University of Pavia
Department of Molecular Medicine

PhD course in Translational Medicine

XXXIV cycle

PhD thesis on

A new single-domain intrabody against TDP-43: selection, modelling and characterization

Tutor:
Dr. Hugo de Jonge
Prof. Annalisa Pastore

Candidate:
Dr. Martina Gilodi

Academic year 2020-2021

Dedicated to my beloved dad.
A great warrior.
My guardian angel.

Index

Index of figures

Index of Tables

Index of abbreviations

This is a list of the useful abbreviations I used in my thesis:

β-gal	β-galactosidase
3AT	3-amino-1,2,4-triazole
Abs	Antibodies
AD	Activation domain
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
ATP	Adenosine triphosphate
Amp	Ampicillin
BBB	Blood-Brain Barrier
BLAST	Basic Local Alignment Search Tool
BS^3	Bis[sulfosuccinimidyl]suberate
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CDRs	Complementarity-Determining Regions
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
CL	Constant domain of the light chain
CH	Constant domain of the heavy chain
CNS	Central Nervous System
CTD	C-terminal domain
dAbs	domain antibodies
DBD	DNA-binding domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ELISA	Enzyme-Linked Immunosorbent Assay

Fab	Fragment antigen-binding domain
Fc	Fragment crystallizable
FOR	Forward
FPLC	Fast Protein Liquid Chromatography
FRs	Framework regions
FTD	Frontotemporal Dementia
Fvs	variable domain
GaroS	Gly-aromatic-Ser-rich region
HCs	Heavy chains
HD	Huntington's Disease
hnRNPs	ribonucleoproteins
HRP	Horseradish Peroxidase
HSQC	Heteronuclear Single Quantum Correlation
IACT	Intracellular Antibody Capture Technology
Ig	Immunoglobulin
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVEM	In Vivo Epitope Mapping
Kan	Kanamycin
LB	Luria-Bertani medium
LCs	Light chains
LiAc	Lithium Acetate
LLPS	Liquid-liquid phase separation
LMN	Lower motor neuron
MCS	Multiple cloning site
MW	Molecular weight
NES	Nuclear export signal

NHS	N-hydroxysuccinimide
Ni-NTA	Ni-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
NLS	Nuclear localization signal
NTD	N-terminal domain
OD	Optical Density
PBS	Phosphate-Buffered Saline
PBST	Phosphate Buffered Saline Tween 20
PCR	Polymerase Chain Reaction
PD	Parkinson's Disease
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PelB	Periplasmic pectate lyase
PNK	Polynucleotide Kinase
REase	Restriction endonucleases
REV	Reverse
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RRMs	RNA-recognition motifs
scFVs	single-domain variable fragments
sdAb	single-domain antibody
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SGs	Stress Granules
SOC	Super Optimal Broth
STD	Salmon Testis DNA
SUMO	Small Ubiquitin like Modifier

TBST	Tris Buffered Saline Tween
TDP-43	TAR DNA binding protein-43
TE	Tris- ethylene-diamine-tetraacetic acid buffer
TEV	Tobacco Etch Virus
TMB	3,3',5,5'-tetramethylbenzidine
Tris	Tris(hydroxymethyl)aminomethane
UMN	Upper Motor Neuron
VH	Variable domain of the heavy chain
VHH	Single variable domain of the heavy chain antibody
VL	Variable domain of the light chain
WB	Western Blot
X-gal	5-Bromo-4-Chloro-3-Indolyl β-D- Galactopyranoside
Y2H	Yeast Two-Hybrid
YPA	Yeast Extract Peptone Adenine
YPD	Yeast Extract Peptone Dextrose
YPAD	Yeast Extract Peptone Dextrose Adenine
YNB	Yeast Nitrogen Base

Abstract

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder associated to deteriorating motor and cognitive functions, and short survival. The disease is caused by neuronal death which results in progressive muscle wasting and weakness, ultimately leading to lethal respiratory failure. The misbehaviour of a specific protein, TDP-43, which aggregates and becomes toxic in ALS patients' neurons, is supposed to be one of the causes. TDP-43 is a DNA/RNA-binding protein involved in several functions related to nucleic acid metabolism. Sequestration of TDP-43 aggregates is a possible therapeutic strategy that could alleviate or block pathology. In this thesis I describe how I selected and characterized a new intracellular antibody (intrabody) against TDP-43 from a llama nanobody library. The structure of the selected intrabody was predicted *in silico* and the model was used to suggest mutations that enabled to improve its expression yield, facilitating its experimental validation. I showed how coupling experimental methodologies with *in silico* design may allow us to obtain an antibody able to recognize the RNA binding regions of TDP-43. These findings illustrate a strategy for the mitigation of TDP-43 proteinopathy in ALS and provide a potential new tool for diagnostics.

1. Introduction

1.1 Neurodegenerative diseases

"Neurodegeneration" is a commonly used word that refers to a variety of conditions that modify the neuronal normal functions in the human brain and lead to a progressive and consistent neuronal loss. Neurons are the main cellular components of the central nervous system (CNS). They have complex and extended morphologies compared to other cell types, within the CNS. They can vary greatly in their properties and are characterized by the inability to reproduce or replace themselves. In this way, when damage threatens, the loss of cells will be permanent. Neurodegenerative diseases involve a multitude of diseases which share common features such as misfolded proteins, oxidative stress, inflammation, excitotoxicity and, obviously, neuronal loss. Involvement of functional systems differs between disorders and is associated with a wide spectrum of clinical presentations. This group of incurable diseases include Alzheimer's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis, Frontotemporal Dementia, Prion diseases, and polyQ diseases such as, Huntington's disease and Spinocerebellar Ataxias.

The most consistent risk factor for developing these pathologies is increasing age. Over the past century, the growth rate of the population aged 65 and beyond in industrialized countries has far exceeded that of the whole population. Thus, it can be anticipated that, over the next generations the proportion of elderly citizens will double, and, with this, possibly the proportion of people suffering from some kind of neurodegenerative disorder (1). Patients with these diseases display serious neurological disabilities, such as memory impairment and motor problems, for which there are no cure.

Neurodegenerative diseases are different for transmission, clinical and progression, however, they share common features, such as old age-onset and the presence of protein toxicity. These diseases imply the change in the conformation of a physiologic protein changes, resulting in an altered function or potentially toxic intra- or extracellular accumulation. Both, fibrils and oligomers formed during the aggregation pathways have been found to induce toxicity. Protein toxicity in affected neurons may result in cellular defects such as transcriptional alteration, mitochondrial dysfunction, and an impaired protein/RNA quality control system, all of which critically contribute to the initiation and progression of neurodegenerative diseases (2-4). Despite the large number of studies, the role of toxic species and the equilibria between oligomers and fibrils remains to be clarified, both in terms of assembly of oligomers into fibrils and potential dissolution of fibrils into

oligomers. Many controversial results exist in literature and no clear conclusions can be drawn due to the great variety and heterogeneity of different experimentally observed states (5).
The study and understanding of the mechanisms and factors underlying aggregation processes can allow the development of new therapeutic strategies thus opening a way to possible horizons of treatment.

1.2 Amyotrophic Lateral Sclerosis

The term “Amyotrophic Lateral Sclerosis” was coined by the French neurologist Jean-Martin Charcot in the 1800s: “Amyotrophic” refers to muscular atrophy, and “Lateral Sclerosis” describes the scarring or hardening of tissues in the lateral spinal cord (6).
Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disease, characterized by progressive degeneration of both the upper and lower motor neurons, which display cytoplasmic inclusions. The Upper Motor Neuron (UMN) also called first motor neuron, central, superior, or cortical motor neuron, is located in the cerebral cortex and carries the nerve signal through the axon to the spinal cord. The Lower Motor Neuron (LMN) also called second motor neuron or lower peripheral spinal motor neuron, is located in the brain stem and spinal cord and carries the nerve signal form the spinal cord to the muscles (**Figure 1.1**) (7, 8). The degradation of the Upper Motor Neurons leads to spasticity and hyper-excitability, while the death of the Lower Motor Neurons causes weakness, fasciculations and eventually muscular atrophy followed by progressive paralysis (9).

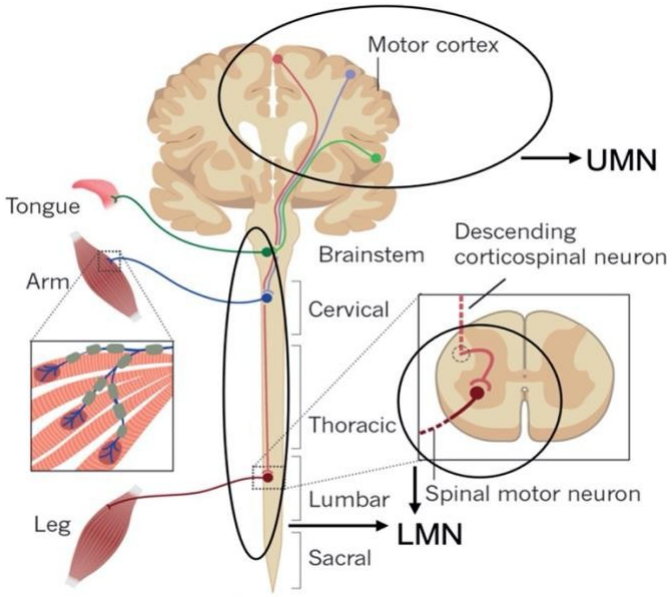


Figure 1.1 Route of the corticospinal tract involved in ALS. *Upper motor neurons (UMN) from the motor cortex and lower motor neurons (LMN) projecting from the brain stem and the spinal cord to the skeletal muscles are affected in ALS. Adapted from Taylor et al., 22016 (7).*

The selective degeneration of motor neurons typically appears in middle-aged patients (average age 55 years) and progresses to muscle atrophy followed by complete paralysis. Death is caused by respiratory failure and typically occurs within 3 to 5 years from diagnosis. The disease has a prevalence of ~5 individuals out of 100,000 each year worldwide. Men are slightly more frequently affected than women, with a male:female incidence rate ratio of 1.4:1 (10). While the majority of the ALS cases (~90–95%) are considered as sporadic (sALS) with unknown cause, ~5–10% cases involve Mendelian pattern of inheritance of familial gene mutations and are known as familial ALS (fALS) (7, 11). The first genetic mutations found to cause ALS, reported in 1993, affected the gene *SOD1* (12) and several potential ALS genes have been identified since then. These genes can be grouped into different categories: genes that alter proteostasis and protein quality control, genes that perturb aspects of RNA stability, function, and metabolism and genes that alter cytoskeletal dynamics in the motor neuron axon and distal terminal. The mutations involved are mostly missense substitutions. Pathogenic mutations in other ALS genes, include *TARDBP, FUS, C9orf72, SQTM1, VCP, OPTN* and *PFN1* (7) .

Similar to most neurodegenerative diseases, ALS starts focally and spreads involving different sets of motor neurons or different regions of the body: symptoms that start as subtle cramping or weakness in the limbs or bulbar muscles progress to the paralysis of almost all skeletal muscles.

Some subsets of motor neurons, including those that innervate the extraocular muscles or sphincters, are spared until late in the progression of the disease (7).

Importantly, ALS shows clinical overlap with several other adult-onset degenerative disorders, most frequently frontotemporal dementia (FTD), a type of dementia that involves impaired judgment and executive skills. In FTD, the loss of cortical motor neurons is accompanied by loss of neurons in the frontal and temporal cortices, which correlates clinically with the symptoms of FTD. This may constitute a clinical spectrum with 50% of ALS patients presenting cognitive impairment (15%–20% recognized as FTD) and 15% of FTD patients having motor impairments. The relationship between ALS and FTD has been confirmed through genetic studies, and these two conditions are now considered to be at opposite ends of the same disease continuum (13-15). Multiple pathogenic pathways have been identified to be implicated in the motor neuron degeneration of ALS, such as excitotoxicity, neuroinflammation, proteolysis impairment, oligodendrocyte dysfunction, mitochondrial dysfunction, and disruption of RNA homeostasis (16-18). Emerging evidence suggests that the accumulation of various disease-related misfolded proteins may underlie these pathogenic processes. Despite advances in understanding the pathogenesis of ALS have increased significantly in recent years, major gaps in our knowledge remain and there is no curative or effective control therapy.

1.3 TAR DNA binding protein-43 (TDP-43)

One of the main pathological characteristics of ALS is the presence of insoluble protein inclusions in the soma of motor neurons. TAR DNA binding protein-43 (TDP-43) is one of the proteins implicated in this disease and it represents the major component of the inclusions (13, 19) in almost all (~97%) ALS patients and ~50% of FTD patients (13, 19-21). TDP-43 belongs to the family of the heterogeneous ribonucleoproteins (hnRNPs), sharing with them several structural and functional features (22). It is abundantly expressed in nearly all tissues and is well conserved among mammals and invertebrates (23).

1.3.1 TDP-43 structure

TDP-43 protein contains 414 amino acids and the encoding gene *TARDBP* is located on the chromosome number 1. Structural analysis has identified a well folded N-terminal domain (NTD, aa 1–105) with a nuclear localization signal (NLS, aa 82–98), two RNA-binding RRM tandem

domains, named RRM1 (aa 106–177) and RRM2 (aa 192–259), a nuclear export signal (NES, aa 239–250), and an unstructured C-terminus domain (CTD, aa 260– 414) contains a so-called prion-like glutamine/asparagine-rich (Q/N) domain (aa 345–366) and a glycine-aromatic -Ser rich region (aa 273 – 317, 368– 414) (24-26). It has been shown that the CTD harbours the majority of ALS-associated mutations (**Figure 1.2**).

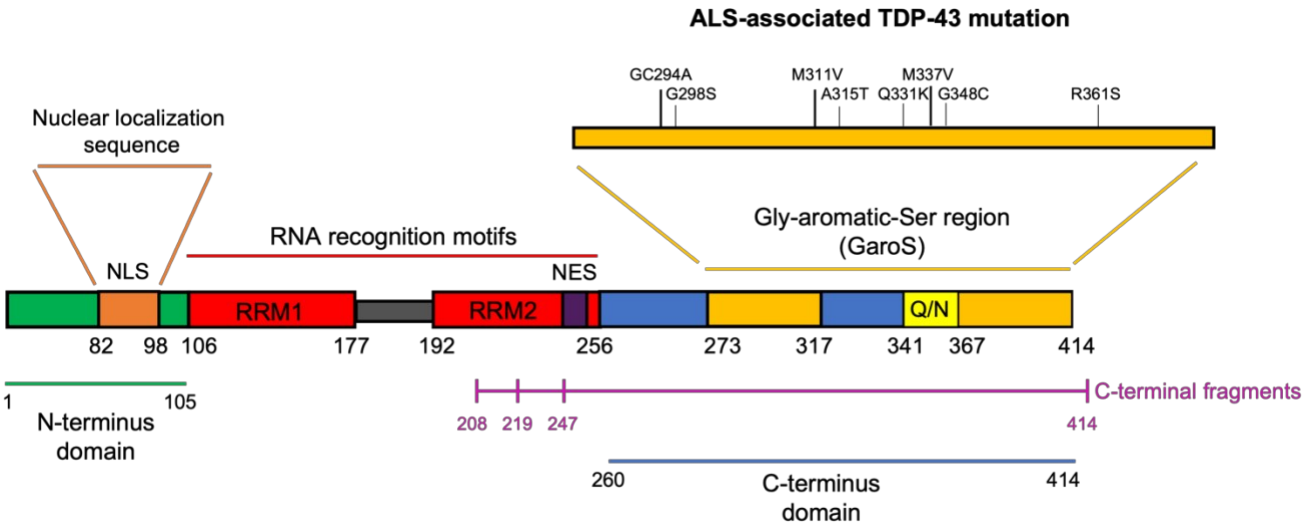


Figure 1.2 Schematic TDP-43 structural features. *TDP-43 contains a N-terminus domain (in green), a nuclear localization signal (in orange), two RNA-binding motifs (in red), a nuclear export sequence (in purple), a C-terminus domain (in blue), two GaroS region (in ochre) and a Q/N domain (in yellow). The precise fragments length of the TDP-43 C-terminal truncated protein, observed in the insoluble aggregates from FTLD-TDP brain tissue, are indicated (in magenta). Adapted from Cohen et al., 2011 (24).*

TDP-43 is localized preferentially in the nucleus and the NLS sequence allows the protein to shuttle between this compartment and the cytoplasm. The NLS is located in the linker region between the NTD and the RRMs while the putative NES sequence is located within the RRM2. Recently, it has been observed that the two NES sequences are not functional, and that TDP-43 exits the nucleus by passive diffusion, as other RNA-binding proteins, like FUS/TLS (27). TDP-43's NTD and NLS have a role as structural support for nucleic acid binding and might contribute to specificity toward certain nucleic acid sequences (28, 29).

N-terminal domain (NTD). The folded N- terminal region (aa 1-105) shows an ubiquitin-like fold, which consists of one α-helix and six β-strands in the β1-β2-α1-β3-β4-β5-β6 arrangement (26, 29, 30) (**Figure 1.3 A**). It has been observed that TDP-43 is natively dimeric or exists in a monomer-

dimer equilibrium under normal physiological conditions. TDP-43 dimer shows a head-to-head interaction of two NTDs which allows its physiological functions like RNA splicing, enhance its solubility and protect against the formation of the cytoplasmic TDP-43 inclusions (31, 32). The NTD undergoes a reversible oligomerization which enhances the tendency of the intrinsically disordered C-terminal region to aggregate (33).

RNA-binding domains. RNA-recognition motifs of TDP-43 contain two highly conserved RRM domains: RRM1 (aa 106-177) and RRM2 (aa 192-259) which are separated by 15 amino acids (25, 30, 34). These RRM domains fold into five β-strands and two α-helices arranged in the β1-α1-β2-β3-α2-β4-β5 pattern (**Figure 1.3 B**). Thanks to its RRM domains, TDP-43 can bind to both DNA and RNA molecules, with preferential specificity toward UG/TG-rich sequences. Contrary to typical tandem RRMs, it binds nucleic acid in the 5′ to 3′ direction (25, 35). The RRMs can bind 3-4 UG repeats *per* TDP-43 molecule. Structural analysis of tandem RRMs in complex with UG-rich RNA highlighted a unique mode of RNA binding for TDP-43, characterized by the formation of a continuous RNA binding surface, allowing the binding of longer RNA sequences (22, 25). The RRMs are connected by a highly flexible loop (aa 178–191), which confer adaptability to different nucleic acid partners by allowing different orientations of the RRM domains. Some mutations in the RRMs disrupt the RNA binding capability while not significantly interfering with the RNA recognition (25). It has been shown the presence of amyloidogenic cores in these motifs and the ability of these domains to misfold and participate in either nucleation or propagation of TDP-43 aggregation in ALS (36, 37).

C-terminal domain. The C-terminal region of TDP-43 (aa 260–414) is highly disordered and aggregation- prone (38). It harbours most of the ALS-associated *TARDBP* mutations and phosphorylation sites. This region contains distinct domains: two Gly-aromatic-Ser-rich (GaroS) regions and an amyloidogenic core divided into a hydrophobic region and a Q/N-rich region (26, 30) (**Figure 1.3 C**). These unusual sequences resemble those of the prion-like domain. Prion-like domains are low-complexity sequences enriched in uncharged polar amino acids (asparagine, glutamine, and tyrosine) and glycine, and their sequence properties are similar to those of yeast prion proteins. Prion-like domain plays a key role in regulating the solubility and folding of proteins (39, 40).

The GaroS regions (residues 273–317 and 368–414) are similar to FUS/TLS domains (41) proposed to interact within RNA granules (26) and contributing to the formation of hydrogels (42). TDP-43 C-terminal region likely adopts a range of transient well-ordered shapes, many of which

are capable of self-association (43-46). Mutations and post-translational modifications may change the kinetics of these states, influencing the equilibrium between fibril formation and dissolution (30, 46). The C-terminus has multiple functions and has been shown to mediate protein-protein interaction, to participate in the stabilization of protein- RNA interactions and to be necessary for TDP-43 splicing activity (22, 23, 47, 48), including autoregulation (49). TDP-43 C-terminal region can also undergo liquid-liquid phase separation (LLPS) to form dynamic protein droplets. Within these droplets, the C-terminal residues show mild transient interactions, that are crucial for stress granule formation and to potentially allow an increase in protein concentration at specific locations inside the cell (50).

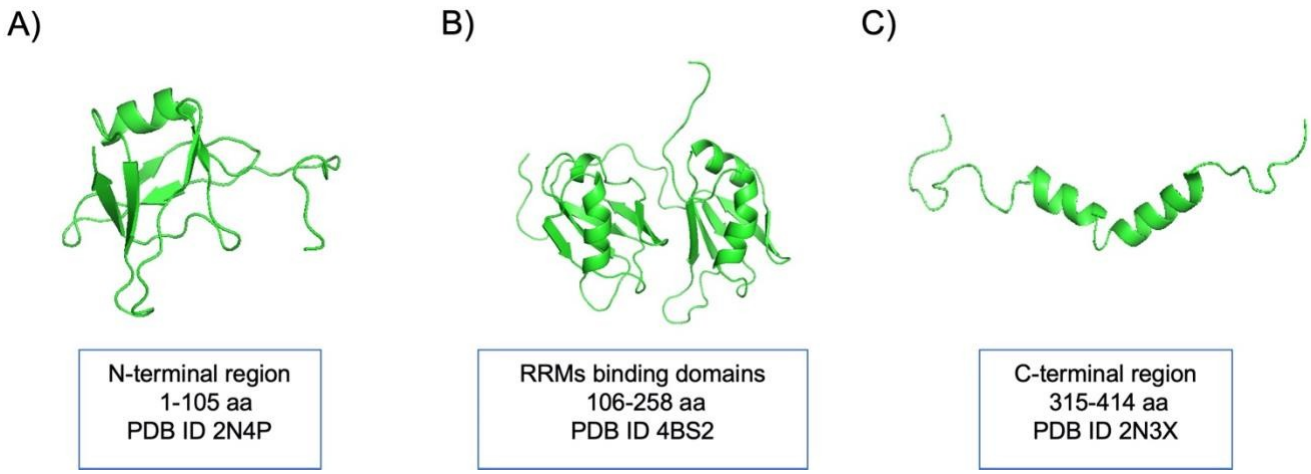


Figure 1.3 PDB structures of TDP-43 fragments. *(A) N-terminal region (aa 1-105) (PDB ID 2N4P), (B) RRM-binding motifs (aa 106-258) (PDB ID 4BS2) and (C) C-terminal region (aa 315-414) (PDB ID 2N3X). Adapted from Prasad et al., 2019 (9).*

1.3.2 TDP-43 functions

TDP-43 was first identified as repressor protein associated with HIV-1 transcription, which binds to the trans-active response element (TAR) DNA sequence of the viral genome and is critical for the regulation of the viral gene expression (51). It was later shown that the protein is not capable in modifying HIV-1 replication in human cells (52). However, the protein maintained the name derived from that first description (TAR DNA-binding protein of 43 kDa).

TDP-43 is a modular protein which has a plethora of roles related to the RNA metabolism (**Figure 1.4**), including transcription, splicing, transport, and scaffolding (24, 53, 54). Alterations in such processes due to its loss or dysfunction, can favour ALS related toxicity in different ways (55).

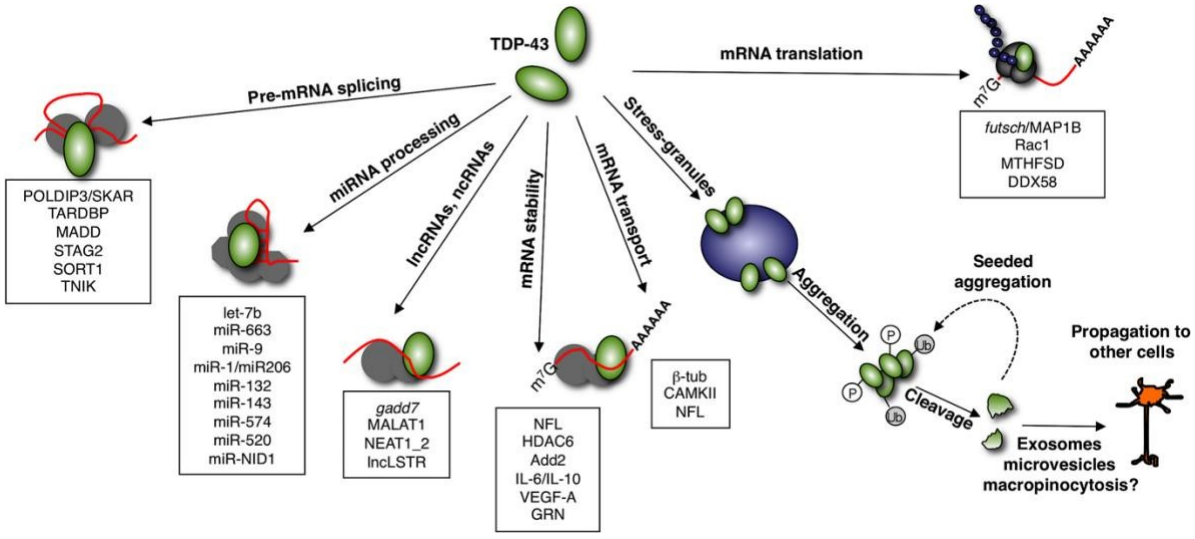


Figure 1.4 Schematic diagram of the TDP-43 functions. *Representation of the known TDP-43 regulated cellular functions in the nuclear and cytoplasmic compartments (51).*

TDP-43 has been shown to modulate the splicing patterns of transcripts of other genes (55). The first described target of TDP-43 regulatory function on alternative splicing has been the cystic fibrosis transmembrane conductance regulator (CFTR), whose regulation is mediated in cooperation with other hnRNPs, like the Serine-Arginine (SR) proteins (56).

While TDP-43 binds with high specificity to the UG-rich sequences of RNAs, it mostly binds to the 3' UTRs of mRNAs/pre-mRNAs when localized to the cytoplasm. This suggests a wide role of TDP-43 in maintaining mRNA stability, maturation, and transport (57, 58).

TDP-43 also promote biogenesis and processing of the non-coding RNAs, such as microRNA (miRNA) (59). Recent studies have confirmed the interactions of TDP-43 with the Drosha and Dicer complexes (59, 60). The nuclear Drosha complex cleaves pri-miRNAs to release intermediate precursors that are termed pre-miRNAs. Pre-miRNAs are then transported by Exportin-5 into the cytoplasm where they are cleaved further by the Dicer complex to generate mature miRNAs (59). TDP- 43 associates with the nuclear Drosha complex and binds directly to the primary miRNAs to facilitate the production of a subset of precursor miRNAs (pre-miRNAs) (59, 61).

As mentioned previously, another function of TDP-43 is its ability to phase separate into liquid droplets facilitating its recruitment into stress granules which are formed thanks to its C-terminal low complexity domain and its RNA binding propensity (62).

Finally, TDP-43 is also capable of assembling into stress granules (SGs), the membrane-less cytoplasmic foci of sizes ≤5 μm, indicating its protective role against cellular insults (63, 64). In fact, TDP-43 is involved in both assembly and maintenance of stress granules, and it also regulates the expression of key SG nucleating proteins.

1.3.3 TDP-43 pathology in ALS

Most neurodegenerative diseases are characterized by the aggregation of specific proteins. In ALS that protein is TDP-43. Post-translational modification is an important regulatory mechanism to maintain protein homeostasis of TDP-43. The pathological hallmarks of TDP-43 proteinopathies are depletion from the nucleus and accumulation as cytoplasmic aggregates where the protein can be found ubiquitinated, phosphorylated or cleaved (65).

It is accepted that the cytoplasmic accumulation and the aggregation of TDP-43 into inclusion bodies confer both a loss-of-function as well as a gain-of-toxic-function (66, 67). The precise sequence of events leading to the pathological TDP-43 mislocalization is a source of discussion, however, nuclear TDP-43 depletion appears to precede the inclusion body formation (68, 69). The abnormal localization of TDP-43 in the cytoplasm in affected neurons in ALS and FTD, irrespective of the presence of a genetic mutation, suggests a pathogenic mechanism associated with the loss of the normal nuclear TDP-43 function in regulating transcription, splicing and mRNA stability (53, 70). In addition, loss of TDP-43 in human cells has been shown to induce morphological nuclear defects and increased apoptosis (71). Alternatively, sequestration of TDP-43 in cellular inclusions may induce a toxic gain-of-function, independent of the basic biological role of TDP-43. Factors that affect the normal intracellular trafficking of TDP-43, between the cytoplasm and nucleus, may predispose to both the formation of abnormal aggregates (inclusions) and the loss of nuclear localization (72).

TDP-43 can also be found N-terminally truncated in the cytoplasm of neurons and glia cells. The corresponding C-terminal fragments are predominantly found in the cortex of ALS patients (73), but how these fragments affect TDP-43 pathology has only been studied in cellular overexpression models (74). Overexpression of C-terminal fragments, that still include RRM2 and the glycine-rich-

domain, promotes the formation of cytoplasmic aggregates with post-translationally modified fragments. In the presence of these aggregates, the alternative splicing of the *CFTR* gene transcript, which is regulated by TDP-43, becomes dysregulated and is an indicator of abnormal function of TDP-43.

In disease, TDP-43 inclusions have varying morphologies, ranging from neuronal cytoplasmic inclusions with compact rounded or skein-like appearance, to short or long dystrophic neurites, to rare intranuclear inclusions or glial (primarily oligodendrocytic) inclusions (13, 75). The presence of various forms of TDP-43 pathology in both ALS and FTLD suggests that these diseases may occur as manifestations of a spectrum of TDP-43-related disease (76, 77).

TDP-43 cytoplasmic aggregates have been also observed in other neurological diseases, such as Alzheimer's disease (AD), Parkinson's disease (PD), and Huntington's disease (HD) (78, 79).

Clinical mutations of TDP-43 are rare and seem to occur mainly, but not exclusively, in the C-terminus of the protein (80, 81). This observation had originally suggested that this region is the main cause of protein aggregation and misfolding. Recently, it has been proved that TDP-43 fragments are prone to misfolding and promote the formation of aggregates also in the absence of the prion-like domain (82, 83): constructs containing only the RRM domains or the whole region from the N-terminus to the end of RRM2 form aggregates with amyloid features (83-85) indicating that TDP-43 contains multiple aggregation-prone hotspots. Accordingly, clinically relevant mutations occurring in the two RRM domains have been described (85).

1.3.3.1 Stress granules formation

The response to stress by the RNA metabolism machinery has been strongly highlighted in the pathology of neurodegeneration, particularly in ALS. Stress Granules (SGs) are cytoplasmic membrane-less organelles belonging to the wider class of ribonucleoprotein (RNP) granules that are enriched in RNA and proteins (86). SGs are an important part of RNA metabolism during cellular stress (87). Defects in SG assembly and disassembly have already been linked to a number of neurodegenerative disorders including ALS (87-89). Typically, SGs form in response to some sort of cellular stress, such as oxidative stress, and disassemble once conditions return to normal. TDP-43 protein is able to undergo liquid-liquid phase separation resulting in the formation of membrane-less granules. SGs are meant to be transient structures, but chronic stresses can cause persistent SGs. These long-term SGs appear to function as a core for the aggregation of pathological

proteins. It is hypothesized that aging, among other things, places cells under chronic stress (90). When this occurs in a patient with non-ideal RNA metabolism machinery, this could lead into a neurodegenerative disease (91). This hypothesis is founded on the fact that the pathological TDP-43 inclusions found in ALS patient samples also colocalize with markers for SGs in the cytoplasmic inclusions (87). Many of the mutations known to contribute to ALS disrupt TDP-43 shuttling between the nucleus and cytoplasm or increase their aggregation ability. Moreover, TDP-43 mislocalization affects the entire cells nucleo-cytoplasmic transport capability (92). Either of these disruptions can lead to toxic inclusions (88). Overexpression of TDP-43 or expression of the truncated form normally found in ALS patients will trigger colocalization with SG markers even in the absence of stress (88). If the presence of SGs is sufficient to cause pathology, then a malfunction in the disassembly of SGs would be an additional source of pathology.

The identification of TDP-43 pathology is very important in the context of targeted drug development and clinical trials design. Indeed, the high failure rate of clinical trials for neurodegenerative diseases could be explained, at least in part, by the high degree of pathological heterogeneity of these disorders.

1.4 Antibodies

Despite the advancements made in understanding TDP-43 aggregation, too many details of the mechanism remain unclear due to the lack of adequate research tools able to accurately probe aggregation. In this regard, antibodies constitute a ductile means widely used in research and in clinics, thanks to their high binding affinity and specificity. Antibody applications extend from quantitative *in vitro* measurements to *in vivo* studies. When expressed as intrabodies inside cells (93, 94), they can for instance be used to sequester protein aggregates reducing cell toxicity (95). They are also great assets in diagnostics and basic science as they may be used in super-resolution microscopy, allowing visualization of protein aggregates at the nanoscale as in the recently developed DNA-PAINT methodology (96-98).

Our knowledge of the three-dimensional structure of antibodies has arisen from crystallographic studies described at the beginning of 1970s (99, 100). Now, the Protein Data Bank (PDB) (101) contains numerous structures of antibody fragments (Fabs, Fvs, scFvs, and Fcs), as well as a small

number of intact antibody structures. The overall structure of antibodies has been the subject of several reviews (102-104).

Antibodies, or Immunoglobulins (Ig), are symmetrical multi-domain protein. They play a fundamental role in the immune system, by being among the most important effectors of adaptive immunity (105). Human immunoglobulins are large molecules, having a molecular weight of approximately 150 kDa, and Y-shaped proteins composed of two different kinds of polypeptide chain. One of approximately 50 kDa, is termed the heavy or H chain (HCs), and the other, of 25 kDa, is termed the light or L chain (LCs). Each molecule consists of two identical heavy chains and two identical light chains. In natural system, the two heavy chains are linked to each other by disulfide bonds and each heavy chain is linked to a light chain by a disulfide bond to form the intact immunoglobulin. Two types of light chains, termed lambda (λ) and kappa (κ), are found in antibodies. Although slightly different in length, they both have the same structural motifs, and both similarly form disulfide bonds with the heavy chain, but the loci are quite different at the genetic level (106). A given immunoglobulin either has κ chains or λ chains, never one of each. No functional difference has been found between antibodies having λ or κ light chains (105, 106). By contrast, the structure of the heavy chain defines the effector function of the antibody (107). There are five main heavy-chain classes some of which have several subtypes, and these determine the functional activity of an antibody molecule. The five major classes of immunoglobulins are IgA, IgD, IgE, IgG and IgM, each with its own class of heavy chain -α, δ, ε, γ and μ, respectively (108). The heavy chains give distinctive conformations to the hinge and tail regions of antibodies, so that to each class correspond characteristic properties (109). Either type of light chain may be found in antibodies of any of the five major classes (107). In each chain, two functionally distinct domains are present, which are the variable region (V) and the constant region (C) (**Figure 1.5**). The constant domains of the heavy chains (CH2 and CH3) form the Y-stem and are referred to as fragment crystallizable (Fc). The glycosylated Fc region binds to a variety of receptor molecules providing the effector function profile that dictates how the antibody interacts with other components of the adaptive and humoral immune system. Each arm of the Y-shape is instead formed by the pairing of the variable domain and the constant domain (VL and CL) of the LCs with the variable domain and the constant domain (VH and CH1) of the HCs and is termed Fragment antigen binding (Fab). The two Fabs have identical antigen-binding sites (or antigen-combining sites) for binding to a specific target antigen (100). These two regions are linked with the hinge region, a short sequence that enable flexibility to the protein (110).

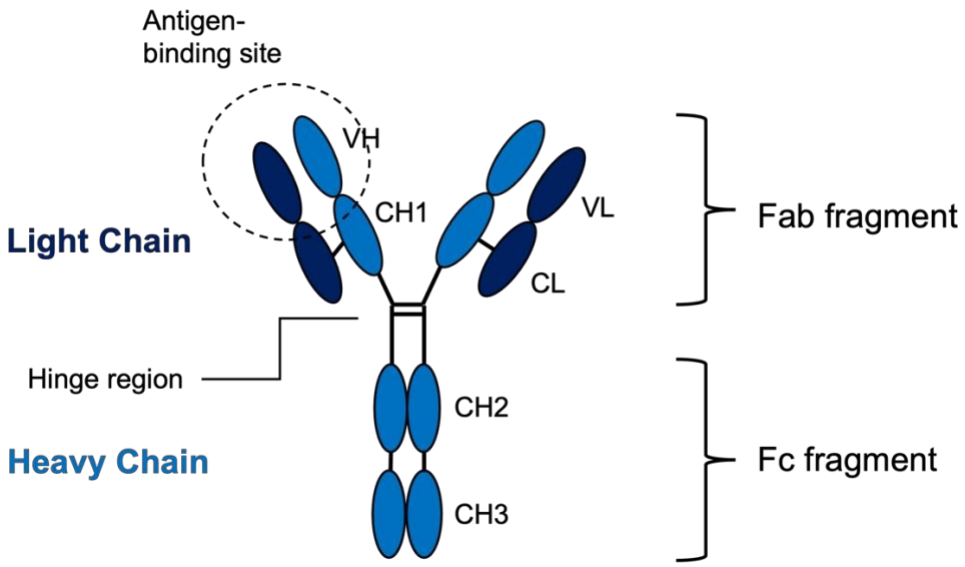


Figure 1.5 Basic antibody structure. *Schematic representation of an antibody molecule made up of two identical heavy chains (light blue) and two identical light chains (dark blue). The black lines signify the disulfide bonds. The antigen binding site is circled.*

The VH and VL domains which interact with the target antigen, are at the N-termini of the HCs and LCs.

In any immunoglobulin molecule, the two heavy chains and the two light chains are identical, giving to the antibody molecule two identical antigen-binding sites, and thus the ability to bind simultaneously to two identical structures (105).

Each variable domain is composed of three hypervariable regions, called complementarity-determining regions (CDRs), which are responsible for antigen recognition. In total, the 6 CDR loops, of which three reside within the VL chain and three reside with the VH chain, form the binding surface that characterizes the affinity and specificity between antibody and antigen (111, 112).

In the VH and VL CDR regions, the amino acid residues which establish the direct contact with the antigen are called paratope, while the surface of the antigen in direct contact with the antibody is called epitope (105, 107, 113).

1.5 Single domain antibodies (sdAb or Nanobodies)

Among the natural antibody scaffolds, variable domains of the heavy-chain antibody (HCAb) [also named nanobodies or single-domain antibodies (sdAb) or VHHs] offer specific advantages over normal antibodies but also respect to single chain Fv fragments (114) or domain antibodies (dAbs) (115) or other antibody mimetics. Natural VHHs were first identified in *Camelidae* (116), including Vicugna, Alpaca, Llama, Camel and Dromedary. As other mammals, they express conventional antibodies composed of two heavy chains and two light chains and in addition, they also express non-conventional antibodies, the heavy-chain antibodies. These are typically single variable heavy chain domains of ca. 110 amino acids that are derived from heavy-chain-only antibodies, (VH), devoid of the light chain partners. These heavy-chain antibodies also lack the CH1 domain which in a conventional antibody, associates with the light chain and to a lesser degree interacts with the VH domain (**Figure 1.6**).

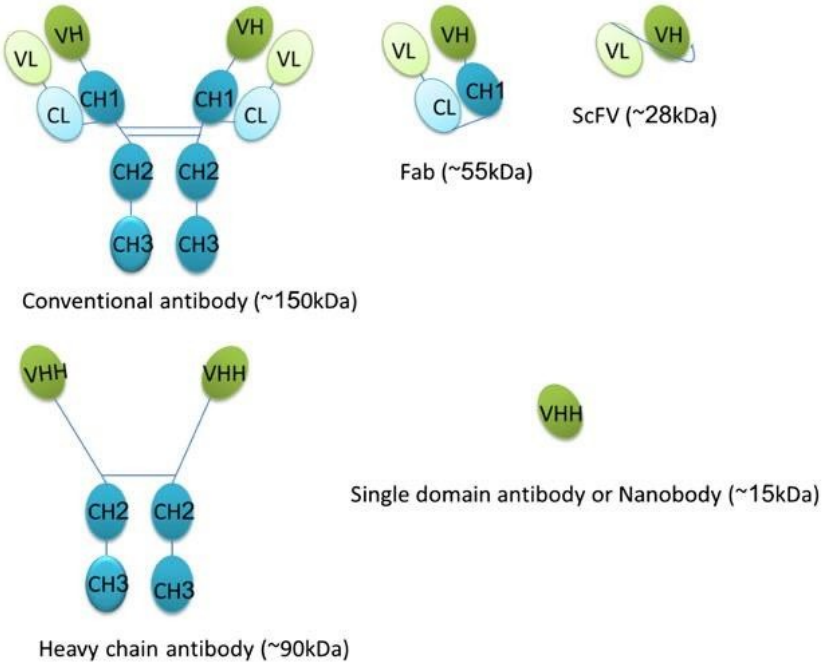


Figure 1.6 Overview of different antibody formats. *Schematic representation of conventional, heavy chain and single domain antibody (113). A conventional immunoglobulin (IgG) antibody is composed of the two heavy and two light chains. The heavy chains consist of one variable domain (VH) and three constant domains (CH1, CH2, CH3). The light chains have one variable domain (VL) and one constant domain (CL). The constant domains are joined by disulfide binding. The antigen binding fragment (Fab) is the region of an antibody that binds to antigens and is composed of one constant and one variable domain, each with a heavy and a light chain. Single-chain variable fragment (scFv) is comprised of the shortest variable-region fragment (Fv) of the VH and VL joined together by a protein loop (linker). Single domain antibodies (sdAb) are often referred to as*

nanobodies due to their size (~15kD). Nanobodies, such as camelid antibodies only have the heavy chains, with a single variable domain (VHH) (107).

Sequence analysis (117-120) and crystal structures (121, 122) has revealed that VHHs adopt a normal immunoglobulin fold consisting of nine β-strands spread over two β-sheets packed against each other and linked with a conserved disulfide bond (121) (**Figure 1.7**). The organization of the Cα-atoms within the β-strands scaffold of the nanobody superimposes nicely with that in a human or mouse VH (118). Similar to conventional VH domains, VHHs contain four conserved framework regions (FRs) that form the core structure of the immunoglobulin domain and three complementarity-determining regions (CDRs) that are involved in antigen binding and represent the paratope of the nanobody. Nevertheless, CDR3 is the main contributor for antigen recognition and specificity, whereas CDR1 and CDR2 assist in the binding strength (123-125). The CDR3 of nanobodies is on average 18 amino acids long (126, 127).

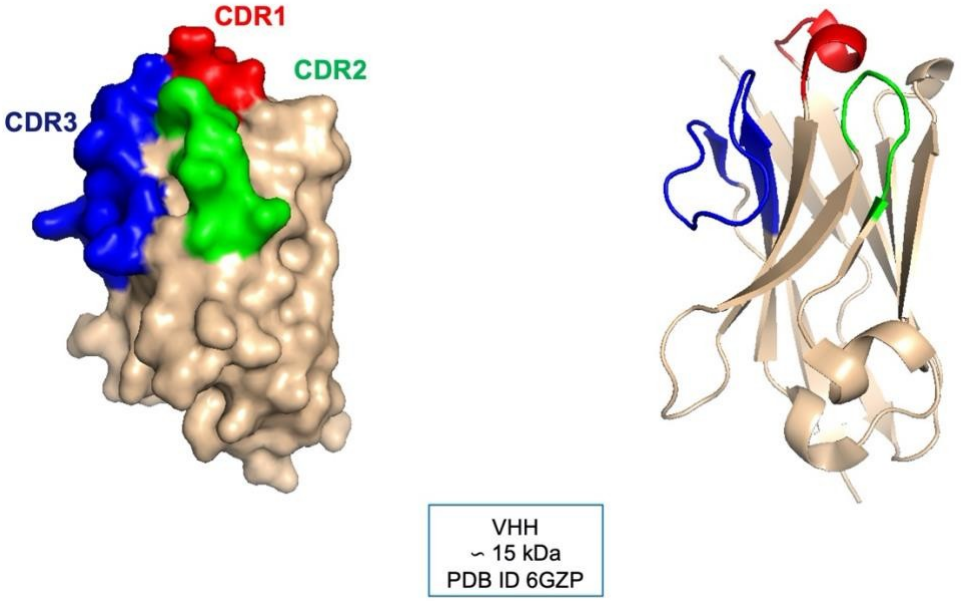


Figure 1.7 PDB structure of single-domain antibody.*Surface (left) and ribbon (right) representation of the structure of a single-domain antibody. Framework region is highlighted in white, CDR1 is highlighted in red, CDR2 is highlighted in green while CDR3 is highlighted in blue.*

These VHHs with a MW of only 15 kDa, which is at least half size of the intact antigen-binding site of a conventional antibody (i.e., the VH–VL pair), are the smallest, intact antigen-binding fragments derived from a functional immunoglobulin.

A major advantage of camelid VHHs, with respect to immunoglobulin-derived dAbs (24), is their ability to specifically recognize antigens with affinities similar to those obtained by whole antibodies despite their smaller size and the absence of the hydrophobic VH-VL interface. VHHs are also usually more stable, with melting temperatures as high as 90°C, low pH and higher resilience to detergents and denaturants (128-131). Given their small size, good tissue penetration, and low immunogenicity, VHHs have been developed for different neurodegenerative disorders such as AD, Lewy body disease, PD, and HD, in the attempt to block or prevent aggregation (132-135).

VHHs can be simply expressed in bacteria, yeasts or in other hosts as recombinant (136). Because of potential biophysical and pharmaceutical properties such as small molecule size, they have wide application in development of immunotherapeutic and diagnostic methods.

Methods to isolate antigen-specific VHHs from immune (137, 138), nonimmune (139-141), or semisynthetic (142) libraries using phage, yeast, or ribosome display are now well established. Various VHHs have been selected and produced for diagnostic and research applications, as well as for clinical applications (143).

1.5.1 Selection and production of recombinant VHHs

One of the main challenges of VHHs production is the selection and identification of the best technology for their preparation. Many recombinant expression systems have been established and optimized for this purpose, especially in yeast, bacteria, or mammalian cells. Three types of VHHs banks can be employed to retrieve antigen specific VHHs, the so-called immune, naïve and synthetic library (144).

Immune libraries are the most common option to produce VHHs (144). They require an active immunization of Camelidae animals such as camels, dromedaries, llamas, or alpacas. Immune cells containing the heavy chain-only antibodies of interest are then harvested from a simple blood draw from the animal. Because the cDNAs of the VHH are well characterized, mRNA from these cells can be isolated and PCR amplified to generate sequences for the VHH only. These can then be cloned and screened using a recombinant antibody display technology (phage, bacterial, yeast,

etc.) with resulting clones transferred to bacterial or mammalian expression vectors for larger scale production. The resulting constructs enable further manipulation such as the addition of affinity tags or cellular targeting sequences can be produced (145) (**Figure 1.8**).

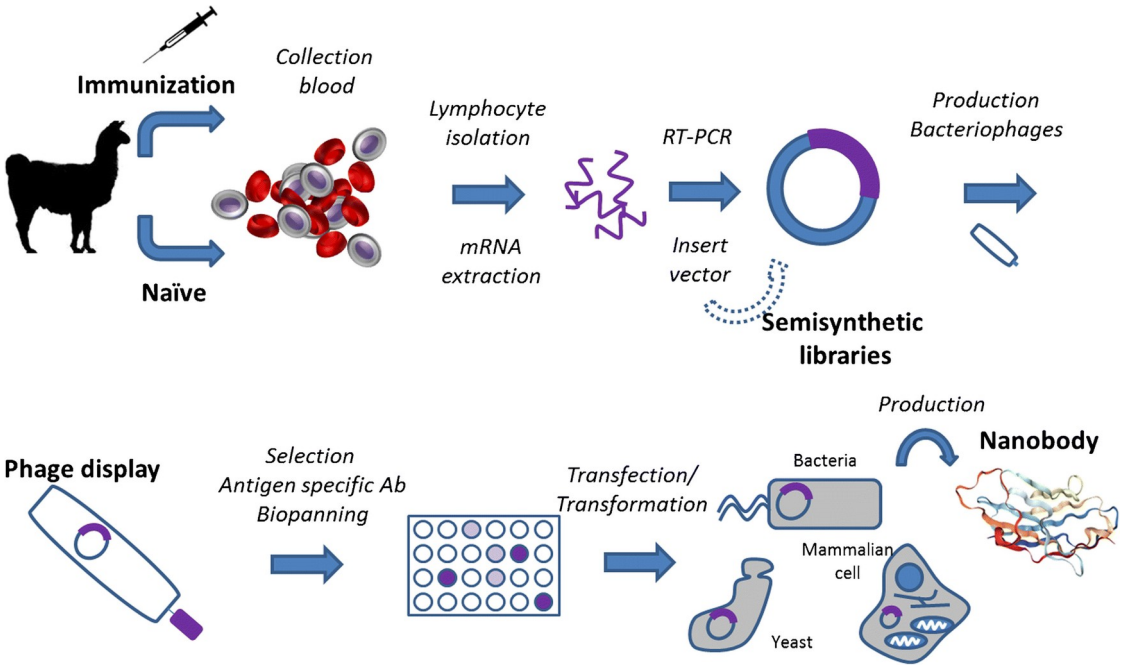


Figure 1.8 Schematic overview of the strategies for selecting and producing VHHs*. Nanobodies can be obtained by immunization using an immunogen or using non-immunized animals by collecting blood. Other format is the semisynthetic/synthetic libraries which is allows to increase the degree of diversity and the binding capability. After the selection, the nanobody can be produce in large scale through an expression vector (145).*

VHHs selection from naïve libraries takes advantage of the natural immunological diversity of the host animal, without immunization against any particular antigen. High diversity allows one naïve library to be a source of multiple nanobodies. This versatility eliminates the need to generate a new immune library for every new antigen being targeted (146, 147). Therefore, a large amount of blood is required to increase the genetic diversity. The preparation of naïve libraries does not significantly differ from the procedure used for developing immune libraries (148). Clearly, the naïve libraries are well suited to produce specific VHHs against antigens that are either non-immunogenic or highly toxic. In addition, this is a less directed method than the previous one but more universal, allowing the production of VHHs from less antigenic compounds.

Another strategy is relying on the use of semisynthetic/synthetic libraries, based mainly on keeping the common conserved structure of the VHHs and randomly varying the corresponding CDR sequences to enhance the binding capability. This format has a high level of diversity, and these types of libraries are considered a promising alternative to the conventional method (149).
A wide range of different expression models can be used for VHHs production, including organisms such as bacteria, yeast, fungi, insect cells, mammalian cells, or even plant hosts (150). The most widely used expression system is *Escherichia coli*, which expresses proteins in different cellular compartments.

1.5.2 Nanobodies as therapeutic tools

Treating central nervous system (CNS) disorders remains one of the greatest challenges in modern medicine. CNS diseases are among the most difficult to treat not only because therapeutic targets (e.g., misfolded proteins, ion channels and G-protein coupled receptors) are very complex, but also because they are 'hidden' behind brain barrier and are thus difficult to access.
While the use of monoclonal antibodies (mAbs) in targeted therapy remains limited, in recent years, a rapid progress has been made in engineering and selection of VHH domains for therapeutic purposes.
Over the decades, VHHs, have emerged as a credible next-generation antibody- derived biologics thanks to their peculiar properties, including nanoscale size, robust stability, water solubility, high affinity and specificity, deep tissue penetration (135, 151).
Currently, the Therapeutic Antibody Database lists more than 31 VHHs-based agents ranging from preclinical to clinical development. VHHs therapeutic interests include, but are not limited to, oncology, inflammation, anti-viral, neurology, anti-infection agents, and rare haematological disorders.
In a study (152) was shown that VHHs are able to cross the blood-brain barrier (BBB). The main role of the BBB is the regulation of homeostasis and protection of the brain from damaging agents. For therapeutic applications it becomes an obstacle as it prevents the passage of high molecular weight drugs. Most proteins, including antibodies, cannot be used as a shuttle to deliver drugs to targeted cells within the central nervous system. VHHs are superior to currently used antibodies for transport through the BBB because they show better penetration and more homogeneous distribution (151, 153). Once inside the cell they can alter protein folding and/or interactions.

Another important property is their application in drug delivery to access intracellular protein targets by penetrating cells (154).

New strategies aim at the production of scaffolds of smaller size that can be easily selected and produced. Thus, VHHs are particularly well suited for drug development given their biophysical and biochemical characteristics (145).

2. Aims

The main focus of this project was to select and characterize a single-domain antibody (nanobody or VHH) against TDP-43 protein, which aggregates and becomes toxic in ALS patients' neurons. This strategy allows testing how intracellular antibody can mitigate TDP-43 proteinopathy in ALS and provide a potential new tool for diagnostics.
I developed a multiplexed experimental approach, based on molecular and structural biology methods and spectroscopic techniques, that allows observing different aspects of the selected intrabody targeting TDP-43.

The **first objective** of this project was the selection of anti-TDP-43 nanobodies through the Intracellular Antibody Capture Technology (IACT) system based on the Yeast-Two Hybrids (Y2H) technique, in living yeast cells, from a naïve library of llama nanobody.

The **second objective** was the cloning, production, and characterization of the selected nanobody (VHH5). The structure of the selected intrabody was predicted *in silico* and the model was used to suggest mutations that enabled the improvement of its expression yield, facilitating its experimental biochemical validation. I observed how coupling experimental methodologies with *in silico* design may allow to obtain an antibody able to recognize specific regions of TDP-43.

The **third objective** of this project was the characterization of the antigen-antibody interaction site. Biochemical techniques and In Vivo Epitope Mapping (INVEM) were performed to characterize the epitope of TDP-43 recognized by the selected nanobody.

3. Materials and Methods

3.1 Materials

3.1.1 DNA Cloning

Restriction Enzymes and Buffers from New England BioLabs (NEB), KAPA HiFi PCR Kit (BioSystem), Q5 High-Fidelity 2X Master Mix from New England Biolabs (NEB), Agarose (SIGMA), 6x Loading Sample Buffer (SIGMA), SYBR Safe DNA Gel Stain (Invitrogen by Thermo Fisher Scientific), PowerPac™ Basic Power Supply (Bio-Rad), Mini-Sub Cell GT Systems (Bio-Rad), Monarch® PCR & DNA Cleanup Kit from New England Biolabs (NEB), T4 Polynucleotide Kinase from New England Biolabs (NEB), NEBuilder® HiFi DNA Assembly Cloning Kit from New England BioLabs (NEB), Quick Ligation Kit from New England Biolabs (NEB), NEB® 5-alfa Competent *E.coli* (High-Efficency) from New England Biolabs (NEB), Monarch® Plasmid Miniprep Kit from New England Biolabs (NEB), SupremeRun & LightRun Sequencing Services (Eurofins Genomics), pMicBD1 vector from Antonino Cattaneo's group (Scuola Normale Superiore), pET-17b vector from Dr. Francesco Aprile (Cambridge), pET-SUMO vector from Annalisa Pastore's group (King's College London), Eppendorf Mastercycler® X50 family (Eppendorf), NanoDrop™ 2000/2000c Spectrophotometers (ThermoFisher Scientific).

3.1.2 Yeast Two-Hybrid

Analytical grade chemicals, Chemidoc XRS (Biorad), GoTaq polymerase (Promega), Restriction enzymes from New England Bioscience, Wizard® *Plus* SV Minipreps DNA Purification System (Promega).

3.1.3 Expression of Recombinant Protein

BL21(DE3) Competent *E.coli* from New England Biolabs (NEB), BL21(DE3)pLysS Competent *E.coli* from Agilent Technologies, Overnight Express™ Instant TB Medium (Merck Millipore), LB (Luria Bertani) broth, Miller – Granulated (VWR CHEMICALS), New Brunswick™ Innova® 42 Incubators, Antibiotics (Sigma-Aldrich), Ultrapure IPTG (generon), Varian Cary 50 UV-Vis Spectrophotometer (Agilent), Avanti J-26XP High-Performance Centrifuge (Beckman Coulter), NuPAGE™ 12% Bis-Tris Gel (Invitrogen), SeeBlue® Marker Pre-Stained Standard 1x (Invitrogen), Sample Buffer Laemmli

2x Concentrate (Sigma), InstantBlue® Coomassie Protein Stain (abcam), Electrophoresis Power Supply- EPS 601 (GE Healthcare), Ultrasonic Sonifier® 250 (Branson) , analytical grade chemicals, autoclave machine (Systec VX-150).

3.1.4 Protein Purification

HPLC system (AKTA PURE, GE Healthcare), Hiload 16/60 Superdex 75 pg (Cytiva), Spectra/Por® 1 Dialysis Membrane MWCO 6-8 kDa (SPECTRUMLABS), Superdex 75 Increase 10/300 GL (Cytiva), Super Ni-NTA agarose resin (Generon), NuPAGE™ 12% Bis-Tris Gel (Invitrogen), SeeBlue® Marker Pre-Stained Standard 1x (Invitrogen), Sample Buffer Laemmli 2x Concentrate (Sigma), InstantBlue® Coomassie Protein Stain (abcam), Electrophoresis Power Supply- EPS 601 (GE Healthcare), NanoDrop™ 2000/2000c Spectrophotometers (ThermoFisher Scientific), VivaSpin Concetrators 5000 MWCO (Generon), analytical grade chemicals.

3.1.5 Protein Analysis and Characterization

NuPAGE™ 12% Bis-Tris Gel (Invitrogen), SeeBlue® Marker Pre-Stained Standard 1x (Invitrogen), Sample Buffer Laemmli 2x Concentrate (Sigma), InstantBlue® Coomassie Protein Stain (abcam), Electrophoresis Power Supply- EPS 601 (GE Healthcare), Nitrocellulose Membrane Filter Paper Sandwich 0.2 µm Pore Size (ThermoFisher Scientific), Pierce™ Power Blot Cassette (Thermo Scientific), 1X casein blocking buffer (Sigma), mouse anti-6X Histag (GeneTex), Alexa Fluor 680 goat anti-mouse IgG antibody (LI-COR), Odissey CLx (LI- COR), JASCO-1100 spectropolarimeter, quartz cuvette (type S3/Q/1; Starna Scientific), DichroWeb software, 800 MHz spectrometer (Avance Bruker).

3.1.6 Antigen-antibody interaction measurements

96 well-plates Nunclon™ Delta Surface, Pierce™ (Thermo Scientific, Cat.no.168055), Micro-Assay-Plate Chimney 96 well-plates (Greiner bio one, Cat.no. 655096), TMB Substrate Kit (ThermoFisher Scientific), rabbit TDP-43 Polyclonal antibody (Proteintech), goat anti-rabbit IgG HRP-linked antibody (CellSignaling Technology), rabbit anti-camelid VHH antibody [HRP] (GenScript), CLARIOstar® Plus plate reader (BMG Labtech), BS^3 Crosslinker (Thermo Scientific), NuPAGE™ 12% Bis-Tris Gel (Invitrogen), SeeBlue® Marker Pre-Stained Standard 1x (Invitrogen), Sample

Buffer Laemmli 2x Concentrate (Sigma), InstantBlue® Coomassie Protein Stain (abcam), Electrophoresis Power Supply- EPS 601 (GE Healthcare), Novex™ WedgeWell 10-20% Tris-Glycine Gel (Invitrogen), Nitrocellulose Membrane Filter Paper Sandwich 0.2 µm Pore Size (ThermoFisher Scientific), 1x casein blocking buffer (Sigma), 1-Step Ultra TMB-Blocking Solution (ThermoFisher Scientific).

3.1.7 Prediction/Modelling

PDB (Protein Data Bank), SWISS-MODEL (BIOZENTRUM, University of Basel), ABodyBuilder (SAbPred), Sphinx (SAbPred), ClusPro, PyMOL, AGGRESCAN.

3.2 Methods

3.2.1 DNA Cloning

Molecular cloning allowed the development of modern biology. Molecular cloning is the set of experimental techniques used to generate a population of organisms carrying the same molecule of recombinant DNA. The DNA molecule containing the sequence of interest can be isolated from a mixture and propagated in a suitable host cell that can direct its replication in coordination with its growth. This is usually achieved in an easy-to-grow, non-pathogenic laboratory bacterial strain of *Escherichia coli*. A single modified *E. coli* cell carrying the desired recombinant DNA can easily be grown in an exponential fashion to generate virtually unlimited identical copies of this DNA. The universality of DNA as the genetic material (except the viruses which contain RNA) means that DNA from one organism can be copied accurately and maintained in another (155). The DNA sequence encoding for the protein of interest, cDNA, is isolated and amplified by Polymerase Chain Reaction (PCR). This is a widely used molecular biology technique developed in 1983 by Kary Mullis, which allows specific amplification of a target DNA sequences by a series of repeated temperature changes, called thermal cycles (156). Once the cDNA is isolated and amplified, it is inserted into an expression vector which is usually either plasmid (circular DNA molecules that can replicate independently of the bacterial chromosome) or bacteriophages, which are essentially bacterial viruses. The vectors contain DNA sequences that enable it and the accompanying DNA to be copied (replicated) after entering suitable host cell. The host cell provides the enzymatic machinery necessary for generating copies of the recombinant DNA molecule (155). For this purpose, expression vectors must provide an inducible promoter activable by a chemical inducer, the origin of replication, the multiple cloning site (MCS) and a selection marker (generally an antibiotic resistance gene) (157, 158). Furthermore, expression vectors are often modified to introduce a nucleotide sequence encoding a fused tag or a protein upstream or downstream of the sequence encoding the protein of interest. This facilitates the purification or detection of the final protein product and/or improves the yield production of recombinant protein (158). The method of insertion of the cDNA into the vector depends on the chosen cloning strategy. Of the many molecular cloning protocols that have been developed, the following are the main techniques currently used for routine cloning: restriction digestion- and ligation-based cloning (159), Gateway cloning (160), Gibson assembly (161), and Golden Gate cloning (162). Each of these methods have specific limitations. Finally, the expression vector containing the cDNA is transferred, by

electroporation or by chemical treatment, into the host cell which can be prokaryotic or eukaryotic. The choice of the host is made to ensure the maximal amount of expression, the proper folding of the product, and to sustain the protein in the intact and functional state.

Primer design

To achieve successful DNA amplification is essential design proper primers, also referred to as oligonucleotides. A primer is a short, single-stranded DNA sequence used in the PCR technique. These primers are defined reverse and forward, depending on whether they are complementary to the filament 5' to 3' or to the inverse 3' to 5'. Generally, the criteria to be consider when designing primers are: an optimal length between 20-24 nt, the shorter the primers are, the more efficiently they will bind or anneal to the target; a melting temperature (Tm) between 65-75 °C, within 5 °C of difference between each other; GC content between 40-60 %, with the 3' end of the primer in CG to promote binding; 3 to 5 extra nucleotides at the 5' end of the primer if a restriction site is inserted, to allow efficient digestion. The annealing temperature (Ta) is generally calculated to be 3- 5 °C lower than the Tm of the primers used. However, the optimal temperature for PCR needs to be empirically determined.
The primer design allows to introduce desired modifications, such as point mutations, insertions, deletions, and recombinant Tag sequences, into the original sequence, such as the insertion of specific restriction sites at the 5' and 3' terminus of the insert sequence.

In this thesis, different cloning approaches were used such as restriction digestion- and ligation-based cloning, Gibson Assembly, deletion of DNA sequences and site-directed mutagenesis.
Each method is described for its features, and the protocols of the cloning steps are reported as follow:

- Primer design
- PCR
- Vector-Insert Ligation
- Bacterial transformation

3.2.1.1 Restriction digestion- and ligation-based cloning

Traditional (or conventional) cloning also denoted to "cut-and-paste" usually refers to the use of restriction endonucleases to generate DNA fragments with specific complementary end sequences that can be joined together with a DNA ligase, prior to transformation. This typically involves preparing both a DNA fragment to be cloned (insert) and a self-replicating DNA plasmid (vector) by cutting with two unique restriction enzymes that flank the DNA sequence and are present at the preferred site of insertion of the vector, often called the multiple cloning site (MCS).

Restriction enzymes, a class of endonucleases (REase, restriction endonucleases) are bacterial enzymes that recognize specific 4 to 8 base pairs (bp) sequences, called *restriction sites,* and then cleave both DNA strands at this site. Since these enzymes cleave DNA within the molecule, they are also called *restriction endonucleases* to distinguish them from exonucleases, which digest nucleic acids from an end. Based on their composition and structure, cofactor requirements, the nature of the target sequence and the position of the cleavage site, there are four groups of restriction enzymes: type I, type II, type III, and type IV. Type I and type III require ATP as a coenzyme for cleavage and can also catalyze DNA modification reactions such as methylation. While type I performs DNA hydrolysis at random distance from the recognition site, type III endonucleases recognize the target site and cut outside of their recognition site. These types of REase are not useful tool for molecular cloning. Type II not require ATP for their function and the cut occurs at a very specific sequence, adjacent or inside their recognition sites, producing restriction fragments of predictable length. These are the proper enzyme used in several applications. Type IV recognize instead modified, typically methylated, DNA.

Restriction endonucleases generate either "sticky ends", in which the DNA fragment has a single-stranded overhang, or "blunt ends", in which no overhang is present at the extremities of the restriction fragment. The use of two different REase leads to the formation of sticky ends, thus promoting the insert to be directionally cloning, and preventing the re-annealing of the vector alone without the insert. The restriction endonuclease pair used is chosen based on the characteristic of MCS of the vector and the absence of these sites in the sequence of the insert. The same REase can be used as a post-cloning validation means.

Primer pair for the restriction digestion- and ligation-based cloning, reported in **Table 3.1**, were designed, and used for cloning TDP-43 protein fused in-frame to LexA DNA Binding Domain (DBD)

into pMIC-BD1 expression vector. pMIC-BD1 vector was kindly provided by the Prof. Cattaneo's research group (Scuola Normale Superiore di Pisa, Italy). The primers were ordered and purchased form IDT. BamHI and PstI restriction sites were introduced respectively in the forward and reverse primer.

Table 3.1 PCR primers for cloning TDP-43 into pMicBD1

Primer name	Primer sequence	T_m
BamHI Forw	5'- cggatccgcatgtctgaatatattcgg -3'	59 °C
PstI stop Rev	5'- ttggctgcagCTAcattccccagccagaaga-3'	61 °C

Table 3.1 *Primer sequences used to insert BamHI and PstI restriction sites in TDP-43 sequence. Highlighted in green the sequence recognized by BamHI enzyme, and in cyan the sequence recognized by PstI enzyme. In red the additional base pairs necessary to allow an efficient cleavage. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

3.2.1.2 Gibson Assembly Cloning

Gibson Assembly, developed by Dr. Daniel Gibson, is a robust exonuclease-based method that permits to insert one or more DNA fragments into a vector without the use of restriction enzymes, thanks to combined action of an exonuclease. The method is carried out under isothermal condition using three enzymatic activities in one single reaction. Firstly a 5' exonuclease cuts the 5' ends of the double stranded DNA fragments generating single-stranded 3' overhangs. These overhangs facilitate the annealing of complementary fragments (overlap region). Secondly a DNA polymerase fills the gaps and thirdly the DNA ligase seals the fragments in the assembled DNA (161). In this thesis, was used the NEBuilder® HiFi DNA Assembly Cloning Kit (NEB).

Primer design is important for a successful Gibson Assembly reaction. For an optimal assembly to join two or more fragments (e.g., vector and insert), is necessary to use fifteen- to forty-base pair overlaps of the complementary strands that exhibit a melting temperature greater than 48 °C. It is important to underline that the forward primers share a region of complementarity with the reverse primers. While the 3' end of the primer must contain a sequence that is specific to the target gene.

An online tool (NEBuilder® Assembly Tool, nebuilder.neb.com) was used to design primers for the cloning of pET-17b_VHH5 (**Table 3.2**), pMIC-BD1_N-terminus, pMIC-BD1_RRM1, pMIC-BD1_RRM, pMIC-BD1_RRM1-2 (**Table 3.3**), pET-17b_SUMOtag+TEV_VHH5 (**Table 3.4**), pET-SUMO_VHH5(mut) (**Table 3.5**) and pET-SUMO_VHH5 (**Table 3.6**). The primers were ordered and purchased from IDT.

Table 3.2 PCR primers for cloning VHH5 into pET-17b

Primer name	Primer sequence	T_m
Plasmid _PelB FOR	5'- GCTGCACACCACCACCAT -3'	58 °C
Plasmid_PelB REV	5'- CGCCATCGCTGGTTGCGC -3'	63 °C
VHH_antiTDP43 FOR	5'- cggcgcaaccagcgatggcgATGAAGGTGCAATTA GTCG -3'	70 °C
VHH_antiTDP43 REV	5'- tgatggtggtggtgtgcagcCTATGATGCAGGCTGA GG -3'	70 °C

Table 1 *Primers sequences used to insert VHH5 into pET-17b expression vector. Highlighted in red and green the complementary region between forward and reverse primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

Table 3.3 PCR primers for cloning TDP-43 fragments into pMIC-BD1

Primer name	Primer sequence	T_m
pMIC-BD1 FOR	5'- GAATTCCCGGGGATCCGTCG -3'	61 °C
pMiC-BD1 REV	5'- CAGCCAGTCGCCGTTGCG -3'	65 °C
N-term_TDP43 FOR	5'- ttcgcaacggcgactggctgATGTCTGAATATATTCG GG -3'	70 °C
N-term_TDP43 REV	5'-cgacggatccccgggaattcCTAATCGGATGTTTTCT GG -3'	71 °C
RRM1_TDP43 FOR	5'- ttcgcaacggcgactggctgATGTTAATAGTGTTGGG TC -3'	71 °C
RRM1_TDP43 REV	5'- cgacggatccccgggaattcCTATTTGCAGTCACACC ATC -3'	71 °C
RRM2 FOR	5'- ttcgcaacggcgactggctgATGAAAGTGTTTGTGG	73 °C

	GG -3’	
RRM2 REV	5’- cgacggatccccgggaattcCTAGGATATATGAACG CTG -3’	70 °C
RRM1-2 FOR	5’- ttcgcaacggcgactggctgATGTTAATAGTGTTGG GTC -3’	72 °C
RRM1-2 REV	5’- cgacggatccccgggaattcCTAGGATATATGAACG CTG -3’	74 °C

Table 3.3 *Primer sequences used to insert TDP-43 fragments into pMIC-BD1 expression vector. Highlighted in red and green the complementary region between forward and reverse primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

Table 3.4 PCR primers for cloning SUMOtag+TEV into pET-17b_VHH5

Primer name	Primer sequence	T_m
pET-17b_VHH FOR	5’- ATGAAGGTGCAATTAGTCGAAAGCGG -3’	69 °C
pET-17b_VHH REV	5’- CGCCATCGCTGGTTGCGC -3’	74 °C
SUMOtag_TEV FOR	5’- cggcgcaaccagcgatggcgATGTCGGACTCAGAA GTC -3’	61 °C
SUMOtag_TEV REV	5’- tcgactaattgcaccttcatTCCCTGGAAGTACAGG TTC -3’	63 °C

Table 3.4 *Primer sequences used to insert SUMOtag+TEV into pET-17b_VHH5 expression vector. Highlighted in red and green the complementary region between forward and reverse primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

Table 3.5 PCR primers for cloning VHH5(mut) into pET-SUMO

Primer name	Primer sequence	T_m
pET-SUMO_TEV FOR	5'- TCCATGGATCTCGAGTAAAG -3'	61 °C
pET-SUMO_TEV REV	5'- TCCCTGGAAGTACAGGTTC -3'	63 °C
VHH mut FOR	5'- agaacctgtacttccagggaATGAAGGTGCAATTAG TCG -3'	59 °C
VHH mut REV	5'- ctttactcgagatccatggaTTATGATGCAGGCTGAGG -3'	61 °C

Table 3.5 *Primer sequences used to insert mutagenized VHH5 into pET-SUMO expression vector. Highlighted in red and green the complementary region between forward and reverse primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

Table 3.6 PCR primers for cloning VHH5 into pET-SUMO

Primer name	Primer sequence	T_m
pET-SUMO FOR	5'- AGACAAGCTTAGGTATTTATTCGG -3'	62 °C
pET-SUMO REV	5'- GCCCTGGAAGTACAGGTTC -3'	65 °C
VHH_IDT FOR	5'- agaacctgtacttccagggcATGAAGGTGCAATTAG TCG -3'	59 °C
VHH_IDT REV	5'- ataaatacctaagcttgtctCTATGATGCAGGCTGAGG -3'	62 °C

Table 3.6 *Primer sequences used to insert VHH5 into pET-SUMO expression vector. Highlighted in red and green the complementary region between forward and reverse primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

3.2.1.3 Deletion of DNA sequences using a PCR-based approach

This is a simple, rapid, and reliable method by using PCR which takes advantage of the circularity of plasmid DNA to amplify the entire plasmid except the region that is to be deleted. Oligonucleotide primers are designed in inverted tail-to-tail directions to amplify the expression vector together with the target sequence (163). The extraordinary simplicity of this technique makes it a valuable tool to generate DNA deletions, insertions, and base-substitution in plasmids.

Primers were constructed to amplify the entire sequence of the plasmid excluding the specific region that is to be deleted. The 5' ends of the primers are facing each other, and 3' ends oriented so that the extension will amplify the entire plasmid.

Starting from a plasmid containing the entire sequence of TDP-43, two constructs were obtained using primers reported in **Table 3.7**. C-term FOR and REV were used to remove the C-terminus domain, while N-term FOR and REV were used to remove the N-terminus domain. **Table 3.8** shows the primers used for removing Histag(6x) from pET-17b_VHH5 (mut) construct. The primers were ordered and purchased from IDT.

Table 3.7 PCR primers for cloning TDP-43 fragments into pMicBD1

Primer name	Primer sequence	T_m
C-term FOR	5'- tagctgcagccaagctaattccgggcgaatttc -3'	68 °C
C-term REV	5'-ggatatatgaacgctgattcctttaatgatcaagtcctctcc -3'	64 °C
N-term FOR	5'- aatgccgaacctaagcacaatagca -3'	61 °C
N-term REV	5'- catgcggatccccgggaattccagccag -3'	65 °C

Table 3.7 *Primer sequences used to remove C-terminus and N-terminus from TDP-43 full-length into pMIC-BD1. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

Table 3.8 PCR primers for removing Histag from pET-17b_VHH5(mut)

Primer name	Primer sequence	T_m
HisTag FOR	5'- taactcgagcagatccggctgc -3'	61 °C
HisTag REV	5'- tgatgcaggctgaggctttgggg -3'	64°C

Table 3.8 *Primer sequences used to remove HisTag(6X) in the construct pET-17b_VHH5(mut). Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

3.2.1.4 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) is an important tool to modify DNA sequences in molecular biological studies and genetic engineering. The mutation may be a substitution of bases or an insertion or deletion of nucleotides with respect to the original sequence. It is widely used to understand the protein structure–function relationships (164, 165). To introduce targeted mutations, several PCR-based methods were developed, such as overlap extension (166), inverse PCR (167), and megaprimer PCR (168, 169). The mutation can be introduced using primers that carry the desired mutation in a PCR which amplifies the entire plasmid template.

Point mutations are generated by designing a mismatch in the mutagenic primer. The length of the correctly matched sequence in the mutagenic primers should be 24–30 nt. The desired mutation should be in the middle of the primer with 10–15 perfectly matched nucleotides on each side. Several point mutations can be introduced in the same primer. The primers used are phosphorylated non-overlapping primers, and only one of them carrying the required mutation. They anneal to opposite strands on the target vector and the whole plasmid is amplified. **Table 3.9** shows the primers used. The primers were ordered and purchased from IDT.

Table 3.9 PCR primers for mutagenesis into VHH5

Primer name	Primer sequence	T_m
I15A FOR	5'- gagcgttcaaGCcggcggcagctta -3'	67 °C
I15A REV	5'- cccccgccgctttcgactaattgca -3'	65 °C
M74K FOR	5'- ttacctccaaatggattctctgacaccggatga-3'	61 °C
M74K REV	5'- gctgtgttatccTtgcgactgatagtgaaacg -3'	63 °C

Table 3.9 *Primer sequences used to mutagenize VHH5 cloned into pET17b+SUMOtag+TEV. Highlighted in red the mismatch in the mutagenic primer. Last column reports the estimated melting temperature of each primer (influencing the annealing temperature).*

PCR

PCR is a revolutionary method developed by Kary Mullis in the 1980s. PCR is based on using the ability of DNA polymerase to synthesize new strand of DNA complementary to the target template strand. A PCR reaction relies on thermal cycling: three steps with different temperatures (T) which are repeated in series (cycles). The first step is a denaturation step to amplify DNA. The two strands of the DNA template must be separated. This occurs by heating (94-98 °C) the dsDNA template to a point where the hydrogen bonds break between the base pairs opening into single strand structures. The second step consist in annealing the pairing of short DNA oligonucleotides to complementary regions of the template strands, decreasing temperature. In this step, temperature should be carefully regulated, as a too high temperature does not allow the primer annealing, while a too low temperature could result in unspecific primer pairing. Generally, to establish the best annealing temperature, a gradient of ten degrees is performed. In the third and last step the DNA polymerase synthesizes and elongates, in the 5' to 3' direction, the new DNA strand complementary to the DNA template strand by adding free triphosphate nucleotides (dNTPs). The temperature of this step is adjusted according to the features of DNA polymerase used.
This process of denaturation, annealing and extension constitute a single cycle. Multiple cycles (25-35) are required to amplify the DNA of interest. This entire reaction used to be done in a programmed thermocycler.

Importantly, before performing a cloning experiment it is always recommended to perform an *in-silico* simulation of the procedure using dedicated software for DNA sequence manipulation (several free and commercial options are available). This same software is also useful to align DNA sequences and create publication-quality plasmid maps. In this thesis, it has been used the SnapGene software.

PCR Experimental Settings

Restriction digestion- and ligation-based cloning

The PCR reaction used the KAPA Taq DNA polymerase, a B-family DNA polymerase, engineered to have an increased affinity for DNA without the need for accessory proteins or DNA binding domains. The mixture composition for the PCR reaction is reported in **Table 3.10**, while thermo-cycling conditions are reported in **Table 3.11**. For each pair of primers, it has been performed a

gradient of ten degrees to identify the best temperature annealing. Then, the PCRs were repeated at the chosen temperature. Elongation and Final extension were adjusted according to the length of the plasmid.

Table 3.10 PCR reaction mixture

Component	Final Concentration
milliQH_2O	N/A
KAPA HiFi Buffer (Fidelity or GC)	5X
KAPA dNTP Mix	10mM
Primer FOR	10µM
Primer REV	10 µM
Template DNA	1.5ng
KAPA HiFi DNA Polymerase	1U/µL

Table 3.11 PCR cycling

Step	Temperature (° C)	Time	No. of Cycle
Initial denaturation	95	2min	1
Denaturation	98	20s	35
Annealing	#	15s	
Elongation	72	*	
Final extension	72	**	1

Temperature of annealing was adjusted according to the pair of primers used.
* The time of the elongation was calculated on the length of the DNA sequence. 30 sec each 1 kb were considered.
** The time of the final extension was twice of the elongation.

After thermal-cycling reactions, the product from the PCR was analysed by 2% agarose gel electrophoresis to confirm the success of the amplification. It was incubated with DpnI enzyme for 1 hour at 37 °C, to remove the methylated DNA template. Confirmed PCR product was then purified and used as template for the restriction digestion. Empty pMIC-BD1 expression vector and TDP-43 PCR product were then digested with BamHI-HF and PstI-HF restriction enzymes to generate sticky ends.

Restriction digestion experimental setting

Several key factors must be considered when setting up a restriction endonuclease digestion. To achieve an optimal digestion, it is necessary to use the proper amounts of DNA, enzyme, and buffer components in the correct reaction volume. Each restriction enzyme has specific features. The restriction digestion reaction mixture used is reported in **Table 3.12**. DNA digestion was performed for 2 hours at 37 °C. Digestion mixtures were analysed through 1% agarose gel electrophoresis.

Table 3.12 Restriction digestion reaction

Component	Final Concentration
milliQH_2O	N/A
REase 1	1µL
REase 2	1µL
10x CutSmart Buffer	1X
DNA	1µg

Gibson Assembly Cloning

The PCR reactions used the Q5 ® High-Fidelity DNA Polymerase. Q5 DNA Polymerase is composed of a novel polymerase that is fused to the processivity-enhancing Sso7d DNA binding domain, improving speed, fidelity, and reliability of performance. The mixture composition for the PCR reaction is reported in **Table 3.13**, while thermo-cycling conditions is reported in **Table 3.14**.
For each construct a gradient of ten degrees was performed to identify the best temperature for the primers annealing. Then, the PCRs were repeated at the chosen temperature. Elongation and Final extension were adjusted according to the length of the plasmid.

Table 3.13 PCR reaction mixture

Component	Final Concentration
milliQH_2O	N/A
Primer FOR	0.5µM
Primer REV	0.5µM
Template DNA	1ng
Q5 High-Fidelity 2X Master Mix	1X

Table 3.14 PCR cycling

Step	Temperature (°C)	Time	No. of Cycle
Initial denaturation	98	1min	1
Denaturation	98	10s	30
Annealing	#	15s	
Elongation	72	*	
Final extension	72	**	1

Temperature of annealing (T_a) was adjusted according to the pair of primers used.
* The time of the elongation was calculated on the length of the DNA sequence. 30 sec each 1 kb were considered.
** The time of the final extension was twice of the elongation.

After thermal-cycling reactions, the products from the PCR were analysed by 1% and 2% agarose gel electrophoresis to confirm the success of the amplification. They were incubated with DpnI enzyme for 1 hour at 37 °C, to remove the methylated DNA template. Confirmed PCR products were then purified, and the resulting products were incubated together at 50 °C for 1 hour, as reported in **Table 3.15.** The reaction was set up in ice.

Table 3.15 NEBuilder HiFi DNA Assembly

Component	Final Concentration
milliQH_2O	N/A
Insert	250ng
Vector	100ng

NEBuilder® HiFi DNA Assembly Master Mix	1X

Deletion of DNA sequences using a PCR-based approach

In this cloning approach was used the KAPA Taq DNA polymerase as in the traditional cloning method reported above. The mixture composition for the PCR reaction is the same as in **Table 3.10**, while thermo-cycling conditions are the same reported in **Table 3.11**.

After thermal-cycling reactions, the products from the PCR were analysed by 1% agarose gel electrophoresis to confirm the success of the amplification. They were incubated with DpnI enzyme for 1 hour at 37 °C, to remove the methylated DNA template. Confirmed PCR products were then purified. Generally, amplification by PCR does not use phosphorylated primers. Therefore the 5' ends of the amplicon are non-phosphorylated and need to be treated by a kinase to introduce the 5' phosphate. The resulting products were incubated with the T4 Polynucleotide Kinase (PNK) for 30 minutes at 37 °C, as reported in **Table 3.16**.

Table 3.16 T4 Polynucleotide Kinase reaction

Component	Final Concentration
Amount of DNA	5µL
T4 Polynucleotide Kinase (PNK)	1µL
Quick Ligase Buffer	4µL

Site-Directed Mutagenesis

In this cloning method was used the Q5® High-Fidelity DNA Polymerase as in the Gibson Assembly cloning. The mixture composition for the PCR reaction is the same reported in **Table 3.13**, while thermo-cycling conditions are the same as in **Table 3.14**.

After thermal-cycling reactions, it was following the same protocol reported above for the deletion of DNA sequences using a PCR-based approach. The mixture composition for adding the 5' phosphate is the same as in **Table 3.16**.

Vector-Insert Ligation

The final step in the construction of a recombinant DNA molecule is the connection of a linear DNA fragment with appropriately linearized cloning vector. This reaction, called ligation, is generally performed by the T4 DNA ligase enzyme. The DNA ligase catalyses the formation of covalent phosphodiester linkages, which permanently join the nucleotides together.

In this experimental work were performed different types of ligations. The protocols used for each cloning method are reported below.

Ligation Experimental Setting

Restriction digestion- and ligation-based cloning

Insertion of the digested TDP-43 sequence into the digested vector was performed thanks to the action of T4 DNA ligase enzyme, which allows the formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another. The reaction mixture, reported in **Table 3.17**, was incubated 10 minutes at room temperature.

Table 3.17 Ligation reaction

Component	Final Concentration
milliQH_2O	N/A
Insert	37,5ng
Vector	50ng
10X T4 DNA Ligase Buffer	1X

Gibson Assembly

As described in paragraph 3.2.1.2, with the activities of three different enzymes involved in the reaction reported in **Table 3.15**, the product result from a Gibson Assembly reaction is a fully ligated double-stranded DNA molecule.

Deletion of DNA sequences using a PCR-based approach and Site-directed mutagenesis

After cooling the mixtures treated with PNK, the amplified linear DNAs were self-ligated using Quick Ligase enzyme. 1.5 µL were added at the mixtures and were incubated 10 minutes at room temperature.

Bacterial cells transformation

After ligation, the insert DNA is physically attached to the backbone and the complete plasmid can be transformed into bacterial cells for propagation. The protocol used is shown below and is the same for each method of cloning used.

The ligation mixture was transformed into competent *E. coli* NEB 5-alpha cells, using the heat shock method. This method involves a sudden increase in temperature which creates pores in the plasma membrane of the bacteria and allows the foreign DNA to enter the bacterial cells. 1-5 µl containing 1 pg-100 ng of plasmid DNA was added into a tube containing 10 µl of thawed *E. coli* competent cells. The mixtures were incubated on ice for 20 min, then heat-shocked using a thermoblock at 42 °C for 45 s and then placed back in ice for 5 minutes. 500 µl of LB liquid media was added and the transformed cells were incubated at 37 °C for 45 minutes. This step allows the production of the factor involved in antibiotic resistance in positively transformed cells. Under sterile conditions, recovered cells were plated onto LB Agar selective plates and incubated at 37 °C overnight.

DNA Sequencing

A DNA sequence analysis is generally performed to verify the correctness of DNA sequences following the cloning procedure. In this thesis, the DNA sequencing of each construct was performed using the SupremeRun & LightRun Sequencing Services, Eurofins Genomics (Ebersberg, Germany). The sequencing results were then analysed performing a sequence alignment aligned with the DNA sequence of the recombinant construct using the SnapGene software.

Expression vector

Three different expression vectors were used to obtain the constructs necessary for carrying out this experimental work.

pMIC-BD1

The pMIC-BD1 bait expression vector is used for the selection of antibodies against target antigen through IACT, a Yest Two Hybrid-based technique modified from Cattaneo et al. (170).

pMIC-BD1, used as bait vector, was designed to contain bacterial chloramphenicol resistance, TRP1 gene (which allows yeast containing this plasmid to grow in minimal medium lacking tryptophan) and the 2μ origin of replication. This plasmid contains the entire region of the *E. coli* LexA protein, whose expression is under the control of the yeast alcohol dehydrogenase I (ADH1) promoter, followed by a polylinker for cDNA insertion, to generate in-frame fusions to lexA.

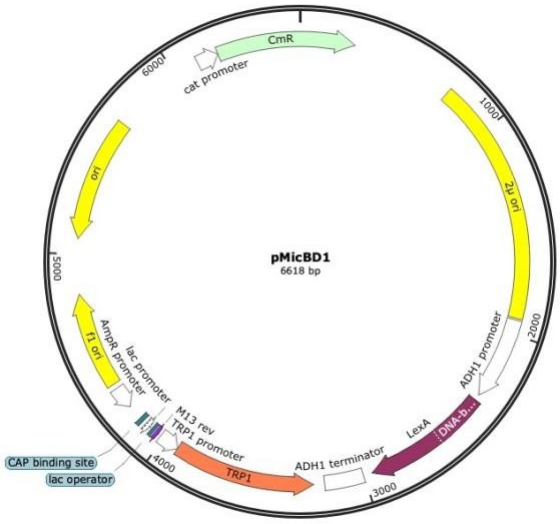


Figure 3.1 pMIC-BD1 vector map *(SnapGene).*

pET-17b

pET-17b is part of the pET vector system, a powerful and widely used system for the cloning and *in vivo* expression of recombinant proteins in *E. coli*. Activation of expression of target protein is achieved under the control of the T7 promoter. When the system is completely induced, nearly all the cell's resource are implying to expressing the gene of interest. This plasmid contains bacterial ampicillin resistance and pelB leader sequence. This leader sequence consists in a sequence of 22

N-terminal amino acid which when attached to the protein, directs the protein to the periplasmic membrane of *E. coli* where the signal sequence is removed by pelB peptidase. This expression vector was used to take advantages from the oxidative environment of the periplasmic space required for a correct protein folding and thus avoid the formation of inclusion bodies.

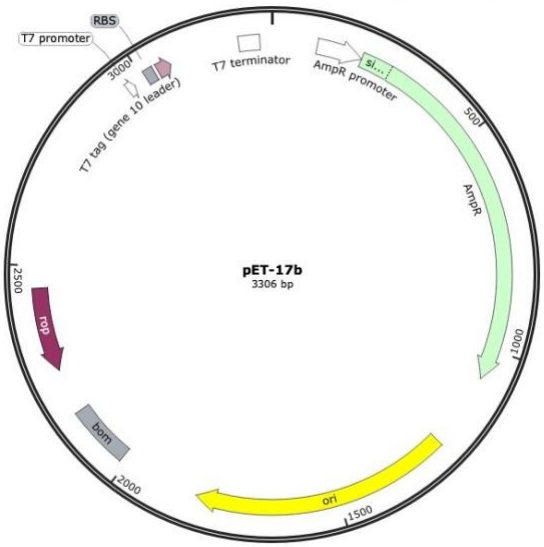


Figure 3.2 pET-17b vector map *(SnapGene).*

pET-SUMO

pET-SUMO is part of the pET vector system. This plasmid produces the highest levels of soluble protein in *E. coli*. It uses a small ubiquitin-like modifier (SUMO) fusion, belonging to the family of ubiquitin-related proteins, to enhance the solubility of expressed recombinant protein. In contrast to ubiquitin, SUMO is involved in the stabilization and localization of proteins *in vivo*. SUMO can be later removed from the protein of interest using a protease (e.g., Tobacco Etch Virus, TEV). This plasmid contains a bacterial kanamycin resistance and the promoter from the Lac operon to control the expression of transgenes. This plasmid was used to yield the protein soluble and prevent its accumulation in inclusion bodies.

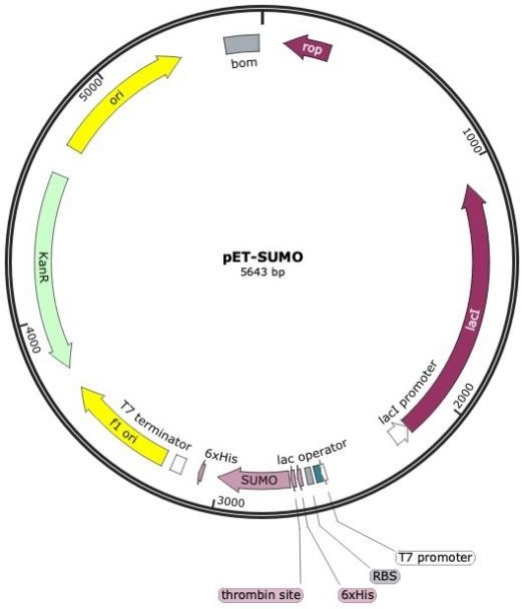


Figure 3.3. pET-SUMO vector map *(SnapGene).*

3.2.2 Yeast Two-Hybrid

The Yeast Two-Hybrid (Y2H) technique allows detection of interacting proteins in living yeast cells. Particularly, the interaction between two proteins, called bait (protein of interest) and prey (interaction partner), respectively activates reporter genes that allow the growth on specific media or the development of a colorimetric reaction. In 1989, Fields and Songs revolutionized protein interaction analysis by describing this system to detect direct protein-protein interactions in *Saccharomyces cerevisiae*. They studied the modular properties of the native GAL4 protein. It is a potent activator of transcription when yeasts are grown in galactose media and it contains both domains: N-terminal DNA binding domain (DBD) and C-terminal (transcriptional) activation domain (AD) (171). Both domains resulted to individually maintain their function independent of the presence of the other. The basic idea of Y2H is to fuse the two proteins of interest, X (called bait) and Y (called prey), to DBD and AD of Gal4, respectively. Expression of both fusion proteins in yeast and interaction between bait and prey indeed reconstituted a functional Gal4 transcription factor from the two separate polypeptides. Gal4 then recruited RNA polymerase II, leading to transcription of a *GAL1-lacZ* fusion gene. This reporter gene encodes the enzyme β-galactosidase which labels the yeast cell when using a colorimetric substrate.

Methodologically, any Y2H selection implies that yeast cells are transformed with bait and prey cDNAs cloned into different vectors under the regulation of yeast promoters. After expression in the cytosol, bait and prey must be able to enter the nucleus for the transcription of reported genes to be activated. Suitable reporter genes are the *lacZ* gene, which generates a colorimetric reaction, and auxotrophic markers (e.g., LEU2, HIS3, ADE2, URA3, LYS2) which enable growth on minimal media. More than one reporter gene is assayed in parallel to increase the accuracy of Y2H screens. In fact, one of the common problems of Y2H is the generation of false positives due to non-specific interactions (172, 173).

3.2.2.1 IACT technology

The Intracellular Antibody Capture Technology (IACT) is an in vivo selection procedure to obtain functional intracellular antibodies using a two-hybrid approach. This procedure was recently developed and is a rapid and effective selection step for the isolation of antibodies able to fold correctly and to bind antigens under condition of intracellular expression (170, 174).

For the antibody-antigen screening, expression vectors are obtained by cloning the sequences coding for target antigen at the 3′ end of the DNA binding domain (DBD) of the *Escherichia coli* protein LexA (175) and cloning recombinant antibody domains (VHH) fragments sequences at the 5′ end of the activation domain (AD) of the herpes virus 1 VP16 transcription factor (176). These vectors are cotransfected in yeast cells lacking histidine (H) production (cannot grow in the absence of histidine) but that carry the *HIS3* gene under the control of a minimal promoter fused to multimerized LexA binding sites. If the VHH fragment binds to the antigen target, a complex is formed that can bind to the *HIS3* promoter and activate transcription. This activity will restore yeast growth independently of histidine. In addition, yeast cells carry a *lacZ* gene controlled by a minimal promoter GAL1 with a LexA DNA binding site (**Figure 3.4**). The positive interaction in the transfected yeast cells provides the expression of lacZ, which encodes the β-galactosidase enzyme, providing a colorimetric assay, resulting in blue colonies in the presence of X-gal (5-Bromo-4-Chloro-3-Indolyl β-D- Galactopyranoside). The selection of positive transformants takes place in a selective media lacking TRP and LEU. L40 yeast strain once transformed with plasmids bearing TRP1 and LEU2 genes, is able to grow onto appropriate selective media lacking tryptophan (W) and leucine (L) (auxotrophic phenotype).

Finally, two nuclear localization signals are located on the VHH-VP16 fusion product, while the antigen bait has none. Therefore, the interaction between the antigen and the VHH must take place in the cytoplasm before the complex is translocated to the nucleus and activates transcription.

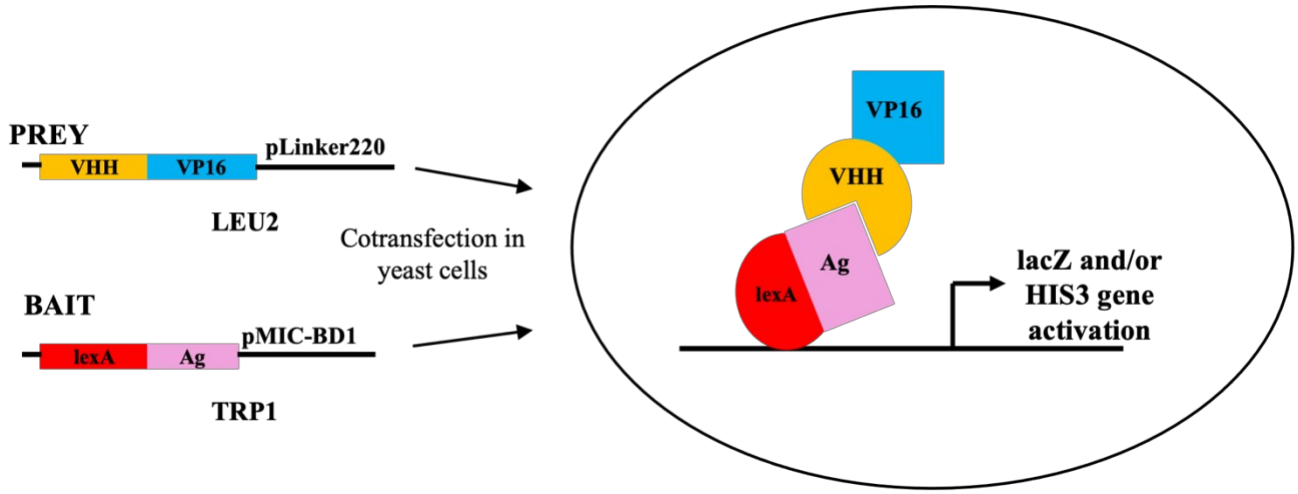


Figure 3.4. Intracellular Antibody Capture technology (IACT) schematic view*. The llama nanobody library is screened in yeast with the antigen "bait". Only those VHHs that retain the specific binding ability in vivo can activate the reporter genes, His3 and lacZ.*

Antibody Screening

Features of target/bait vectors

In this work, the target protein (TDP-43 antigen) fused in-frame to LexA DBD and llama-derived nanobodies library (from a naïve repertoire with a density of 10^6) fused to VP16 AD were co-expressed in the *Saccharomyces cerevisiae* L40 reporter strain (177). pMIC-BD1 vector was kindly provided by the Prof. Cattaneo's research group (Scuola Normale Superiore di Pisa, Italy) and has been used as a vector for the LexA-TDP-43 cloning. This vector, previously described, is designed to contain bacterial chloramphenicol resistance, TRP1 gene, which allows yeast containing this plasmid to grow in minimal medium lacking tryptophan, and the 2μ origin of replication. It also contains the entire region of the *E. coli* LexA protein, whose expression is controlled from the yeast alcohol dehydrogenase I (ADH1) promoter, followed by a polylinker for cDNA insertion, to generate in-frame fusions to LexA.

The vector for the antibody library expression was the pLinker220. This is a derivative of VP16* vector (174, 178) and contains two nuclear localization signals (NLS), the SV40 Large T antigen

followed by in-frame fusion of the VP16 acidic activation domain, whose expression is regulated again by yeast ADH1 promoter. It also contains the *LEU2* gene as nutritional selection marker, the 2µ origin of replication, a bacterial origin of replication *ColE1*, and the ß- lactamase gene for prokaryotic selection (170).

L40 yeast strain phenotype

Genotype of L40 host strain: mat-a hisΔ200trp1-901 leu2-3,112 ade2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ Gal4 (177).

L40 is an auxotroph strain for W, H and L amino acids. The bait plasmid (pMIC-BD1) contains the TRP1 gene, which allows yeast to grow in minimal medium lacking tryptophan (W). On the other side, the library plasmid (pLinker220) contains the LEU2 gene, which allows yeast to grow in the absence of leucine (L). Furthermore, antibody-antigen (prey-bait) interaction activates the reporter gene HIS3, meaning that, when the interaction occurs, cells can grow in the absence of histidine (H). However, a bait construct could transactivate leading to unspecific activation of reporter genes. For this reason, a competitive inhibitor of *HIS3* gene product, 3-Amino-1,2,4-triazole (3AT), is added into the selective media in case of transactivation of the LexA-bait protein. Cells will be able to grow in presence of 3AT only if the level of the *HIS3* gene product is sufficient to sustain the inhibitory effect of 3AT to produce enough histidine and allow cells survival. The presence of high level of HIS3 depends on the strength of the interaction between bait and prey. An increasing concentration of 3AT can therefore be used to increase the stringency of the selection.

-W	Selective for bait positive transformants
-L	Selective for prey positive transformants
-WL	Selective for bait+prey positive transformants
-WH	Selective for transactivation
-WHL	Selective for positive interactors

Media

Yeast strain, such as *S. cerevisiae* L40, are generally grown on nutrient-rich media at 30 °C, to minimize the selection of mutants.

The complete medium for the growth of yeast is represented by Yeast Extract Peptone Dextrose (YPD), as reported in **Table 3.18**. Addition of adenine hemisulfate results in Yeast Extract Peptone Dextrose Adenine (YPAD) and removing glucose results in Yeast Extract Peptone Adenine (YPA). Selective media, YNB (Yeast Nitrogen Base), were complemented with an amino acidic mixture and a salt mixture as reported in **Table 3.19**.

Table 3.18 Yeast strain media

	YPD	**YPAD**	**YPA**
Yeast extract	10 gr/L	10 gr/L	10 gr/L
Bacto-Peptone	20 gr/L	20 gr/L	20 gr/L
D-glucose	20 gr/L	20 gr/L	/
Adenine hemisulfate	/	0.1 gr/L	0.1 gr/L

*Adjust pH to 5.8. Autoclave at 120 °C for 20 minutes and then cool it down to room temperature.

Table 3.19 Media for auxotrophic S.Cerevisiae L40

	YNB	**Salts**	**AA mix**	**W (5 g/L)**	**H (10 g/L)**	**L (10 g/L)**	**H_2O**
-W	800 mL	150 mL	10 mL	/	10 mL	10 ml	20 mL
-L	800 mL	150 mL	10 mL	10 mL	10 mL	/	20 mL
-WL	800 mL	150 mL	10 mL	/	10 mL	/	30 mL
-WH	800 mL	150 mL	10 mL	/	/	10 mL	30 mL
-WHL	800 mL	150 mL	10 mL	/	/	/	40 mL

* Add 20 g/L of bacto agar for solid plates.

**If there is a transactivation, 3AT is added using a 2 M stock.

Antibody Screening Experimental Setting

Bait validation

Before it can be used, the bait must be validated. The process that leads to validation includes a test for toxicity, a test for expression (Western blot) and a test for spontaneous activation of reporters (3AT test). A validated bait is a bait that expresses the fusion protein, is not lethal for the host organism and does not spontaneously activate the selection markers. The transformation of

the bait is unique and is used for all subsequent tests. It is performed on a small scale following the protocol below.

Small Scale LiAc Transformation

S. cerevisiae L40 yeast strain was transformed with the TDP-43 bait or with an unrelated bait (LexA-Synuclein) to be used in the secondary screening as negative control of interaction. Few colonies of L40 were inoculated in 6 ml of YPD medium and incubated at 30 °C, 220 rpm overnight. The next morning, the overnight pre-inoculum was diluted into YPD to an OD_{600} = 0.3, and incubated at 30 °C, 225 rpm up to OD_{600} = 0.6. The culture was then centrifuged at 3500 rpm at room temperature (22-25 °C) for 5 min and the pellet was washed with 10 ml of sterile H_2O. The pellet was centrifugated at 3000 rpm for 5 min and then resuspended into 1.5 ml of 1xTE/1xLiAc solution. At this point, 100 µl of resuspended pellet were transferred into a tube containing 0.2 µg of the pMicBD1-TDP43 (selection bait) or pMicBD1-Synuclein (unrelated bait for secondary screening) and 0.1 mg of salmon testis DNA (STD) (previously denaturised). 600 µl of 50% PEG/1xTE/1xLiAc solution were added into each tube. Cells were mixed and incubated at 30 °C, 225 rpm for 30 min. 70 µl of sterile DMSO was added and the cells were heat shocked at 42 °C for 15 min under gentle mixing. Cells were then incubated in ice for 2 min, pelleted by centrifugation at 14000 rpm for 30 s, and resuspended into 100 µl of sterile 1xTE. Cells from each transformation were then plated onto -W plates, to monitor the transformation efficacy, and -WH plates, to monitor transactivation of the bait plasmid.

Screening

The screening procedure consists in a Maxi-Scale LiAc transformation of the TDP-43 bait with the Llama VHH library (naïve, 10^6).

First day. Some colonies of the strain expressing the LexA-TDP-43 bait were grown overnight at 30 °C, 2225 rpm into 50 ml of -W selective medium, to obtain the culture in a mid-log phase in the next day.

Second day. In the morning, the OD_{600} of the culture was measured (3.15) and the temperature of incubation was lowered to 25 °C for the whole day. In the late afternoon, the culture was diluted to an OD_{600} = 0.2 into 150 ml of -W selective media, and incubated overnight, 220 rpm at 30 °C.

Third day. The culture was diluted to OD_{600} = 0.3 into 1 L of YPAD and incubated 3 h, 220 rpm at 30 °C. When the OD_{600} was between 0.6-0.7, cells were centrifugated at 3000 rpm at room temperature (22-25 °C) for 5 min. The pellet was washed several times in a total amount of 500 ml of 1xTE. After that, the pellet was resuspended into 15 ml of 1xLiAc/1xTE. At this point, 10 mg of salmon testis DNA (STD) (previously denaturised) and 250 µg VHH llama DNA library cloned in the pLinker220 prey plasmid were added. The mixture was transferred in a flask with 140 ml of 50% PEG/1xTE/1xLiAC and incubated 30 min, 150 rpm at 30 °C. 17,6 ml of sterile DMSO was added and the cells were heat shocked at 42 °C for 15 min under gentle mixing. The flask was then put in ice for 5 min and the cells were washed three times with a total amount of 1 L of YPA (previously cooled on ice) and centrifugated at 4500 rpm at room temperature (22-25 °C) for 5 min, and recovered in 1 L of YPAD (previously heated at 30 °C) for 1 h at 30 °C.

1 ml of cells was centrifuged and resuspended into 1 ml of -WL medium to make serial dilutions for the estimation of transformation efficiency (10^{-3},10^{-4}, 10^{-5},10^{-6}, 10^{-7}, 10^{-8}). After the dilutions, the cells were plated on -WL Petri dishes and incubated at 30 °C.

The culture was split and processed as follows.

Three quarter of the culture was immediately processed. The cells were harvested at room temperature (22-25 °C), 4500 rpm for 5 min, and washed three times in a total amount of 450 ml of -WHL. The pellet was then resuspended in 5 ml of -WHL and plated on fifteen -WHL+ 2mM 3AT Petri dishes (350µl per plate). Plates were incubated at 30 °C for 4-5 days. The remaining culture was pelleted at room temperature (22-25 °C), 4500 rpm for 5 min and washed three times in a total amount of 400 ml of -WL. The pellet was then resuspended in 20 ml of -WL, were added 200 ml of -WL and it was grown over night at 30 °C at 225 rpm, to allow to the slower clones, referred to as “doublings”, to grow. The next morning “doublings” were processed in the same way and plated on five -WHL+ 2mM 3AT petri dishes. As described before, were also made serial dilutions for the estimation of transformation efficiency of “doublings” (10^{-3},10^{-4}, 10^{-5},10^{-6}, 10^{-7}, 10^{-8}).

99 clones between pre-doublings and post-doublings were picked and re-streaked onto a -WHL and -WL plates and incubated at 30 °C.

ß-gal Assay

Most yeast two-hybrid systems take advantage of the *Escherichia coli lacZ* gene encoding β-galactosidase (β-gal) as a colorimetric reporter gene, since it is very robust, and a variety of qualitative and quantitative assays are available (179, 180).

In this thesis a liquid β-galactosidase (β-gal) assay, adapted from Mockli et al., 2004 (180), was performed using a 96-well plate. A small amount of the biomass from single colonies (99 clones) was resuspended and incubated in 50 µl of lysis buffer made of 20 mM Tris HCL pH 7.5 containing 333 U/ml lyticase and incubated at 37 °C, 150 rpm for 2 h. Then, 50 µl of Z buffer, a solution made of 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7.0, X-gal at 20 mg/ml (170 µl), and β-mercaptoethanol (30 µl), was added to each well and incubated at 37 °C for 2 h. Strong prey–bait interactions were identified by the development of blue color.

Colony PCR and Fingerprint

Colony PCR and fingerprint analysis were performed only on double positive colonies (His+/LacZ+). The clones were lysed using 10 µl of buffer 20 mM Tris HCl pH 7.5, 300 U/ml Lyticase and incubated for 45 min at 37 °C, followed by Lyticase inactivation at 95 °C for 5 min. Then, the VHH of each clone was amplified by PCR. The primers were pL220 Fw (5'-AAG CTT ATT TAG GTG ACA CTA TAG-3') and pL220 Rev (5'- CTT CTT CTT GGG TGC CAT G-3'). The PCR master mix reaction was prepared as described in **Table 3.20**. The PCR reaction was performed as follows: 3 min at 95 °C, followed by 30 cycles at 95 °C for 30 s, 50 °C for 30 s and 72 °C for 40 s, 5 min at 72 °C and then 4 °C to store. PCR were analysed using 1.3% agarose gel, followed by EtBr (10 mg/ml stock, DIL 1:16000) staining. Images were acquired using Chemidoc XRS (Biorad).

Table 3.20 PCR reaction mixture

Component	Final Concentration
milliQH2O	N/A
5X Green Buffer	2x
10 mM dNTPs	0.4 mM
Primer For	1 µM
Primer Rev	1 µM

5U/ul GoTaq polymerase	1.6 U/ul

Then, the PCR mixture (8 µl/20 µl) was digested with the restriction enzymes NlaIV and AluI as reported in **Table 3.21**, for 2 h at 37 °C, to identify a specific pattern for each isolated VHH. Digested fragments were resolved using 8% polyacrylamide gel electrophoresis, followed by EtBr (10 mg/ml stock, 1:10000) staining. Images were acquired using Chemidoc XRS (Biorad).

Table 3.21 Restriction digestion mixture

Component	Final Concentration
milliQH_2O	N/A
5X Green Buffer	0.5X
AluI	0.25 µl
NlaIV	0.25 µl
10X CutSmart Buffer	1X
DNA	8 µl

DNA extraction from yeast and bacterial transformation

Once the different patterns were highlighted, six individual clones were selected to extract the prey DNA from yeast. They were inoculated in 3 ml of selective -L media and incubated at 30 °C, 225 rpm overnight. Next morning cells were pelleted by centrifugation at 13000 rpm for 2 min and lysed into 250 µl of Lyticase Buffer (2 M Sorbitol, 1 M Tris pH 8, 1 M CaCl2, 10 % v/v β-mercaptoethanol, 95 U/ml Lyticase) at 37 °C for 1 h. Plasmid DNA was then extracted following the protocol of Wizard® *Plus* SV Minipreps DNA Purification System (Promega) and quantified via nanodrop.

Then, each plasmid DNA was transformed by electroporation, using DH5α Emax cells into bacteria to obtain a pure and monoclonal preparation. ~ 100 ng of DNA were added to 100 µl of competent *E. coli* DH5a cells, transferred into an electroporation cuvette, and shocked at ~2000 V for ~2 ms. Electroporated cells were added with 900 µl of SOC medium, transferred into a snap cap tube and incubated at 37 °C, 250 rpm for 1 h to recover. 60 µl of cells were then plated onto LB-Amp (100 µg/ml) plates and incubated at 37 °C overnight.

Secondary screening

To confirm the antigen-nanobody (bait-prey) interaction of the selected candidates, and to exclude their affinity for LexA (TDP-43 fusion partner), their purified plasmid DNA were re-transformed in parallel into *S. cerevisiae* L40 TDP-43 bait and into an *S. cerevisiae* L40 unrelated bait (LexA-Synuclein), along with a prey negative control (anti-pTAU VHH), a prey positive control (anti-LexA Y1 VHH) and a transformation negative control (STD). Small-scale transformations were performed following the protocol reported above. Transformants were plated onto -WL and -WHL plates. Growth onto selective -WHL plates was expected to occur only when the candidates were transformed into L40 TDP-43 bait. When a growth on such plates is observed also for the transformed unrelated L40 bait, the possibility of a bait-prey interaction with LexA should be considered.

In vivo Epitope Mapping (IVEM) of the anti-TDP-43 VHH5

To characterize the epitope recognized by the anti-TDP-43 VHH5 the original LexA-TDP-43 bait was truncated in two fragments named LexA-N-term+RRM1-2 (residues 1-258) and LexA-C-term (residues 259-414) and transformed in L40 yeast as described above. These strains were then transformed with the pLinker220 plasmid carrying the VHH5 with the same protocol described so far and plating the cells on SD-WL or SD-WHL. To further narrow down the region carrying the epitope a second IVEM was done, splitting the region found positive (1-258) into four smaller baits, the N-terminus (1-105), RRM1 (106-176), RRM2 (192-258), and a fragment of RRMs (160-208) which contains the linker between RRM1 and RRM2. The anti-TDP-43 VHH5 was transformed in L40 yeast strains individually carrying one of the smaller baits.

3.2.3 Expression of recombinant protein

Several methods for protein expression are available. A successful expression strategy consists in the production of proteins with functional and biochemical characteristics corresponding to the biological and native state. The prokaryotic expression system *Escherichia Coli* (*E. coli*), the most

widely used (181), resulted suitable to produce the two recombinant proteins used in this experimental work: VHH5(mut) and VHH5.

There are several advantages of using *E. coli* as the host organism. (i) It has a fast growth kinetics as, in optimal environmental conditions, its doubling time is about 20 min. (ii) It easily achieved a high cell density culture. (iii) *E. coli* suitable media contain canonical and inexpensive components. (iv) Exogenous recombinant DNA can be transformed into *E. coli* competent cells.

In this thesis, two *E. coli* strains were used for protein expression: BL21(DE3) and BL21(DE3)pLysS.

BL21(DE3) competent cells do not contain the lon protease. They are also deficient in the outer membrane protease OmpT. The DE3 designation means that respective strains contain the λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter, allowing expression of the T7 RNA polymerase to be induced with IPTG.

BL21(DE3)pLysS competent cells provide tighter control of toxic protein expression and are resistant to chloramphenicol. With respect to BL21(DE3), they also contain a plasmid, pLysS, which carries the gene encoding T7 lysozyme. T7 lysozyme lowers the background expression level of target genes under the control of the T7 promoter but does not interfere with the level of expression achieved following induction by IPTG.

The bacterial expression vectors used in this experimental work were all pET (Plasmid for Expression by T7 RNA polymerase) (182) containing the T7 bacteriophage promoter and the gene for the resistance for ampicillin (pET-17b) and kanamycin (pET-SUMO). pET-17b was provided by Dr. Francesco Aprile from University of Cambridge while pET-SUMO was provided by Prof. Annalisa Pastore from King's College London. They were engineered to obtain the recombinant proteins fused to a poly-histidine tag (6xHis-tag) which allowed protein purification by Immobilized Metal Affinity Chromatography (IMAC) (183). To enhance the protein solubility and facilitate the purification, expression vectors were also modified to generate the recombinant proteins fused with the highly soluble Small Ubiquitin-like Modifier (SUMO) (184). In both two vectors, the sequence encoding for the cleavage site of Tobacco Etch Virus (TEV) protease was also inserted downstream the sequence encoding for the 6xHis-tag (185). This allowed the enzymatic removal of the SUMOtag and/or the 6xHis-tag from the purified recombinant products since they might interfere with the biophysical and structural studies (186, 187).

3.2.3.1. Small scale Expression

The best conditions of recombinant protein expression were evaluated performing expression tests in small scale for each host and recombinant protein of interest by changing some parameters of cell growth (e.g., media, temperature, time of induction, concentration of inducing agent).

Small scale Expression VHH5

One colony of pET-17b and pET-17b_SUMOtag+TEV *E. coli* strain of interest was inoculated into 10 ml of Luria Bertani medium (LB) supplemented with 50 μg/ml of ampicillin at 37 °C, 190 rpm overnight. The day after cells were diluted 1:50 into Overnight Express™ Instant TB Medium supplemented with 50 μg/ml of ampicillin and grown at 30 °C 190 rpm for 48 h. The Overnight Express™ Autoinduction medium enable regulated protein expression in *E. coli*, without monitoring the culture or adding IPTG during cell growth. The method uses media components that are metabolized differentially to promote growth to high cell density and automatically induce protein expression from *lac* promoters. Then, the cells were collected by centrifugation at 5000 rpm for 20 min at 4 °C and resuspended in lysis buffer [20 mM Tris/HCl pH 8, 150 mM NaCl, supplemented with 1 mg/ml lysozyme, a cOmplete™ EDTA-free Protease Inhibitor tablet (Roche), and 1 μg/ml DNase I], and then lysed by sonication. A SDS-PAGE was run at each step.

Small scale Expression VHH5(mut) and VHH5

One colony of each recombinant pET-SUMO *E. coli* strain of interest was inoculated into 10 ml of Luria Bertani medium (LB) supplemented with 50 μg/ml of kanamycin at 37 °C, 190 rpm overnight. The day after cells were diluted 1:50 into Luria Bertani medium (LB) supplemented with 50 μg/ml of kanamycin and grown to an OD_{600} of 0.6, before adding 0.5 mM Isopropyl-β-D-1-tiogalattopiranoside (IPTG) to induce protein expression. Two different induction methods were test, 18 °C under shaking overnight or 37 °C under shaking for 4 h. Cells were collected by centrifugation at 4000 rpm for 20 min at 4 °C and resuspended in lysis buffer [10 mM potassium phosphate buffer at pH 7.2, 150 mM KCl, 5 mM imidazole, 5% v/v glycerol supplemented with 1 mg/ml lysozyme, a cOmplete™ EDTA-free Protease Inhibitor tablet (Roche), and 1 μg/ml DNase I], then lysed by sonication. A SDS-PAGE was run at each step.

Small scale Expression ^{15}N-labeled VHH5(mut)

The same protocol used for the VHH5(mut) and VHH5 described above was followed to express the single labelled (^{15}N) VHH5(mut) protein. Cells were grown in M9 minimal medium (**Table 3.22**), using ($^{15}NH_4$)SO_4 as a sole source of nitrogen.

Table 3.22 Composition of M9 medium (1 L)

Salt solution: *	
Na_2HPO_4	6.0 gr
KH_2PO_4	3.0 gr
NaCl	0.5 gr
$^{15}NH_4(SO_4^{2-})_2$	1.0 gr
20% (w/v) D-glucose	15 ml
1 M $MgSO_4$	2.0 mM
1M $CaCl_2$	0.1 mM
D-biotin (1 mg/ml)	1.0 mg
Thiamine-HCl	10.0 μM
Trace elements solution (100x)	10.0 ml

* Mix first the salt solution with water and autoclave. Add the other components when the solution cools down.

3.2.3.2 Large scale Expression

Once the optimal cell growth parameters for recombinant protein expression have been established through small scale experiments, large scale production can be carried on.

Large scale Expression VHH5(mut) and VHH5

In details, each of the two constructs were transformed with plasmids using the standard heat shock protocol (188). One colony of pET-SUMO_VHH5(mut) and pET-SUMO_VHH5 was incubated

into 20 ml of Luria Bertani medium (LB) supplemented with 50 μg/ml of kanamycin at 37 °C, 190 rpm overnight. The day after cells were diluted 1:50 into 1 L of Luria Bertani medium (LB) supplemented with 50 μg/ml of kanamycin and grown to an OD_{600} of 0.6, before adding 0.5 mM Isopropyl-β-D-1-tiogalattopiranoside (IPTG) to induce protein expression for 4 h at 37 °C under shaking. Cells were collected by centrifugation at 4000 rpm for 20 min at 4 °C and resuspended in lysis buffer [10 mM potassium phosphate buffer at pH 7.2, 150 mM KCl, 5 mM imidazole, 5% v/v glycerol supplemented with 1 mg/ml lysozyme, a cOmplete™ EDTA-free Protease Inhibitor tablet (Roche), and 1 μg/ml DNase I], and stored at -20 °C overnight.

Large scale Expression ^{15}N VHH5

The single labelled (^{15}N) VHH5(mut) protein was produced following the same protocol used for the VHH5(mut) and VHH5 described above. M9 minimal media (**Table 3.22**) was used to grow the cells.

3.2.4 Protein Purification

3.2.4.1 Chromatography

Chromatography is a biophysical technique that enables the separation, identification, and purification of the components of a mixture for qualitative and quantitative analysis. Proteins can be purified based on features such as size and shape, total charge, hydrophobic groups present on the surface, and binding capacity with the stationary phase. Chromatography is based on the principle for which molecules in a mixture are dissolved into a fluid called mobile phase, which carries it through a system (e.g., column, capillary tube) which is called stationary phase (189). Several chromatography methods have been developed. Some of them include column chromatography, thin-layer chromatography (TLC), paper chromatography, gas chromatography, ion exchange chromatography, gel permeation chromatography, high-pressure liquid chromatography, and affinity chromatography (190-197).

The chromatographic equipment used in the experiments reported in this thesis is a FPLC system (ÄKTA pure, GE Healthcare), controlled by Unicorn 7.3 software, which also allows data evaluation. The absorbance of the UV detector was set at 280 nm, which corresponds to the wavelength at which proteins show a sub-maximum of absorbance. Generally, fractions of purified protein were

aliquoted, flash-frozen and stored at -20 °C. In this thesis, two different types of chromatography were used: immobilized-metal affinity chromatography and size exclusion chromatography, which differ based on their separation mechanisms and type of stationary phase.

Immobilized- Metal Affinity Chromatography (IMAC)

The concept of immobilized-metal affinity chromatography (IMAC) has first been formulated and its feasibility shown by Porath et al. 1975 (198). It is based on the affinity of transition metal ions such as Zn^{2+}, Cu^{2+}, Ni^{2+}, and Co^{2+} towards histidine and cysteine in aqueous solutions (199). This was extended to the design supports with immobilized metal ions to purify proteins in solutions (200). The affinity of the protein of interest for nickel mediated by the HisTag protein at N- or C-terminus usually underlies the technique. Proteins without HisTag protein are unable to bind to the nickel resin and are eluted during the loading. Resin-bound proteins are eluted by Imidazole which competes with the tagged proteins for binding to nickel. IMAC may be performed as a procedure with consecutive operations which are: column equilibration, sample introduction, removal of unbound fraction (washing), elution. The elution fractions are collected and analysed by SDS-PAGE. A second step of purification may be required according to the purification degree of the protein of interest. In this thesis, IMAC was used to purify recombinant protein with a histidine tag [VHH5(mut) and VHH5] by a nickel affinity chromatography (Super Ni-NTA agarose resin, Generon). Size exclusion chromatography was carried out as a final purification step.

Size Exclusion Chromatography

Size exclusion chromatography (SEC) or gel filtration appeared in late 1950s. SEC is an important protein biochemistry tool commonly used as a fractionation step to remove proteins and other materials that are either smaller or larger than the target protein. Size exclusion columns is filled with porous beads. The porosity of the beads is the critical parameter for protein separation, and it varies depending on the formulation. When a protein sample passed through the column, components move down the column at different rates depending on their size. Particles with small hydrodynamic volumes, as they enter into pores, have a longer path length compared to those with larger hydrodynamic volumes. This results in their separation (201, 202).

In this work, VHH5(mut) and VHH5 purified from IMAC were analysed and further purified using a Hi-Load 16/600 Superdex 75 pg GE Helthcare column. The sample was previously concentrated down to 5 ml. SEC run was performed at 1 ml/min flowrate and using the filtered SEC buffer (10 mM KPi pH 7.2 and 15 mM KCl) as the only mobile phase. Fractions were collected with 1 ml volumes and analysed through SDS-PAGE. Protein concentration was quantified using a nanodrop (Extinction coefficient: 24075 M^{-1} cm^{-1}).

VHH5 was further purified using a Superdex 75 Increase 10/300 GL (Cytiva) column. The sample was previously concentrated down to 500 µl. SEC run was performed at 0.5 ml/min flowrate and using the filtered SEC buffer (10 mM KPi pH 7.2 and 15 mM KCl) as the only mobile phase. Fractions were collected with 0.5 ml volumes and analysed through SDS-PAGE. Protein concentration was quantified using a nanodrop (Extinction coefficient: 24075 M^{-1} cm^{-1}).

3.2.4.2 VHH5 purification protocol

VHH5(mut) purification

Cells has been thawed and lysed by sonication. Sonication is a technique which uses mechanical vibration transmitted by the liquid phase to disrupt cell wall by shear stress. This results in the release of intracellular proteins upon cell cavitation. Sonication was performed on ice, at continuous duty cycle according to experimental needs, using Ultrasonic Sonifier® 250 (Branson) sonicator. Serial cycles of sonication were performed, generally consisting of 2 min of sonication and 5 min of rest. Cell extract was then clarified by centrifugation to remove cell debris. The soluble protein was recovered in the supernatant by centrifugation at 20,000 rpm for 50 min at 4 °C and purified by nickel affinity chromatography (Super Ni-NTA agarose resin, Generon) at 4 °C. His-tagged proteins were retained by the column, and then eluted with 10 mM potassium phosphate buffer at pH 7.2, 150 mM KCl with 250 mM imidazole. Before elution, 20 column volumes (CV) of 10 mM potassium phosphate buffer at pH 7.2, 150 mM KCl, 5 mM imidazole, 5% v/v glycerol were passed through the column to wash out unspecific-bounded components. Eluted fractions were collected and analysed by denaturing SDS-PAGE. Fractions containing the target protein were then pooled together and digested with (TEV) (1:5 protein construct/tobacco etch virus molar ratio) to remove the N-terminal tag. Digestion was performed while dialyzing the mixture overnight at 4 °C, with 10 mM potassium phosphate at pH 7.2, 15 mM KCl. Digested proteins were then separated from tags and TEV with a second nickel column at 4 °C. The protein was isolated with an incubation

for 45 min at 4 °C with high-salt phosphate buffer (10 mM potassium phosphate at pH 7.2, 1M KCl). Then, it was dialyzed overnight at 4 °C against 10 mM potassium phosphate buffer at pH 7.2 and 15 mM KCl. Protein was subjected to a final step of Size Exclusion Chromatography (SEC) [123] using a Hi-Load 16/600 Superdex 75 pg GE Helthcare column equilibrated with 10 mM potassium phosphate buffer at pH 7.2 and 15 mM KCl. The use of SEC as final purification step allowed the removal of both protein aggregates and degradation products resulting in high grade purity proteins samples suitable for the biophysical studies later performed. Protein concentration was quantified using a nanodrop (Extinction coefficient: 24075 M^{-1} cm^{-1}). Pure VHH5(mut) was aliquoted, flash-frozen and stored at -20 °C. The protein purity was assessed by SDS-Page.

VHH5 purification

The same protocol used for the VHH5(mut) described above was followed to produce the VHH5 protein. In this case, after the first round of SEC was performed an additional step of purification was performed. It was used a Superdex 75 Increase 10/300 GL (Cytiva), suitable for small scale preparative purification (µg-mg) as a final polishing step. The column was equilibrated with 10 mM potassium phosphate buffer at pH 7.2 and 15 mM KCl. The sample was previously concentrated down to 500 µl. SEC run was performed at 0.5 ml/min flowrate and using the filtered SEC buffer (10 mM KPi pH 7.2 and 15 mM KCl) as the only mobile phase. Fractions were collected with 0.5 ml volumes and analysed through SDS-Page. Protein concentration was quantified using a nanodrop (Extinction coefficient: 24075 M^{-1} cm^{-1}). Pure VHH5 was aliquoted, flash-frozen and stored at -20 °C.

^{15}N VHH5(mut) purification

The same protocol used for the VHH5(mut) and VHH5 described above was followed to produce the single labelled (^{15}N) VHH5 protein. In this case, the SEC step was avoided as the protein was sufficiently pure. Protein concentration was quantified using a nanodrop (Extinction coefficient: 24075 M^{-1} cm^{-1}). Pure ^{15}N VHH5 was aliquoted, flash-frozen and stored at -20 °C. The protein purity was assessed by SDS-Page.

3.2.5 Protein analysis and characterization

3.2.5.1 Western Blot

Western Blot is an important technique used in research to separate and identify proteins. A mixture of proteins is first separated based on molecular weight through gel electrophoresis and transferred to a nitrocellulose or polyvinylidene fluoride (PVDF) membrane where they appear as bands. The membrane is then incubated with labels antibodies specific to the protein of interest. The unbound antibody is washed off. The antibody specifically bound to the protein is detected by developing the membrane. The thickness of the band corresponds is directly correlated to the amount of protein present that can be quantified, provided a standard curve is performed (203, 204).

In this experimental work, the western blot was performed to verify the expression of the protein of interest, VHH5. Samples containing proteins were denatured for 1 min at 100 °C in 2X Laemmli sample buffer, and subjected to SDS- PAGE analysis, using 12% polyacrylamide gels (Invitrogen). The protein was then transferred from the gel to a nitrocellulose membrane (ThermoFisher Scientific). An electric field with constant voltage (25 V) was applied in a semi-dry blotting device (Thermo Scientific™ Pierce™ Power Blotter) for 10 min, to allow proteins migration from the gel to the membrane. The membrane was blocked for 1 hour in 1X casein blocking buffer (Sigma), to saturate all non-specific binding sites reducing the background. After three washes of 5 min each into 1X TBS supplied with 0.05% (v/v) Tween- 20 (TBST) the membrane was incubated overnight at 4 °C with a mouse anti-6X Histag (GeneTex) at a 1:500 dilution. The primary antibody is usually coupled with signal-inducing secondary antibody. The following day, after three washes into TBST, the membrane was incubated with a proper secondary antibody, Alexa Fluor 680 goat anti-mouse IgG antibody (LI-COR) at a 1:20000 dilution, for 1h at room temperature. After three washes into TBST, the secondary antibody conjugated with fluorophores was detected by fluorescence detection system Odissey CLx (LI-COR).

3.2.5.2 Circular Dichroism

CD spectroscopy is an absorption spectroscopy method, in the UV-vis range, which quantifies the unequal absorption of left-handed and right-handed circularly polarized light by chiral molecule. CD is an excellent method of determining the secondary structure of proteins (205). A CD spectrum

is obtained when the dichroism is measured as a function of wavelength (206). As the different secondary structure elements show specific fingerprints in the CD, a CD protein spectrum information on its secondary structure composition (% α-helix, β-sheet, turns, etc.) (**Figure 3.5**), and on conformational changes occurring during processes as aggregation or unfolding (206-208). In particular, α-helical proteins are characterised by two negative peaks centred at about 220 nm and 208 nm and a positive contribution at 193 nm (209). Proteins with well-defined antiparallel β- pleated sheets present a large peak centred at 218 nm and positive bands at 195 nm (210), while disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm (211). To quantify the relative proportion of each associated secondary structure contained in a protein sample, the resulting CD spectrum acquired between wavelengths of 190 to 240 nm is typically analytically interpreted as a sum of fractional multiples of reference spectra for each type of secondary structure (205). This process is conducted using a variety of mathematical tools (212) along with reference datasets of highly resolved protein structures (i.e., protein structures from X-ray crystallography and NMR spectroscopy) (206).

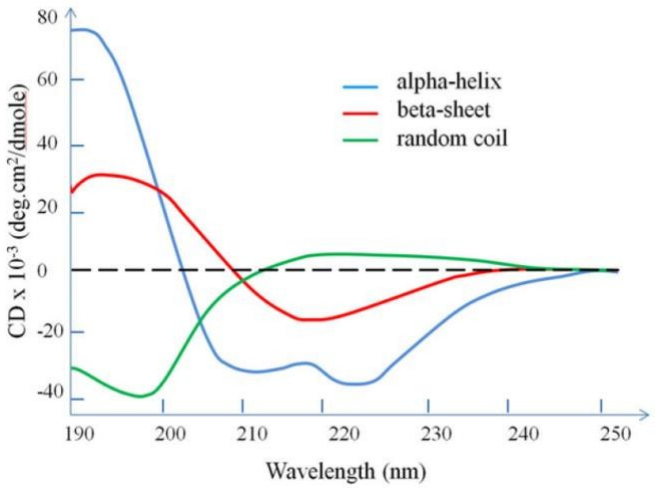


Figure 3.5. Standard CD spectra of proteins*. Typical Far-UV CD spectra indicating α-helix protein (in blue), a β-sheet protein (in red) and a random coil protein (in green) (201).*

In this work, the Far-UV CD spectra of VHH5(mut) (50 μM) was acquired at 25 °C in 10 mM potassium phosphate buffer at pH 7.2 and 15 mM KCl. CD spectra were recorded on a JASCO-1100 spectropolarimeter equipped with a temperature control system, averaged over 10 scans and deconvoluted with the online analysis software DichroWeb (213). Measurements were carried out

in 1 mm path-length quartz cuvettes (type S3/Q/1; Starna Scientific), applying a constant N_2 flush at 4.0 l/min.

3.2.5.3 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful tool for understanding the structure, dynamics, and interactions of biological macromolecules. NMR is a recognized method in a variety of scientific fields such as physics, chemistry, biology, and medicine (214). The discovery of NMR was made independently by two groups of scientists, Felix Bloch and Edward Purcell around 1946. The Nuclear Magnetic Resonance is based on the excitation of a sample upon introduction of it into a magnetic field and irradiation by a radio frequency. The irradiation excites the nuclear spins from the ground to an excited state. The signal obtained can be recorded as a plot of the energy emitted as a function of time. This signal, which contains all the information needed, can successively be Fourier transformed to a plot which reports the energy emitted as a function of frequency and corresponds to the traditional NMR spectrum. A direct-extension of this mono-dimensional spectrum is the two-dimensional one, in which the signal is recorded as a function of two times variables and the resulting data are Fourier transformed twice to yield a spectrum which is a function of two frequency variables. Apart from protons, proteins contain other magnetically active nuclei, such as ^{15}N and ^{13}C. ^{15}N-Heteronuclear Single Quantum Correlation (HSQC) experiments correlate the nitrogen atom of amines with the directly attached proton, so each signal in a HSQC spectrum represent a proton that is bound to a nitrogen (215). Since there is only one H-N per amino acid each HSQC signal represent one single amino acid. In the case of Asparagine, Glutamine and of the aromatic Tryptophane and Histidine, the HSQC spectrum also contains the signal from the side chains.

In this experimental work, a 1D and a HSQC NMR spectra of VHH5(mut) were carried out at 800 MHz on an Avance Bruker spectrometer equipped with a cryogenic probe. The sample (160 μM) was in 10 mM phosphate buffer at pH 7.2 and 15 mM KCl and 10% D_2O. 1D spectra was acquired at 25 °C.

To obtained ^{15}N single labelled protein, it was expressed in M9 medium supplemented with 1 g/L of ammonium sulfate as a sole source of nitrogen and purified as previous described. In this case, the SEC was not performed as the protein was already sufficiently pure.

3.2.6 Antigen-antibody interaction measurements

3.2.6.1 ELISA

Enzyme-linked immunosorbent assay (ELISA), the gold standard of immunoassays, is a quantitative analytical method that shows antigen–antibody reactions through a colour change obtained by using an enzyme-linked conjugate and enzyme substrate to identify the presence and concentration of molecules in biological fluids (216, 217). ELISA permits the highly sensitive and selective quantitative/qualitative analysis of antigens, including proteins, peptides, nucleic acids, and hormones.

There are different types of ELISA. (i) Direct ELISA. The surface of the plate is coated directly with the antigen and a tagged primary antibody binds to the protein of interest and enables the detection. (ii) Indirect ELISA. Requires two antibodies: a primary capture antibody that binds to the protein of interest and a secondary enzyme-linked antibody, complementary to the primary antibody, for the detection. (iii) Sandwich ELISA. It is termed a "sandwich" because the antigens are sandwiched between two layers of antibodies (capture and detection antibodies). The sandwich ELISA has the highest sensitivity among all the ELISA types (216).

In this work both Sandwich ELISA and Indirect ELISA approaches (**Figure 3.6**) were used to verify the affinity of VHH5(mut) and VHH5 to TDP-43 fragments (**Table 3.22**).

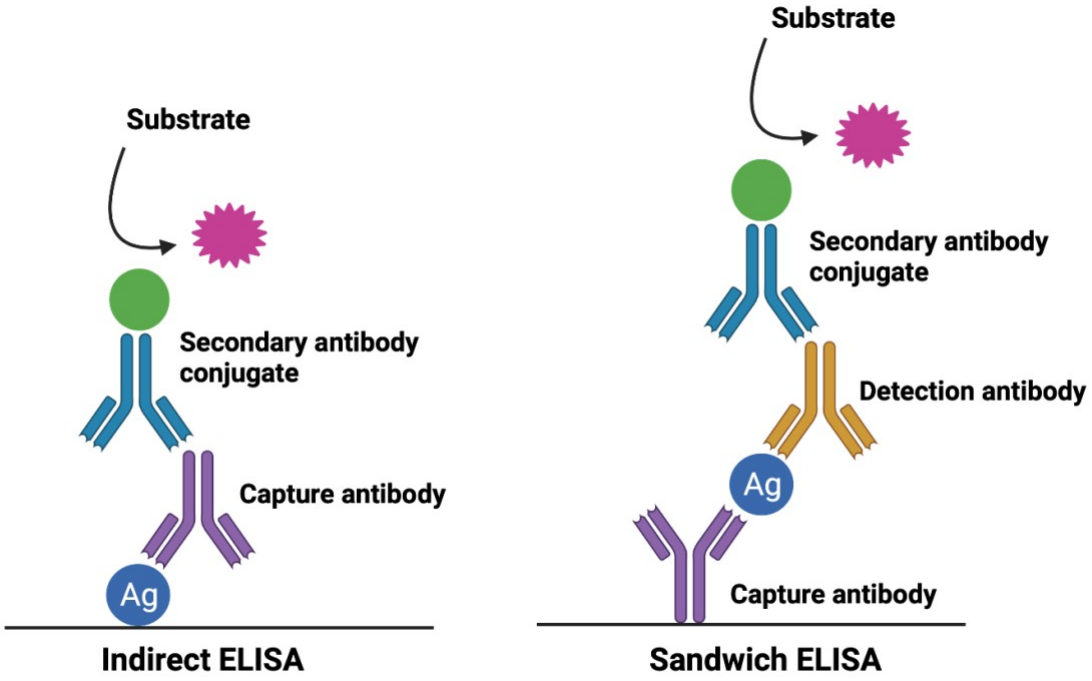


Figure 3.6. ELISA assays schematic view. *Representation of both Indirect and Sandwich ELISA highlighting the components involved in the assays. The chromogenic substrate results in a coloured reaction product that absorbs light in the visible range.The absorbance of the reaction product is typically proportional to the amount of analyte being measured.*

Table 3.23 ELISA Assay

Indirect ELISA

	Antigen	Antibody	Primary Ab	Secondary Ab
Positive control	RRM1-2, RRM1, RRM2	/	1:2000 anti-TDP-43 (Proteintech)	1:2000 anti-rIgG [HRP] (Cell Signaling)
Negative control	/	VHH5(mut)	/	1:5000 anti-camelid VHH [HRP] (GenScript)
Samples	RRM1-2, RRM1, RRM2	VHH5(mut)	/	1:5000 anti-camelid VHH [HRP] (GenScript)

Sandwich ELISA

	Capture Ab	Antigen	Primary Ab	Secondary Ab
Positive control	VHH5(mut)	/	/	1:5000 anti-camelid VHH [HRP] (GenScript)
Negative control	/	RRM1-2, RRM1, RRM2	1:2000 anti-TDP-43 (Proteintech)	1:2000 anti-rIgG [HRP] (Cell Signaling)
Samples	VHH5(mut)	RRM1-2, RRM1, RRM2	1:2000 anti-TDP-43 (Proteintech)	1:2000 anti-rIgG [HRP] (Cell Signaling)

For the **Indirect ELISA**, purified RRM1-2, RRM1 and RRM2 were coated in triplicates in a 96-well plate at a concentration of 1 µM (corresponding to 10 µg/µl), left overnight at 4 °C in carbonate buffer at pH 9.6. After coating, the reaction was blocked for 2 h at room temperature by PBS/BSA at 1% and pH 7.4. Purified VHH5(mut) (1 µM, 3 µM, 5 µM and 10 µM, corresponding to 15-150 µg/ml) was used to capture the antigen by a 2 h incubation at room temperature. Detection of VHH5(mut) and VHH5 was performed with rabbit anti-camelid VHH [HRP] antibody (GenScript) at a 1:5000 dilution with 2 h of incubation at room temperature. The chromogenic reaction was obtained by adding 3,3′,5,5′Tetramethylbenzidine (TMB) (ThermoFisher, cat. No.34021) to each well after 15-30 min incubation. Finally, it was added an equal volume of stopping solution (2 M H2SO4) and the absorbance at 450 nm was detected. The antibody dilutions were in PBS/BSA 1%, pH 7.4. The wells were washed three times between steps with PBS Tween 20 at 0.05% and pH 7.4. Wells that did not contain the antigen (TDP-43 fragments), but all the other components, were used as negative controls.

For the **Sandwich ELISA**, purified VHH5(mut) and VHH5 were coated in triplicates onto a 96-well plate at concentrations of 1 µM, 3 µM, 5 µM and 10 µM (corresponding to 15-150 µg/ml), left overnight at 4 °C, in carbonate buffer at pH 9.6. After coating, 2 h blocking at room temperature was performed in PBS/BSA at 1% and pH 7.4. Purified RRM1-2, RRM1 and RRM2 (10 µg/ml) were used to capture the VHH5 prey. The solution was incubated for 2 h at room temperature, followed by a further 2 h incubation in the presence of rabbit anti-TDP-43 polyclonal antibodies (Proteintech) at a 1:2000 dilution. Detection of the retained antigen was performed with goat anti-rIgG [HRP] antibody (Cell Signaling) at a 1:2000 dilution with 2 h of incubation at room

temperature. The chromogenic reaction was obtained by adding 3,3′,5,5′Tetramethylbenzidine (TMB) (ThermoFisher, cat. No.34021) to each well after 15-30 min incubation. Finally, it was added an equal volume of stopping solution (2 M H_2SO_4) and the absorbance at 450 nm was detected. Antibody dilutions were in PBS/BSA 1%, pH 7.4. The wells were washed three times between steps with PBS Tween 20 at 0.05% and pH 7.4. Wells that did not contain VHH5(mut) and VHH5, but all the other components, were used as negative controls.

3.2.6.2 Cross-linking

Cross-linking is the process of chemically joining two or more molecules through a covalent bond. Cross-linking reagents contain ends reactive to specific functional groups (such as primary amines or sulfhydryls) on proteins or other molecules. The availability of several chemical groups in proteins and peptides make them targets for conjugation and the use of cross-linking methods (218). Two amino acids are cross-linked when the linker length correlates with the distance between these two amino acids. Zero-length cross-linkers, consequently, do not introduce a linker and only link reactive groups in proximity. Of the available reagents, N-hydroxysuccinimide (NHS) esters are most used. One prominent, water-soluble example, used also in this experimental work, is bis(sulfosuccinimidyl)suberate (BS^3), which targets primary amines of lysine residues or the proteins' N-termini (219).

In this work, different protein ratio antigen:nanobody (1:1; 1:5 and 1:10) were used. Samples containing RRM1-2, RRM1 and RRM2 at a concentration of 10 µM were mixed with VHH5(mut) in conjugation buffer. Then the crosslinker, BS^3, was added at concentration of 2 mM. The reaction mixture was incubated at room temperature for 30 min. The reaction was quenched with the addition of Tris to a final concentration of 40 mM, for 15 min at room temperature. The species in cross-linking reaction mixture were detected by SDS-PAGE.

3.2.6.3 Far-Western Blot

Protein denaturation during SDS-PAGE prevents the detection of interactions with proteins due to the loss of secondary and tertiary structure. On the contrary, native PAGE analysis can be very useful because it allows the conservation of protein structure and the detection of a protein within

a complex. However, it presents some limitations as some protein interactions are poorly detected and large protein complexes are difficult to solve in a native gel (220). Far-Western blotting is a technique used to assay interactions that occur between natively, structured proteins. It can be used to specifically detect interactions between the protein of interest (prey) and the bait proteins immobilized on a solid support membrane, such as nitrocellulose. This is useful not only for the detection of interaction between full-length purified proteins, but also for identifying interaction between purified proteins which may be truncated or have amino acid changes (221, 222). Far-western blotting can be performed both in native conditions and under reducing and denaturing conditions followed by renaturation process (222).

In this experimental work the far-western blot was performed in native conditions to verify the interaction between TDP-43 fragments and VHH5(mut). Samples containing RRM1-2, RRM1 and RRM2 (200 µg/µl) were separated by size using 10-20% Tris-Glycine polyacrylamide gels (Invitrogen). The proteins were then transferred from the gel to a nitrocellulose membrane (ThermoFisher Scientific) in 1X transfer buffer (0.25M Tris, 1.92M glycine, 20% methanol), using a wet blotting system. The wet western blotting system, placed in an ice bath, was carried out applying a constant voltage (60 V) for 5 hours, to allow proteins migration from the gel to the membrane. The membrane was blocked for 1 hour in 1X casein blocking buffer (Sigma), to saturate all non-specific binding sites reducing the background. After three washes of 5 min each into 1X TBS supplied with 0.05% (v/v) Tween- 20 (TBST) the membrane was incubated overnight at 4°C with the purified recombinant VHH5(mut) (1 mg/ml). The following day, after three washes into TBST, the membrane was incubated with a proper secondary antibody, rabbit anti-camelid VHH [HRP] antibody (GenScript) at a 1:5000 dilution, for 1h at room temperature. After three washes into TBST, the secondary antibody conjugated with horseradish peroxidase (HRP) was detected incubating the membrane with 1-Step Ultra TMB-Blocking Solution (ThermoFisher Scientific) for 30 min at room temperature. BSA 2% was used as negative control.

3.2.7 Prediction/Modelling

3.2.7.1 Homology Modelling

The three-dimensional structure of a protein provides important information for understanding its biochemical function and interaction properties in molecular details (223). However, the

number of known protein sequences is much larger compared to the numberer of experimentally solved protein structures. All the experimentally determined protein structures are deposited in the Protein Data Bank (PDB) (224). Homology modelling (or comparative protein structure modelling) techniques have been developed to build three-dimensional models of a protein (target) from its amino-acid sequence based on an alignment with a similar protein with known structure (template). In case no suitable template structure can be identified, *de novo* (a.k.a. *ab initio*) structure prediction methods could be used to generate three-dimensional protein models without relying on a homologous template structure. However, *de novo* predictions are limited in terms of accuracy compared with comparative models. Therefore, homology modelling is the method of choice to build reliable three-dimensional *in silico* models of a protein in all cases where template structures can be identified (223). Homology modelling consists of four main steps: (i) identifying evolutionarily related proteins with experimentally solved structures that can be used as template(s) for modelling the target protein of interest; (ii) plotting corresponding residues of target sequence and template structure(s) by means of sequence alignment methods and manual adjustment; (iii) building the three-dimensional model for the target protein based on the alignment; and (iv) evaluating the quality of the predicted model (225).
In this work the homology modelling techniques was used to predict the structure of VHH5 and understand its biochemical functions and interaction properties.

Initial model generation

The structure of the antibody main scaffold, that corresponds to the β-sandwich that holds the antigen recognizing CDR loops, can be easily predicted as this region is highly conserved. The most suitable template was identified by submitting the sequence of the target protein to the BLAST search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) over the PDB database. Models were built both by the SWISS-MODEL (226) and the ABodyBuilder (227) servers.
SWISS-MODEL is a server homology modelling pipeline (226) which relies on ProMod3, an in-house comparative modelling engine based on OpenStructure (228). The target protein was submitted as amino acid sequence in FASTA format. When the search was completed, templates and target sequences were aligned, and they were filtered to remove redundancy. A set of templates was chosen according to a score which combines sequence coverage and sequence similarity. The top-ranked templates were ranked according to expected quality of the resulting

models, based on Global Model Quality Estimate (GMQE) (229) and Quaternary Structure Quality Estimate (QSQE) (230). For each selected template, a 3D protein model was automatically generated by first transferring conserved residues as characterized by the target-template alignment.

ABodyBuilder is an antibody modelling pipeline that uses the increasing knowledgebase of antibody structures to guide decision-making in modelling antibodies (227). It selects template structures based on sequence identity, and, if necessary, predict the antibody's orientation based on the ABangle parameters (231). All six CDR loops are then predicted by FREAD using a CDR-specific database. The model's side chains are then completed by SCWRL4 (232). ABodyBuilder then annotates the 'confidence' of the model as the probability that a component of the antibody (e.g., a loop or a strand) is modelled within a RMSD threshold. ABodyBuilder also highlights structural motifs within the model antibody that are known to hinder *in vitro* development (233). Finally, ABodyBuilder is the only publicly available software that is able for modelling nanobodies (e.g., camelid VHH antibodies) (227).

Loop generation

The challenge in antibody structure prediction is the design of the CDR loops, in particular the H3 loop. Modelling of the complementarity-determining region (CDR) H3 loop was carried out using the Sphinx algorithm (234), a combination of FREAD (235, 236), a knowledge-based method, and an *ab initio* algorithm. The input to Sphinx is a protein structure or a model (in PDB format), the location and sequence of the loop to be modelled. In this work the best SWISS-MODEL structure was used to model the loop region comprising residues 94 to 114. Once a complete set of decoys was generated, a statistical potential was used to reduce the set to only 500 structures, which were then scored using SOAP-Loop (237) to produce a ranking. SOAP-Loop was assessed by the average $RMSD_{global}$ of the top ranked model for each loop. From the ranking that was generated based on the frequency of how often similar conformations were selected and the energy of single conformations, were selected ten models for the loop which were used as a mould to perform the docking between the nanobody and TDP-43 (238).

Docking

Antigen-antibody binding was carried out based on the NMR structure of human TDP-43 tandem RRM1-2 in a complex with a UG-rich RNA (PDB code 4bs2) from which the RNA molecule was removed. Molecular docking was performed by using the ClusPro software (239). The standard inputs of ClusPro are two PDB files, one denoted as the ligand, the other one as the receptor. To influence docking, an attractive force was set on the residues of H3 using default parameters. The calculations were repeated on each of the ten best structures obtained by Sphinx. Cluster selection was made to exclude solutions that did not show any contact between the CDR loops and the TDP-43 ligand. An additional filtering step was included to remove all the solutions in which less than ten CDR residues were involved in molecular interactions with the antigen. A residue was defined as interacting if any of its atoms was at less than 4 Å distance from any antigen atom. Similarly, each solution was annotated based on the number of contacts with the first and second domain in the TDP-43 structure. All the representative structures from the ten ClusPro runs were then pooled together and analysed to identify conserved interaction patterns with the antigen. The interface RMSD (iRMSD) between each pair of solutions was then computed, by superimposing the antigen structure, and measuring the RMSD of the Cα atoms in the CDR regions of the respective interacting antibody. A clustering of the solution was then performed on the complete distance matrix, by using the DBScan algorithm from the Python package SciKit-Learn (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html), using the parameters eps=9 and min_clust=3 (240). The clustering results were then visualised by transforming the distance matrix to a two-dimensional space using the t-SNE algorithm in SciKit-Learn (241) (https://scikit-learn.org/stable/auto_examples/index.html). The models (238) were visualised by the Pymol software.

Sequence analysis

Protein aggregation constitutes the main bottleneck in protein production. The ability to predict the aggregation propensity of a protein from its sequence is much value for example in the discovery of more soluble variants of proteins. AGGRESCAN (242) was used to predict the aggregation properties of VHH5. The standard input for AGGRESCAN is the polypeptide sequence(s) consistent with FASTA format. In the output, the regions of the sequence with the

highest predicted aggregation propensity are highlighted in red in the peptide sequence column and appear as peaks in the profile graphs. The position of the CDR loops was obtained by the http://cao.labshare.cn/abRSA/abrsa.php server (243).

4. Results

In this section, experimental results obtained in this work will be described. The aims of the project "A new single-domain intrabody against TDP-43: selection, modelling and characterization" were to select, test, produce, and characterize a new single-domain intrabody able to recognize selectively the TDP-43 protein and to identify the epitope of the antibody interacting with the antigen in solution.

The results reported in this chapter were published in the paper: Gilodi M, Lis S, F. Dudás E, Fantini M, Puglisi R, Louka A, Marcatili P, Cattaneo A, Pastore A. *Selection and modelling of a new single-domain intrabody against TDP-43.* Front. Mol. Biosc. 2022. (238)

4.1 Antibody selection

The yeast two hybrid based IACT system was used to select intracellular specific antibodies against TDP-43 from the VHH SPLINT library (154, 158). IACT screening works by exploiting yeast L40 strains co-transformed with antigen-bait/antibody-prey pairs, in which the antigen-bait is fused to a DNA binding domain (LexA-DBD) that is challenged with a library of natural recombinant antibody domains fused to the VP16 activation domain (prey).

A collaboration with Professor Antonino Cattaneo's research group (Scuola Normale Superiore di Pisa, Pisa, Italy) was established to perform this screening. They have extensive expertise in yeast two hybrid based IACT system and have nanobodies libraries available for screening.

The TDP-43 gene (amino acids 1-414) was cloned into pMIC-BD1 vector (kindly provided by Prof. Antonino Cattaneo's research group) in fusion with LexA using the traditional (or conventional) cloning technique as described in paragraph 3.2.1.1.
Once the construct was obtained, it was shipped to Pisa where it has been transformed into L40 *S. cerevisiae* as previously described (paragraph 3.2.2.1). L40 cells containing TDP-43 bait were analysed for LexA-TDP-43 construct expression through WB analysis (**Figure 4.1, Panel A, B**). Panel A shows the detection of a band (MW ~ 67 kDa) corresponds to LexA-TDP-43 construct by using anti-LexA antibody. As reported in WB analysis, LexA-Tau-GSK3β (MW ~120KDa) was used

as positive anti-LexA control. The three bands of the selected clones show a different intensity of protein expression. The different levels of TDP-43 expression were normalized on the housekeeping PGK protein level used as loading control (Panel B) resulting in equal expression of the protein target, bait (LexA-TDP-43) in the three clones.

In addition, a toxicity test, and a test to verify the spontaneous activation of reporters were performed by growing cells on a media lacking Tryptophan, Histidine and Leucine amino acids (SD-WHL plates) (**Figure 4.1 Panel C, D**). Panel C, on the left-hand side, highlights the expression of LexA-TDP43 in L40 strain on a media lacking Tryptophan (SD-W), selective for positive bait transformants. In the middle panel is shown the growth of the construct on a media lacking Tryptophan and Histidine (SD-WH) selective for the transactivation. As shown in the picture, the bait transactives. Consequently, the 3AT test was performed. 3AT, a competitive inhibitor of *HIS3* gene product, decreases the cell growth background as it slows down the synthesis rate of *HIS3* gene product to a suitable amount to produce enough histidine and allow cells survival. The 3AT test identified 2 mM as suitable amount to switch off the bait transactivation. Finally, on the right, is shown the growth of three single clones of LexA-TDP-43 re-streak on a SD-WH plates.

Panel D highlights the controls used for the expression of LexA-TDP-43 construct on selective SD-WL and SD-WHL. The construct was transformed with Salmon Testis DNA (STD) as transformation negative control, anti-Y1 VHH as interactor positive control, and anti-pTAU VHH as interactor negative control. On the left, the SD-WL plate (selective for bait+prey positive transformants) shows the negativity for STD and the positivity for the two interaction controls. As expected, the negativity for the interactor negative control was confirmed by SD-WHL plate (selective for positive interactors), shown on the right.

Based on this analysis, the protein was successfully expressed. It was observed a bait transactivation and it was switched off adding 2 mM 3AT to the media. The bait was confirmed to be suitable as such for the screening method.

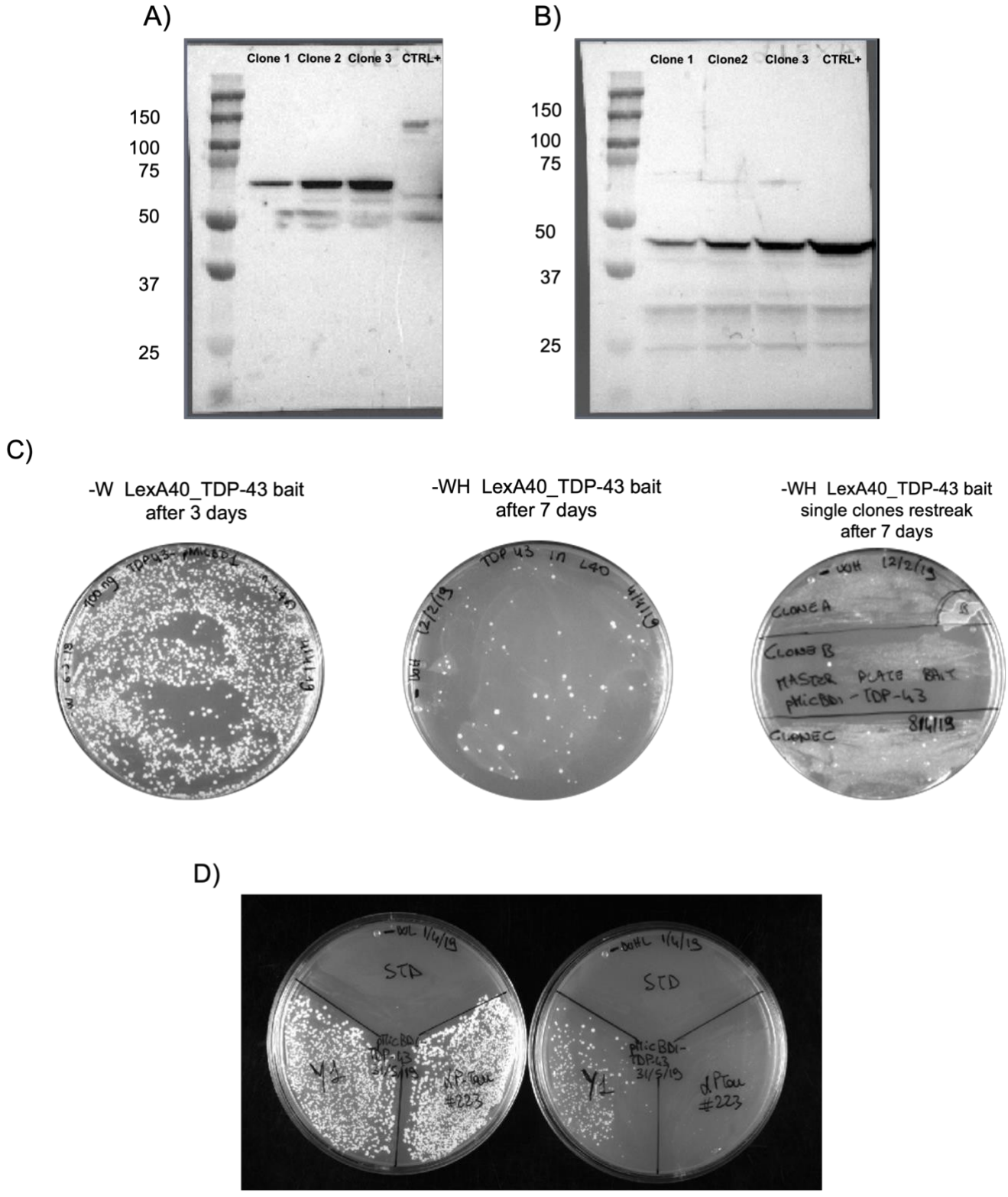


Figure 4.1 LexA-TDP-43 bait controls. *A) anti-LexA Western blot on L40 TDP-43 bait (protein extraction); clone 1, 2, 3: LexA-TDP-43 (~* 67 kDa*); CTRL+: positive pMIC-BD1_LexA-Tau-GSK3β (120 kDa). B) anti-PGK Western blot on L40 TDP-43* bait and *CTRL+* protein extracts *(PGK: 42 kDa). C) L40 S. cerevisiae transformed with pMIC-BD1_lexA_TDP-43 construct plated onto -W plate (selective for transformants), -WH (selective for interactors), 3 clones re-streaked onto -WH plate (selective for transactivation). D) pMIC-BD1_TDP-43*

construct transformed on selective SD-WL and SD-WHL (selective for interactors) with Salmon Testis DNA (STD) as transformation negative control, anti-LexA Y1 VHH as interactor positive control (interaction) and anti-pTAU VHH as interactor negative control.

I spent some months in Catteneo's research group at Scuola Normale Superiore di Pisa to carry out the screening of a VHH llama-derived nanobody library (naïve, 10^9) co-expressed into L40 *S. cerevisiae* reporter strain with the target protein TDP-43 (bait) (157) as previously described (paragraph 3.2.2.1).

The selection was performed using a SD-WHL plates (lacking Tryptophan, Histidine and Leucine amino acids) with the addition of 2 mM 3AT, as previously described. A quarter of cells were immediately processed while the remaining culture, called doubling clones, was grown overnight (n° of duplications: 8.26). 99 clones between pre-doublings and post-doublings were picked and re-streaked onto a SD-WHL and SD-WL plates. A positive interaction between a prey and the bait activates transcription of the *HIS3* gene, allowing survival in selective media (SD-WHL) and of the *LacZ* gene a second marker of interaction. A liquid β-galactosidase (β-gal) assay, adapted from Mockli et al., 2004 (180), was performed to identify positive prey-bait interactions by the development of a blue colour as previously described in paragraph 3.2.2.1 (**Figure 4.2, Panel A**). Panel A shows the β-gal assay in a 96 well-plate of the 99 clones. The positive interaction is highlight in almost all clones, both pre-doublings (red) and post-doublings (yellow). Colony PCR and a fingerprint analysis were performed to identify clones' different patterns and then single clones were isolated (paragraph 3.2.2.1) (**Figure 4.2, Panel B, C**). The VHH present in each clone was amplified as reported in Panel B. DNA fingerprint, shown in panel C, was performed to identify for each isolated VHH (clone) a specific pattern. Six clones were selected (**Figure 4.2, Panel C**). Among them, five different patterns were identified as positive clones in the β-gal assay except one which was tested negative (clone #93) and used as negative control.

After a primary selection, a secondary screening was performed to verify the positivity and to point out a candidate (VHH) as a positive TDP-43 interactor (paragraph 3.2.2.1).

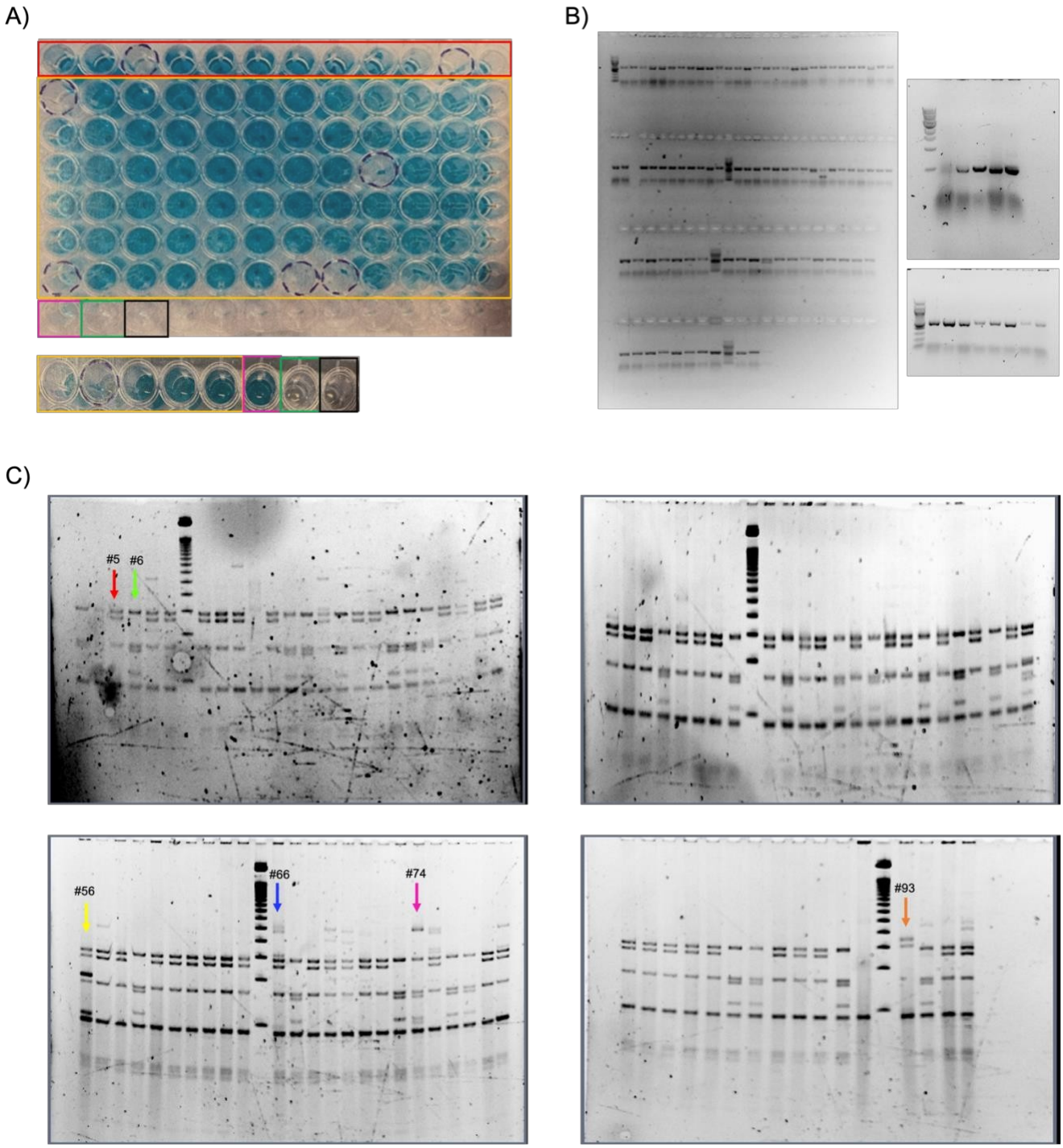


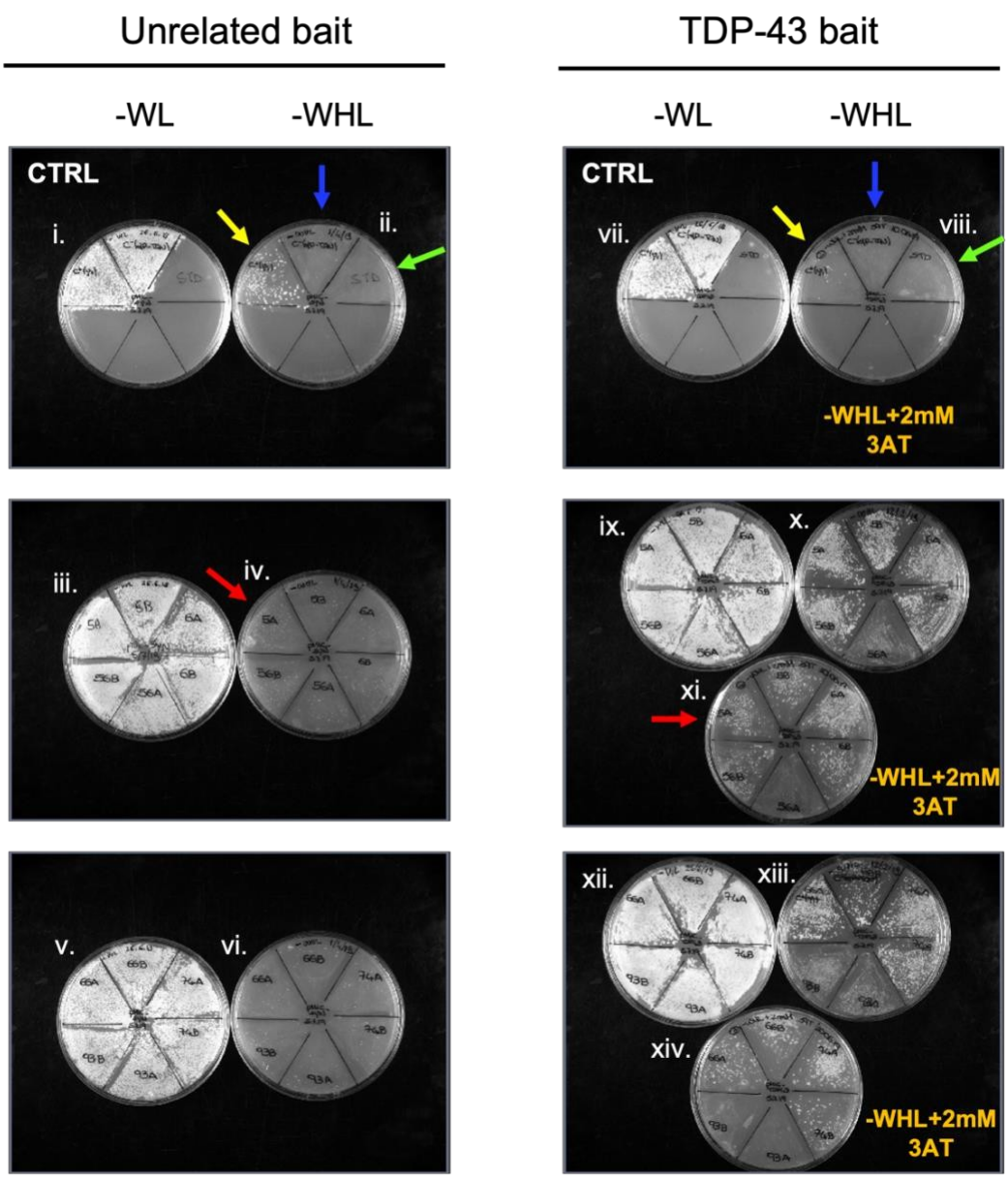
Figure 4.2 Selection of VHH. *A) ß-gal assay (liquid). Red: pre-doubling clones; yellow: post-doubling clones; magenta: positive control; green: negative control; black: blank. B) Colony PCR. C) DNA fingerprint. Arrows: red: first pattern (#5); green: second pattern (#6); yellow: third pattern (#56); blue: fourth pattern (#66); magenta: fifth pattern (#74); orange: sixth pattern (#93).*

A secondary screening was performed on selective media (SD-WHL) using either the screening bait (LexA-TDP-43) or an unrelated bait (LexA-Synuclein) to exclude interactions between VHHs and the LexA domain of the fusion protein bait (**Figure 4.3, Panel A, B**). As reported in panel A, two distinct clones of each six selected patterns have been taken into consideration. On the left, is shown a negative interaction of the selected VHHs with the unrelated bait, excluding any interaction with the LexA domain. Contrary, on the right, using SD-WHL + 2 mM 3AT plates the interaction between the selected VHHs and TDP-43 protein is shown and confirmed. The six selected clones were sequenced and found to be sequentially identical to each other. The presence of different patterns was probably due to the polyclonal nature of clones. The different patterns in clones #56, #66 and #74 were explained by the presence of two antibodies in the cell. Unlike to bacteria, in yeast cells more than one plasmid carrying an antibody could enter. The different patterns observed with the fingerprint in the clones #5 and #6 were due to the presence of an additional *Alu* cleavage site in clone #6. However, the sequence of the two clones was the same and the selected antibody was only one and will from now on be called VHH5.

Panel B shows the zoom of the positive candidate, VHH5, interacting both with TDP-43 and the unrelated bait.

Activation of the second reporter marker *LacZ* was assessed in a liquid β-gal assay. VHH5 interaction with LexA-TDP-43 gave positive β-gal assay as compared to the positive control of the assay (interaction of LexA-TDP-43 with the Y1 anti-LexA nanobody) and the negative control (interaction of LexA-TDP-43 with a scFv anti p-Tau) (**Figure 4.3, Panel C**).

A)


Unrelated bait
TDP-43 bait
-WL
-WHL
-WL
-WHL
CTRL
i.
ii.
CTRL
vii.
viii.
-WHL+2mM
3AT
iii.
iv.
ix.
x.
xi.
-WHL+2mM
3AT
v.
vi.
xii.
xiii.
xiv.
-WHL+2mM
3AT

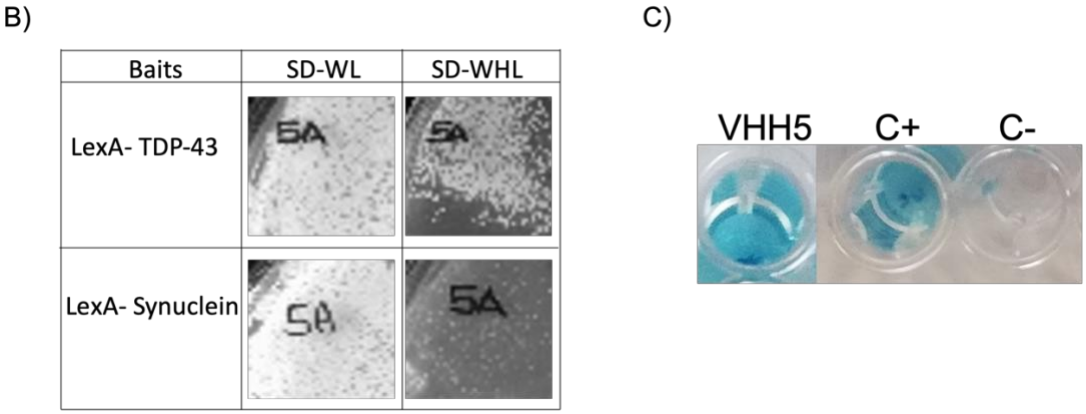
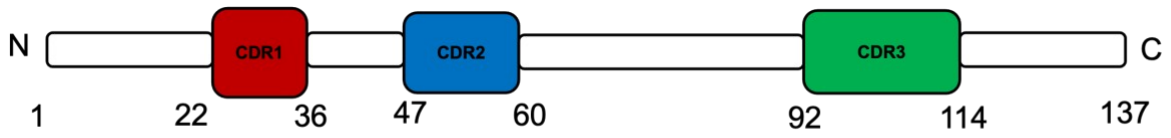


Figure 4.3 Secondary screening. *A)Growth on selective plate SD-WHL (and SD-WHL+2 mM 3AT) of the VHH5 co-transformed with the LexA-TDP-43 bait (right column) and an unrelated LexA-Synuclein bait (left column). L40 unrelated bait transformed with selected clones (six) plated on -WL (selective for transformants) (i. iii, v) and -WHL (selective for interactors) (ii, iv, vi) plates. L40 CD19 bait transformed with selected clones (six) plated on -WL (selective for transformants) (vii, ix, xii), -WHL and -WHL+ 2 mM 3AT (selective for interactors) (viii, x, xi, xiii, xiv) plates. Arrows: blue: interaction negative control; yellow: interaction positive control; green: transformation negative control; red: candidate (VHH5). The images were acquired using Chemidoc XRS (Biorad). B) Zoom on the plates of the positive candidate, VHH5. C) Liquid β-gal assay of yeast co-expressing the LexA-TDP-43 bait and the VHH5 intrabody, C+: LexA-TDP-43+ Y1, an anti-LexA intrabody, C-: LexA-TDP-43+scFv anti-pTau. The images were acquired using HUAWEI Mate 10 lite.*

A single anti-TDP-43 nanobody was selected through yeast two hybrid IACT based technology: VHH5.

Analysis of the selected nanobody revealed a short charged CDR1 (H1) loop (22-36), a shorter CDR2 (H2) loop (47-60) containing a Trp in position 53, and a rather long CDR3 (H3) loop (92-114), comprising 17 residues according to Chotia and Lesk numbering system (112). This loop is about ten residues longer than the average H3 in antibodies, but within average for intrabodies (**Figure 4.4**). It does however contain many degrees of freedom, making prediction of its structure not straightforward.



VHH5 anti-TDP-43 aminoacid sequence

KVQLVESGGGSVQIGGSLRLSCRISEGSGNKYVMAWFRQAPGQEREFVGVISWSGTRTHYADSVRGRFT
ISRMDNTAYLQMDSLTPDDSGVYYCAASVSVPPFLELTAAHFGSWGQGTQVRVGSSPEPKTPKPQPAS

Figure 4.4 Overview of VHH5 anti-TDP-43. *Schematic representation of VHH5 with the position of the CDRs and amino acid sequence, as defined using the Chothia and Lesk numbering scheme (112) in the http://cao.labshare.cn/AbRSA/abrsa.php server (243).*

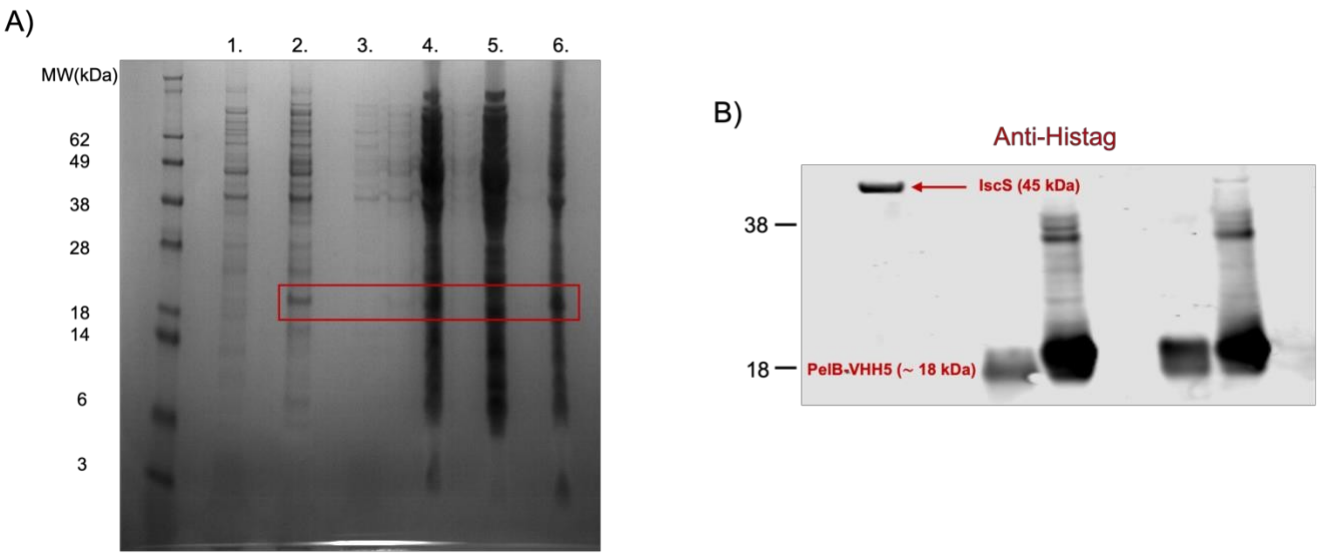
4.2 Attempts to characterize recombinant VHH5 by E. coli overexpression

With the purpose of characterizing the anti-TDP-43 VHH5, the expression and purification in *E. coli* was attempted. VHH5 was first inserted into a pET-17b expression vector (kindly provided by Dr. Francesco Aprile, University of Cambridge, United Kingdom) fused with the PelB leader sequence that directs proteins to the periplasmic space allowing disulfide bridge formation. The cloning was made using the Gibson Assembly cloning technique as described in paragraph 3.2.1.2. The construct was first transformed in *E. coli* BL21(DE3) and then in BL21(DE3)pLysS cells, as described in paragraph 3.2.3.1, for small-scale expression test. The absence of protein expression in BL21(DE3) may be due to the plasmid instability in this strain or because the target overexpression is toxic. This led to a change in *E. coli* strain to BL21(DE3)pLysS. These cells provide a tighter control of the expression of toxic proteins. Cells were grown in an Overnight Express Instant TB Medium which enables a regulated protein expression in *E. coli* without needing to monitor the culture or adding IPTG during cell growth (paragraph 3.2.3.1). As reported in the SDS-Page analysis (**Figure 4.5, Panel A**), the protein (MW ~ 18 kDa) resulted poorly overexpressed. Despite the weak yield, the protein expression was confirmed by western blot (**Figure 4.5, Panel B**). Panel B shows a band (MW ~ 18 kDa) that corresponds to the PelB+VHH5 construct, detected by using a mouse anti-6X Histag antibody (GeneTex) as described in paragraph 3.6.1.2As reported in WB analysis, IscS (45 kDa) was used as positive control.

Using the Gibson Assembly cloning technique (as described in paragraph 3.2.1.2), a SUMOtag downstream to the PelB sequence, to increase the protein solubility, and a $(His)_7$-tag downstream to the nanobody sequence were added. The construct was transformed in *E. coli* BL21(DE3)pLysS

cells that were then grown in an autoinduction medium as described in paragraph 3.2.3.1. The expression yield was appreciably increased, however the highly expressed protein accumulated in the cytoplasm as inclusion bodies as shown in the SDS-Page (**Figure 4.5, Panel C**). The protein expression was confirmed by western blot (**Figure 4.5, Panel D**). WB analysis shows the detection of a band (MW ~ 26 kDa) that corresponds to the PelB+SUMO+VHH5 construct, detected by using a mouse anti-6X Histag antibody (GeneTex) as described in paragraph 3.2.5.1. As before, IscS (45 kDa) was used as positive control.

All attempts to avoid precipitation failed, including changes of the induction temperature. Inclusion bodies formation resulted from the conflict between aggregation and protein fold. This is a well-known impediment particularly in antibody production (244).



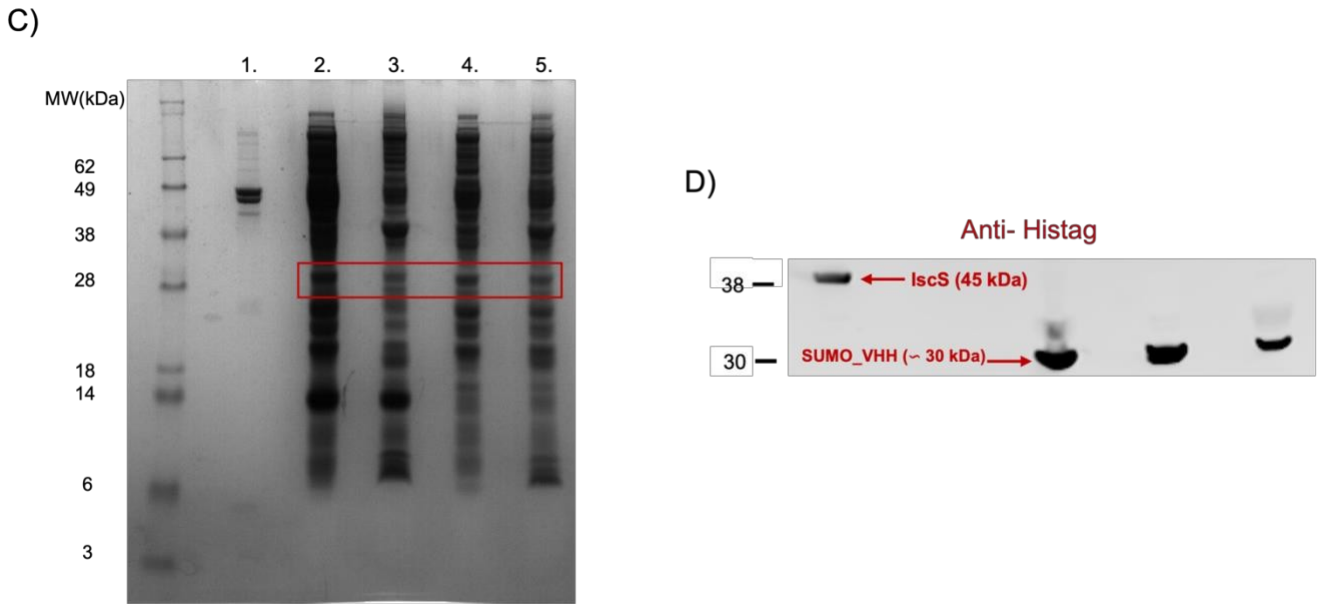


Figure 4.5. Attempts to produce recombinant VHH5 in E. coli. *A) SDS-PAGE analysis of PelB+VHH5 (18 kDa) in BL21(DE3)pLysS showing poor expression of the construct. Each column corresponds to a sample as following: lane 1, first day of growth; lane 2, second day of growth; lane 3, pre-lysis supernatant; lane 4, pre-lysis pellet; lane 5, post-lysis supernatant; lane 6, post-lysis pellet. B) Western Blot analysis of PelB+ +VHH5(18 kDa) using an Anti-Histag antibody. IscS protein (45 kDa) was used as positive control. C) SDs-PAGE analysis of PelB+SUMOtag+VHH5 (26 kDa) in BL21(DE3)pLysS showing an increase in the expression. Each column corresponds to a sample as following: lane 1, IscS, protein; lane 2, pre-lysis supernatant; lane 3, pre-lysis pellet; lane 4, post-lysis supernatant; lane 5, post-lysis pellet. D) Western Blot analysis of PelB+SUMOtag+VHH5(26 kDa) using an Anti-Histag antibody. IscS protein (45 kDa) was used as positive control.*

To predict which residues/regions of the protein could contribute to aggregation, the sequence was analysed by AGGRESCAN (242). This is a web-based software that allows prediction of the aggregation properties of a protein on the basis of its sequence as described in paragraph 3.2.7.1. The aggregation properties of proteins strongly depend on specific sequence regions whose aggregation propensity is exceptionally high. **Figure 4.6** shows the aggregation profile of VHH5 protein sequence. Several regions are found to be aggregation prone, some of which in the CDR loops. The aggregation prone residues are distributed along the entire protein sequence. The peaks above the hotspot threshold line are supposed to promote aggregation. The higher peaks correspond to the CDR loops regions essential for the antigen recognition and binding. Instead, the smaller peaks correspond to regions of the backbone. The modification of their amino acids sequence could reduce the aggregation propensity of VHH5.

I also reasoned that a 3D model of the structure of the intrabody would give an independent insight and a more solid idea of the expected structural features. I thus model the structure of the VHH5 by comparative modelling.

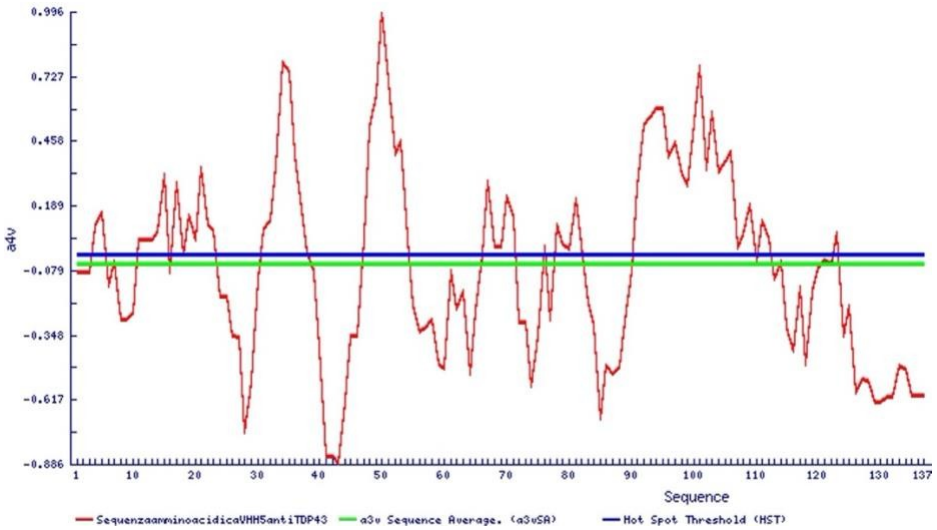


Figure 4.6 Aggregation-prone regions profile of VHH5. *Plot of the aggregation-prone regions profile along the VHH5 sequence as predicted by AGGRESCAN. The red line represents the aggregation profile of the VHH5 (137 amino acids). The blue line indicates the hot spot threshold, consistent with the aggregation propensity of each single natural amino acids and their frequency in protein (245). The green line corresponds to the average aggregation propensity of the protein of interest.*

4.3 Modelling of antibody scaffold

The structure of the antibody main scaffold, corresponding to the β-sandwich that holds the antigen recognizing CDR loops, can be easily predicted as this region is highly conserved amongst antibodies and their derivatives (246). The most suitable template was identified by submitting the sequence of the target protein (VHH5) to the BLAST search (https://balst.ncbi.nlm.nih.gov/Blast.cgi) over the PDB database (paragraph 3.2.7.1). The search identified several different possible templates. Among them *5wcc* was the closest sequence-wise template for comparative modelling. This is the crystal structure of the broadly neutralizing Influenza A antibody VRC 315 02-IF07 Fab. Both the SWISS-MODEL (226) and the ABodyBuilder (227) servers were used in parallel for the prediction, as described in paragraph 3.2.7.1. Different models were obtained and closely evaluated. The two energetically best structures from each of the two programs could be superposed with a RMSD of 0.45 Å (**Figure 4.7**). The template and the

target structures were of similar lengths with two one-residue insertions in the H2 and H3 CDR loops and a deletion in the other loop H1.

The selected SWISS-MODEL (green) and ABodyBuilder (red) structures were then used as the basis for further modelling.

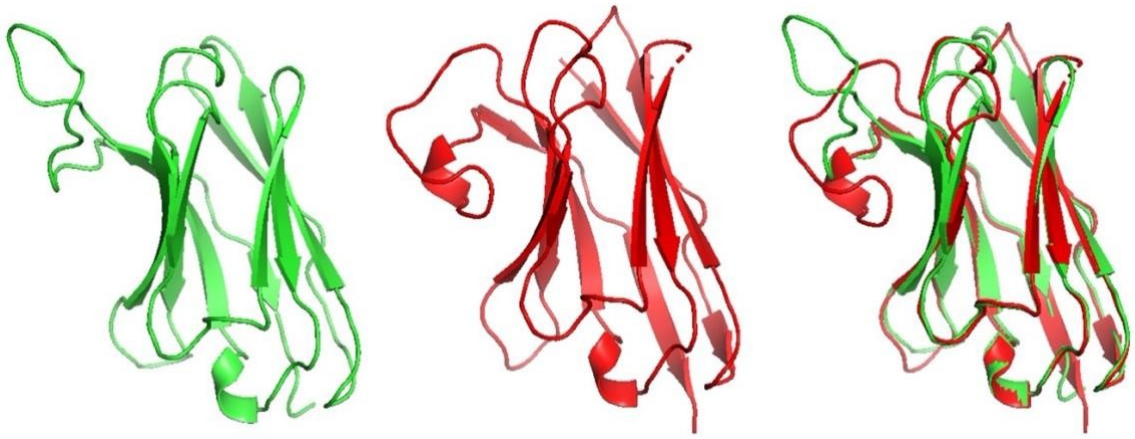


Figure 4.7 Modelling of the intrabody scaffold. *The two best models according to SwissModel (left) and ABodyBuilder (middle) and their superposition (right). The two models were superposed on the backbone atoms of the scaffold.*

4.4 H3 loop modelling

The challenge in antibody structure prediction is the conformation of the CDR loops. Of the three loops, H1 and H2 can easily be classified according to the canonical structures first described in 1987 by Chothia and Lesk and their structures can confidently be predicted (247). On the contrary, H3 loop appears to be problematic because of the high variability of its sequence, length, and conformation that makes difficult to build a high-quality structure with ordinary modelling techniques.

The modelling of the H3 loop (residues 94-114) was carried out using the Sphinx algorithm (234), as described in paragraph 3.2.7.1. Given the overall similarity between the two VHH5 structural bundle predicted from SWISS-MODEL and ABodyBuilder and to reduce the number of the structures to analyse, the prediction was restricted only to the best structure obtained from SWISS-MODEL. The Sphinx algorithm gave a bundle of 500 structures from which the 10 energetically best structures were selected. These structures are shown in **Figure 4.8.** In most of them, the loop turned out not to contain any regular structural element with the loop mostly protruding out from

the rest of the molecule. Only in one model (the first one in the upper left, Figure 4.8), the loop contains a short 1-turn helical element. In seven out of ten structures, the first two residues of the loop pair with a close-by strand.

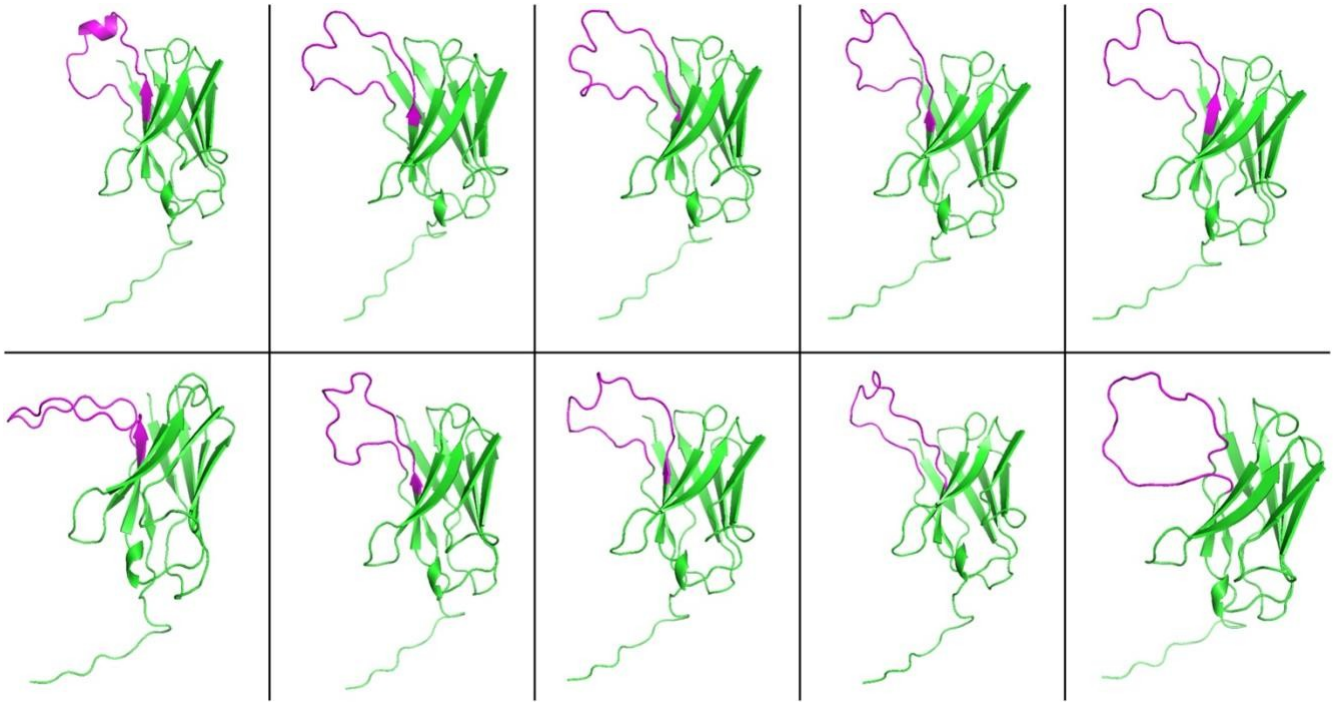


Figure 4.8 Modelling of the H3 loop. *The energetically best ten structures obtained by the Sphinx program are displayed. The structures were first superposed on the backbone atoms of the scaffold residues and then translated. The H3 loop is shown in magenta. Arrows and ribbons indicate beta strands and helical elements.*

Figure 4.9 reported the Ramachandran plot of VHH5 predicted from SWISS-MODEL. The dihedral angles of amino acid residues appear as dots in the plot. The white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their van der Waals radii. These regions are sterically forbidden for all amino acids except glycine as it lacks a side chain. The regions delimitated with cyan and blue lines represent the favoured and allowed regions respectively. The favoured regions include the dihedral angles typical of the alpha-helical and beta-sheet conformations. They correspond to the "core" regions representing the most favourable combination of phi/psi values. The allowed regions, located around the "core" regions or unassociated with these regions, contain fewer data points than the favoured regions (248).

The predicted model was validated by PROCHECK (PDBSum) (249, 250). According to this analyser, the Ramachandran plot contained 90% of the residues in the most favoured regions and 10% in additional/generously allowed regions, as shown in **Figure 4.9**. Gly and Pro residues were also located in allowed regions. The G-factors on dihedral angles, that provide a measure of how unusual, or out-of-the-ordinary, a property is, were all above the -0.5 threshold or positive. Values below the threshold indicate regions of concern. The overall average value was -0.14.

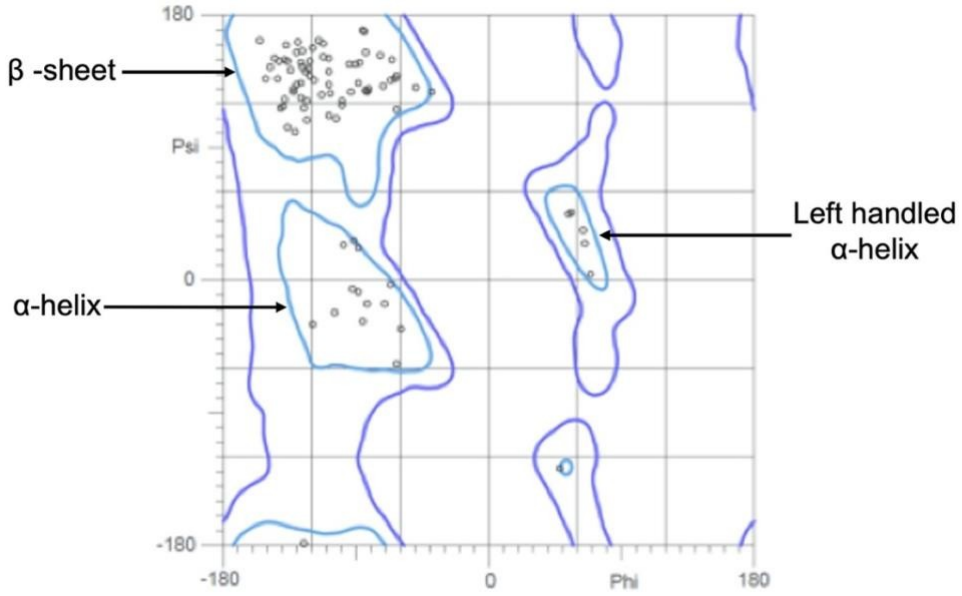


Figure 4.9 Ramachandran plot of VHH5*. Ramachandran plot of VHH5 from SWISS-MODEL after modelling of H3 with Sphinx. The plot shows the distribution of the Phi/Psi angles in the VHH5. The Ramachandran plot is a visualization produced by MolProbity website (http://molprobity.biochem.duke.edu) (251, 252).*

4.5 Optimization of VHH5(mut) production

The predicted structures were used to analyse the protein surface and identify exposed hydrophobic residues not contributing to the hydrophobic cores or to the CDR loops that could be mutated to reduce the risk of the proteins to be in inclusion bodies. The models were inspected, and the coordinates were analysed with the DSSP software which provides per residue accessible surface areas. As the result of this analysis, it was found that the regions mainly responsible to promote aggregation could be the H3 loop which is indeed rather hydrophobic. However, this region cannot be mutated as it may be essential for epitope recognition. Additionally, seven other residues were found to potentially interfere with the inclusion bodies formation. They are more

exposed hydrophobic residues and not contributing to the loops. The software AGGRESCAN (242) not only allowed to the identification of residues, but also test suggested specific substitutions to select the amino acids that lead to a decrease in aggregation.

According to the AGGRESCAN profile (**Figure 4.6**) and the hot spot area (**Figure 4.10, Panel A**), eight hotspots have been recognized. Panel A shows the area plots of VHH5 (top) and VHH5-I15A_M74K (bottom) with highest aggregation propensity. In the AGGRESCAN output, the sequence stretches with highest predicted aggregation are shown in red in the protein sequence column and appear as peaks in the profile plot. The figure shows how the I15A and M74K mutations strongly decrease the aggregation of VHH5, resulting in the loss of the two highlighted hot spots. I thus decided to mutate I15 to alanine and M74 to lysine creating the double mutant VHH5-I15A_M74K [later called VHH5(mut)] and I attempted to express it in *E. coli* (**Figure 4.10, Panel B**). Panel B shows the VHH5 structure with the two selected hydrophobic residues (I15 and M74) highlighted in blue.

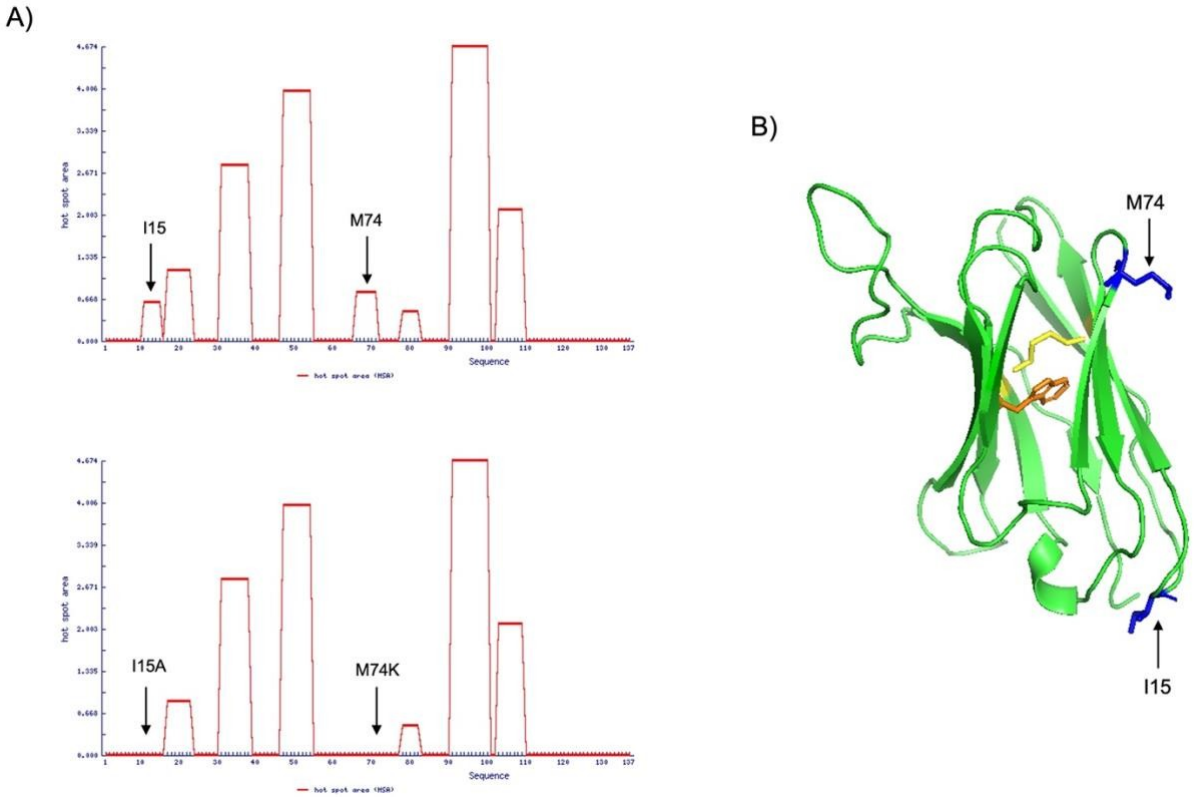


Figure 4.10 Residues prone to aggregation. *A) AGGRESCAN analysis. Hot spot area plots of VHH5 (top) and VHH5-I15A_M74K (bottom). Peaks are supposed to promote aggregation. Two hydrophobic residues were identified and substituted, I15A and M74K, which interfere with the inclusion bodies formation. B) VHH5*

structure. The disulfide bond is highlighted in yellow, and the side chain of the tryptophan is in orange. The two hydrophobic residues, I15 and M74, that were hypothesized to help inclusion body formation are highlighted in blue.

This construct was created in two steps. First a sited-directed mutagenesis cloning method allowed the substitution of the two selected amino acids, as described in paragraph 3.2.1.4. The Gibson Assembly cloning method was then used to insert the mutated construct, VHH5-I15A_M74K, into pET-SUMO expression vector (kindly provided by Prof. Annalisa Pastore, King's College London, United Kingdom), as fusion protein with a N-terminal SUMO solubilization domain, to enhance protein solubility, and a $(His)_6$-tag, as described in paragraph 3.2.1.2. The construct was first transformed in *E. coli* BL21(DE3) and tested for expression in small-scale experiments (paragraph 3.2.3.1) giving positive results but unsatisfactory. I changed the expression cells to BL21(DE3)pLysS, and tried two different induction modality (0.5 mM IPTG at 37 °C for 4 hours and 0.5 mM IPTG at 18 °C overnights) in a small-scale experiments (paragraph 3.2.3.1), to identify which one was the most suitable for obtaining a good level of protein expression (**Figure 4.11, Panel A**). Panel A compares the different conditions used for the protein expression. An increase in expression yield was observed using BL21(DE3)pLysS cells compared to BL21(DE3) and also with increasing temperature and decreasing time of incubation. In conclusion, it was observed that BL21(DE3)pLysS cells and protein induction using 0.5 mM IPTG at 37 °C for 4 hours (paragraph 3.2.3.2) were the most suitable conditions to obtain a satisfactory expression. The protein production switched from being all in the inclusion bodies to being mostly soluble (**Figure 4.11 Panel B**). Panel B highlights the VHH5(mut) expression as soluble fraction, as it was detected in the supernatant.

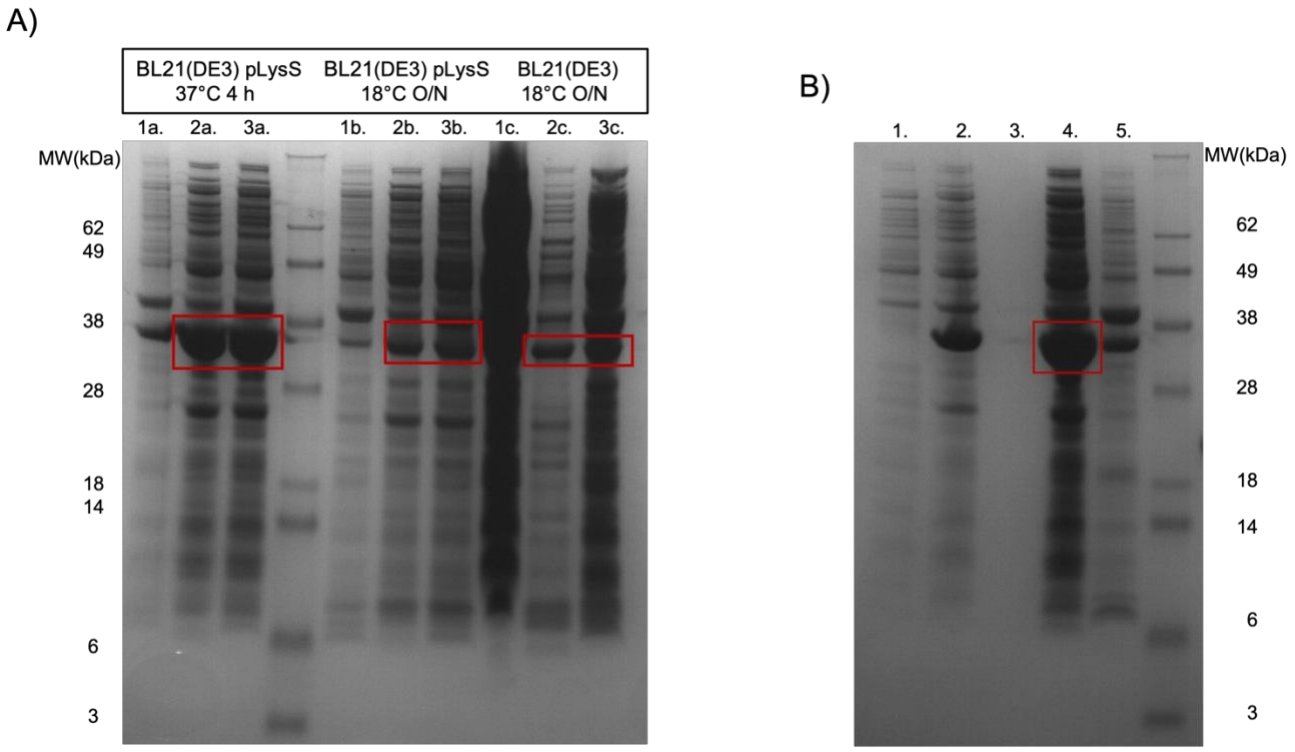
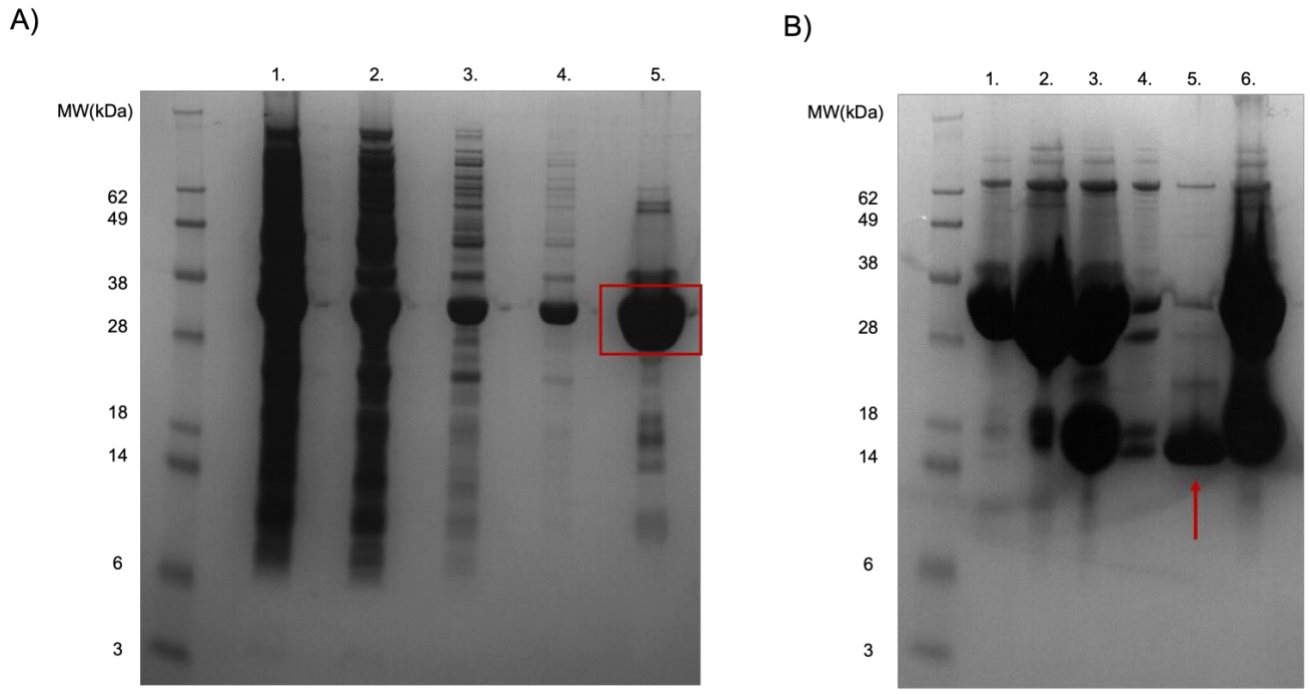


Figure 4.11 Expression of VHH5(mut). *A) Expression and induction tests of SUMOtag+VHH5(mut). SDS-PAGE analysis shows the expression yield using different E. Coli cell expression [BL21(DE3) or BL21(DE3)pLysS] and induction methods (37 °C 4 h or 18 °C O/N). Each column corresponds to a sample as following: 1a. post-lysis pellet, 2a. post-lysis supernatant, 3a. pre-lysis pellet, [BL21(DE3)pLysS, 0.5 mM IPTG at 37 °C for 4 h]; 1b. post-lysis pellet, 2b. post-lysis supernatant, 3b. pre-lysis pellet, [BL21(DE3)pLysS, 0.5 mM IPTG at 18 °C overnight]; 1c. post-lysis pellet, 2c. post-lysis supernatant, 3c. pre-lysis pellet, [BL21(DE3), 0.5 mM IPTG at 18 °C overnight]. B) Overexpression of SUMOtag+VHH5(mut) in E. Coli BL21(DE3)pLysS cells as a soluble protein. SDS-PAGE analysis of SUMOtag+VHH5(mut) (29kDa) shows the soluble protein and a high yield of expression. Each column corresponds to a sample as following: 1. pre-induction, 2. after induction with IPTG, 3. pre-lysis supernatant, 4. post-lysis supernatant, 5. post-lysis pellet.*

The purification of VHH5(mut) from the BL21(DE3)pLysS supernatant was performed *via* two steps of Immobilized Metal Affinity Chromatography (IMAC) using a Nickel column (Super Ni-NTA agarose resin, Generon. 5 ml) and a phosphate buffer (pH 7.2), as described in paragraph 3.2.4.2. Both steps were performed in a cold room (4 °C) because of the high tendency of the protein to precipitate. A SDS-PAGE analysis (**Figure 4.2, Panel A**) of the first step showed the high expression yield of SUMOtag+VHH5(mut) construct. The removal of the SUMOtag from the N-terminal was obtained by digesting the protein with the TEV protease. A high molar ratio (1:5 protein construct/TEV) was used as it was found that a low molar ratio was inefficient. The cleavage site was probably hidden and therefore difficult to identify. A SDS-PAGE analysis reported in **Figure 4.12 Panel B,** shows the second step. It was necessary to use a high salt phosphate buffer (pH 7.2)

to allow the elution of the protein (Panel B, fraction 5) as it remained attached to the column. A size exclusion chromatography (paragraph 3.2.4.2) using a Hiload 16/60 Superdex 75 pg (Cytiva), was performed as a final purification step allowing the removal of both protein aggregates and degradation products resulting in high grade purity protein samples. The chromatogram profile of VHH5(mut), reported in **Figure 4.12, Panel C**, shows a single peak with an elution volume around 77 ml. Fractions corresponding to the peak were analysed through SDS-PAGE (**Figure 4.12, Panel D**). The protein resulted to be present in fractions 30-42 after Gel Filtration. They were pooled, quantified (1.96 mg/ml), and stored at – 20 °C.



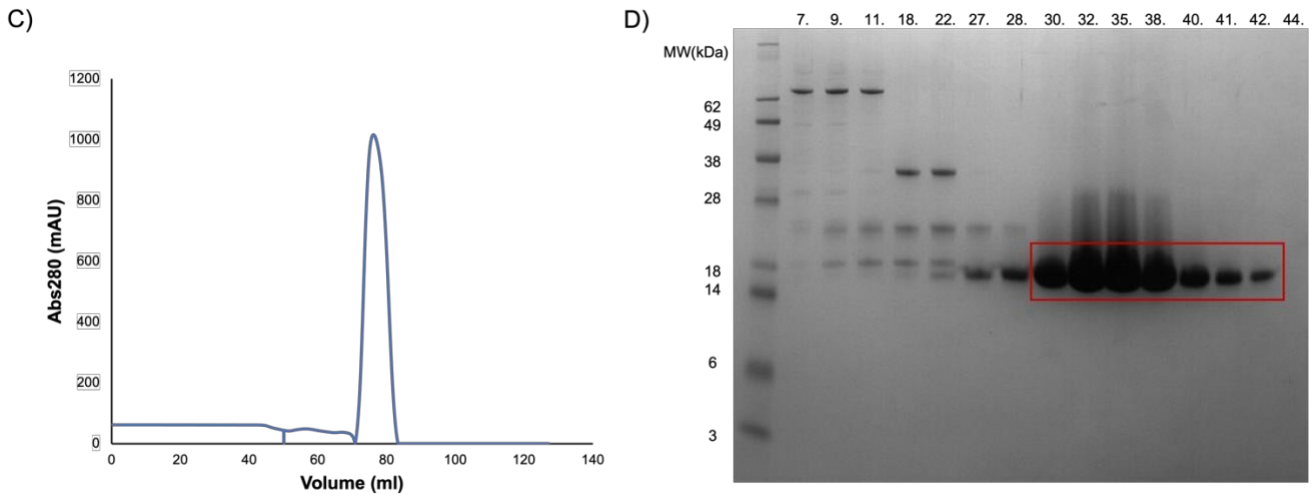


Figure 4.12 Purification of VHH5(mut). *A) Affinity Chromatography (IMAC) of SUMOtag+VHH5 (29kDa) step 1. SDS-PAGE analysis shows the eluted fraction corresponding to the construct SUMOtag+VHH5. Each column corresponds to a sample as following: 1. flow-through, 2. wash1, 3. wash 2, 4. wash 3, 5. elution. B) Affinity Chromatography (IMAC) of VHH5 (14.8 kDa) step 2. SDS-PAGE analysis shows a band corresponding to VHH5 after 1 h incubation with 10 mM potassium phosphate buffer (pH 7.2), 1 M KCl. Each column* corresponds to a *sample as following: 1. pre-TEV, 2. TEV+VHH5 (time 0), 3. after-TEV (overnight), 4. flow-through, 5. wash 1 M KCl, 6. elution. C) SEC Chromatography profile. D) SDS-PAGE analysis of SEC shows the eluted fractions (30-42) corresponding to the protein, VHH5(mut) (14.8 kDa). All SEC fractions resulted high-grade pure and where all used for identity and epitope mapping experiments.*

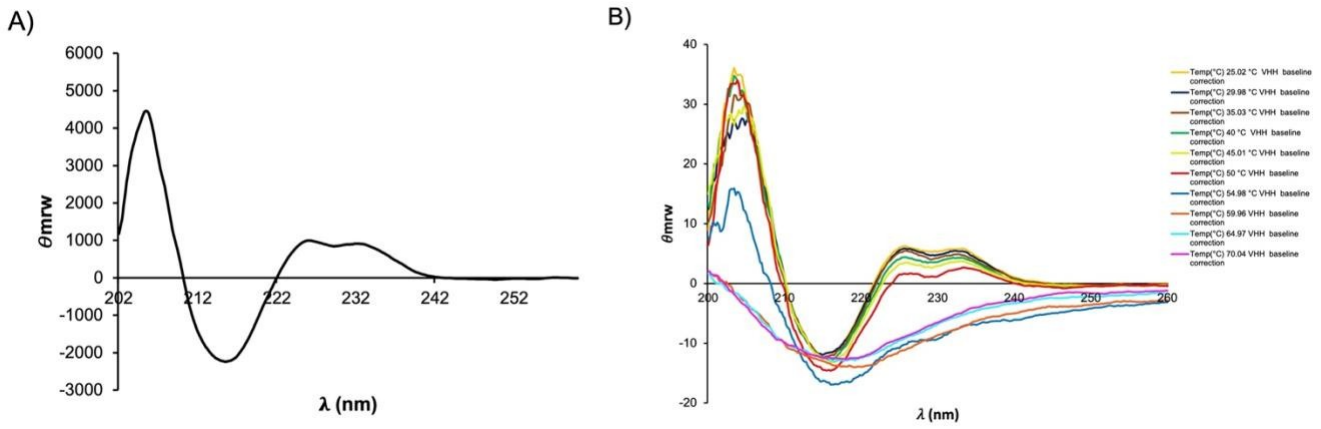
This strategy allowed to obtain VHH5-I15A_M75K in suitable quantities for further studies, as we typically obtained ca. 13 ml (1.96 mg/ml or 132 µM) of >98% pure protein after cleaving it from the tag.

4.6 VHH5(mut) identity

Mass spectrometry confirmed the protein identity and disulfide formation. The experiments were kindly performed by Prof. Dal Piaz Fabrizio from the University of Salerno, Italy. The state of fold was confirmed by far-UV circular dichroism (CD), a technique able to detect the secondary structure of proteins, as described in paragraph 3.2.5.2. The CD spectrum of VHH5(mut) was recorded at room temperature (25 °C). It showed a maximum at 205 nm and a single minimum around 215 nm which are features typical of the β-sheet conformation expected for an antibody (**Figure 4.13, Panel A**). The positive contribution at 225-235 nm is usually diagnostic of the presence of stacking interactions between aromatic residues (253). According to the deconvolution analysis the structure composition of VHH5(mut) comprises: 46% β-sheet, 22.4%

β-turn and 29.6% random coiled. No alpha helix features were detected. In a different experiment, I observed the sample behaviour at increasing temperatures. I gradually heated up the sample from 25 °C to 70 °C (1 °C/min) and recorded the complete spectra every minute, as shown in **Figure 4.13, Panel B**. The protein spectra conserved the typical β-sheet pattern of an antibody until ~55 °C (cyan) when the maximum at 205 nm slightly decreased in intensity and the positive contribution at 225-235 nm was lost. At this temperature the protein started to unfold and lose its secondary structure. This was due to precipitation and aggregation.

VHH5(mut) protein folding, and stability were also investigated by using NMR as described in paragraph 3.2.5.3. The mono-dimensional and the HSQC NMR spectra of the single labelled protein with ^{15}N-single labelled protein were acquired at 25 °C and presented a well dispersed resonances as expected for a folded monomeric protein of the size of VHH5(mut) (**Figure 4.13, Panel C, D**). It was thus concluded that the protein purified was folded and well-behaved.



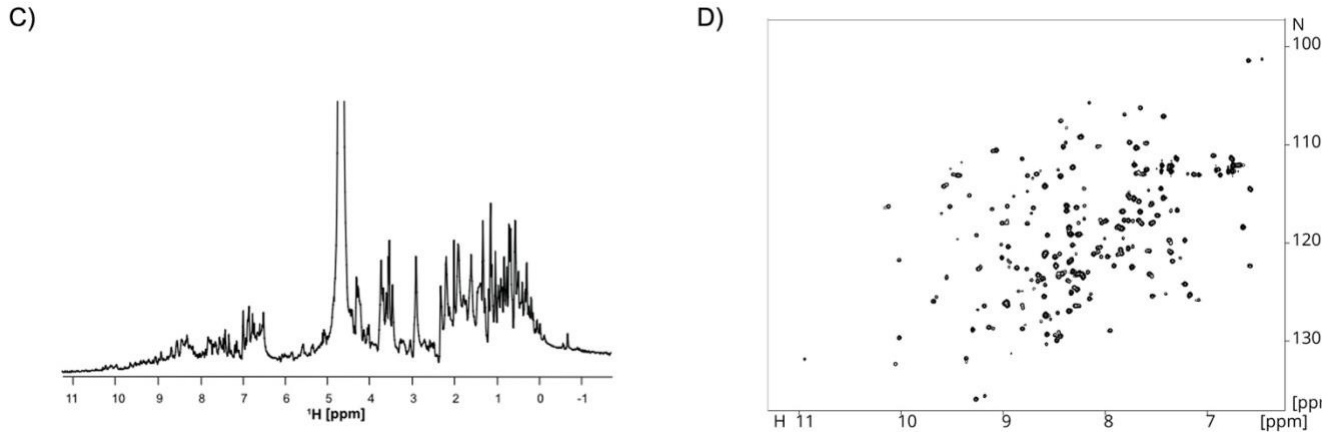


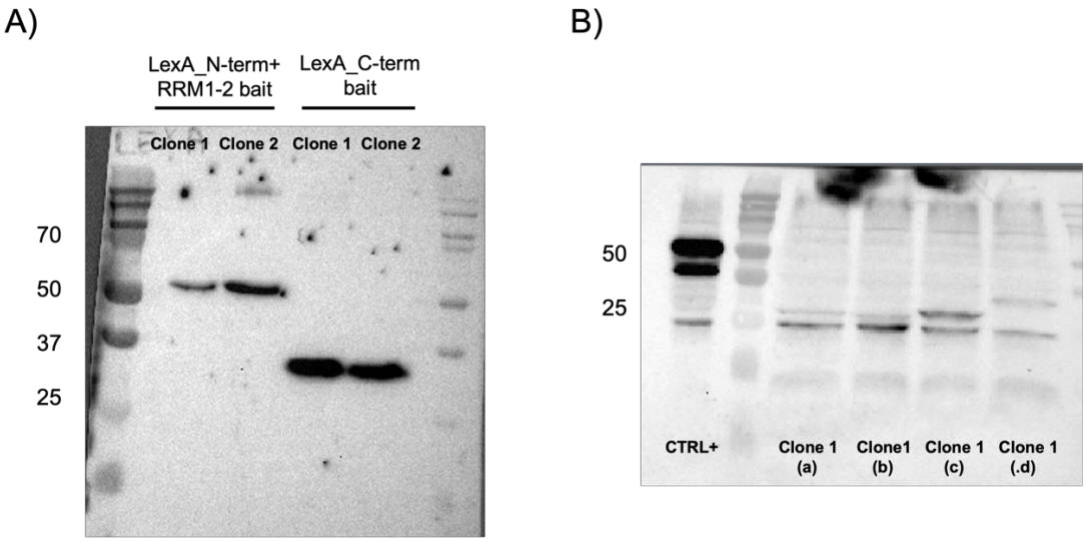
Figure 4.13 Structure identity of VHH5(mut). *A) CD spectrum of 50 μM of VHH5(mut) in 10 mM KPi pH 7.2 and 15 mM KCl recorded at room temperature, B) CD spectra of a temperature scans (1 degree/min) of 50 μM of VHH5(mut) in 10 mM KPi pH 7.2 and 15 mM KCl recorded heating up sample 25-70 °C and C) ^{1}H NMR spectra of VHH5(mut) (200 μM) in 10 mM KPi pH 7.2 and 15 mM KCl recorded at room temperature. D) HSQC-NMR spectrum of ^{15}N-labelled VHH5(mut) (200 μM) in 10 mM KPi pH 7.2 and 15 mM KCl recorded at room temperature*

4.7 Epitope mapping

The characterization of the epitope of TDP-43 recognized by VHH5, was first determined with In Vivo Epitope Mapping (IVEM) experiment in yeast (174). The IVEM was performed in collaboration with Prof. Antonino Cattaneo's research group (Scuola Normale Superiore di Pisa). The original LexA-TDP-43 bait was initially truncated into two fragments: LexA-N-term+RRM1-2 (residues 1-258) and LexA-C-term (residues 259-414), yielding two TDP-43 fragment baits. The cloning was performed by a deletion of DNA sequences using a PCR-based approach as described in paragraph 3.2.2.1. Once the constructs were obtained, they were transformed into L40 *S. cerevisiae* as previously described (paragraph 3.3.1.1). L40 cells containing TDP-43 fragment baits were analysed for LexA-TDP-43 fragments expression through WB analysis (**Figure 4.14, Panel A**). Panel A shows the detection of bands correspond to LexA-N-term+RRM1-2 construct (MW ~52 kDa) and LexA-C-term construct (MW ~38 kDa) by using anti-LexA antibody. In addition, a test to evaluate the level of transactivation was performed by growing them on plates lacking Tryptophan and Histidine amino acids (SD-WH plates, selective for transactivation). 5 mM 3AT was added to the medium for the LexA-C-term bait while LexA-N-term+RRM1-2 did not show any transactivation. These baits were then transformed with the pLinker220 plasmid carrying the VHH5 and plating the cells on SD-WL or SD-WHL with the same protocol described in paragraph

3.2.2.1. The epitope recognized by the VHH5 resulted to be located in the N-terminal half of the protein as shown in **Figure 4.14, Panel C**.

To further narrow down the region carrying the epitope a second cycle of IVEM was performed, splitting the N-terminus+RRM1-2 (1-258) into four smaller baits, the N-terminus (1-105), RRM1 (106-176), RRM2 (192-258), and a fragment of RRMs (160-208) which contains the linker between RRM1 and RRM2. The cloning was performed using the Gibson Assembly method as described in paragraph 3.2.1.2. L40 cells containing the smaller baits were analysed for LexA-TDP-43 fragments expression through WB analysis (**Figure 4.14, Panel B**). Panel B shows the detection of a band corresponds to LexA-N-term short construct (MW 34 kDa), LexA-RRM1 construct (MW 30.8 kDa), LexA-RRM2 construct (MW 30.1 kDa) and LexA-RRMs construct (MW 28.3 kDa) by using anti-LexA antibody. As reported in WB analysis, LexA- Tau (MTBD) protein extract was used as positive control (MW 50 kDa). Then, the baits were transformed individually in L40 yeast strains with the prey, VHH5, as described in paragraph 3.2.2.1. As shown in **Figure 4.14, Panel C**, the epitope seemed to be mainly located in RRM2, since growth on SD-WHL plates was detected both with the LexA-RRM2 and the LexA-RRMs baits. No colonies were reported in LexA-N-term short SD-WHL plates.



C)

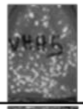
Baits	Growth on -WHL
LexA-TDP-43	
LexA-N-term+RRMs	
LexA-N-term short	
LexA-RRM1	
LexA-RRM2	
LexA-RRMs	

Figure 4.14 Epitope mapping of VHH5 on TDP-43. *A) anti-LexA Western blot on L40 TDP-43 baits (protein extraction); LexA-N-term-RRM1-2: (52 kDa); LexA-C-term: (38 kDa); CTRL+: positive control [LexA- Tau (MTBD) (50KDa)]. B) anti-LexA Western blot on L40 TDP-43 baits (protein extraction); clone 1(a): LexA-N-term short (34 kDa); clone 1 (b): LexA-RRM1 (30.8 kDa); clone 1 (c): LexA-RRM2: (30.1 kDa); clone 1 (d): LexA-RRMs (28.3 kDa); CTRL+: positive control (pMIC-BD1_LexA). C) In vivo Epitope Mapping. The VHH5 was transfected in L40 yeast strain expressing the baits LexA-TDP-43 full length (residues 1-414), LexA- N-Term+RRM1-2 (1-258), N-term short (1-105), RRM1 (106-176), RRM2 (192-258), RRM1-2 (160-208). Interaction is detected by growth on –WHL plates.*

A cross-linking experiment was carried out, as described in paragraph 3.2.6.2, to evaluate the epitope mapping results. The purified recombinant VHH5(mut) (MW 14.8 kDa) was mixed with TDP-43 fragments (RRM1-2: 19 kDa; RRM1: 10.5 kDa; RRM2: 9 kDa), kindly provided by Prof. Annalisa Pastore's research group (King's College London). In this assay, a cross-linking agent was added to link covalently two proteins located in a close spatial distance. The protein-protein interaction was thus captured in native conditions, limiting the risk of false positive and the loss of complex stability. A SDS-PAGE analysis was performed to identify species in the cross-linking reaction mixture (**Figure 4.15, Panel A**). I used the bis[sulfosuccinimidyl]suberate (BS^3), a cross-linking agent that reacts with primary amino groups. The mixture contained VHH5(mut) and RRM1-2, RRM1 and RRM2 at different molar ratios (1:1, 1:5, 1:10). The reaction produced

complexes with molecular weights of 33.8 kDa [VHH5(mut)+RRM1-2], 25.3 kDa [VHH5(mut)+RRM1], and 23.8 kDa [VHH5(mut)+RRM2]. The interaction seemed to involve both domains. As shown in Panel A, the intensity of the interactions increases as the protein ratios increased.

To substantiate these results with further evidence, both sandwich and indirect ELISA assays were performed, as described in paragraph 3.2.6.1. In these assays, an enzyme-linked conjugate and enzyme substrate (chromogen) were used to identify the antigen-antibody interaction. For the indirect ELISA (**Figure 4.15, Panel B**) the purified recombinant TDP-43 fragments (RRM1-2, RRM1 and RRM2) were coated and then the VHH5(mut) was added. On the contrary, for sandwich ELISA (**Figure 4.15, Panel C**) the purified recombinant VHH5(mut) was coated and then the TDP-43 fragments were added. Different molar ratio (1:1, 1:3, 1:5, 1:10) of VHH5(mut) and TDP-43 fragments were used in both assays. As shown in figure, in both cases, response to RRM1 (106-176), RRM2 (192-258), and RRM1-2 (106-258), were observed, indicating that the epitope involves both domains.

VHH5(mut) recognises each of the repeats which share some homology. However, while the homology is fairly high the sequence identity is only 26%. It was reasoned that the presence of two independent epitopes one in each repeat is fairly unlikely. It is more likely that the epitope is conformational and involves both domains. A conformational (discontinuous) epitope consists in a bunch of residues that are far away from each other in the primary sequence but are located to spatial proximity as a result of the protein folding. It was also noticed that only the indirect ELISA showed a dependence on the antibody to protein ratio. This could be explained by considering the experiment methodology: in the indirect ELISA the intrabody is added at increasing concentrations while the target protein (the TDP-43 fragments) is fixed and may be partially masked. On the contrary, the absence of concentration dependence in the sandwich ELISA assay could easily be explained by the assumption that when the intrabody is fixed it could adopt a conformation that makes it more competent for binding. This means that the detected affinity can be different in the two cases, and the signal could be saturated in the sandwich ELISA (**Figure 4.15, Panel C**) but not in the indirect ELISA (**Figure 4.15, Panel B**) done with the intrabody in solution.

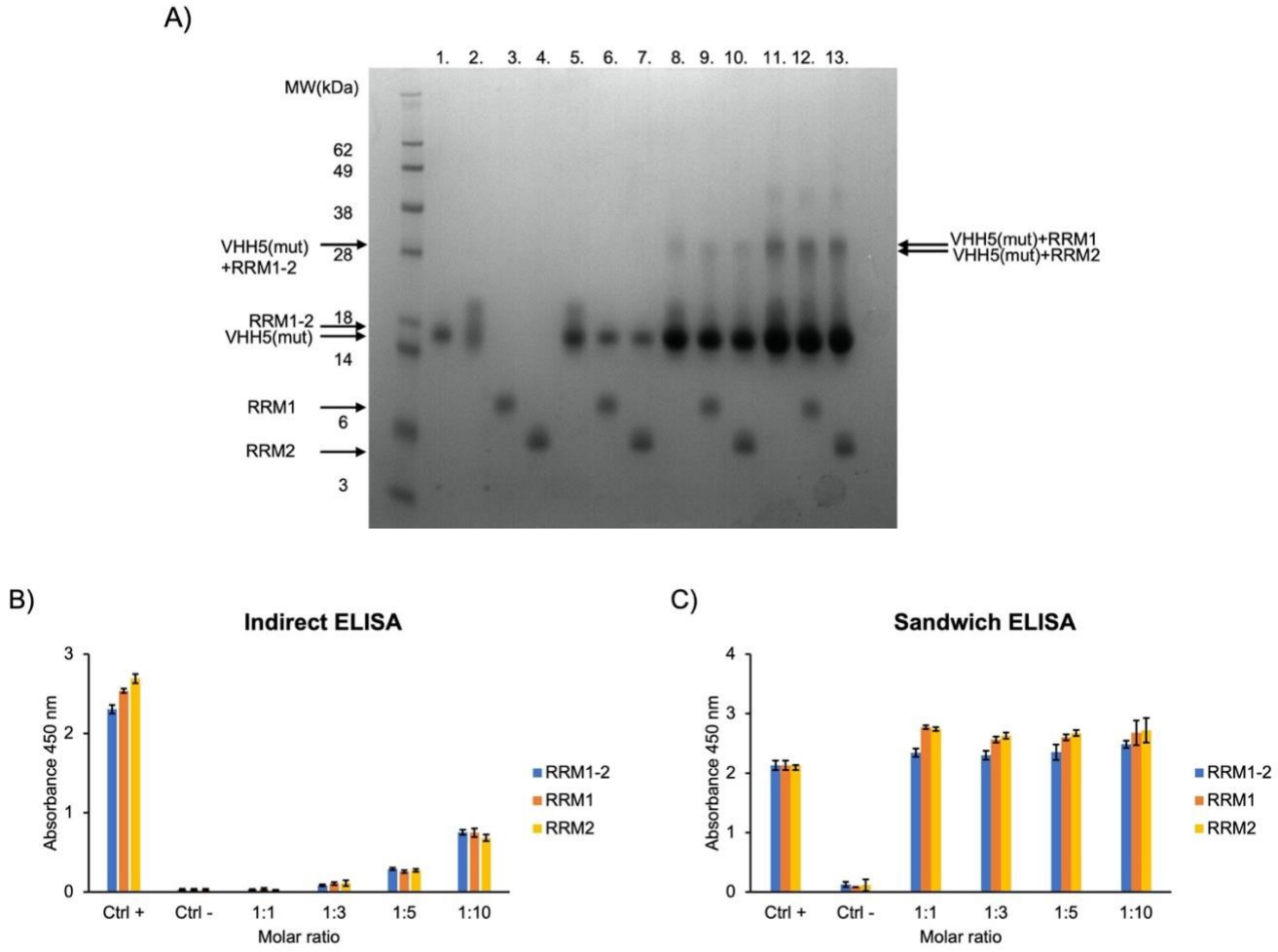


Figure 4.15 Epitope mapping of VHH5(mut) on TDP-43*. A) Cross-linking analysis. Each column corresponds to a sample as following: 1. VHH5(mut) (14.8 kDa), 2. RRM1-2 (19 kDa), 3. RRM1(9 kDa), 4. RRM2 (7.7 kDa), 5-7 RRM1-2+VHH5(mut) (33.8 kDa), 8-10. RRM1+VHH5(mut) (25.3 kDa), 11-13. RRM2+VHH5(mut) (23.8 kDa). VHH5(mut): 10 µM, molar ratio 1:1, 1:5, 1:10. Antigen-antibody interaction increases as the molar ratio increases. B) Indirect ELISA assay. Coating antigen: RRM1-2, RRM1 and RRM2 (1 µM); Binding antibody: VHH5(mut) (molar ratio 1:1, 1:3, 1:5, 1:10); Detection: anti-VHH-HRP. The assay shows an interaction of VHH5(mut) with all the TDP-43 fragments. The interaction increases as the molar ratio increases. C) Sandwich ELISA assay. Coating antibody: VHH5(mut) (final molar ratio coating antibody: binding antigen 1:1, 1:3, 1:5, 1:10); Binding antigen: RRMs, RRM1 and RRM2 (1 µM); Detection: anti-TARDBP and then anti-hIgG-HRP. The assay shows an interaction of VHH5(mut) with all the TDP-43 fragments.*

A far-western blot analysis was performed in native conditions (paragraph 3.2.6.3) preserving the conformation of the interaction site, to add more evidence to confirm the binding. The purified recombinant VHH5(mut) was used as primary antibody to detect the TDP-43 fragments (RRM1, RRM2 and RRM1-2) and then a rabbit anti-camelid VHH [HRP] antibody (GenScript) was used to highlight the complex. As shown in native gel **(Figure 4.16)**, the data were inconclusive due to the

high tendency of TDP-43 fragments to aggregate. The fragments were completely aggregated in the wells. Therefore, the results obtained could not be take in consideration and could not confirm what was previously observed about the epitope mapping.

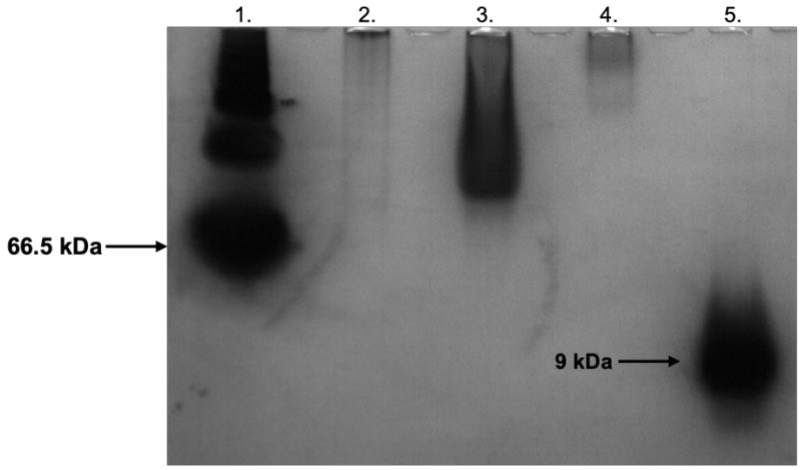


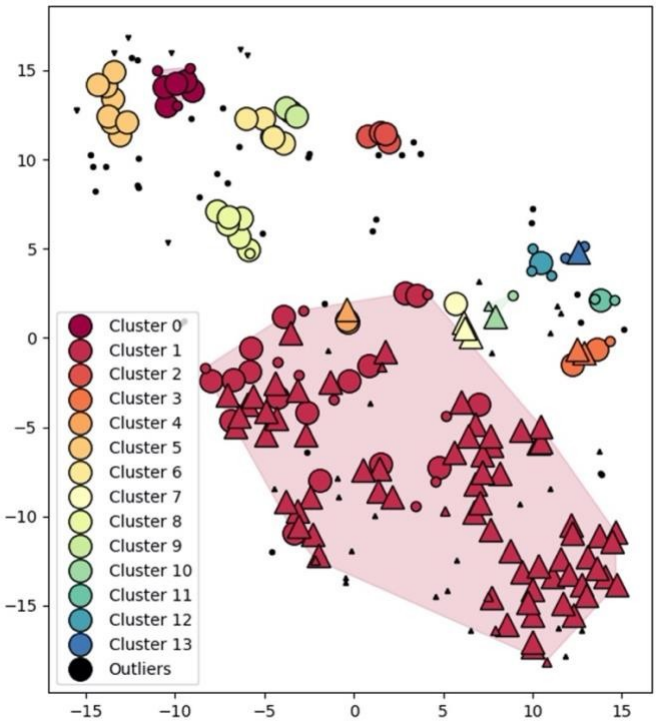
Figure 4.16 Native gel of VHH5(mut) on TDP-43. *Far-western blot analysis. Native gel shows the aggregated TDP-43 fragments. Each column corresponds to a sample as following: 1. 1% BSA (66.5 kDa), 2. N-term RRM1-2 (29 kDa), 3. RRM 1-2 (19 kDa), 4. RRM1 (10.5 kDa), 5. RRM2 (9 kDa).*

4.8 VHH5(mut) model

Molecular docking was performed as described in paragraph 3.2.7.1 using the information described above. Models of the antigen-antibody complex were generated by the ClusPro software using each of the ten energetically best Sphinx structures described in section 3.2.7.1 and the NMR structure of the putative antigen (PDB 4bs2). To influence docking, an attractive force was set on the residues of H3 loop using default parameters. This calculation resulted in 228 models which were further analysed using an ad hoc written script (paragraph 3.2.7.1). The filtering procedure, as described in the Materials and Methods section, was carried out in close collaboration with Professor Paolo Marcatili (Technical University of Denmark, Denmark), and identified a total of 14 clusters (**Figure 4.17, Panel A**). The complex structures with the lowest score and lowest binding free energy were selected and analysed (**Figure 4.17, Panel B**) as described in paragraph 3.2.7.1. Cluster 1 contained most of the solutions, in which the antibody only interacts with a single domain of the antigen. However, upon closer inspections, I realised that these solutions were likely the result of an artefact of the docking procedure: the H3 loop of the antibody would encircle the C-terminus of the antigen, in a configuration that would result in a knot or a lasso in the complete

antigen. Excluding these docking solutions, cluster 0, 2, 5, 6, 8, 9, 11, 12, and 13 mainly contained solutions in which the interaction involved both domains. In total, 51 out of the 61 solutions that were not outliers nor part of cluster 1, contained interactions to both domains (**Figure 4.17, Panel B**). These models, that were only indicative and low resolution, will need experimental validation through fine epitope-mapping at the level of the individual residues.

A)



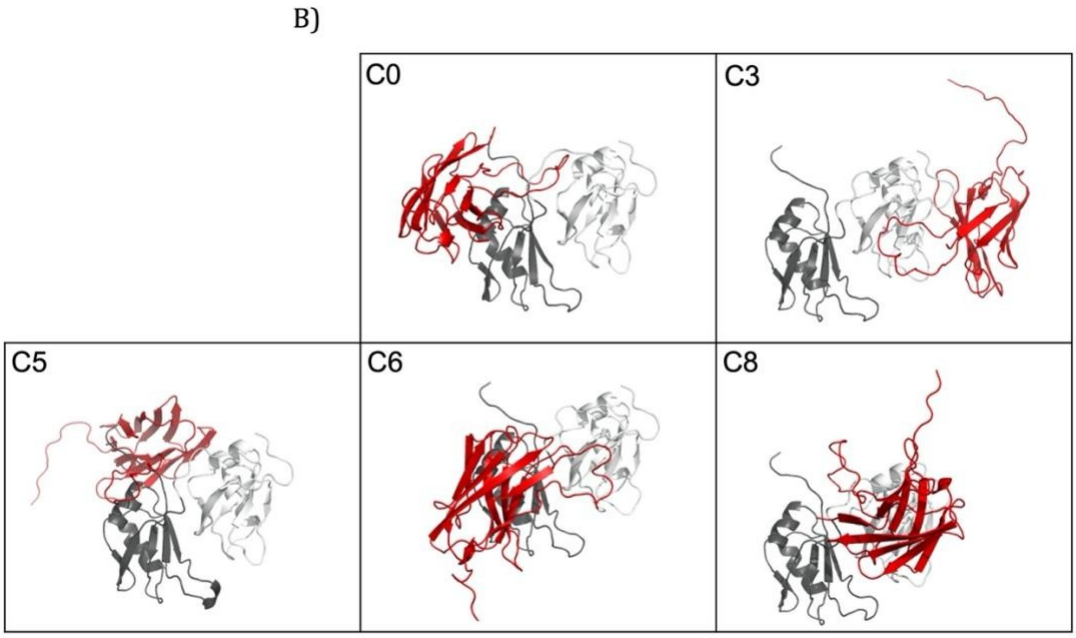


Figure 4.17 Clustering and structure of the docking solutions. *A) Clustering is represented as a 2D map that preserves local similarity. Each dot corresponds to a docking solution and is coloured according to the cluster it belongs to. Dots depicted as upward triangles, downward triangles, and circles represent solutions where the antibody interacts with the first (RRM1), second (RRM2), or both antigen domains (RRM1-2), respectively. Solutions depicted in black are considered outliers by the clustering algorithm, small dots and large dots are core and reachable elements, respectively. B) Representative solutions from clusters with more than five elements, excluding cluster 1. The antibody is represented in red, and the first and second antigen domains in dark grey and white, respectively.*

4.9 Attempts to express and produce VHH5

Will the wild-type nanobody interact with the antigen similarly to the mutant or will the interaction be stronger? Did the mutations affect the intensity of the antigen-antibody interaction? These were the questions that led me to start the optimization of the expression and purification protocol for the wild-type selected nanobody. The first attempts of VHH5 expression were carried out using the pET-17b vector, and subsequently adding the SUMOtag. The protein had an unsatisfactory yield as properly folded protein in the periplasm (254). If the protein still contains the leader sequence, it probably has not been exported and aggregates in the cytoplasm. As described in paragraph 4.2 the protein accumulated as inclusion bodies.

Based on the results obtained for the VHH5(mut), I decided to use the same expression vector to increase the protein solubility. VHH5 was inserted into pET-SUMO expression vector (kindly provided by Prof. Annalisa Pastore, King's College London, United Kingdom) as fusion protein with

N-terminal SUMO solubilization domain, to enhance protein solubility, and a $(His)_6$-tag. The cloning was made using the Gibson Assembly cloning technique as previously described (paragraph 3.2.1.2). The construct was transformed in *E. coli* BL21(DE3)pLysS cells for small-scale expression test, as described in paragraph 3.2.3.1. A SDS-PAGE was performed to confirm the VHH5 expression (**Figure 4.18**). The figure highlights the protein as present, before lysis, in the pellet.

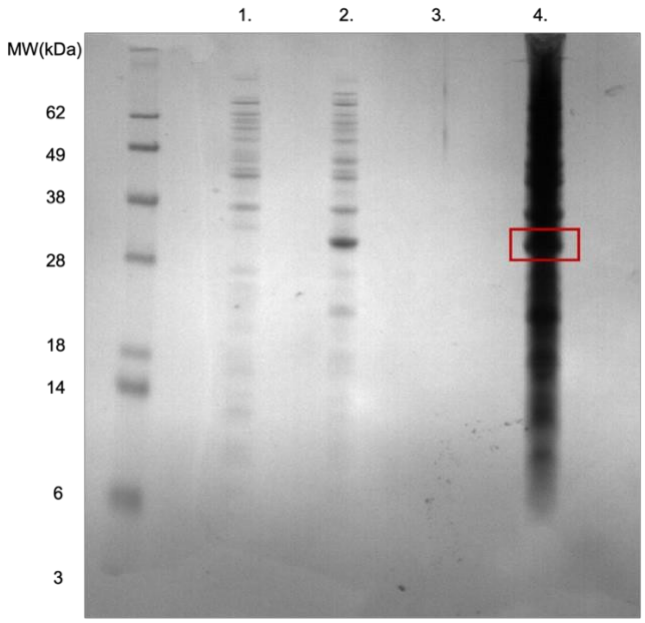
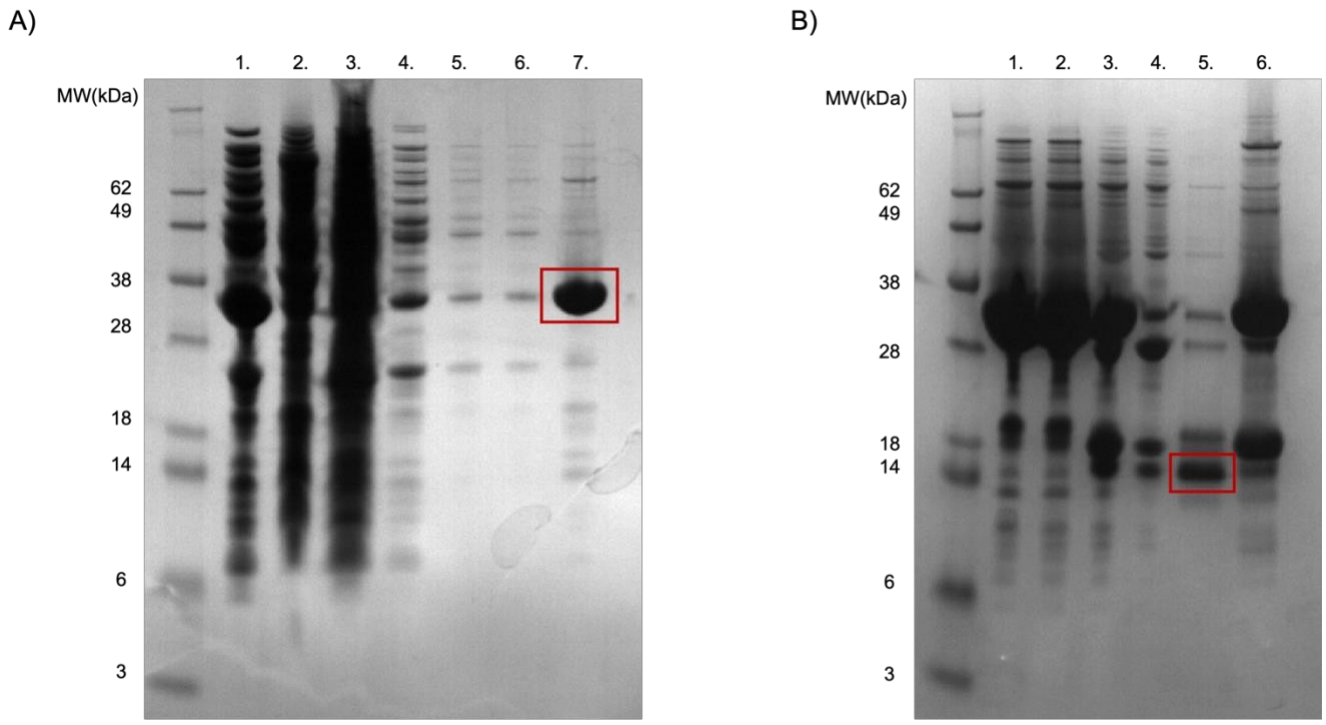


Figure 4.18 Expression of VHH5. *Overexpression of SUMOtag+VHH5 in E. Coli BL21(DE3)pLysS cells as a soluble protein. SDS-PAGE analysis of SUMOtag+VHH5 (29kDa) shows the soluble protein and a high yield of expression. Each column corresponds to a sample as following: 1. pre-induction, 2. after induction with IPTG, 3. pre-lysis supernatant, 4. pre-lysis pellet.*

The purification from the BL21(DE3)pLysS supernatant was performed following the same protocol as VHH5(mut). In summary, two steps of IMAC chromatography using a Nickel column were performed followed by a size exclusion chromatography (paragraph 3.2.4.2) to remove both protein aggregates and degradation products. As shown in **Figure 4.19, Panel A** (column 1), the protein production was mostly soluble as, after lysis, it appears to be in the supernatant. SDS-PAGE analysis (**Figure 4.19, Panel A**) of the first step shows the high expression yield of SUMOtag+VHH5 construct. The removal of the SUMOtag from the N-terminal was carried out by digesting the protein with the TEV protease (1:5 protein construct/TEV). **Figure 4.19 Panel B** shows a SDS-

PAGE analysis of the second step. The protein elution was performed using a high salt phosphate buffer (pH 7.2) (column 5) as it remained attached to the column. The chromatogram profile of VHH5 reported in **Figure 4.19, Panel C**, shows a single peak with an elution volume around 62 ml. Fractions corresponding to the peak were analysed through SDS-PAGE **(Figure 4.19, Panel D)**. Two bands (fractions 26-39) were observed in the SDS-Page indicating a non-separation of the protein from other products. The bands' identity was confirmed by mass spectrometry experiment (kindly performed by Prof. Dal Piaz Fabrizio from the University of Salerno, Italy). The lower band corresponds to VHH5 (MW 14.8 kDa) while the upper band to the SUMOtag. Although the theoretical molecular weight of the SUMO proteins is approximately 11 kDa, on SDS-PAGE it runs at higher molecular weights typically in the range of 15-17 kDa (255).



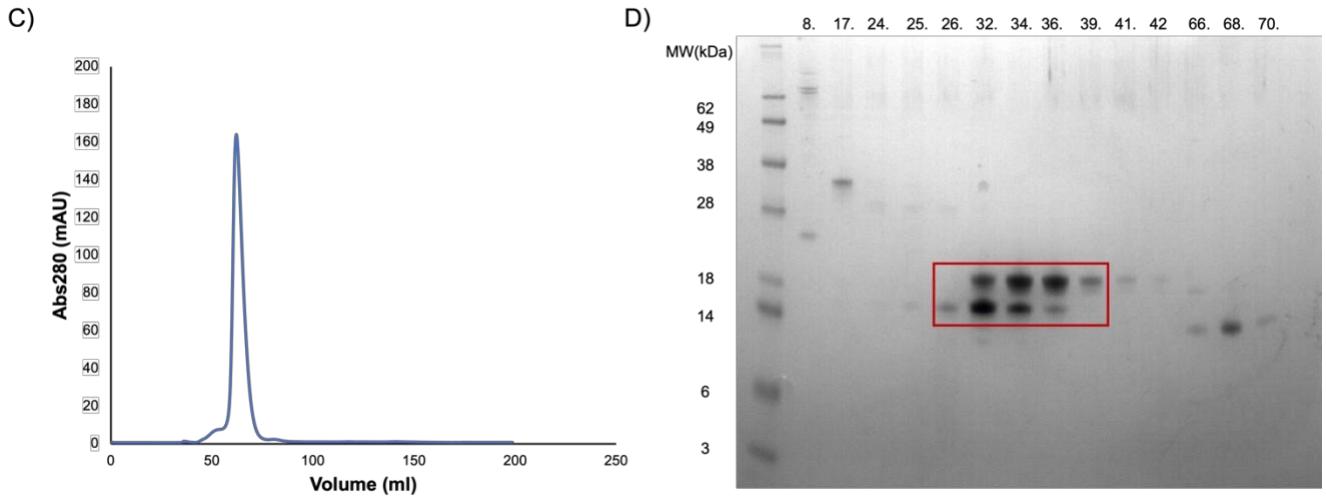


Figure 4.19 Purification of VHH5. *A) Affinity Chromatography (IMAC) of SUMOtag+VHH5 (29kDa) step 1. SDS-PAGE analysis shows the eluted fraction corresponding to the construct SUMOtag+VHH5. Each column corresponds to a sample as following: 1. post-lysis supernatant, 2. post-lysis pellet, 3. flow-through, 4. wash1, 5. wash 2, 6. wash 3, 7. elution. B) Affinity Chromatography (IMAC) of VHH5 (14.8 kDa) step 2. SDS-PAGE analysis doesn't show any band corresponding to VHH5 after 1 h incubation with 10 mM potassium phosphate buffer (pH 7.2), 1 M KCl. Each column corresponds to a sample as following: 1. TEV, 2. pre-TEV, 3. TEV+VHH5 (time 0), 4. after-TEV (overnight), 5. flow-through, 6. wash 1 M KCl, 7. elution. C) SEC Chromatography profile. Hiload 16/60 Superdex 75 pg (Cytiva). D) SDS-PAGE analysis of SEC shows the eluted fractions (26-39) corresponding to a complex formed by VHH5 (14.8 kDa, lower band) and SUMOtag (11 kDa, upper band).*

In the attempt to separate the two bands, a Superdex 75 Increase 10/300 GL (Cytiva), suitable for small scale preparative purification (µg-mg) was used as a final polishing step. The chromatogram profile of VHH5 reported in **Figure 4.20, Panel A**, shows two peaks with an elution volume around 15 ml and 16 ml respectively. Fractions corresponding to the peaks were analysed through SDS-PAGE **(Figure 4.20, Panel B)**. The protein resulted to be present in the first peak (fractions 3, 4, 5, 6 and 7) after Gel filtration. The fractions were pooled and concentrated to a fourth of the initial volume before protein quantification, which demonstrated a very low concentration (0.355 mg/ml). The protein was then stored at -20 °C.

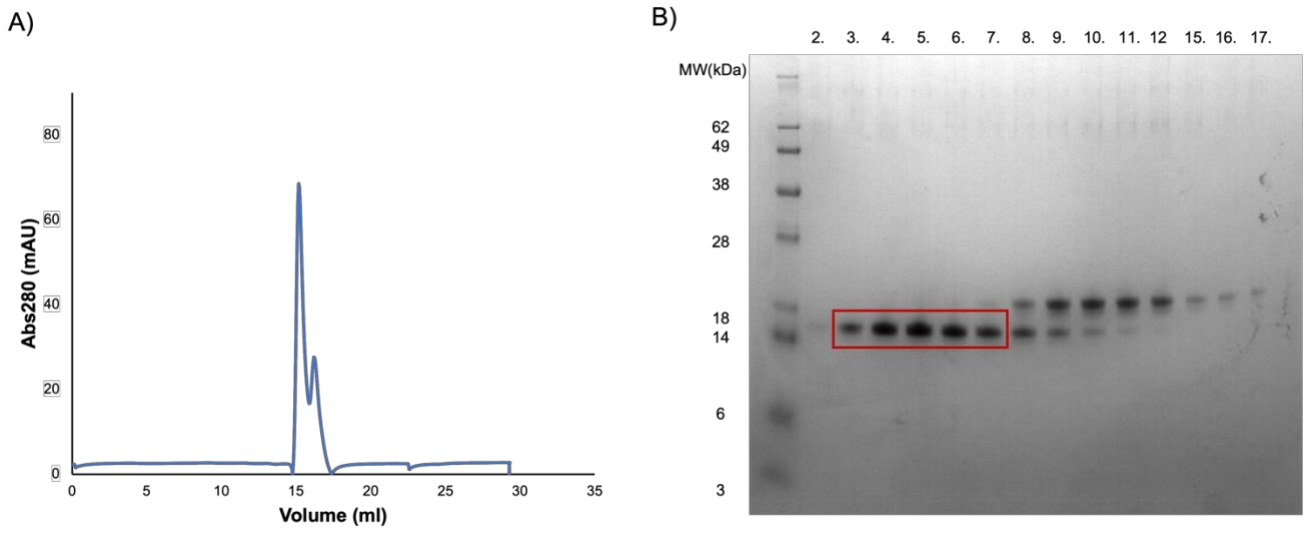


Figure 4.20 Attempt to separate VHH5 and SUMOtag. *A) SEC Chromatography profile. Superdex 75 Increase 10/300 GL (Cytiva). B) SDS-PAGE analysis of SEC shows the eluted fractions (3-7) of peak 1 corresponding to VHH5 (14.8 kDa). The eluted fractions (8-12) of peak 2 correspond to a complex formed by VHH5 (14.8 kDa, lower band) and SUMOtag (11 kDa, upper band). The eluted fractions (15-17) of the final part of the peak 2 correspond to SUMOtag (11 kDa).*

This strategy allowed to obtain very small quantities of VHH5 as around 1 ml (0.355 mg/ml or 24 µM) of >98% pure protein after cleaving it from the tag.

4.10 Preliminary binding test

The small quantity of VHH5 obtained was sufficient to perform a preliminary antigen-antibody binding test. The interaction was observed using the purified VHH5 with both sandwich and indirect ELISA assays as described in paragraph 3.2.6.1. For the indirect ELISA (**Figure 4.21, Panel A**) the purified recombinant TDP-43 fragments (RRM1-2, RRM1 and RRM2) were coated and then the VHH5 was added. On the contrary, for sandwich ELISA (**Figure 4.21, Panel B**) the purified recombinant VHH5 was coated and then the TDP-43 fragments were added. Different molar ratio (1:1, 1:3, 1:5, 1:10) of VHH5 and TDP-43 fragments were used in both assays. In both cases, response to RRM1 (106-176), RRM2 (192-258), and RRM1-2 (106-258) were observed, indicating that the epitope involves both domains. As detect for the mutant, only the indirect ELISA showed a dependence on the antibody to protein ratio. The signal appeared saturated in the sandwich ELISA (**Figure 4.21, Panel B**) but not in the indirect ELISA where the intrabody is in solution and the epitope may be partially masked. Further tests should be done to confirm these results.

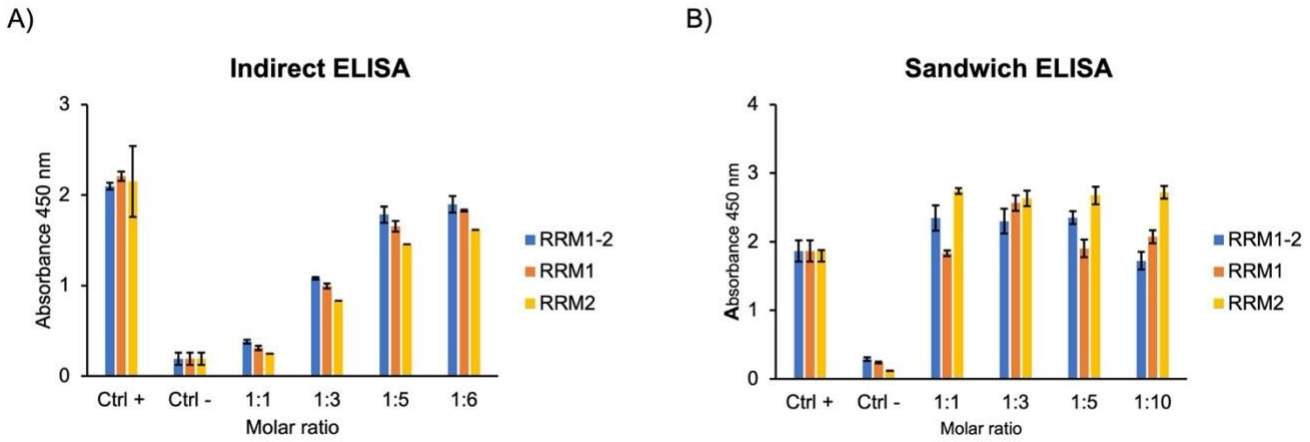


Figure 4.21 ELISA assays of VHH5-TDP-43 fragments interaction. *A) Indirect ELISA assay. Coating antigen: RRM1-2, RRM1 and RRM2 (1 µM); Binding antibody: VHH5 (molar ratio 1:1, 1:3, 1:5, 1:10); Detection: anti-VHH-HRP. The assay shows an interaction of VHH5 with all the TDP-43 fragment. The interaction increases as the molar ratio increases. B) Sandwich ELISA assay. Coating antibody: VHH5 (final molar ratio coating antibody: binding antigen 1:1, 1:3, 1:5, 1:10); Binding antigen: RRMs, RRM1 and RRM2 (1 µM); Detection: anti-TARDBP and then anti-hIgG-HRP. The assay shows an interaction of VHH5 with all the TDP-43 fragments.*

5. Discussion

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by the selective loss of motor neurons resulting in mortality within an average of 2-5 years (8). Emerging evidence suggests that the accumulation of various disease-related misfolded proteins may underlie these pathogenic processes. Despite advances in understanding ALS pathogenesis, however, there is no curative or effective control therapy.

The use of antibodies (Abs) in misfolding diseases is in principle a flexible and powerful strategy to control protein aggregation, because, by binding to a monomeric protein, they prevent self-assembly by steric hindrance. Abs are characterized by highly target specificity that make them potential tools to act on identified key pathogenetic targets. There are now several different strategies that allow screening (256-259), ab initio design (260, 261) or evolutionary selection of antibodies and smaller derivatives (170). A problem remains however the production of the antibody by bacterial expression once a potentially effective sequence has been identified. Unfortunately, the large molecule weight (typically ~ 150 000 Da) and hetero- tetrameric composition of antibodies with two different polypeptides (a heavy and a light chain) and a total of up to 15 disulfide bridges make difficult, when not prohibitive, their production in bacteria or in the cytoplasm of eukaryotic cells. Therefore, scFv fragments, that contain only one copy of the variable domains of immunoglobulin motif, offer undiscussable advantages. However, also in this case, it difficult to predict *a priori* whether an intrabody obtained by library screening can easily be produced in *E. coli* and problems in successfully refolding the intrabody from inclusion bodies have been described (262, 263).

In this work, it was used a composite approach in which an intrabody for TDP-43 recognition was screened, produced in bacteria and characterized it for epitope recognition.

The first aim of this thesis was focused on the selection of a new anti-TDP-43 VHH, named VHH5, directly from the TDP43 cDNA. The VHHs screening was performed in Prof. Antonino Cattaneo's research group at the CNR Neuroscience Institute of Scuola Normale Superiore di Pisa.

Usually, VHH libraries are obtained from immunized animals, and are used in different display platforms (phage, yeast, ribosomal, etc.), that require the immunizing protein for antibody detection from the library. Antonino Cattaneo's research group constructed instead a *llama glaba* naïve VHH library in the SPLINT (Single Pot Library of Intracellular Antibodies) format in yeast, followed by antibody selection with the two-hybrid-based Intracellular Antibody Capture

Technology (IACT) (174, 178, 264). A significant advantage of SPLINT-derived antibodies, as the anti-TDP-43 VHH5 described in this work, is that the genes coding for the antibody domains are by definition well validated as intrabodies, since the IACT selection is performed under conditions of intracellular expression in yeast cells. SPLINT-derived antibody domains are well suited to be used as intrabodies (93), possibly coupled to effector domains for targeted degradation (265, 266) or imaging purpose. This approach allows direct selection of antibodies from antigen cDNA, with no need to express and purify the protein antigen (267). A positive interaction between a prey (VHHs) and the bait (TDP-43) activated transcription of the *HIS3* gene and of a second marker of interaction, the *LacZ* gene, allowing the selection of the anti-TDP-43 VHH5. The selected antibody comprises three CDRs loops according to Chothia and Lesk numbering system (112). CDR3 loop is longer than the average for antibodies, but within average for intrabodies. Its structure prediction is not straightforward as it contains many degrees of freedom.

The second aim was the production and characterization of the selected VHH5. This experimental work was performed in Prof. Annalisa Pastore's research group at the Basic & Clinical Neuroscience Department of King's College London. Several attempts were made to express the protein. As first attempt, I tried a periplasmic expression. By targeting the proteins to the periplasm, it is possible to take advantage of the oxidizing environment and the resident disulfide bond formation (Dsb)-system to facilitate the proper disulfide bonds (268-270). Although using the periplasm is more ideal for disulfide bond containing recombinant proteins, the bottlenecks associated with the targeting across the cytoplasmic membrane can substantially limit periplasmic yields. Since the periplasmic expression shown a suboptimal yields and incomplete removal of peptide leader sequences, a second attempt was performed adding the SUMOtag for increasing the expression levels of recombinant proteins and to significantly improve protein solubility. As shown in paragraph 4.2 the expression yield was appreciably increased but the highly expressed protein accumulated in the cytoplasm as inclusion bodies. Inclusion bodies formation has been proven to result from the conflict between aggregation and protein fold and it is a well-known impediment particularly in antibody production (244).

Based on the amino acid sequence deduced from the DNA sequence of the selected VHH5 intrabody, it was performed an *in silico* prediction of the antibody structure. The antibody main scaffold, a β-sandwich structure, could be easily predicted (246) while the challenge was the prediction of the CDR loops. According to the canonical structures, H1 and H2 loops can be easily classified and predicted (247). However, the high variability of the H3 loop sequence, length, and

conformation made it difficult to build a high-quality structure with ordinary modelling techniques. Using the Sphinx algorithm (234) it was possible to model the H3 loop. Several structures were selected and subsequently used for the complex (VHH5+TDP-43) structure prediction.

The resulting model was used to identify possible mutations for increasing the expression of VHH5 in bacterial cells, enabling the experimental biochemical validation of the intrabody. The model suggested seven exposed hydrophobic residues that could be mutated to reduce the risk of inclusion body formation. It was found that the mutations of two exposed hydrophobic residues (I15A and M75K) was sufficient to have a soluble protein that could be purified in suitable amounts for proper direct characterization. The protein was expressed in a pET-SUMO vector and switched from being all in the inclusion bodies to being mostly soluble. Then, it was demonstrated by far-UV CD and NMR studies that the protein was well folded and monomeric and that it had all the features expected for the expected β-rich structure.

The characterization of epitope recognition was first carried out coarsely, in collaboration with Prof. Antonino Cattaneo's research group. The In vivo Epitope Mapping (IVEM) was performed in two steps. The first cycle of IVEM showed the epitope recognized by the VHH5 located in the N-terminal half of the protein. For a more detailed analysis, a second cycle of IVEM was performed by splitting this region into four smaller baits. Results showed the epitope mainly located in RRM2, as it was detected both in the LexA-RRM2 and the RRMs baits. Then, it was demonstrated by ELISA experiments that the double mutant was still able to recognise the TDP-43 epitope. This conclusion was far from being obvious, since it is known that regions outside the CDR loops can contribute to epitope recognition of intrabodies (271).

More specifically, the epitope binding anti-TDP-43 VHH5(mut) intrabody was mapped by cross-linking and ELISA experiments with individual or tandem domains of TDP-43 and it was found to bind both RRM1 and RRM2. This was in agreement with structural studies that have revealed that VHHs often tend to recognize concave surfaces of their antigens with high shape-complementarity. To substantiate these findings with further evidence it was performed a far-western blot analysis. However, the results obtained were inconclusive because of the aggregation of the TDP-43 fragments in native conditions.

Based on these experimental results, the interaction was modelled by *in silico* docking. Although structures provided by docking, in the absence of experimental restrains, had a low resolution and were very limited in reliability, they could provide a visual impression of the epitope binding and

inform future studies. Despite their overall diversity, structural studies have revealed that VHHs tend to recognize concave surfaces of their antigens with high shape-complementarity. CDR3 of VHH often protrudes from the body of the antibody and docks into clefts located on the surface of antigens with high shape complementarity, in a manner that mimics some human and bovine antibodies. It is believed that in this way, VHHs compensate for the limitations of their small size, while maintaining the high affinity and specificity that constitute the hallmarks of antibodies. This mechanism has been observed in a structural study that compared the binding mode of VHH with that of Fvs using hen egg lysozyme (HEL) as a model antigen (272). Several more studies revealed additional details of how VHHs target the concave surface of antigen molecule (273-275). Indeed, in most of our structures I observed that CDR3 inserts in the groove formed by the interface between the two domains.

It is interesting to compare this intrabody with previously developed anti-TDP-43 antibodies. A systematic survey in 2015 revealed the existence of 29 antibodies, many of which were generated in house (276). Amongst the ten highest-ranking primary antibodies, one has two distinct epitopes, that recognize TDP-43 N-terminus and RRM2. Two other antibodies are directed at RRM2, and three have epitopes in the C-terminus of TDP-43. The remaining four antibodies also map in the C-terminus but are specific for phosphorylated serine residues. The majority of these antibodies are polyclonal and therefore their genes cannot be available for further downstream engineering. A single chain antibody against RRM1 was generated in 2019 (277). Two more monoclonal antibodies, recently described, were raised against an epitope within the RRM2 domain of TDP-43 (residues 198-216) (278).

To the best of my knowledge, this is the first detailed description of an anti-TDP-43 intrabody. The novel antibody will aid in diagnostic and research efforts within the context of TDP-43 proteinopathies. Availability of this new VHH opens new avenues for diagnostic and treatment of ALS, to interfere with protein aggregation and for imaging applications by super-resolution microscopy (98, 279).

Based on the results obtained for the double mutant, preliminary studies are being carried out on the wild-type single-domain intrabody. This experimental work is performed in Prof. Annalisa Pastore's research group. Several optimizations are necessary to produce the wild-type and to demonstrate the TDP-43 epitope recognition by VHH5. These studies are in progress and will benefit of the large amount of evidence and protocols background produced so far.

6. Conclusion and future perspective

In conclusion, in this thesis, a novel single-domain intrabody targeting TDP-43 was selected and characterized from a llama nanobody library.
Moreover, structure prediction was demonstrated to be a powerful tool to guide carefully planned mutagenesis that can facilitate soluble intrabody production.
The idea to use single-domain intrabody to interfere with protein aggregation provide a new approach for the mitigation of TDP-43 proteinopathy and represent a new potential tool for diagnostic. The novel intrabody opens new avenues for the diagnosis and treatment of ALS.

Based on this work, the study and characterization of VHH5 wild-type was planned. This will allow to accurately delineate the protein-protein interaction profile and verify if the mutations introduced have negatively or positively influenced the binding.
Several steps will be necessary to optimize the protocols of VHH5 wild-type production. Strategies for further characterization of VHH5-TDP-43 interaction will be investigated using Isothermal Titration Calorimetry (ITC) and NMR. The nanobody structure derived from NMR data and assignment will allow the identification of precise amino acids involved in the binding and its characterization.
Meanwhile, a cellular study will be launched in collaboration with Antonino Cattaneo's research group. The aim is to investigate the VHH5 expression in mammalian cells and investigate the effects of the nanobody on the aggregation processes of TDP-43. The reduction of cellular aggregates would represent a possibility for the treatment of ALS patients.
The work will be conducted by Annalisa Pastore's research group and related collaborators.

7. Bibliography

1. Przedborski S, Vila M, Jackson-Lewis V. Neurodegeneration: what is it and where are we? J Clin Invest. 2003;111(1):3-10.
2. Chung CG, Lee H, Lee SB. Mechanisms of protein toxicity in neurodegenerative diseases. Cell Mol Life Sci. 2018;75(17):3159-80.
3. Kovacs GG. Concepts and classification of neurodegenerative diseases. Handb Clin Neurol. 2017;145:301-7.
4. Taylor JP, Hardy J, Fischbeck KH. Toxic proteins in neurodegenerative disease. Science. 2002;296(5575):1991-5.
5. Verma M, Vats A, Taneja V. Toxic species in amyloid disorders: Oligomers or mature fibrils. Ann Indian Acad Neurol. 2015;18(2):138-45.
6. Goetz CG. Amyotrophic lateral sclerosis: early contributions of Jean-Martin Charcot. Muscle Nerve. 2000;23(3):336-43.
7. Taylor JP, Brown RH, Jr., Cleveland DW. Decoding ALS: from genes to mechanism. Nature. 2016;539(7628):197-206.
8. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers. 2017;3:17085.
9. Prasad A, Bharathi V, Sivalingam V, Girdhar A, Patel BK. Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis. Front Mol Neurosci. 2019;12:25.
10. Logroscino G, Traynor BJ, Hardiman O, Chio A, Mitchell D, Swingler RJ, et al. Incidence of amyotrophic lateral sclerosis in Europe. J Neurol Neurosurg Psychiatry. 2010;81(4):385-90.
11. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014;17(1):17-23.
12. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 1993;362(6415):59-62.
13. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006;314(5796):130-3.
14. Burrell JR, Halliday GM, Kril JJ, Ittner LM, Gotz J, Kiernan MC, et al. The frontotemporal dementia-motor neuron disease continuum. Lancet. 2016;388(10047):919-31.
15. Devenney E, Vucic S, Hodges JR, Kiernan MC. Motor neuron disease-frontotemporal dementia: a clinical continuum. Expert Rev Neurother. 2015;15(5):509-22.
16. Barmada SJ. Linking RNA Dysfunction and Neurodegeneration in Amyotrophic Lateral Sclerosis. Neurotherapeutics. 2015;12(2):340-51.
17. Fogarty MJ. Driven to decay: Excitability and synaptic abnormalities in amyotrophic lateral sclerosis. Brain Res Bull. 2018;140:318-33.
18. Ruegsegger C, Saxena S. Proteostasis impairment in ALS. Brain Res. 2016;1648(Pt B):571-9.
19. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 2006;351(3):602-11.
20. Mackenzie IR, Bigio EH, Ince PG, Geser F, Neumann M, Cairns NJ, et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann Neurol. 2007;61(5):427-34.
21. Le NT, Chang L, Kovlyagina I, Georgiou P, Safren N, Braunstein KE, et al. Motor neuron disease, TDP-43 pathology, and memory deficits in mice expressing ALS-FTD-linked UBQLN2 mutations. Proc Natl Acad Sci U S A. 2016;113(47):E7580-E9.

22. Afroz T, Perez-Berlanga M, Polymenidou M. Structural Transition, Function and Dysfunction of TDP-43 in Neurodegenerative Diseases. Chimia (Aarau). 2019;73(6):380-90.
23. Ayala YM, Pantano S, D'Ambrogio A, Buratti E, Brindisi A, Marchetti C, et al. Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. J Mol Biol. 2005;348(3):575-88.
24. Cohen TJ, Lee VM, Trojanowski JQ. TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. Trends Mol Med. 2011;17(11):659-67.
25. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. Nat Struct Mol Biol. 2013;20(12):1443-9.
26. Mompean M, Romano V, Pantoja-Uceda D, Stuani C, Baralle FE, Buratti E, et al. The TDP-43 N-terminal domain structure at high resolution. FEBS J. 2016;283(7):1242-60.
27. Ederle H, Funk C, Abou-Ajram C, Hutten S, Funk EBE, Kehlenbach RH, et al. Nuclear egress of TDP-43 and FUS occurs independently of Exportin-1/CRM1. Sci Rep. 2018;8(1):7084.
28. Chang CK, Wu TH, Wu CY, Chiang MH, Toh EK, Hsu YC, et al. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. Biochem Biophys Res Commun. 2012;425(2):219-24.
29. Qin H, Lim LZ, Wei Y, Song J. TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. Proc Natl Acad Sci U S A. 2014;111(52):18619-24.
30. Francois-Moutal L, Perez-Miller S, Scott DD, Miranda VG, Mollasalehi N, Khanna M. Structural Insights Into TDP-43 and Effects of Post-translational Modifications. Front Mol Neurosci. 2019;12:301.
31. Afroz T, Hock EM, Ernst P, Foglieni C, Jambeau M, Gilhespy LAB, et al. Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. Nat Commun. 2017;8(1):45.
32. Wang A, Conicella AE, Schmidt HB, Martin EW, Rhoads SN, Reeb AN, et al. A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. EMBO J. 2018;37(5).
33. Tsoi PS, Choi KJ, Leonard PG, Sizovs A, Moosa MM, MacKenzie KR, et al. The N-Terminal Domain of ALS-Linked TDP-43 Assembles without Misfolding. Angew Chem Int Ed Engl. 2017;56(41):12590-3.
34. Kuo PH, Doudeva LG, Wang YT, Shen CK, Yuan HS. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. Nucleic Acids Res. 2009;37(6):1799-808.
35. Kuo PH, Chiang CH, Wang YT, Doudeva LG, Yuan HS. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. Nucleic Acids Res. 2014;42(7):4712-22.
36. Shodai A, Ido A, Fujiwara N, Ayaki T, Morimura T, Oono M, et al. Conserved acidic amino acid residues in a second RNA recognition motif regulate assembly and function of TDP-43. PLoS One. 2012;7(12):e52776.
37. Shodai A, Morimura T, Ido A, Uchida T, Ayaki T, Takahashi R, et al. Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). J Biol Chem. 2013;288(21):14886-905.
38. Santamaria N, Alhothali M, Alfonso MH, Breydo L, Uversky VN. Intrinsic disorder in proteins involved in amyotrophic lateral sclerosis. Cell Mol Life Sci. 2017;74(7):1297-318.
39. Patel BK, Gavin-Smyth J, Liebman SW. The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. Nat Cell Biol. 2009;11(3):344-9.
40. Liebman SW, Chernoff YO. Prions in yeast. Genetics. 2012;191(4):1041-72.

41. Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell. 2017;171(3):615-27 e16.
42. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149(4):753-67.
43. Jiang LL, Che MX, Zhao J, Zhou CJ, Xie MY, Li HY, et al. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. J Biol Chem. 2013;288(27):19614-24.
44. Mompeán M, Hervás R, Xu Y, Tran TH, Guarnaccia C, Buratti E, et al. Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. The journal of physical chemistry letters. 2015;6(13):2608-15.
45. Guenther EL, Cao Q, Trinh H, Lu J, Sawaya MR, Cascio D, et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. Nat Struct Mol Biol. 2018;25(6):463-71.
46. Cao Q, Boyer DR, Sawaya MR, Ge P, Eisenberg DS. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. Nat Struct Mol Biol. 2019;26(7):619-27.
47. Buratti E, Dork T, Zuccato E, Pagani F, Romano M, Baralle FE. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. EMBO J. 2001;20(7):1774-84.
48. Buratti E, Brindisi A, Giombi M, Tisminetzky S, Ayala YM, Baralle FE. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. J Biol Chem. 2005;280(45):37572-84.
49. Ayala YM, De Conti L, Avendano-Vazquez SE, Dhir A, Romano M, D'Ambrogio A, et al. TDP-43 regulates its mRNA levels through a negative feedback loop. EMBO J. 2011;30(2):277-88.
50. Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure. 2016;24(9):1537-49.
51. Ou SH, Wu F, Harrich D, Garcia-Martinez LF, Gaynor RB. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. J Virol. 1995;69(6):3584-96.
52. Nehls J, Koppensteiner H, Brack-Werner R, Floss T, Schindler M. HIV-1 replication in human immune cells is independent of TAR DNA binding protein 43 (TDP-43) expression. PLoS One. 2014;9(8):e105478.
53. Buratti E, Baralle FE. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. Front Biosci. 2008;13:867-78.
54. Liu EY, Cali CP, Lee EB. RNA metabolism in neurodegenerative disease. Dis Model Mech. 2017;10(5):509-18.
55. Ratti A, Buratti E. Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. J Neurochem. 2016;138 Suppl 1:95-111.
56. Buratti E, Baralle FE. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. J Biol Chem. 2001;276(39):36337-43.
57. Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci. 2011;14(4):452-8.
58. Colombrita C, Onesto E, Megiorni F, Pizzuti A, Baralle FE, Buratti E, et al. TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. J Biol Chem. 2012;287(19):15635-47.
59. Kawahara Y, Mieda-Sato A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc Natl Acad Sci U S A. 2012;109(9):3347-52.

60. Ling SC, Albuquerque CP, Han JS, Lagier-Tourenne C, Tokunaga S, Zhou H, et al. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc Natl Acad Sci U S A. 2010;107(30):13318-23.
61. Buratti E, De Conti L, Stuani C, Romano M, Baralle M, Baralle F. Nuclear factor TDP-43 can affect selected microRNA levels. FEBS J. 2010;277(10):2268-81.
62. Dewey CM, Cenik B, Sephton CF, Dries DR, Mayer P, 3rd, Good SK, et al. TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. Mol Cell Biol. 2011;31(5):1098-108.
63. Colombrita C, Zennaro E, Fallini C, Weber M, Sommacal A, Buratti E, et al. TDP-43 is recruited to stress granules in conditions of oxidative insult. J Neurochem. 2009;111(4):1051-61.
64. Aulas A, Vande Velde C. Alterations in stress granule dynamics driven by TDP-43 and FUS: a link to pathological inclusions in ALS? Front Cell Neurosci. 2015;9:423.
65. Bentmann E, Neumann M, Tahirovic S, Rodde R, Dormann D, Haass C. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). J Biol Chem. 2012;287(27):23079-94.
66. Vanden Broeck L, Kleinberger G, Chapuis J, Gistelinck M, Amouyel P, Van Broeckhoven C, et al. Functional complementation in Drosophila to predict the pathogenicity of TARDBP variants: evidence for a loss-of-function mechanism. Neurobiol Aging. 2015;36(2):1121-9.
67. Ederle H, Dormann D. TDP-43 and FUS en route from the nucleus to the cytoplasm. FEBS Lett. 2017;591(11):1489-507.
68. Lee EB, Lee VM, Trojanowski JQ. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. Nat Rev Neurosci. 2011;13(1):38-50.
69. Xu Z-S. Does a loss of TDP-43 function cause neurodegeneration? Molecular neurodegeneration. 2012;7(1):1-10.
70. Strong MJ, Volkening K, Hammond R, Yang W, Strong W, Leystra-Lantz C, et al. TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. Mol Cell Neurosci. 2007;35(2):320-7.
71. Ayala YM, Misteli T, Baralle FE. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. Proc Natl Acad Sci U S A. 2008;105(10):3785-9.
72. Winton MJ, Igaz LM, Wong MM, Kwong LK, Trojanowski JQ, Lee VM. Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. J Biol Chem. 2008;283(19):13302-9.
73. Igaz LM, Kwong LK, Xu Y, Truax AC, Uryu K, Neumann M, et al. Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Am J Pathol. 2008;173(1):182-94.
74. Igaz LM, Kwong LK, Chen-Plotkin A, Winton MJ, Unger TL, Xu Y, et al. Expression of TDP-43 C-terminal Fragments in Vitro Recapitulates Pathological Features of TDP-43 Proteinopathies. J Biol Chem. 2009;284(13):8516-24.
75. Neumann M, Kwong LK, Lee EB, Kremmer E, Flatley A, Xu Y, et al. Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. Acta Neuropathol. 2009;117(2):137-49.
76. Geser F, Lee VM, Trojanowski JQ. Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of TDP-43 proteinopathies. Neuropathology. 2010;30(2):103-12.
77. Ling H, Lenz TL, Burns TL, Hilleman DE. Reducing the risk of obesity: defining the role of weight loss drugs. Pharmacotherapy. 2013;33(12):1308-21.
78. Gao J, Wang L, Huntley ML, Perry G, Wang X. Pathomechanisms of TDP-43 in neurodegeneration. J Neurochem. 2018.

79. Buratti E, Baralle FE. The molecular links between TDP-43 dysfunction and neurodegeneration. Adv Genet. 2009;66:1-34.
80. Barmada SJ, Skibinski G, Korb E, Rao EJ, Wu JY, Finkbeiner S. Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. J Neurosci. 2010;30(2):639-49.
81. Pesiridis GS, Lee VM, Trojanowski JQ. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. Hum Mol Genet. 2009;18(R2):R156-62.
82. Zacco E, Martin SR, Thorogate R, Pastore A. The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. Front Mol Neurosci. 2018;11:372.
83. Zacco E, Grana-Montes R, Martin SR, de Groot NS, Alfano C, Tartaglia GG, et al. RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. J Mol Biol. 2019;431(8):1671-88.
84. Budini M, Romano V, Quadri Z, Buratti E, Baralle FE. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. Hum Mol Genet. 2015;24(1):9-20.
85. Chen HJ, Topp SD, Hui HS, Zacco E, Katarya M, McLoughlin C, et al. RRM adjacent TARDBP mutations disrupt RNA binding and enhance TDP-43 proteinopathy. Brain. 2019;142(12):3753-70.
86. Protter DSW, Parker R. Principles and Properties of Stress Granules. Trends Cell Biol. 2016;26(9):668-79.
87. Wolozin B, Ivanov P. Stress granules and neurodegeneration. Nat Rev Neurosci. 2019;20(11):649-66.
88. Fan AC, Leung AK. RNA Granules and Diseases: A Case Study of Stress Granules in ALS and FTLD. Adv Exp Med Biol. 2016;907:263-96.
89. Samir P, Kesavardhana S, Patmore DM, Gingras S, Malireddi RKS, Karki R, et al. DDX3X acts as a live-or-die checkpoint in stressed cells by regulating NLRP3 inflammasome. Nature. 2019;573(7775):590-4.
90. Khalfallah Y, Kuta R, Grasmuck C, Prat A, Durham HD, Velde CV. TDP-43 regulation of stress granule dynamics in neurodegenerative disease-relevant cell types. Scientific reports. 2018;8(1):1-13.
91. Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. Orphanet J Rare Dis. 2009;4:3.
92. Rossi S, Rompietti V, Antonucci Y, Giovannini D, Scopa C, Scaricamazza S, et al. UsnRNP trafficking is regulated by stress granules and compromised by mutant ALS proteins. Neurobiology of disease. 2020;138:104792.
93. Biocca S, Neuberger MS, Cattaneo A. Expression and targeting of intracellular antibodies in mammalian cells. EMBO J. 1990;9(1):101-8.
94. Cattaneo A, Chirichella M. Targeting the Post-translational Proteome with Intrabodies. Trends Biotechnol. 2019;37(6):578-91.
95. Meli G, Lecci A, Manca A, Krako N, Albertini V, Benussi L, et al. Conformational targeting of intracellular Abeta oligomers demonstrates their pathological oligomerization inside the endoplasmic reticulum. Nat Commun. 2014;5:3867.
96. Sograte-Idrissi S, Oleksiievets N, Isbaner S, Eggert-Martinez M, Enderlein J, Tsukanov R, et al. Nanobody Detection of Standard Fluorescent Proteins Enables Multi-Target DNA-PAINT with High Resolution and Minimal Displacement Errors. Cells. 2019;8(1).
97. Oi C, Mochrie SGJ, Horrocks MH, Regan L. PAINT using proteins: A new brush for super-resolution artists. Protein Sci. 2020;29(11):2142-9.
98. Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O, et al. Super-resolution microscopy demystified. Nat Cell Biol. 2019;21(1):72-84.

99. Schiffer M, Girling RL, Ely KR, Edmundson AB. Structure of a lambda-type Bence-Jones protein at 3.5-A resolution. Biochemistry. 1973;12(23):4620-31.
100. Poljak RJ, Amzel LM, Avey HP, Chen BL, Phizackerley RP, Saul F. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2,8-A resolution. Proc Natl Acad Sci U S A. 1973;70(12):3305-10.
101. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
102. Amzel LM, Poljak RJ. Three-dimensional structure of immunoglobulins. Annu Rev Biochem. 1979;48:961-97.
103. Davies DR, Metzger H. Structural basis of antibody function. Annu Rev Immunol. 1983;1:87-117.
104. Saul FA, Poljak RJ. Structure and specificity of antibody molecules. Ann Inst Pasteur Immunol (1985). 1985;136C(2):263-7.
105. Janeway Jr CA, Travers P, Walport M, Shlomchik MJ. The structure of a typical antibody molecule. Immunobiology: The Immune System in Health and Disease 5th edition: Garland Science; 2001.
106. Smith K, Shah H, Muther JJ, Duke AL, Haley K, James JA. Antigen nature and complexity influence human antibody light chain usage and specificity. Vaccine. 2016;34(25):2813-20.
107. Schroeder HW, Jr., Cavacini L. Structure and function of immunoglobulins. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S41-52.
108. Cohen S, Milstein C. Structure and biological properties of immunoglobulins. Adv Immunol. 1967;7:1-89.
109. Alberts B, Johnson A, Lewis J, Walter P, Raff M, Roberts K. Molecular Biology of the Cell 4th Edition: International Student Edition: Routledge; 2002.
110. Deveuve Q, Gouilleux-Gruart V, Thibault G, Lajoie L. [The hinge region of therapeutic antibodies: major importance of a short sequence]. Med Sci (Paris). 2019;35(12):1098-105.
111. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol. 1987;196(4):901-17.
112. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, et al. Conformations of immunoglobulin hypervariable regions. Nature. 1989;342(6252):877-83.
113. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL. Antibody Structure and Function: The Basis for Engineering Therapeutics. Antibodies (Basel). 2019;8(4).
114. Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, et al. Single-chain antigen-binding proteins. Science. 1988;242(4877):423-6.
115. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. Nature. 1989;341(6242):544-6.
116. Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, et al. Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol. 2005;352(3):597-607.
117. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. Nature. 1993;363(6428):446-8.
118. Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng. 1994;7(9):1129-35.
119. Vu KB, Ghahroudi MA, Wyns L, Muyldermans S. Comparison of llama VH sequences from conventional and heavy chain antibodies. Mol Immunol. 1997;34(16-17):1121-31.

120. Harmsen MM, Ruuls RC, Nijman IJ, Niewold TA, Frenken LG, de Geus B. Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. Mol Immunol. 2000;37(10):579-90.
121. Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, Hamers R, et al. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol. 1996;3(9):803-11.
122. Spinelli S, Frenken L, Bourgeois D, de Ron L, Bos W, Verrips T, et al. The crystal structure of a llama heavy chain variable domain. Nat Struct Biol. 1996;3(9):752-7.
123. Zavrtanik U, Lukan J, Loris R, Lah J, Hadzi S. Structural Basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. J Mol Biol. 2018;430(21):4369-86.
124. Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. Protein Eng Des Sel. 2018;31(7-8):267-75.
125. Mitchell LS, Colwell LJ. Comparative analysis of nanobody sequence and structure data. Proteins. 2018;86(7):697-706.
126. Kolkman JA, Law DA. Nanobodies–from llamas to therapeutic proteins. Drug discovery today: technologies. 2010;7(2):e139-e46.
127. Muyldermans S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, Kinne J, et al. Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol. 2009;128(1-3):178-83.
128. Dumoulin M, Conrath K, Van Meirhaeghe A, Meersman F, Heremans K, Frenken LG, et al. Single-domain antibody fragments with high conformational stability. Protein Sci. 2002;11(3):500-15.
129. Dolk E, van Vliet C, Perez JM, Vriend G, Darbon H, Ferrat G, et al. Induced refolding of a temperature denatured llama heavy-chain antibody fragment by its antigen. Proteins. 2005;59(3):555-64.
130. Hussack G, Arbabi-Ghahroudi M, van Faassen H, Songer JG, Ng KK, MacKenzie R, et al. Neutralization of Clostridium difficile toxin A with single-domain antibodies targeting the cell receptor binding domain. J Biol Chem. 2011;286(11):8961-76.
131. Hussack G, Hirama T, Ding W, Mackenzie R, Tanha J. Engineered single-domain antibodies with high protease resistance and thermal stability. PLoS One. 2011;6(11):e28218.
132. Harmsen MM, De Haard HJ. Properties, production, and applications of camelid single-domain antibody fragments. Appl Microbiol Biotechnol. 2007;77(1):13-22.
133. Messer A, Butler DC. Optimizing intracellular antibodies (intrabodies/nanobodies) to treat neurodegenerative disorders. Neurobiol Dis. 2020;134:104619.
134. Khodabakhsh F, Behdani M, Rami A, Kazemi-Lomedasht F. Single-Domain Antibodies or Nanobodies: A Class of Next-Generation Antibodies. Int Rev Immunol. 2018;37(6):316-22.
135. Hoey RJ, Eom H, Horn JR. Structure and development of single domain antibodies as modules for therapeutics and diagnostics. Exp Biol Med (Maywood). 2019;244(17):1568-76.
136. Rahbarizadeh F, Ahmadvand D, Sharifzadeh Z. Nanobody; an old concept and new vehicle for immunotargeting. Immunol Invest. 2011;40(3):299-338.
137. Arbabi Ghahroudi M, Desmyter A, Wyns L, Hamers R, Muyldermans S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS Lett. 1997;414(3):521-6.
138. van der Linden R, de Geus B, Stok W, Bos W, van Wassenaar D, Verrips T, et al. Induction of immune responses and molecular cloning of the heavy chain antibody repertoire of Lama glama. J Immunol Methods. 2000;240(1-2):185-95.
139. Tanha J, Dubuc G, Hirama T, Narang SA, MacKenzie CR. Selection by phage display of llama conventional V(H) fragments with heavy chain antibody V(H)H properties. J Immunol Methods. 2002;263(1-2):97-109.

140. Yau KY, Groves MA, Li S, Sheedy C, Lee H, Tanha J, et al. Selection of hapten-specific single-domain antibodies from a non-immunized llama ribosome display library. J Immunol Methods. 2003;281(1-2):161-75.
141. Verheesen P, Roussis A, de Haard HJ, Groot AJ, Stam JC, den Dunnen JT, et al. Reliable and controllable antibody fragment selections from Camelid non-immune libraries for target validation. Biochim Biophys Acta. 2006;1764(8):1307-19.
142. Goldman ER, Anderson GP, Liu JL, Delehanty JB, Sherwood LJ, Osborn LE, et al. Facile generation of heat-stable antiviral and antitoxin single domain antibodies from a semisynthetic llama library. Anal Chem. 2006;78(24):8245-55.
143. Revets H, De Baetselier P, Muyldermans S. Nanobodies as novel agents for cancer therapy. Expert Opin Biol Ther. 2005;5(1):111-24.
144. Liu W, Song H, Chen Q, Yu J, Xian M, Nian R, et al. Recent advances in the selection and identification of antigen-specific nanobodies. Mol Immunol. 2018;96:37-47.
145. Salvador JP, Vilaplana L, Marco MP. Nanobody: outstanding features for diagnostic and therapeutic applications. Anal Bioanal Chem. 2019;411(9):1703-13.
146. Wang Y, Fan Z, Shao L, Kong X, Hou X, Tian D, et al. Nanobody-derived nanobiotechnology tool kits for diverse biomedical and biotechnology applications. Int J Nanomedicine. 2016;11:3287-303.
147. Woods J. Selection of Functional Intracellular Nanobodies. SLAS Discov. 2019;24(7):703-13.
148. Olichon A, de Marco A. Preparation of a naive library of camelid single domain antibodies. Methods Mol Biol. 2012;911:65-78.
149. Sevy AM, Chen MT, Castor M, Sylvia T, Krishnamurthy H, Ishchenko A, et al. Structure- and sequence-based design of synthetic single-domain antibody libraries. Protein Eng Des Sel. 2020;33.
150. Liu Y, Huang H. Expression of single-domain antibody in different systems. Appl Microbiol Biotechnol. 2018;102(2):539-51.
151. Jovcevska I, Muyldermans S. The Therapeutic Potential of Nanobodies. BioDrugs. 2020;34(1):11-26.
152. Li T, Bourgeois JP, Celli S, Glacial F, Le Sourd AM, Mecheri S, et al. Cell-penetrating anti-GFAP VHH and corresponding fluorescent fusion protein VHH-GFP spontaneously cross the blood-brain barrier and specifically recognize astrocytes: application to brain imaging. FASEB J. 2012;26(10):3969-79.
153. Rissiek B, Koch-Nolte F, Magnus T. Nanobodies as modulators of inflammation: potential applications for acute brain injury. Front Cell Neurosci. 2014;8:344.
154. Bathula NV, Bommadevara H, Hayes JM. Nanobodies: The Future of Antibody-Based Immune Therapeutics. Cancer Biother Radiopharm. 2021;36(2):109-22.
155. Miles JS, Wolf CR. Principles of DNA cloning. BMJ. 1989;299(6706):1019-22.
156. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239(4839):487-91.
157. Studier FW. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol. 1991;219(1):37-44.
158. Overton TW. Recombinant protein production in bacterial hosts. Drug Discov Today. 2014;19(5):590-601.
159. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A. 1973;70(11):3240-4.
160. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. Genome Res. 2000;10(11):1788-95.
161. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343-5.

162. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One. 2008;3(11):e3647.
163. Imai Y, Matsushima Y, Sugimura T, Terada M. A simple and rapid method for generating a deletion by PCR. Nucleic Acids Res. 1991;19(10):2785.
164. Zheng L, Baumann U, Reymond JL. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 2004;32(14):e115.
165. Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A. 1985;82(2):488-92.
166. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene. 1989;77(1):51-9.
167. Landt O, Grunert HP, Hahn U. A general method for rapid site-directed mutagenesis using the polymerase chain reaction. Gene. 1990;96(1):125-8.
168. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ. A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res. 1989;17(16):6545-51.
169. Qi D, Scholthof KB. A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. J Virol Methods. 2008;149(1):85-90.
170. Visintin M, Quondam M, Cattaneo A. The intracellular antibody capture technology: towards the high-throughput selection of functional intracellular antibodies for target validation. Methods. 2004;34(2):200-14.
171. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340(6230):245-6.
172. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci. 2009;10(6):2763-88.
173. Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, et al. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev. 1993;7(4):555-69.
174. Visintin M, Settanni G, Maritan A, Graziosi S, Marks JD, Cattaneo A. The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. J Mol Biol. 2002;317(1):73-83.
175. Hartley DA. Cellular interactions in development: a practical approach1993.
176. Hollenberg SM, Sternglanz R, Cheng PF, Weintraub H. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. Mol Cell Biol. 1995;15(7):3813-22.
177. Schiestl RH, Gietz RD. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr Genet. 1989;16(5-6):339-46.
178. Visintin M, Tse E, Axelson H, Rabbitts TH, Cattaneo A. Selection of antibodies for intracellular function using a two-hybrid in vivo system. Proc Natl Acad Sci U S A. 1999;96(21):11723-8.
179. Serebriiskii IG, Golemis EA. Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system. Anal Biochem. 2000;285(1):1-15.
180. Mockli N, Auerbach D. Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. Biotechniques. 2004;36(5):872-6.
181. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.
182. Tabor S. Expression using the T7 RNA polymerase/promoter system. Curr Protoc Mol Biol. 2001;Chapter 16:Unit16 2.
183. Gaberc-Porekar V, Menart V. Perspectives of immobilized-metal affinity chromatography. J Biochem Biophys Methods. 2001;49(1-3):335-60.
184. Panavas T, Sanders C, Butt TR. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. Methods Mol Biol. 2009;497:303-17.

185. Blommel PG, Fox BG. A combined approach to improving large-scale production of tobacco etch virus protease. Protein Expr Purif. 2007;55(1):53-68.
186. Wu J, Filutowicz M. Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. Acta Biochim Pol. 1999;46(3):591-9.
187. Perron-Savard P, De Crescenzo G, Moual HL. Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent. Microbiology (Reading). 2005;151(Pt 12):3979-87.
188. Chan WT, Verma CS, Lane DP, Gan SK. A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. Biosci Rep. 2013;33(6).
189. Coskun O. Separation techniques: Chromatography. North Clin Istanb. 2016;3(2):156-60.
190. Harwood LM, Moody CJ. Experimental organic chemistry: Blackwell Scientific; 1989.
191. Biosciences A. Ion Exchange chromatohgraphy, Principles and methods, Amercham Pharmacia. Biotech SE. 2002;751.
192. Firer MA. Efficient elution of functional proteins in affinity chromatography. J Biochem Biophys Methods. 2001;49(1-3):433-42.
193. Das M, Dasgupta D. Pseudo-affinity column chromatography based rapid purification procedure for T7 RNA polymerase. Prep Biochem Biotechnol. 1998;28(4):339-48.
194. Helmut D. Gel Chromatography, Gel Filtration, Gel Permeation, Molecular Sieves: a Laboratory Handbook. Springer-Verlag; 1969.
195. Eiceman GA, Gardea-Torresdey J, Overton E, Carney K, Dorman F. Gas chromatography. Anal Chem. 2004;76(12):3387-94.
196. Regnier FE. High-performance liquid chromatography of proteins. Methods Enzymol. 1983;91:137-90.
197. Sherma J. Thin-layer and paper chromatography. Anal Chem. 1988;60(12):74R-86R.
198. Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature. 1975;258(5536):598-9.
199. Hearon JZ. The configuration of cobaltodihistidine and oxy-bis (cobaltodihistidine). J Natl Cancer Inst. 1948;9(1):1-11.
200. Cheung RC, Wong JH, Ng TB. Immobilized metal ion affinity chromatography: a review on its applications. Appl Microbiol Biotechnol. 2012;96(6):1411-20.
201. Barth HG, Boyes BE, Jackson C. Size exclusion chromatography. Anal Chem. 1994;66(12):595R-620R.
202. Burgess RR. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. Protein Expr Purif. 2018;150:81-5.
203. Dennis-Sykes CA, Miller WJ, McAleer WJ. A quantitative Western Blot method for protein measurement. J Biol Stand. 1985;13(4):309-14.
204. Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. N Am J Med Sci. 2012;4(9):429-34.
205. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc. 2006;1(6):2876-90.
206. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci. 2000;1(4):349-84.
207. Juszczyk P, Kolodziejczyk AS, Grzonka Z. Circular dichroism and aggregation studies of amyloid beta (11-8) fragment and its variants. Acta Biochim Pol. 2005;52(2):425-31.
208. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc. 2006;1(6):2527-35.
209. Holzwarth G, Doty P. The Ultraviolet Circular Dichroism of Polypeptides. J Am Chem Soc. 1965;87:218-28.

210. Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry. 1969;8(10):4108-16.
211. Venyaminov S, Baikalov IA, Shen ZM, Wu CS, Yang JT. Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. Anal Biochem. 1993;214(1):17-24.
212. Wallace BA, Lees JG, Orry AJ, Lobley A, Janes RW. Analyses of circular dichroism spectra of membrane proteins. Protein Sci. 2003;12(4):875-84.
213. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers. 2008;89(5):392-400.
214. Marion D. An introduction to biological NMR spectroscopy. Mol Cell Proteomics. 2013;12(11):3006-25.
215. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ. Protein NMR Spectroscopy: Principles and Practice: Elsevier Science; 1995.
216. Alhajj M, Farhana A. Enzyme Linked Immunosorbent Assay. StatPearls. Treasure Island (FL)2021.
217. Sakamoto S, Putalun W, Vimolmangkang S, Phoolcharoen W, Shoyama Y, Tanaka H, et al. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. J Nat Med. 2018;72(1):32-42.
218. Tang X, Bruce JE. Chemical cross-linking for protein-protein interaction studies. Methods Mol Biol. 2009;492:283-93.
219. Barth M, Schmidt C. Quantitative Cross-Linking of Proteins and Protein Complexes. Methods Mol Biol. 2021;2228:385-400.
220. Wittig I, Schagger H. Advantages and limitations of clear-native PAGE. Proteomics. 2005;5(17):4338-46.
221. Walsh BW, Lenhart JS, Schroeder JW, Simmons LA. Far Western blotting as a rapid and efficient method for detecting interactions between DNA replication and DNA repair proteins. Methods Mol Biol. 2012;922:161-8.
222. Machida K, Mayer BJ. Detection of protein-protein interactions by far-western blotting. Methods Mol Biol. 2009;536:313-29.
223. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc. 2009;4(1):1-13.
224. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res. 2007;35(Database issue):D301-3.
225. Kopp J, Schwede T. Automated protein structure homology modeling: a progress report. Pharmacogenomics. 2004;5(4):405-16.
226. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.
227. Leem J, Dunbar J, Georges G, Shi J, Deane CM. ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation. MAbs. 2016;8(7):1259-68.
228. Biasini M, Schmidt T, Bienert S, Mariani V, Studer G, Haas J, et al. OpenStructure: an integrated software framework for computational structural biology. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 5):701-9.
229. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42(Web Server issue):W252-8.
230. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep. 2017;7(1):10480.

231. Dunbar J, Fuchs A, Shi J, Deane CM. ABangle: characterising the VH-VL orientation in antibodies. Protein Eng Des Sel. 2013;26(10):611-20.
232. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009;77(4):778-95.
233. Jarasch A, Koll H, Regula JT, Bader M, Papadimitriou A, Kettenberger H. Developability assessment during the selection of novel therapeutic antibodies. J Pharm Sci. 2015;104(6):1885-98.
234. Marks C, Nowak J, Klostermann S, Georges G, Dunbar J, Shi J, et al. Sphinx: merging knowledge-based and ab initio approaches to improve protein loop prediction. Bioinformatics. 2017;33(9):1346-53.
235. Choi Y, Deane CM. FREAD revisited: Accurate loop structure prediction using a database search algorithm. Proteins. 2010;78(6):1431-40.
236. Deane CM, Blundell TL. CODA: a combined algorithm for predicting the structurally variable regions of protein models. Protein Sci. 2001;10(3):599-612.
237. Dong GQ, Fan H, Schneidman-Duhovny D, Webb B, Sali A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. Bioinformatics. 2013;29(24):3158-66.
238. Gilodi M, Lisi S, F. Dudás E, Fantini M, Puglisi R, Louka A, et al. Selection and Modelling of a New Single-Domain Intrabody Against TDP-43. Frontiers in Molecular Biosciences. 2022;8.
239. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc. 2017;12(2):255-78.
240. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.
241. Van Der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008;9:2579-625.
242. Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics. 2007;8:65.
243. Li L, Chen S, Miao Z, Liu Y, Liu X, Xiao ZX, et al. AbRSA: A robust tool for antibody numbering. Protein Sci. 2019;28(8):1524-31.
244. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. Trends Biotechnol. 2006;24(4):179-85.
245. Sanchez de Groot N, Pallares I, Aviles FX, Vendrell J, Ventura S. Prediction of "hot spots" of aggregation in disease-linked polypeptides. BMC Struct Biol. 2005;5:18.
246. Narciso JE, Uy ID, Cabang AB, Chavez JF, Pablo JL, Padilla-Concepcion GP, et al. Analysis of the antibody structure based on high-resolution crystallographic studies. N Biotechnol. 2011;28(5):435-47.
247. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol. 1997;273(4):927-48.
248. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins. 1992;12(4):345-64.
249. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996;8(4):477-86.
250. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res. 2001;29(1):221-2.
251. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018;27(1):293-315.
252. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003;50(3):437-50.

253. Budyak IL, Zhuravleva A, Gierasch LM. The Role of Aromatic-Aromatic Interactions in Strand-Strand Stabilization of beta-Sheets. J Mol Biol. 2013;425(18):3522-35.
254. De Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. Microbial cell factories. 2009;8(1):1-18.
255. Park-Sarge OK, Sarge KD. Detection of sumoylated proteins. Methods Mol Biol. 2009;464:255-65.
256. Smith GP, Petrenko VA. Phage Display. Chem Rev. 1997;97(2):391-410.
257. Uchanski T, Zogg T, Yin J, Yuan D, Wohlkonig A, Fischer B, et al. An improved yeast surface display platform for the screening of nanobody immune libraries. Sci Rep. 2019;9(1):382.
258. Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A. 1997;94(10):4937-42.
259. Ho M, Pastan I. Mammalian cell display for antibody engineering. Methods Mol Biol. 2009;525:337-52, xiv.
260. Hardin C, Pogorelov TV, Luthey-Schulten Z. Ab initio protein structure prediction. Curr Opin Struct Biol. 2002;12(2):176-81.
261. Zhu K, Day T. Ab initio structure prediction of the antibody hypervariable H3 loop. Proteins. 2013;81(6):1081-9.
262. Vaks L, Benhar I. Production of stabilized scFv antibody fragments in the E. coli bacterial cytoplasm. Methods Mol Biol. 2014;1060:171-84.
263. Bao X, Xu L, Lu X, Jia L. Optimization of dilution refolding conditions for a camelid single domain antibody against human beta-2-microglobulin. Protein Expr Purif. 2016;117:59-66.
264. Visintin M, Meli GA, Cannistraci I, Cattaneo A. Intracellular antibodies for proteomics. J Immunol Methods. 2004;290(1-2):135-53.
265. Melchionna T, Cattaneo A. A protein silencing switch by ligand-induced proteasome-targeting intrabodies. J Mol Biol. 2007;374(3):641-54.
266. Schapira M, Calabrese MF, Bullock AN, Crews CM. Targeted protein degradation: expanding the toolbox. Nat Rev Drug Discov. 2019;18(12):949-63.
267. Meli G, Visintin M, Cannistraci I, Cattaneo A. Direct in vivo intracellular selection of conformation-sensitive antibody domains targeting Alzheimer's amyloid-beta oligomers. J Mol Biol. 2009;387(3):584-606.
268. Bardwell JC, McGovern K, Beckwith J. Identification of a protein required for disulfide bond formation in vivo. Cell. 1991;67(3):581-9.
269. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol. 2004;22(11):1399-408.
270. Landeta C, Boyd D, Beckwith J. Disulfide bond formation in prokaryotes. Nat Microbiol. 2018;3(3):270-80.
271. Sela-Culang I, Kunik V, Ofran Y. The structural basis of antibody-antigen recognition. Front Immunol. 2013;4:302.
272. Akiba H, Tamura H, Kiyoshi M, Yanaka S, Sugase K, Caaveiro JMM, et al. Structural and thermodynamic basis for the recognition of the substrate-binding cleft on hen egg lysozyme by a single-domain antibody. Sci Rep. 2019;9(1):15481.
273. Kromann-Hansen T, Oldenburg E, Yung KW, Ghassabeh GH, Muyldermans S, Declerck PJ, et al. A Camelid-derived Antibody Fragment Targeting the Active Site of a Serine Protease Balances between Inhibitor and Substrate Behavior. J Biol Chem. 2016;291(29):15156-68.
274. Rossey I, Gilman MS, Kabeche SC, Sedeyn K, Wrapp D, Kanekiyo M, et al. Potent single-domain antibodies that arrest respiratory syncytial virus fusion protein in its prefusion state. Nat Commun. 2017;8:14158.

275. Gulati S, Jin H, Masuho I, Orban T, Cai Y, Pardon E, et al. Targeting G protein-coupled receptor signaling at the G protein level with a selective nanobody inhibitor. Nat Commun. 2018;9(1):1996.
276. Goossens J, Vanmechelen E, Trojanowski JQ, Lee VM, Van Broeckhoven C, van der Zee J, et al. TDP-43 as a possible biomarker for frontotemporal lobar degeneration: a systematic review of existing antibodies. Acta Neuropathol Commun. 2015;3:15.
277. Pozzi S, Thammisetty SS, Codron P, Rahimian R, Plourde KV, Soucy G, et al. Virus-mediated delivery of antibody targeting TAR DNA-binding protein-43 mitigates associated neuropathology. J Clin Invest. 2019;129(4):1581-95.
278. Trejo-Lopez JA, Sorrentino ZA, Riffe CJ, Lloyd GM, Labuzan SA, Dickson DW, et al. Novel monoclonal antibodies targeting the RRM2 domain of human TDP-43 protein. Neurosci Lett. 2020;738:135353.
279. Messer A, Joshi SN. Intrabodies as neuroprotective therapeutics. Neurotherapeutics. 2013;10(3):447-58.

Appendix

VHH5(mut)

Histag-SUMOtag-TEV-VHH5(mut)

```
atgggcagcagccatcatcatcatcatcatggcagcgccctggtgccgcgcggcagcgctagcatgtcggactcagaagtc
M G S S H H H H H H G S G L V P R G S A S M S D S E V
aatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactcacatcaatttaaaggtgttcgatggatcttca
N Q E A K P E V K P E V K P E T H I N L K V S D G S S
gatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaagacag
E G S S D I F F K I K K T T P L R R L M E A F A K R Q
ggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcagaccccctgaagatttggacatg
G K E M D S L R F L Y D G I R I Q A D Q T P E D L D M
gaggataacgatattattgaggctcacagagaacagattggtggtgagaacctgtacttccagggcatgaaggtgcaatta
E D N D I I E A H R E Q I G G E N L Y F Q G M K V Q L
gtcgaaagcggcgggggagcgttcaaGCcggcggcagcttacgtctctcatgtcgtatctctgaaggttccggcaataaa
V E S G G G S V Q A G G S L R L S C R I S E G S G N K
tacgtgatggcttggttccggcaggctccgggccaagagcgggagttcgtgggcgtgatcagctggagcggcacccgtaca
Y V M A W F R Q A P G Q E R E F V G V I S W S G T R T
cactacgcagattcagttcgtggtcgtttcactatcagtcgcaAggataacacagcttacctccaaatggattctctgaca
H Y A D S V R G R F T I S R K D N T A Y L Q M D S L T
ccggatgactcaggggtctactactgtgccgcatcggtttcagttccgccttttctggagttgacagctgcccatttttggg
P D D S G V Y Y C A A S V S V P P F L E L T A A H F G
tcgtggggccaggggacacaagttcgtgtcgggtcatcgccagaaccaaaaaccccaaagcctcagcctgcatcatag
S W G Q G T Q V R V G S S P E P K T P K P Q P A S -
```

	HisTag-SUMOtag-TEV-VHH5(mut)		VHH5(mut)	
Number of amino acids	264		138	
Molecular weight	29097.50		14847.60	
Theoretical pI	6.37		9.03	
Amino acid composition	Ala (A) 16	6.1%	Ala (A) 10	7.2%
	Arg (R) 16	6.1%	Arg (R) 9	6.5%
	Asn (N) 6	2.3%	Asn (N) 2	1.4%
	Asp (D) 14	5.3%	Asp (D) 5	3.6%
	Cys (C) 2	0.8%	Cys (C) 2	1.4%
	Gln (Q) 14	5.3%	Gln (Q) 8	5.8%
	Glu (E) 19	7.2%	Glu (E) 6	4.3%
	Gly (G) 26	9.8%	Gly (G) 16	11.6%
	His (H) 10	3.8%	His (H) 2	1.4%

	Ile (I) 11 4.2% Leu (L) 15 5.7% Lys (K) 14 5.3% Met (M) 8 3.0% Phe (F) 10 3.8% Pro (P) 15 5.7% Ser (S) 29 11.0% Thr (T) 12 4.5% Trp (W) 3 11% Tyr (Y) 7 2.7% Val (V) 17 6.4% Pyl (O) 0 0.0% Sec (U) 0 0.0%	Ile (I) 3 2.2% Leu (L) 7 5.1% Lys (K) 5 3.6% Met (M) 3 2.2% Phe (F) 5 3.6% Pro (P) 9 6.5% Ser (S) 18 13.0% Thr (T) 8 5.8% Trp (W) 3 2.2% Tyr (Y) 5 3.6% Val (V) 12 8.7% Pyl (O) 0 0.0% Sec (U) 0 0.0%
Total number of negatively charged (Asp + Glu)	33	11
Total number of positively charged (Arg + Lys)	30	14
Ext. coefficient at 280 nm (assuming all pairs of Cys residues form cysteines)	27055	24075
Abs 0.1% (=1 g/l) (assuming all pairs of Cys residues form cysteines)	0.930	1.621
Ext. coefficient at 280 nm (assuming all Cys residues are reduced)	26930	23950
Abs 0.1% (=1 g/l) (assuming all Cys residues are reduced)	0.926	1.613

VHH5

Histag-SUMOtag-TEV-VHH5

```
atgggcagcagccatcatcatcatcatcatggcagcgccctggtgccgcgcggcagcgctagcatgtcggactcagaagtc
M G S S H H. H H H H G S G L V P R G S A S M S D S E V
aatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactcacatcaatttaaaggtgttcgatggatcttca
N Q E A K P E V K P E V K P E T H I N L K V S D G S S
gatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaagacag
E G S S D I F F K I K K T T P L R R L M E A F A K R Q
ggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcagacccctgaagatttggacatg
G K E M D S L R F L Y D G I R I Q A D Q T P E D L D M
gaggataacgatattattgaggctcacagagaacagattggtggtgagaacctgtacttccagggcatgaaggtgcaatta
E D N D I I E A H R E Q I G G E N L Y F Q G M K V Q L
gtcgaaagcggcgggggagcgttcaaatcggcggcagcttacgtctctcatgtcgtatctctgaaggttccggcaataaa
V E S G G G S V Q I G G S L R L S C R I S E G S G N K
tacgtgatggcttggttccggcaggctccgggccaagagcgggagttcgtgggcgtgatcagctggagcggcacccgtaca
Y V M A W F R Q A P G Q E R E F V G V I S W S G T R T
cactacgcagattcagttcgtggtcgtttcactatcagtcgcatggataacacagcttacctccaaatggattctctgaca
H Y A D S V R G R F T I S R M D N T A Y L Q M D S L T
ccggatgactcaggggtctactactgtgccgcatcggtttcagttccgccttttctggagttgacagctgcccatttgggg
P D D S G V Y Y C A A S V S V P P F L E L T A A H F G
tcgtggggccaggggacacaagttcgtgtcgggtcatcgccagaaccaaaaaccccaaagcctcagcctgcatcatag
S W G Q G T Q V R V G S S P E P K T P K P Q P A S -
```

	HisTag-SUMOtag-TEV-VHH5		VHH5(mut)	
Number of amino acids	264		138	
Molecular weight	29142.59		14892.70	
Theoretical pI	6.21		8.66	
Amino acid composition	Ala (A) 15	5.7%	Ala (A) 9	6.5%
	Arg (R) 16	6.1%	Arg (R) 9	6.5%
	Asn (N) 6	2.3%	Asn (N) 2	1.4%
	Asp (D) 14	5.3%	Asp (D) 5	3.6%
	Cys (C) 2	0.8%	Cys (C) 2	1.4%
	Gln (Q) 14	5.3%	Gln (Q) 8	5.8%
	Glu (E) 19	7.2%	Glu (E) 6	4.3%
	Gly (G) 26	9.8%	Gly (G) 16	11.6%
	His (H) 10	3.8%	His (H) 2	1.4%
	Ile (I) 12	4.5%	Ile (I) 4	2.9%
	Leu (L) 15	5.7%	Leu (L) 7	5.1%
	Lys (K) 13	4.9%	Lys (K) 4	2.9%
	Met (M) 9	3.4%	Met (M) 4	2.9%

	Phe (F) 10 3.8% Pro (P) 15 5.7% Ser (S) 29 11.0% Thr (T) 12 4.5% Trp (W) 3 11% Tyr (Y) 7 2.7% Val (V) 17 6.4% Pyl (O) 0 0.0% Sec (U) 0 0.0%	Phe (F) 5 3.6% Pro (P) 9 6.5% Ser (S) 18 13.0% Thr (T) 8 5.8% Trp (W) 3 2.2% Tyr (Y) 5 3.6% Val (V) 12 8.7% Pyl (O) 0 0.0% Sec (U) 0 0.0%
Total number of negatively charged (Asp + Glu)	33	11
Total number of positively charged (Arg + Lys)	29	13
Ext. coefficient at 280 nm (assuming all pairs of Cys residues form cysteines)	27055	24075
Abs 0.1% (=1 g/l) (assuming all pairs of Cys residues form cysteines)	0.928	1.617
Ext. coefficient at 280 nm (assuming all Cys residues are reduced)	26930	23950
Abs 0.1% (=1 g/l) (assuming all Cys residues are reduced)	0.924	1.608

PelB-VHH5-Histag

```
atgaaatacttattaccaacggcggcagcaggcttactgctgttagcggcgcaaccagcgatggcgatgaaggtgcaatta
M K Y L L P T A A A G L L L L A A Q P A M A M K V Q L
gtcgaaagcggcgggggagcgttcaaatcggcggcagcttacgtctctcatgtcgtatctctgaaggttccggcaataaa
V E S G G G S V Q I G G S L R L S C R I S E G S G N K
tacgtgatggcttggttccggcaggctccgggccaagagcgggagttcgtgggcgtgatcagctggagcggcacccgtaca
Y V M A W F R Q A P G Q E R E F V G V I S W S G T R T
cactacgcagattcagttcgtggtcgtttcactatcagtcgcatggataacacagcttacctccaaatggattctctgaca
H Y A D S V R G R F T I S R M D N T A Y L Q M D S L T
ccggatgactcaggggtctactactgtgccgcatcggtttcagttccgccttttctggagttgacagctgcccatttgggg
P D D S G V Y Y C A A S V S V P P F L E L T A A H F G
tcgtggggccaggggacacaagttcgtgtcgggtcatcgccagaaccaaaaaccccaaagcctcagcctgcatcagctgca
S W G Q G T Q V R V G S S P E P K T P K P Q P A S A A
caccaccaccatcaccatcattaa
H H H H H H H -
```

	PelB-VHH5-Histag		VHH5(mut)	
Number of amino acids	169		138	
Molecular weight	18205.61		14892.70	
Theoretical pI	9.00		8.66	
Amino acid composition	Ala (A) 18	10.7%	Ala (A) 9	6.5%
	Arg (R) 9	5.3%	Arg (R) 9	6.5%
	Asn (N) 2	1.2%	Asn (N) 2	1.4%
	Asp (D) 5	3.0%	Asp (D) 5	3.6%
	Cys (C) 2	1.2%	Cys (C) 2	1.4%
	Gln (Q) 9	5.3%	Gln (Q) 8	5.8%
	Glu (E) 6	3.6%	Glu (E) 6	4.3%
	Gly (G) 17	10.1%	Gly (G) 16	11.6%
	His (H) 9	5.3%	His (H) 2	1.4%
	Ile (I) 4	2.4%	Ile (I) 4	2.9%
	Leu (L) 13	7.7%	Leu (L) 7	5.1%
	Lys (K) 5	3.0%	Lys (K) 4	2.9%
	Met (M) 6	3.6%	Met (M) 4	2.9%
	Phe (F) 5	3.0%	Phe (F) 5	3.6%
	Pro (P) 11	6.5%	Pro (P) 9	6.5%
	Ser (S) 18	10.7%	Ser (S) 18	13.0%
	Thr (T) 9	5.3%	Thr (T) 8	5.8%
	Trp (W) 3	1.8%	Trp (W) 3	2.2%
	Tyr (Y) 6	3.6%	Tyr (Y) 5	3.6%
	Val (V) 12	7.1%	Val (V) 12	8.7%
	Pyl (O) 0	0.0%	Pyl (O) 0	0.0%
	Sec (U) 0	0.0%	Sec (U) 0	0.0%

Total number of negatively charged (Asp + Glu)	11	11
Total number of positively charged (Arg + Lys)	14	13
Ext. coefficient at 280 nm (assuming all pairs of Cys residues form cysteines)	25565	24075
Abs 0.1% (=1 g/l) (assuming all pairs of Cys residues form cysteines)	1.404	1.617
Ext. coefficient at 280 nm (assuming all Cys residues are reduced)	25440	23950
Abs 0.1% (=1 g/l) (assuming all Cys residues are reduced)	1.397	1.608

PelB-SUMOtag-TEV-VHH5-Histag

```
atgaaatacttattaccaacggcggcagcaggcttactgctgttagcggcgcaaccagcgatggcgatgtcggactcagaa
M K Y L L P T A A A G L L L L A A Q P A M A M S D S E
gtcaatcaagaagctaagccagaggtcaagccagaagtcaagcctgagactcacatcaatttaaaggtgttcgatggatct
V N Q E A K P E V K P E V K P E T H I N L K V S D G S
tcagatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcgctaaaaga
S E G S S D I F F K I K K T T P L R R L M E A F A K R
cagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcagacccctgaagatttggac
Q G K E M D S L R F L Y D G I R I Q A D Q T P E D L D
atgaacctgtacttccagggcatgaaggtgcaattagtcgaaagcggcgggggggagcgtttcaaatcggcggcagcttacg
M N L Y F Q G M K V Q L V E S G G G S V Q I G G S L R
tctctcatgtcgtatctctgaaggttccggcaataaatacgtgatggcttggttccggcaggctccgggccaagagcgggag
L S C R I S E G S G N K Y V M A W F R Q A P G Q E R E
ttcgtgggcgtgatcagctggagcggcacccgtacacactacgcagattcagttcgtggtcgtttcactatcagtcgcatg
F V G V I S W S G T R T H Y A D S V R G R F T I S R M
gataacacagcttacctccaaatggattctctgacaccggatgactcaggggtctactactgtgccgcatcggtttcagtt
D N T A Y L Q M D S L T P D D S G V Y Y C A A S V S V
ccgccttttctggagttgacagctgcccattttgggtcgtggggccaggggacacaagttcgtgtcgggtcatcgccagaa
P P F L E L T A A H F G S W G Q G T Q V R V G S S P E
ccaaaaaccccaaagcctcagcctgcatcagctgcacaccaccaccatcaccatcattaa
P K T P K P Q P A S A A H H H H H H H -
```

	PelB-SUMOtag-TEV-VHH5-Histag		VHH5(mut)	
Number of amino acids	271		138	
Molecular weight	30301.19		14892.70	
Theoretical pI	6.26		8.66	
Amino acid composition	Ala (A) 23	8.4%	Ala (A) 9	6.5%
	Arg (R) 15	5.5%	Arg (R) 9	6.5%
	Asn (N) 6	2.2%	Asn (N) 2	1.4%
	Asp (D) 14	5.1%	Asp (D) 5	3.6%
	Cys (C) 2	0.7%	Cys (C) 2	1.4%
	Gln (Q) 15	5.5%	Gln (Q) 8	5.8%
	Glu (E) 19	6.9%	Glu (E) 6	4.3%
	Gly (G) 23	8.4%	Gly (G) 16	11.6%
	His (H) 11	4.0%	His (H) 2	1.4%
	Ile (I) 12	4.4%	Ile (I) 4	2.9%
	Leu (L) 20	7.3%	Leu (L) 7	5.1%
	Lys (K) 14	5.1%	Lys (K) 4	2.9%
	Met (M 10	3.6%	Met (M) 4	2.9%
	Phe (F) 10	3.6%	Phe (F) 5	3.6%
	Pro (P) 16	5.8%	Pro (P) 9	6.5%
	Ser (S) 24	8.8%	Ser (S) 18	13.0%

	Thr (T) 13 4.7% Trp (W) 3 1.1% Tyr (Y) 8 2.9% Val (V) 16 5.8% Pyl (O) 0 0.0% Sec (U) 0 0.0%	Thr (T) 8 5.8% Trp (W) 3 2.2% Tyr (Y) 5 3.6% Val (V) 12 8.7% Pyl (O) 0 0.0% Sec (U) 0 0.0%
Total number of negatively charged (Asp + Glu)	33	11
Total number of positively charged (Arg + Lys)	29	13
Ext. coefficient at 280 nm (assuming all pairs of Cys residues form cysteines)	28545	24075
Abs 0.1% (=1 g/l) (assuming all pairs of Cys residues form cysteines)	0.942	1.617
Ext. coefficient at 280 nm (assuming all Cys residues are reduced)	28420	23950
Abs 0.1% (=1 g/l) (assuming all Cys residues are reduced)	0.938	1.608

Scientific production arisen from this thesis

Publication

Gilodi M, Lis S, F. Dudás E, Fantini M, Puglisi R, Louka A, Marcatili P, Cattaneo A, Pastore A. *Selection and modelling of a new single-domain intrabody against TDP-43.* Front. Mol. Biosc. 2022.

Acknowledgements

I would like to thank all those who contributed to the realisation of this work.

First of all, I would like to express my gratitude to Prof. Annalisa Pastore who welcome me in her research group. Thank you for the unique opportunity, for your support and your guidance on these years.

Thanks also to Dr.Hugo De Jonge for his encouragement and his help even from a distance.

Then I would like to thank Prof. Antonino Cattaneo and all his research group in particular Dr. Simonetta Lisi who help me with the nanobody selection allowing the start of this project.

Thanks go to Dr. Paolo Marcatili for his help and for sharing his knowledge about modelling.

This work would not have been possible without the members of Pastore's group, both past and present. Their advice, insights and support were a big part of this journey. They have been not only colleagues but above all invaluable friends. In particular, I'm grateful to Elsa Zacco, Rita Puglisi, Giulia Milordini, Chiara Beghè, Anna Fricano e Tommaso Vannocci.
I owe a debt of gratitude to Dr. Rita Puglisi for her support and help given to me over these years. She has been patient and graciously understanding and embracing my days of frustration.

Thank you to my dear friend Francesca. Despite the distance and the time zone she has always been by my side encouraging me and stimulating me to develop my self-belief.
To all my other friends, thank you for being a part of this.

I'm extremely grateful to my family for their unconditional support. My parents are a constant source of inspiration. They taught me the determination and the dedication. Without their love this would not have been possible.

Finally, thanks to Carlo Alberto who has supported me in all my highs and lows. Thank you for always believing in me and helping me to pursuit my own path. Thank you for sharing your life with me.