry into Cells. *Nature Reviews Molecular Cell Biology*. Nature Research January 1, 2022, pp 3–20. https://doi.org/10.1038/s41580-021-00418-x.

(39) Duan, L.; Zheng, Q.; Zhang, H.; Niu, Y.; Lou, Y.; Wang, H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Frontiers in Immunology* **2020**, *11* (October), 1–12. https://doi.org/10.3389/fimmu.2020.576622.

(40) Duan, L.; Zheng, Q.; Zhang, H.; Niu, Y.; Lou, Y.; Wang, H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Frontiers in Immunology*. Frontiers Media S.A. October 7, 2020. https://doi.org/10.3389/fimmu.2020.576622.

(41) Lebeau, G.; Vagner, D.; Frumence, É.; Ah-Pine, F.; Guillot, X.; Nobécourt, E.; Raffray, L.; Gasque, P. Deciphering SARS-CoV-2 Virologic and Immunologic Features. *International Journal of Molecular Sciences* **2020**, *21* (16), 1–40. https://doi.org/10.3390/ijms21165932.

(42) Xia, X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses* **2021**, *13* (1), 1–16. https://doi.org/10.3390/v13010109.

(43) Zhang, J.; Xiao, T.; Cai, Y.; Chen, B. Structure of SARS-CoV-2 Spike Protein. *Current Opinion in Virology*. Elsevier B.V. October 1, 2021, pp 173–182. https://doi.org/10.1016/j.coviro.2021.08.010.

(44) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; Mclellan, J. S. *Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation*; 2019.

(45) Peng, R.; Wu, L. A.; Wang, Q.; Qi, J.; Gao, G. F. Cell Entry by SARS-CoV-2. *Trends in Biochemical Sciences*. Elsevier Ltd October 1, 2021, pp 848–860. https://doi.org/10.1016/j.tibs.2021.06.001.

(46) Gong, Y.; Qin, S.; Dai, L.; Tian, Z. The Glycosylation in SARS-CoV-2 and Its Receptor ACE2. *Signal Transduction and Targeted Therapy*. Springer Nature December 1, 2021. https://doi.org/10.1038/s41392-021-00809-8.

(47) Gorgun, D.; Lihan, M.; Kapoor, K.; Tajkhorshid, E. Binding Mode of SARS-CoV2 Fusion Peptide to Human Cellular Membrane. https://doi.org/10.1101/2020.10.27.357350.

(48) Borkotoky, S.; Dey, D.; Banerjee, M. Computational Insight into the Mechanism of SARS-CoV-2 Membrane Fusion. *Journal of Chemical Information and Modeling* **2021**, *61* (1), 423–431. https://doi.org/10.1021/acs.jcim.0c01231.

(49) Birtles, D.; Lee, J. Identifying Distinct Structural Features of the SARS-CoV-2 Spike Protein Fusion Domain Essential for Membrane Interaction. *Biochemistry* **2021**, *60* (40), 2978–2986. https://doi.org/10.1021/acs.biochem.1c00543.

(50) Lai, A. L.; Freed, J. H. SARS-CoV-2 Fusion Peptide Has a Greater Membrane Perturbating Effect than SARS-CoV with Highly Specific Dependence on Ca2+. *Journal of Molecular Biology* **2021**, *433* (10). https://doi.org/10.1016/j.jmb.2021.166946.

(51) Basso, L. G. M.; Zeraik, A. E.; Felizatti, A. P.; Costa-Filho, A. J. Membranotropic and Biological Activities of the Membrane Fusion Peptides from SARS-CoV Spike Glycoprotein: The Importance of the Complete Internal Fusion Peptide Domain. *Biochimica et Biophysica Acta - Biomembranes* **2021**, *1863* (11). https://doi.org/10.1016/j.bbamem.2021.183697.

(52) Khelashvili, G.; Plante, A.; Doktorova, M.; Weinstein, H. Ca2+-Dependent Mechanism of Membrane Insertion and Destabilization by the SARS-CoV-2 Fusion Peptide. *Biophysical Journal* **2021**, *120* (6), 1105–1119. https://doi.org/10.1016/j.bpj.2021.02.023.

(53) Das, G.; Ghosh, S.; Garg, S.; Ghosh, S.; Jana, A.; Samat, R.; Mukherjee, N.; Roy, R.; Ghosh, S. An Overview of Key Potential Therapeutic Strategies for Combat in the COVID-19 Battle. *RSC Advances* **2020**, *10* (47), 28243–28266. https://doi.org/10.1039/d0ra05434h.

(54) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(55) Jawad, B.; Adhikari, P.; Podgornik, R.; Ching, W. Y. Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. *Journal of Chemical Information and Modeling* **2021**, *61* (9), 4425–4441. https://doi.org/10.1021/acs.jcim.1c00560.

(56) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181* (4), 894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045.

(57) Cerutti, G.; Guo, Y.; Zhou, T.; Gorman, J.; Lee, M.; Rapp, M.; Reddem, E. R.; Yu, J.; Bahna, F.; Bimela, J.; Huang, Y.; Katsamba, P. S.; Liu, L.; Nair, M. S.; Rawi, R.; Olia, A. S.; Wang, P.;

Zhang, B.; Chuang, G. Y.; Ho, D. D.; Sheng, Z.; Kwong, P. D.; Shapiro, L. Potent SARS-CoV-2 Neutralizing Antibodies Directed against Spike N-Terminal Domain Target a Single Supersite. *Cell Host and Microbe* **2021**, *29* (5), 819-833.e7. https://doi.org/10.1016/j.chom.2021.03.005.

(58) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; Yu, C.; Li, J.; Zhou, Q.; Chen, W. *A Neutralizing Human Antibody Binds to the N-Terminal Domain of the Spike Protein of SARS-CoV-2*.

(59) McCallum, M.; de Marco, A.; Lempp, F. A.; Tortorici, M. A.; Pinto, D.; Walls, A. C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; Zepeda, S.; di Iulio, J.; Bowen, J. E.; Montiel-Ruiz, M.; Zhou, J.; Rosen, L. E.; Bianchi, S.; Guarino, B.; Fregni, C. S.; Abdelnabi, R.; Foo, S. Y. C.; Rothlauf, P. W.; Bloyet, L. M.; Benigni, F.; Cameroni, E.; Neyts, J.; Riva, A.; Snell, G.; Telenti, A.; Whelan, S. P. J.; Virgin, H. W.; Corti, D.; Pizzuto, M. S.; Veesler, D. N-Terminal Domain Antigenic Mapping Reveals a Site of Vulnerability for SARS-CoV-2. *Cell* **2021**, *184* (9), 2332-2347.e16. https://doi.org/10.1016/j.cell.2021.03.028.

(60) Otto, S. P.; Day, T.; Arino, J.; Colijn, C.; Dushoff, J.; Li, M.; Mechai, S.; van Domselaar, G.; Wu, J.; Earn, D. J. D.; Ogden, N. H. The Origins and Potential Future of SARS-CoV-2 Variants of Concern in the Evolving COVID-19 Pandemic. *Current Biology*. Cell Press July 26, 2021, pp R918–R929. https://doi.org/10.1016/j.cub.2021.06.049.

(61) Tracking SARS-CoV-2 variants https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed 2022 -01 -29).

(62) Sars-CoV-2 circulating variants https://viralzone.expasy.org/9556 (accessed 2022 -01 -29).

(63) Cosar, B.; Karagulleoglu, Z. Y.; Unal, S.; Ince, A. T.; Uncuoglu, D. B.; Tuncer, G.; Kilinc, B. R.; Ozkan, Y. E.; Ozkoc, H. C.; Demir, I. N.; Eker, A.; Karagoz, F.; Simsek, S. Y.; Yasar, B.; Pala, M.; Demir, A.; Atak, I. N.; Mendi, A. H.; Bengi, V. U.; Cengiz Seval, G.; Gunes Altuntas, E.; Kilic, P.; Demir-Dora, D. SARS-CoV-2 Mutations and Their Viral Variants. *Cytokine and Growth Factor Reviews*. Elsevier Ltd 2021. https://doi.org/10.1016/j.cytogfr.2021.06.001.

(64) Bourgonje, A. R.; Abdulle, A. E.; Timens, W.; Hillebrands, J. L.; Navis, G. J.; Gordijn, S. J.; Bolling, M. C.; Dijkstra, G.; Voors, A. A.; Osterhaus, A. D. M. E.; van der Voort, P. H. J.; Mulder, D. J.; van Goor, H. Angiotensin-Converting Enzyme 2 (ACE2), SARS-CoV-2 and the Pathophysiology of Coronavirus Disease 2019 (COVID-19). *Journal of Pathology* **2020**, *251* (3), 228–248. https://doi.org/10.1002/path.5471.

(65) Gheblawi, M.; Wang, K.; Viveiros, A.; Nguyen, Q.; Zhong, J. C.; Turner, A. J.; Raizada, M. K.; Grant, M. B.; Oudit, G. Y. Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th Anniversary of the Discovery of ACE2. *Circulation Research* **2020**, 1456–1474. https://doi.org/10.1161/CIRCRESAHA.120.317015.

(66) Samavati, L.; Uhal, B. D. ACE2, Much More Than Just a Receptor for SARS-COV-2. *Frontiers in Cellular and Infection Microbiology* **2020**, *10* (June), 1–9. https://doi.org/10.3389/fcimb.2020.00317.

(67) Hamming, I.; Cooper, M.; Haagmans, B.; Hooper, N.; Korstanje, R.; Osterhaus, A.; Timens, W.; Turner, A.; Navis, G.; Goor, H. van. The Emerging Role of ACE2 in Physiology and Disease. *Wiley InterScience* **2008**, 231–241. https://doi.org/10.1002/path.

(68) Ni, W.; Yang, X.; Yang, D.; Bao, J.; Li, R.; Xiao, Y.; Hou, C.; Wang, H.; Liu, J.; Yang, D.; Xu, Y.; Cao, Z.; Gao, Z. Role of Angiotensin-Converting Enzyme 2 (ACE2) in COVID-19. *Critical Care* **2020**, *24* (1), 1–10. https://doi.org/10.1186/s13054-020-03120-0.

(69) Towler, P.; Staker, B.; Prasad, S. G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N. A.; Patane, M. A.; Pantoliano, M. W. ACE2 X-Ray Structures Reveal a Large Hinge-Bending Motion Important for Inhibitor Binding and Catalysis. *Journal of Biological Chemistry* **2004**, *279* (17), 17996–18007. https://doi.org/10.1074/jbc.M311191200.

(70) Lubbe, L.; Cozier, G. E.; Oosthuizen, D.; Ravi Acharya, K.; Sturrock, E. D. ACE2 and ACE: Structure-Based Insights into Mechanism, Regulation and Receptor Recognition by SARS-CoV. *Clinical Science*. Portland Press Ltd November 1, 2020, pp 2851–2871. https://doi.org/10.1042/CS20200899.

(71) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Zhou, Q. Structure of Dimeric Full-Length Human ACE2 in Complex with B0AT1. *bioRxiv* **2020**. https://doi.org/10.1101/2020.02.17.951848.

(72) Gross, L. Z. F.; Sacerdoti, M.; Piiper, A.; Zeuzem, S.; Leroux, A. E.; Biondi, R. M. ACE2, the Receptor That Enables Infection by SARS-CoV-2: Biochemistry, Structure, Allostery and Evaluation of the Potential Development of ACE2 Modulators. *ChemMedChem*. John Wiley and Sons Ltd September 16, 2020, pp 1682–1690. https://doi.org/10.1002/cmdc.202000368.

(73) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. *Biophysical Journal* **2021**, *120* (6), 1072–1084. https://doi.org/10.1016/j.bpj.2020.10.036.

(74) Shajahan, A.; Archer-Hartmann, S.; Supekar, N. T.; Gleinich, A. S.; Heiss, C.; Azadi, P. Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* **2021**, *31* (4), 410–424. https://doi.org/10.1093/glycob/cwaa101.

(75) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. *Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2*; 2020; Vol. 367.

(76) Acharya, A.; Lynch, D. L.; Pavlova, A.; Pang, Y. T.; Gumbart, J. C. ACE2 Glycans Preferentially Interact with SARS-CoV-2 over SARS-CoV. *Chemical Communications* **2021**, *57* (48), 5949–5952. https://doi.org/10.1039/d1cc02305e.

(77) Allen, J. D.; Watanabe, Y.; Chawla, H.; Newby, M. L.; Crispin, M. Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. *Journal of Molecular Biology* **2021**, *433* (4). https://doi.org/10.1016/j.jmb.2020.166762.

(78) Mehdipour, A. R.; Hummer, G. Dual Nature of Human ACE2 Glycosylation in Binding to SARS-CoV-2 Spike. **2021**, *118*, 2022. https://doi.org/10.1073/pnas.2100425118/-/DCSupplemental.

(79) Hadi-Alijanvand, H.; Rouhani, M. Studying the Effects of ACE2 Mutations on the Stability, Dynamics, and Dissociation Process of SARS-CoV-2 S1/HACE2 Complexes. *Journal of Proteome Research* **2020**, *19* (11), 4609–4623. https://doi.org/10.1021/acs.jproteome.0c00348.

(80) Procko, E. The Sequence of Human ACE2 Is Suboptimal for Binding the S Spike Protein of SARS Coronavirus 2. *bioRxiv : the preprint server for biology* **2020**. https://doi.org/10.1101/2020.03.16.994236.

(81) Zhang, H.; Penninger, J. M.; Li, Y.; Zhong, N.; Slutsky, A. S. Angiotensin-Converting Enzyme 2 (ACE2) as a SARS-CoV-2 Receptor: Molecular Mechanisms and Potential Therapeutic Target. *Intensive Care Medicine* **2020**, *46* (4), 586–590. https://doi.org/10.1007/s00134-020-05985-9.

(82) Sivasankarapillai, V. S.; Pillai, A. M.; Rahdar, A.; Sobha, A. P.; Das, S. S.; Mitropoulos, A. C.; Mokarrar, M. H.; Kyzas, G. Z. On Facing the SARS-Cov-2 (COVID-19) with Combination of Nanomaterials and Medicine: Possible Strategies and First Challenges. *Nanomaterials* **2020**, *10* (5), 1–23. https://doi.org/10.3390/nano10050852.

(83) Vanpatten, S.; He, M.; Altiti, A.; F Cheng, K.; Ghanem, M. H.; Al-Abed, Y. Evidence Supporting the Use of Peptides and Peptidomimetics as Potential SARS-CoV-2 (COVID-19) Therapeutics. *Future Medicinal Chemistry* **2020**, *12* (18), 1647–1656. https://doi.org/10.4155/fmc-2020-0180.

(84) Sadremomtaz, A.; Al-Dahmani, Z. M.; Ruiz-Moreno, A. J.; Monti, A.; Wang, C.; Azad, T.; Bell, J. C.; Doti, N.; Velasco-Velázquez, M. A.; de Jong, D.; de Jonge, J.; Smit, J.; Dömling, A.; van Goor, H.; Groves, M. R. Synthetic Peptides That Antagonize the Angiotensin-Converting Enzyme-2 (ACE-2) Interaction with SARS-CoV-2 Receptor Binding Spike Protein. *Journal of Medicinal Chemistry* **2021**. https://doi.org/10.1021/acs.jmedchem.1c00477.

(85) Hunt, A. C.; Case, J. B.; Park, Y.-J.; Cao, L.; Wu, K.; Walls, A. C.; Liu, Z.; Bowen, J. E.; Yeh, H.-W.; Saini, S.; Helms, L.; Zhao, Y. T.; Hsiang, T.-Y.; Starr, T. N.; Goreshnik, I.; Kozodoy, L.; Carter, L.; Ravichandran, R.; Green, L. B.; Matochko, W. L.; Thomson, C. A.; Vögeli, B.; Krüger-Gericke, A.; VanBlargan, L. A.; Chen, R. E.; Ying, B.; Bailey, A. L.; Kafai, N. M.; Boyken, S.; Ljubetič, A.; Edman, N.; Ueda, G.; Chow, C.; Addetia, A.; Panpradist, N.; Gale, M.; Freedman, B. S.; Lutz, B. R.; Bloom, J. D.; Ruohola-Baker, H.; Whelan, S. P. J.; Stewart, L.; Diamond, M. S.; Veesler, D.; Jewett, M. C.; Baker, D. Multivalent Designed Proteins Protect against SARS-CoV-2 Variants of Concern. *bioRxiv : the preprint server for biology* **2021**. https://doi.org/10.1101/2021.07.07.451375.

(86) Cao, L.; Goreshnik, I.; Coventry, B.; Case, J. B.; Miller, L.; Kozodoy, L.; Chen, R. E.; Carter, L.; Walls, A. C.; Park, Y.-J.; Strauch, E.-M.; Stewart, L.; Diamond, M. S.; Veesler, D.; Baker, D. *De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors*.

(87) Jia, H.; Neptune, E.; Cui, H. Targeting ACE2 for COVID-19 Therapy: Opportunities and Challenges. *American Journal of Respiratory Cell and Molecular Biology*. American Thoracic Society April 1, 2021, pp 416–425. https://doi.org/10.1165/rcmb.2020-0322PS.

(88) Ahn, D. G.; Shin, H. J.; Kim, M. H.; Lee, S.; Kim, H. S.; Myoung, J.; Kim, B. T.; Kim, S. J. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). *Journal of Microbiology and Biotechnology* **2020**, *30* (3), 313–324. https://doi.org/10.4014/jmb.2003.03011.

(89) Fischer, W.; Eron, J. J.; Holman, W.; Cohen, M. S.; Fang, L.; Szewczyk, L. J.; Sheahan, T. P.; Baric, R.; Mollan, K. R.; Wolfe, C. R.; Duke, E. R.; Azizad, M. M.; Borroto-Esoda, K.; Wohl, D. A.; Loftis, A. J.; Alabanza, P.; Lipansky, F.; Painter, W. P. Molnupiravir, an Oral Antiviral Treatment for COVID-19. *medRxiv : the preprint server for health sciences* **2021**. https://doi.org/10.1101/2021.06.17.21258639.

(90) Fischer Ii, W. A.; Eron, J. J.; Holman, W.; Cohen, M. S.; Fang, L.; Szewczyk, L. J.; Sheahan, T. P.; Baric, R.; Mollan, K. R.; Wolfe, C. R.; Duke, E. R.; Azizad, M. M.; Borroto-Esoda, K.; Wohl, D. A.; Coombs, R. W.; Loftis, A. J.; Alabanza, P.; Lipansky, F.; Painter, W. P. *A Phase 2a Clinical Trial of Molnupiravir in Patients with COVID-19 Shows Accelerated SARS-CoV-2 RNA Clearance and Elimination of Infectious Virus*; 2022; Vol. 14.

(91) Lee, C. C.; Hsieh, C. C.; Ko, W. C. Molnupiravir—A Novel Oral Anti-SARS-CoV-2 Agent. *Antibiotics* **2021**, *10* (11). https://doi.org/10.3390/antibiotics10111294.

(92) Jayk Bernal, A.; Gomes da Silva, M. M.; Musungaie, D. B.; Kovalchuk, E.; Gonzalez, A.; Delos Reyes, V.; Martín-Quirós, A.; Caraco, Y.; Williams-Diaz, A.; Brown, M. L.; Du, J.; Pedley, A.; Assaid, C.; Strizki, J.; Grobler, J. A.; Shamsuddin, H. H.; Tipping, R.; Wan, H.; Paschke, A.; Butterton, J. R.; Johnson, M. G.; de Anda, C. Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *New England Journal of Medicine* **2021**. https://doi.org/10.1056/nejmoa2116044.

(93) Imran, M.; Kumar Arora, M.; Asdaq, S. M. B.; Khan, S. A.; Alaqel, S. I.; Alshammari, M. K.; Alshehri, M. M.; Alshrari, A. S.; Mateq Ali, A.; Al-Shammeri, A. M.; Alhazmi, B. D.; Harshan, A. A.; Alam, M. T.; Abida. Discovery, Development, and Patent Trends on Molnupiravir: A Prospective Oral Treatment for Covid-19. *Molecules*. MDPI October 1, 2021. https://doi.org/10.3390/molecules26195795.

(94) Interim recommendations for use of the Pfizer–BioNTech COVID-19 vaccine https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccines-SAGE_recommendation-BNT162b2-2021.1 (accessed 2021 -10 -29).

(95) Interim recommendations for use of the Moderna mRNA-1273 vaccine against COVID-19 https://www.who.int/publications/i/item/interim-recommendations-for-use-of-the-moderna-mrna-1273-vaccine-against-covid-19 (accessed 2021 -10 -29).

(96) The Janssen Ad26.COV2.S COVID-19 vaccine: What you need to know https://www.who.int/news-room/feature-stories/detail/the-j-j-covid-19-vaccine-what-you-need-to-know (accessed 2021 -10 -29).

(97) The Oxford/AstraZeneca COVID-19 vaccine: what you need to know https://www.who.int/pt/news-room/feature-stories/detail/the-oxford-astrazeneca-covid-19-vaccine-what-you-need-to-know (accessed 2021 -10 -29).

(98) What are viral vector vaccines and how do they work? https://www.astrazeneca.com/what-science-can-do/topics/covid-19/covid-19-what-are-viral-vector-vaccines-and-how-do-they-work.html (accessed 2021 -10 -29).

(99) Antman, S. S.; Sirovich, L.; Marsden, J. E. *Molecular Modeling and Simulation: An Interdisciplinary Guide*; 2009; Vol. 8.

(100) Leach, A. R. Molecular Modelling: Principles and Applications. 2001.

(101) Heinecke, A.; Eckhardt, W.; Horsch, M.; Bungartz, H.-J. Molecular Dynamics Simulation. In *MDPI - Entropy*; 2014.

(102) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie - International Edition* **2006**, *45* (25), 4064–4092. https://doi.org/10.1002/anie.200502655.

(103) Vanommeslaeghe, K.; Guvench, O.; Jr., A. D. M. Molecular Mechanics.

(104) Doltsinis, N. L. Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum–Classical Approaches. *NIC Series* **2006**, *31*, 389–409.

(105) van Gunsteren, W. F.; Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English* **1990**, *29* (9), 992–1023. https://doi.org/10.1002/anie.199009921.

(106) Chang, C. E. A.; Huang, Y. M. M.; Mueller, L. J.; You, W. Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools. *Catalysts* **2016**, *6* (6). https://doi.org/10.3390/catal6060082.

(107) Hinchliffe, A. *Molecular Modelling for Beginners*; 2008.

(108) GROMACS development team. GROMACS Documentation. *GROMACS Documentation Release 2019* **2019**.

(109) Force fields in GROMACS https://manual.gromacs.org/documentation/2018/user-guide/force-fields.html (accessed 2021 -11 -18).

(110) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(111) Becker, O. M.; Mackerell, A. D.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics*.

(112) Simms, J.; Booth, P. J. Membrane Proteins by Accident or Design. *Current Opinion in Chemical Biology* **2013**, *17* (6), 976–981. https://doi.org/10.1016/j.cbpa.2013.10.005.

(113) Lippow, S. M.; Tidor, B. Progress in Computational Protein Design. *Current Opinion in Biotechnology* **2007**, *18* (4), 305–311. https://doi.org/10.1016/j.copbio.2007.04.009.

(114) Broom, A.; Trainor, K.; MacKenzie, D. W. S.; Meiering, E. M. Using Natural Sequences and Modularity to Design Common and Novel Protein Topologies. *Current Opinion in Structural Biology* **2016**, *38*, 26–36. https://doi.org/10.1016/j.sbi.2016.05.007.

(115) Saven, J. G. Computational Protein Design. *Protein Engineering Handbook, Volume 1 & Volume 2* **2011**, *1*, 325–342. https://doi.org/10.1002/9783527634026.ch12.

(116) Schrödinger https://www.schrodinger.com/ (accessed 2021 -11 -20).

(117) MOLECULAR OPERATING ENVIRONMENT https://www.chemcomp.com/Products.htm (accessed 2021 -11 -20).

(118) Rosetta Commons https://www.rosettacommons.org/software (accessed 2021 -11 -20).

(119) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E. M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; Meiler, J.; Baker, D. Rosettascripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS ONE* **2011**, *6* (6), 1–10. https://doi.org/10.1371/journal.pone.0020161.

(120) Alford, R. F.; Leaver-fay, A.; Jeliazkov, J. R.; Meara, M. J. O.; Dimaio, F. P.; Park, H.; Shapovalov, M. v; Renfrew, P. D.; Mulligan, K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. **2018**, *13* (6), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125.The.

(121) Simons, K. T.; Bonneau, R.; Baker, D. Ab Initio Protein Structure Prediction OfCASP III Targets Using ROSETTA. *Proteins: Structure, Function and Bioinformatics* **1999**, *37* (S3), 171–176.

(122) Kim, D. E.; Chivian, D.; Baker, D. Protein Structure Prediction and Analysis Using the Robetta Server. *Nucleic Acids Research* **2004**, *32* (WEB SERVER ISS.), 526–531. https://doi.org/10.1093/nar/gkh468.

(123) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(124) Hill, T. L. *An Introduction To Statistical Thermodynamics*; 2010. https://doi.org/10.1142/9789812385147.

(125) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *Journal of Chemical Physics* **2007**, *126* (1). https://doi.org/10.1063/1.2408420.

(126) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.

(127) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(128) Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Computer Physics Communications* **2013**, *184* (12), 2641–2650. https://doi.org/10.1016/j.cpc.2013.06.003.

(129) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(130) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. **2009**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.Using.

(131) Greenfield, N. J. Using Circular Dichroism Collected as a Function of Temperature to Determine the Thermodynamics of Protein Unfolding and Binding Interactions. **2009**, *1* (6), 2527–2535. https://doi.org/10.1038/nprot.2006.204.Using.

(132) Kelly, S. M.; Jess, T. J.; Price, N. C. How to Study Proteins by Circular Dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2005**, *1751* (2), 119–139. https://doi.org/10.1016/j.bbapap.2005.06.005.

(133) Bakhtiar, R. Surface Plasmon Resonance Spectroscopy: A Versatile Technique in a Biochemist's Toolbox. *Journal of Chemical Education* **2013**, *90* (2), 203–209. https://doi.org/10.1021/ed200549g.

(134) Zeng, S.; Baillargeat, D.; Ho, H. P.; Yong, K. T. Nanomaterials Enhanced Surface Plasmon Resonance for Biological and Chemical Sensing Applications. *Chemical Society Reviews* **2014**, *43* (10), 3426–3452. https://doi.org/10.1039/c3cs60479a.

(135) Wu, B.; Mathews, N.; Sum, T. C. Surface Plasmon Resonance. *SpringerBriefs in Applied Sciences and Technology* **2017**, *126* (9789811020193), 25–31. https://doi.org/10.1007/978-981-10-2021-6_2.

(136) Tang, Y.; Zeng, X.; Liang, J. Surface Plasmon Resonance: An Introduction to a Surface Spectroscopy Technique. https://doi.org/10.1021/ed100186y.

(137) Zhu, X.; Gao, T. *Spectrometry*; Elsevier Inc., 2018. https://doi.org/10.1016/B978-0-12-815053-5.00010-6.

(138) Wen, J.; Lord, H.; Knutson, N.; Wikström, M. Nano Differential Scanning Fluorimetry for Comparability Studies of Therapeutic Proteins. *Analytical Biochemistry* **2020**, *593* (October 2019), 113581. https://doi.org/10.1016/j.ab.2020.113581.

(139) Chattopadhyay, G.; Varadarajan, R. Facile Measurement of Protein Stability and Folding Kinetics Using a Nano Differential Scanning Fluorimeter. *Protein Science* **2019**, *28* (6), 1127–1134. https://doi.org/10.1002/pro.3622.

(140) Prometheus ™ Series Product Information.

(141) Nie, J.; Li, Q.; Wu, J.; Zhao, C.; Hao, H.; Liu, H.; Zhang, L.; Nie, L.; Qin, H.; Wang, M.; Lu, Q.; Li, X.; Sun, Q.; Liu, J.; Fan, C.; Huang, W.; Xu, M.; Wang, Y. Establishment and Validation of a

Pseudovirus Neutralization Assay for SARS-CoV-2. *Emerging Microbes and Infections* **2020**, *9* (1), 680–686. https://doi.org/10.1080/22221751.2020.1743767.

(142) Tolah, A. M. K.; Sohrab, S. S.; Tolah, K. M. K.; Hassan, A. M.; El-Kafrawy, S. A.; Azhar, E. I. Evaluation of a Pseudovirus Neutralization Assay for Sars-Cov-2 and Correlation with Live Virus-Based Micro Neutralization Assay. *Diagnostics* **2021**, *11* (6). https://doi.org/10.3390/diagnostics11060994.

(143) Chen, M.; Zhang, X. E. Construction and Applications of Sars-Cov-2 Pseudoviruses: A Mini Review. *International Journal of Biological Sciences* **2021**, *17* (6), 1574–1580. https://doi.org/10.7150/ijbs.59184.

(144) Li, Q.; Liu, Q.; Huang, W.; Li, X.; Wang, Y. Current Status on the Development of Pseudoviruses for Enveloped Viruses. *Reviews in Medical Virology* **2018**, *28* (1), 1–10. https://doi.org/10.1002/rmv.1963.

(145) Montoliu-Gaya, L.; Martínez, J. C.; Villegas, S. Understanding the Contribution of Disulfide Bridges to the Folding and Misfolding of an Anti-Aβ ScFv. *Protein Science* **2017**, *26* (6), 1138–1149. https://doi.org/10.1002/pro.3164.

(146) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

(147) Valério, M., Borges-Araújo, L., Melo, M. N., Lousa, D., & Soares, C. M. (n.d.). SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility. https://doi.org/10.1101/2021.11.30.470470.

Supplementary Information

Annex 1 – Solutions

Table A1.1 – Specifications concerning the ingredients, the quantity in grams and the brand of each of the ingredients used to make 800 mL of LB medium.

Luria-Bertani (LB) Medium		
Ingredients	**Quantity (g)**	**Brand**
Tryptone	8	Merck
Yeast extract	4	VWR
Sodium chloride	8	Sigma-Aldrich

Table A1.2 – Specifications concerning the ingredients, the quantity in grams and the brand of each of the ingredients used to make 150 mL of LB agar.

LB agar		
Ingredients	**Quantity (g)**	**Brand**
Tryptone	1.5	Merck
Yeast extract	0.75	VWR
Sodium chloride	1.5	Sigma-Aldrich
Agar	3	Sigma-Aldrich

Table A1.3 – Specifications concerning the ingredients, the quantity in millilitres and the brand of each of the ingredients used to make 100 mL of M9 minimal medium and 1 L of M9 salts.

M9 minimal medium		
Ingredients	**Quantity (mL)**	**Brand**
M9 salts (5x)	20	-
20% Glucose	2	Carl ROTH
1 M Magnesium sulphate	0.2	Sigma-Aldrich
1 M Calcium chloride	0.01	Merck

M9 salts (1 L)		
Ingredients	**Quantity (g)**	**Brand**
Sodium phosphate dibasic heptahydrate	64	Sigma-Aldrich
Monopotassium phosphate	15	Carl ROTH
Sodium chloride	2.5	Sigma-Aldrich
Ammonium chloride	5.0	Sigma-Aldrich

Table A1.4 – Specifications concerning the ingredients, the quantity, and the brand of each of the ingredients used to make 100 mL of BugBuster.

BugBuster		
Ingredients	**Quantity**	**Brand**
50 mM HEPES	1.192 g	Carl ROTH
25% sucrose	25 g	Carl ROTH
5 mM Magnesium chloride	0.048 g	Carl ROTH
1% Triton-X 100	1 mL	Sigma-Aldrich

Table A1.5 – Specifications concerning the ingredients, the quantity, and the brand of each of the ingredients used to make 20 mL of 4× Loading Buffer.

4× Loading Buffer		
Ingredients	**Quantity**	**Brand**
200 mM Tris-HCl pH 7.0	4 mL	Sigma-Aldrich
40% glycerol	8 mL	Merck
500 mM β-mercaptoetanol	0.8 mL	Sigma-Aldrich
0.08% bromophenol blue	0.02 g	Merck
8% SDS	1.6 g	Sigma-Aldrich

Table A1.6 – Specifications concerning the ingredients, the quantity and the brand of each of the ingredients used to make one 15% SDS-PAGE gel.

Separating gel		
Ingredients	**Quantity**	**Brand**
30% Acrylamide	2.5 mL	Bio-Rad Laboratories
1.5 M Tris-HCl pH 8.8	1.25 mL	Sigma-Aldrich
H_2O	1.2 mL	-
10% SDS	50 µL	Sigma-Aldrich
10% APS	50 µL	Sigma-Aldrich
TEMED	2.5 µL	Merck

4% Stacking gel		
Ingredients	**Quantity**	**Brand**
30% Acrylamide	335 µL	Bio-Rad Laboratories
1.5 M Tris-HCl pH 6.8	625 µL	Sigma-Aldrich
H_2O	1.5375 mL	-
10% SDS	25 µL	Sigma-Aldrich
10% APS	25 µL	Sigma-Aldrich
TEMED	2.5 µL	Merck

Table A1.7 – Specifications concerning the ingredients, the quantity in grams and the brand of each of the ingredients used to make 1 L of 10× Tris Buffered Saline.

10x Tris Buffered Saline		
Ingredients	**Quantity**	**Brand**
Trizma base	24	Sigma-Aldrich
Sodium chloride	88	Sigma-Aldrich

Annex 2 – nanoDSF – the different compositions of the tested buffers

Table A2.1 – List of the various buffers used, with sample ID showing their respective compositions.

Capillary	Sample ID
1	Potassium phosphate (KPi), pH 6.5, no salt
2	Bis-Tris propane (BTP), pH 6.5, no salt
3	Ammonium Acetate (AmAce), pH 7.0, no salt
4	MOPS, pH 7.0, no salt
5	Potassium phosphate (KPi), pH 7.0, no salt
6	HEPES, pH 7.5, no salt
7	Tris, pH 7.5, no salt
8	EPPS, pH 8.0, no salt
9	Imidazole, pH 8.0, no salt
10	Bicine, pH 8.5, no salt
11	Tris, pH 8.5, no salt
12	Potassium phosphate (KPi), pH 6.5, 0.5M NaCl
13	Bis-Tris propane (BTP), pH 6.5, 0.5M NaCl
14	Ammonium Acetate (AmAce), pH 7.0, 0.5M NaCl
15	MOPS, pH 7.0, 0.5M NaCl
16	Potassium phosphate (KPi), pH 7.0, 0.5M NaCl
17	HEPES, pH 7.5, 0.5M NaCl
18	Tris, pH 7.5, 0.5M NaCl
19	EPPS, pH 8.0, 0.5M NaCl
20	Imidazole, pH 8.0, 0.5M NaCl
21	Bicine, pH 8.5, 0.5M NaCl
22	Tris, pH 8.5, 0.5M NaCl

Annex 3 – Neutralization assays – cell plates layout

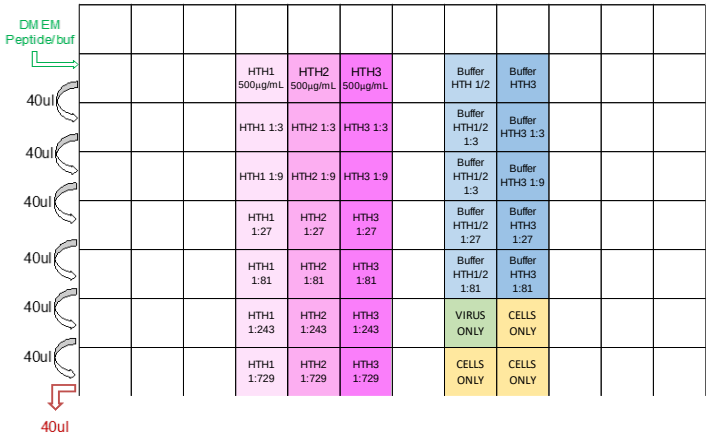

Figure A3.1 – Illustration of the plate layout where serial dilutions were performed. Serial dilutions of 40 µl each were made, starting with a 500 µg/mL solution.

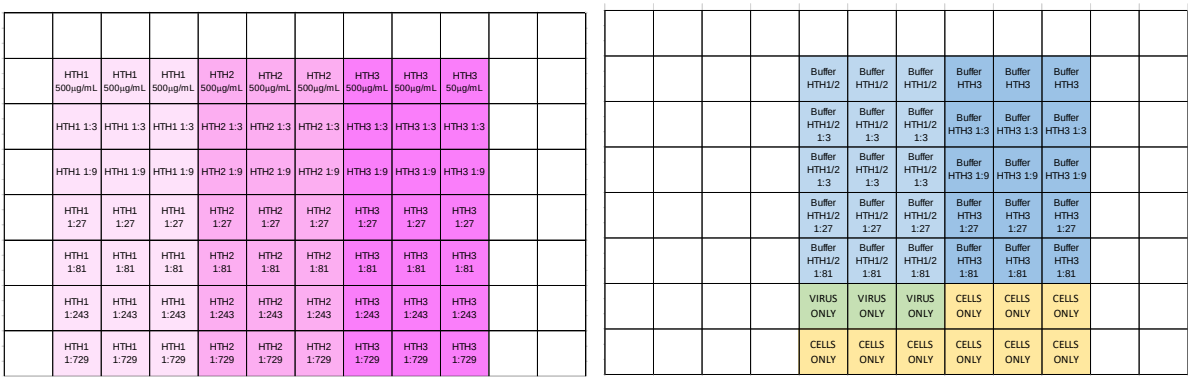

Figure A3.2 – Illustration of the plate layout where the neutralization assays were performed. The plate where these assays were performed for the three proteins is shown on the left, while the plate where the control samples were placed is shown on the right.

Annex 4 – Secondary structure of each HTH for every one of the replicates

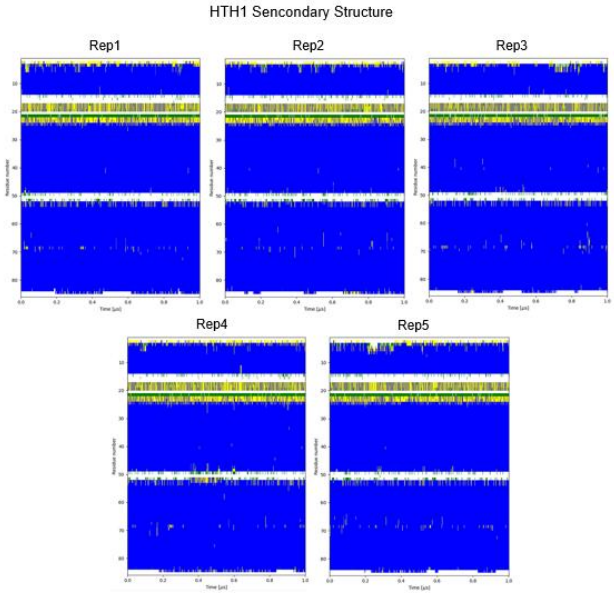

Figure A4.1 – HTH1 secondary structure, free in solution, for the 5 replicates.

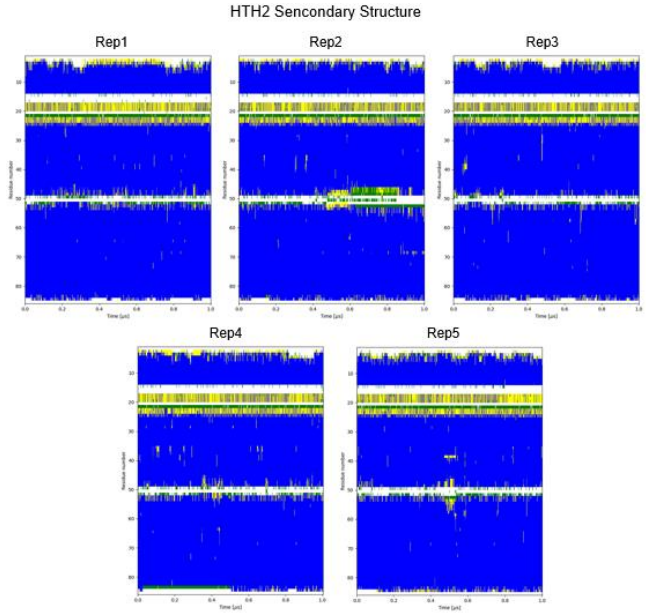

Figure A4.2 – HTH2 secondary structure, free in solution, for the 5 replicates.

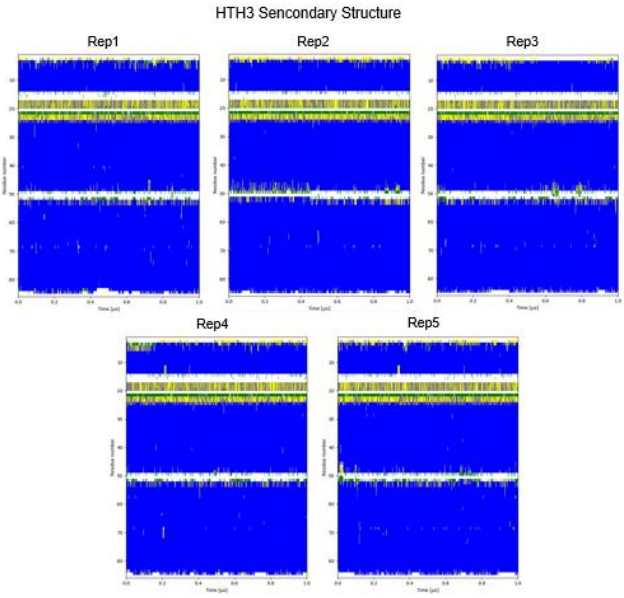

Figure A4.3 – HTH3 secondary structure, free in solution, for the 5 replicates.

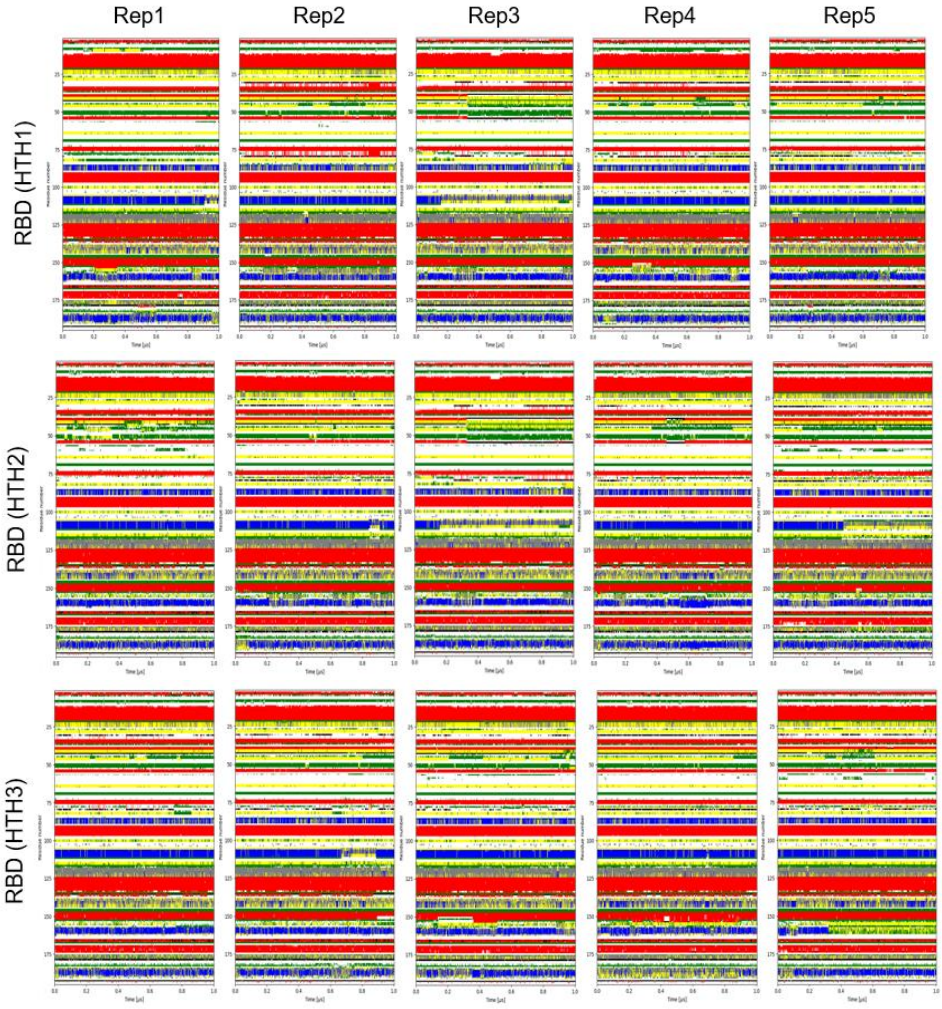

Figure A4.4 – RBD secondary structure in complex with the three HTH, for the 5 replicates.

Annex 5 – Images of the beginning and end of each HTH, free in solution and in complex with the RBD, for each of the replicates

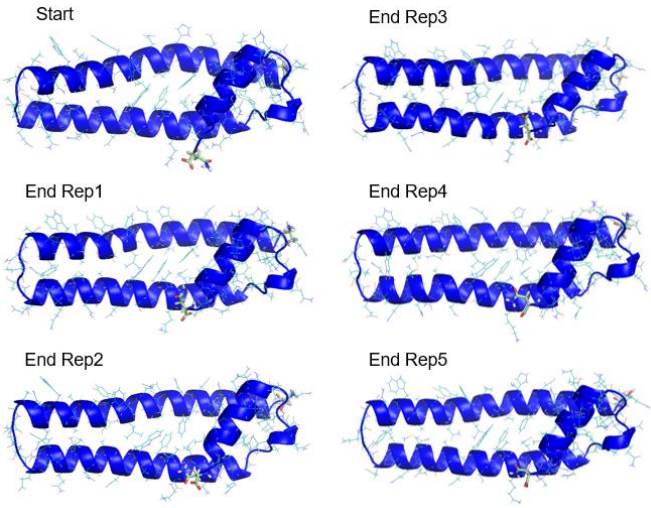

Figure A5.1 – Representation of the start and end of each simulation of HTH1, free in solution.

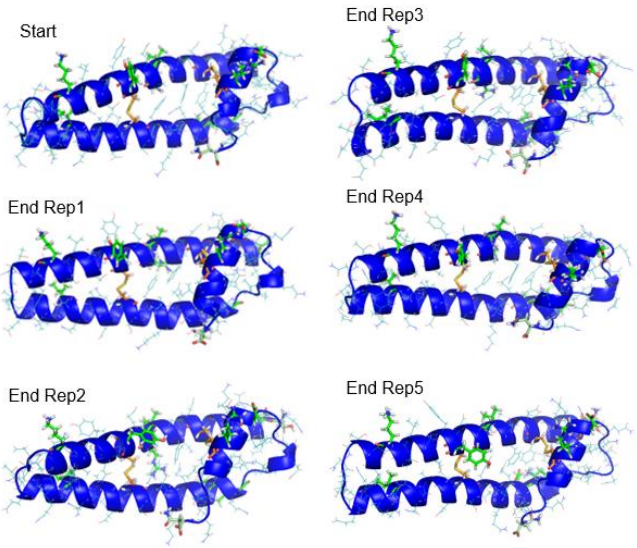

Figure A5.2 – Representation of the start and end of each simulation of HTH2, free in solution.

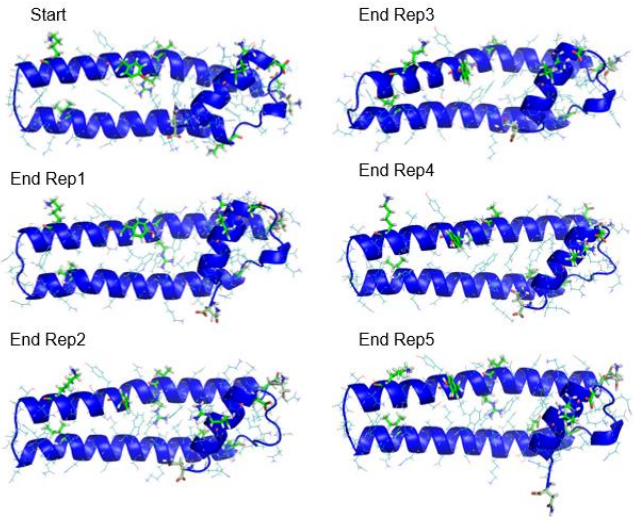

Figure A5.3 – Representation of the start and end of each simulation of HTH3, free in solution.

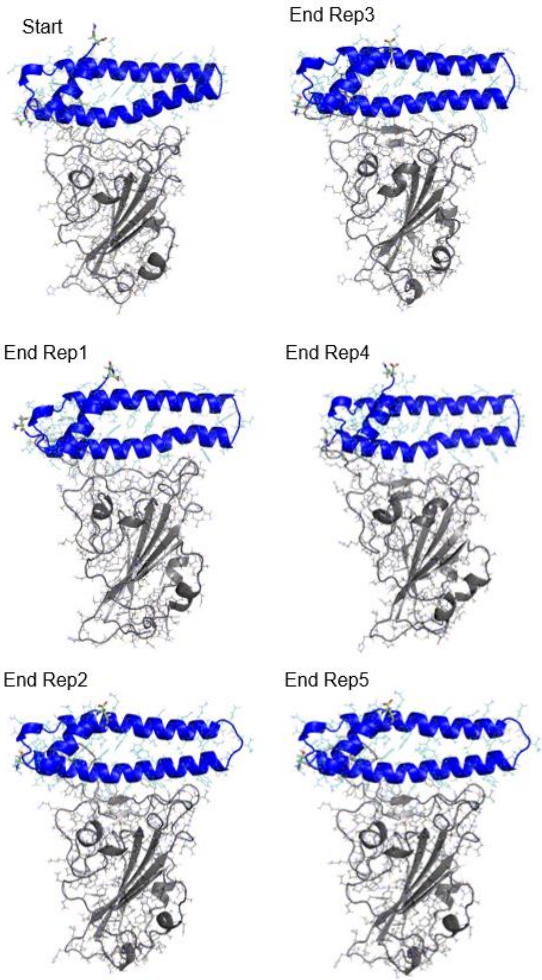

Figure A5.4 – Representation of the start and end of each simulation of HTH1-RBD.

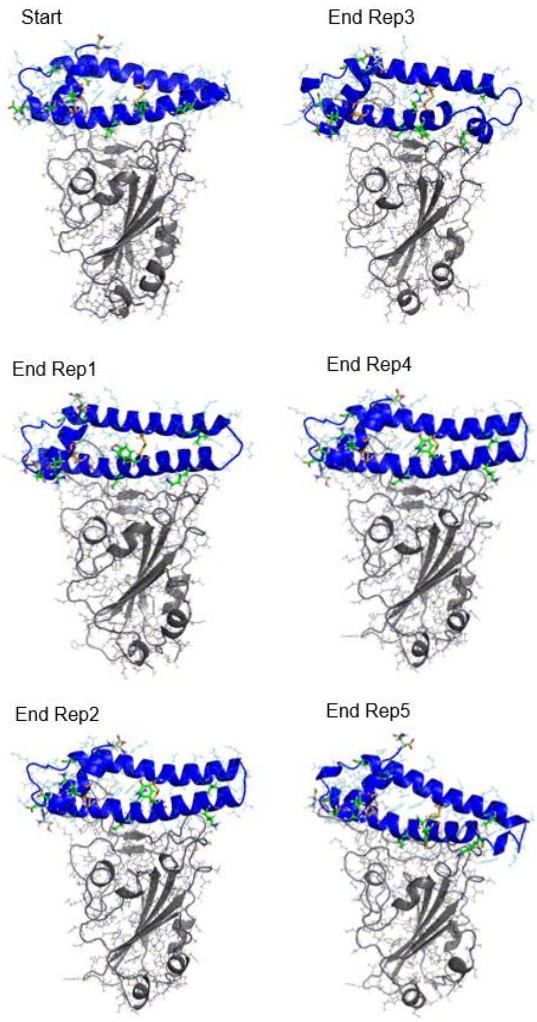

Figure A5.5 – Representation of the start and end of each simulation of HTH2-RBD.

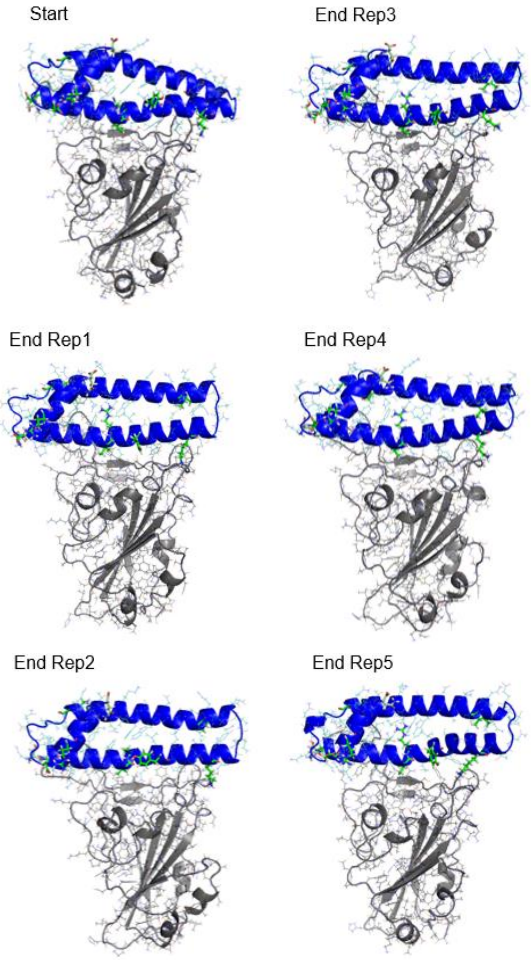

Figure A5.6 – Representation of the start and end of each simulation of HTH3-RBD.