



# **STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF THE HUMAN ROCO PROTEIN MFHAS1**

Thesis submitted for the degree of Doctor of Philosophy

School of Pharmacy

Eloïse Alice Marie Tribollet

September 2023

## **Declaration**

---

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Eloïse A. M. Tribollet

## Acknowledgements

---

The completion of this PhD would not have been possible without the support and help of many people.

To begin, I would like to express my deepest gratitude to all my supervisors, Professor Patrick Lewis, Doctor Eva Kevei, Doctor Maud Dumoux and Doctor Martin Walsh for the opportunity of this PhD and their support, dedication, and expertise.

I would like to thank Professor Patrick Lewis for the interesting discussions and his continuous support and guidance, especially during unforeseen research delays.

I would like to thank Doctor Maud Dumoux for her supervision at DLS and constant support during those not so easy years.

This project would not have been possible without funding from the EPSRC, the University of Reading and Diamond Light Source.

I would like to thank the team members of MPL, Diamond at the RCaH and PPUK, for their help with protein production and support for the first few years of my PhD.

I would like to thank Ben O'Callaghan from UCL for providing the cells lysate of Knock-Down MFHAS1. I would like to thank Nikoleta Vavouraki for her help with the bioinformatic and mental support throughout this journey.

Thanks, should also go to the Kevei lab members at Reading University for their support (and many cakes) during the final year. Special thanks to Susy Cogo for her support and help.

Words cannot express my gratitude to Kunal Rastogi for his unconditional support, encouragement, and patience, especially during the writing part of this thesis.

Finally, I would like to thank my friends and family. En particulier j'aimerais remercier mes parents pour leur soutien et encouragement malgré la distance.

*« Ce n'est pas que les choses sont difficiles que nous n'osons pas, c'est parce que nous n'osons pas qu'elles sont difficiles » Sénèque*

## Abstract

---

The human ROCO protein family consists of four proteins, including Leucine-Rich Repeat Kinase 2 (LRRK2), a widely studied protein for its role in familial Parkinson's disease; Leucine-Rich Repeat Kinase 1 (LRRK1), which is involved in immunity and bone resorption; Death Associated Protein Kinase 1 (DAPK1), which is linked to cancer; and Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1), a protein of unknown function implicated in cancer and immunity. The presence of a ROC-COR supradomain distinguishes this family of proteins. The Ras Of Complex protein (ROC) domain has been described as a small GTPase and is universally followed by a 300-400 amino-acid C-terminal domain of unknown function, termed C terminal Of ROC (COR). Focusing on MFHAS1, the smallest of the ROCO proteins, this project aims to develop a structural and functional understanding of the ROCO proteins.

In this thesis, MFHAS1 was studied with three approaches: structural analysis of MFHAS1 at the Research Complex at Harwell and Diamond Light Source, use of bioinformatic tools, and investigation of the cell biology and pathways surrounding MFHAS1. For the former, MFHAS1 was over-expressed and purified attempted for structural and biophysical characterisation. Bioinformatic analysis was utilised to define a modelled structure of MFHAS1, analyse the post-translational modification presents in MFHAS1 and build a protein-protein interaction network. Finally, cellular studies were used to understand the physiological role of MFHAS1, with the steady state of MFHAS1 and the expression level of surrounding proteins were analysed following artificial mutation of residues subject to putative post translational modification. Immunoprecipitation was employed to determine proteins interacting with MFHAS1, providing evidence that it interacts directly with MAP3K4 in a cellular context. Endogenous MAP3K4 is found to interact with overexpressed wild-type and mutants MFHAS1.

Together, these analyses and experiments give an insight into MFHAS1 protein, illuminating the function of this least studied of the human ROCO proteins.



# Table of Contents

---

<b>CHAPTER 1: INTRODUCTION</b>	<b>1</b>
<b>1.1 Aims</b>	<b>1</b>
<b>1.2 General presentation of ROCO proteins</b>	<b>2</b>
<b>1.3 Phylogenetic origins of the ROCO proteins</b>	<b>3</b>
<b>1.4 Ras proteins</b>	<b>4</b>
<b>1.5 Model organism ROCO</b>	<b>6</b>
1.5.1 <i>Chlorobium tepidum</i>	6
1.5.2 <i>Dictyostelium discoideum</i>	8
1.5.3 <i>Caenorhabditis elegans</i>	8
1.5.4 <i>Drosophila melanogaster</i>	9
1.5.5 <i>Mus musculus</i>	9
<b>1.6 Human ROCO proteins</b>	<b>11</b>
1.6.1 LRRK2	11
1.6.1.1 Pathology	12
1.6.1.2 Cellular Biology	12
1.6.1.3 Expression and localisation in the cells and the body	13
1.6.2 LRRK1	14
1.6.2.1 Pathology	15
1.6.2.2 Cellular Biology	15
1.6.2.3 Expression and localisation in the cells and the body	16
1.6.3 DAPK1	17
1.6.3.1 Pathology	18
1.6.3.2 Cellular Biology	18
1.6.3.3 Expression and localisation in the cells and the body	20
1.6.4 MFHAS1	21
1.6.4.1 Pathologies	21
1.6.4.1.1 Cancer	21
1.6.4.1.2 Sepsis	22
1.6.4.1.3 Diabetes mellitus	23
1.6.4.1.4 Others	23
1.6.4.2 Cellular function	23

1.6.4.3	Expression and localisation in the cells and the body -----	25
<b>1.7</b>	<b>Thesis structure, aims and objectives -----</b>	<b>26</b>
<b>CHAPTER 2: EXPRESSION AND PURIFICATION OF MFHAS1 -----</b>		<b>27</b>
<b>2.1</b>	<b>Introduction -----</b>	<b>27</b>
<b>2.2</b>	<b>Material and methods -----</b>	<b>28</b>
2.2.1	Domain delineation -----	28
2.2.2	Cloning -----	28
2.2.3	Cell culture and protein expression -----	31
2.2.4	Protein extraction -----	32
2.2.5	Protein solubilisation and detergents screen -----	33
2.2.6	Protein purification -----	37
2.2.6.1	Purification using a GFP MFHAS1 construct -----	37
2.2.6.2	Purification using nickel affinity chromatography -----	38
2.2.6.3	Purification using Strep-tag II MFHAS1 construct -----	39
<b>2.3</b>	<b>Results -----</b>	<b>40</b>
2.3.1	Boundary delineation -----	40
2.3.2	Protein expression using vectors containing His and GFP tags -----	42
2.3.3	Protein expression using vectors containing GFP and His tags. -----	46
2.3.4	Protein expression using vectors containing GFP, 2Strep and His tags. -----	54
<b>2.4</b>	<b>Discussion -----</b>	<b>59</b>
<b>CHAPTER 3: BIOINFORMATIC ANALYSIS OF MFHAS1 -----</b>		<b>63</b>
<b>3.1</b>	<b>Introduction -----</b>	<b>63</b>
<b>3.2</b>	<b>Materials and methods -----</b>	<b>66</b>
3.2.1	Protein modelling structure -----	66
3.2.2	Post translational modification. -----	67
3.2.3	Protein network analysis -----	68
<b>3.3</b>	<b>Results -----</b>	<b>69</b>
3.3.1	Protein modelling structure -----	69

3.3.1.1	Assessing the quality of the protein model -----	70
3.3.1.1.1	Comparison of LRRK2 crystal structure with predicted models -----	70
3.3.1.1.2	MFHAS1 predicted structure from Alphafold2 -----	72
3.3.1.1.3	MFHAS1 predicted structure from IntFOLD -----	73
3.3.1.1.4	Comparison predicted model of MFHAS1 from AF2 and IntFOLD -----	74
3.3.1.1.5	Comparison of MFHAS1 predicted structures with CtROCO protein -----	75
3.3.2	Post-translational modification -----	76
3.3.2.1	Phosphorylation site -----	76
3.3.2.1.1	Kinase prediction-----	77
3.3.2.1.2	Sequence alignment -----	79
3.3.2.1.3	Phosphorylation sites in the context of MFHAS1 structure -----	81
3.3.2.2	Ubiquitination site-----	82
3.3.2.2.1	Sequence alignment -----	83
3.3.2.2.2	Ubiquitination sites in in the context of MFHAS1 structure-----	85
3.3.3	Protein-protein interaction network analysis-----	86
3.3.3.1	MFHAS1 protein potential interactors-----	86
3.3.3.1.1	MFHAS1 interaction with the phosphorylation machinery -----	89
3.3.3.1.2	MFHAS1 interaction with the ubiquitination machinery -----	93
3.3.3.1.3	Clusters analysis-----	97
3.3.3.2	Functional categorisation -----	101
3.3.3.3	Gene expression-----	103
<b>3.4</b>	<b>Discussion -----</b>	<b>107</b>
<b>CHAPTER 4:</b>	<b>MFHAS1 SIGNALLING PATHWAYS. -----</b>	<b>112</b>
<b>4.1</b>	<b>Introduction -----</b>	<b>112</b>
<b>4.2</b>	<b>Material and method -----</b>	<b>116</b>
4.2.1	DNA preparation -----	116
4.2.2	Mutagenesis using Agilent kit -----	116
4.2.3	Cell culture -----	118
4.2.4	Protein extract sample preparation -----	119
4.2.5	Determining the protein expression level by immunoblot-----	120
4.2.6	Investigating protein-protein interactions using immunoprecipitation.-----	122
4.2.6.1	Buffer and beads preparation: -----	122
4.2.6.2	Immunoprecipitation -----	122
4.2.7	Statistical analysis -----	123

<b>4.3 Results</b> -----	<b>124</b>
4.3.1 Transfection optimisation-----	124
4.3.2 Effect of ubiquitination site mutation K387A-----	126
4.3.2.1 Protein expression level-----	126
4.3.2.2 Immunoprecipitation-----	128
4.3.3 Effect of phosphorylation site mutation S450A-----	130
4.3.3.1 Protein expression level-----	130
4.3.3.2 Immunoprecipitation-----	132
4.3.4 Understanding the effects of PD-like mutations in MFHAS1-----	133
4.3.4.1 Protein expression level-----	133
4.3.4.2 Immunoprecipitation-----	136
<b>4.4 Discussion</b> -----	<b>138</b>
 <b>CHAPTER 5: CONCLUSIONS AND DISCUSSION</b> -----	 <b>143</b>

## Abbreviation

---

Å: Angstrom

AA: Amino Acid

AD: Alzheimer's Disease

AF2: AlphaFold2

AI: Artificial Intelligence

ANK: Ankyrin Repeat

ARM: Armadillo

ATP: Adenosine triphosphate

AUC: Analytical ultracentrifugation

AURKB: Aurora kinase B

BAC: Bacterial Artificial Chromosome

BMD: bone mineral density

Bp: base pair

*C. elegans*: *Caenorhabditis elegans*

*C. tepidum*: *Chlorobium tepidum*

Ca<sup>2+</sup>: Calcium ion

CaM: calmodulin

CASP: Critical Assessment of techniques for protein Structure Prediction

CB: carbenicillin

CBLB: E3 ubiquitin-protein ligase Casitas B-lineage lymphoma

CD: Circular Dichroism

Cdc42: Cell division control protein 42

CDK: Cyclin-dependent kinase

cGMP: cyclic Guanosine Monophosphate

CHIP: E3 ubiquitin-protein ligase Carboxyl terminal of Hsp70-Interacting Protein

CK1α: Casein Kinase1 α

CLIP 170: Cytoplasmic linker protein 170

CLK1: Dual specificity protein kinase CLK1

Cm: chloramphenicol

cNB: cyclic Nucleotide-Binding

COR: C terminal Of Roc

CRC: Colorectal Cancer

Cryo-EM: cryo-Electron Microscopy

CSK: C- Src Kinase

c-SRC: cellular Rous Sarcoma Oncogene

CtROCO: *C. tepidum* ROCO protein

CV: column bed volume

*D. discoideum*: *Dictyostelium discoideum*

*D. melanogaster*: *Drosophila melanogaster*

DAPK1: Death Associated Protein Kinase 1

DD: Death Domain  
ddH<sub>2</sub>O: double distilled water  
DLBCL: Diffuse Large B-Cell Lymphoma  
DM: Diabetes Mellitus  
DMEM: Dulbecco's Modified Eagle Medium  
DN: Diabetic Nephropathy  
DNA: Deoxyribonucleic acid  
DTT: Dithiothreitol  
ECL: enhanced chemiluminescence  
EGF: Epidermal Growth Factor  
EGFR: Epidermal Growth Factor Receptor  
EPO: Erythropoietin  
ER: Endoplasmic Reticulum  
ERK: Extracellular signal-Regulated Kinase  
F/T: Flow-through  
FBS: foetal bovine serum  
FGFR1: Fibroblast growth factor receptor 1  
FL: Full length  
F-SEC: Fluorescent-Size exclusion chromatography  
GAK: Cyclin-G-associated kinase  
GAP: GTPase activating protein  
GDP: Guanosine Diphosphate  
GEF: guanine nucleotide exchange factor  
GFP: Green fluorescent protein  
GO: Gene Ontology  
GRAM: Glucosyltransferases, Rab-like GTPase Activators and Myotubularins  
GSK3: Glycogen synthase kinase-3 beta  
GTEx: Genotype-Tissue Expression project  
GTP: Guanosine Triphosphate  
GWAS: Genome Wide Association Study  
HA: Hemagglutinin  
HCl: Hydrogen chloride  
HCS: Highly Conserved Sequence  
HIPPIE: Human Integrated Protein-Protein Interaction rEference  
His: Histidine  
HNSCC: Head and Neck Squamous Cell Carcinoma  
HRP: horseradish peroxidase  
HSP60: Heat Shock Protein 60  
HSP70: Heat Shock Protein 70  
HTP: High Throughput  
HUVEC: Human Umbilical Vein Endothelial Cell  
IFN- $\gamma$ : Interferon- $\gamma$

IKKs: I $\kappa$ B kinases  
IL-6: Interleukin-6  
IP: Immunoprecipitation  
IPTG: Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
JNK: C-Jun NH<sub>2</sub>-Terminal Kinase  
KD: Knock-Down  
kDa: KiloDalton  
LB: Lysogeny Broth  
LBA: Lysogeny Broth Agar  
LIMK1: LIM domain kinase 1  
LPS: Lipopolysaccharide  
LRK-1: Leucine-rich repeat serine/threonine-protein kinase 1  
LRR: Leucine-Rich Repeat  
LRRK1: Leucine-Rich Repeat Kinase 1  
LRRK2: Leucine-Rich Repeat Kinase 2  
LTP: Low Throughput  
M1: Macrophages 1  
M2: Macrophages 2  
MAP3K4: Mitogen-activated protein kinase kinase kinase 4  
MAPKKK: Mitogen-Activated Protein Kinase Kinase Kinases  
MASL1: MFH-Amplified Sequences with Leucine-rich tandem repeats 1  
MEK: Mitogen-activated protein kinase kinase  
MFH: Malignant Fibrous Histiocytomas  
MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1  
mL: milli Liter  
mM: milli molar  
MOLT-4: Human acute lymphoblastic leukaemia cell line  
mRNA: messenger RNA  
MS: Mass spectrometry  
MSA: Multiple Sequence Alignments  
NaCl: Sodium chloride  
NaOH: Sodium hydroxide  
NEK11: Serine/threonine-protein kinase Nek11  
NF- $\kappa$  B: Nuclear factor kappa-light-chain-enhancer of activated B cells  
nGEF: neuronal guanine nucleotide exchange factor  
Ni-NTA: Nickel<sup>2+</sup> Nitrilotriacetic Acid  
NMDA: N-Methyl-D-Aspartate  
O/N: Overnight  
OD: optical density  
ORF: open reading frame  
OSCC: Oral Squamous Cell Carcinoma  
OSMD: Osteosclerotic metaphyseal dysplasia

p38d: Mitogen-activated protein kinase 13  
PBS: phosphate-buffered saline  
PBST: phosphate-buffered saline with Tween 20  
PCR: Polymerase Chain Reaction  
PD: Parkinson's Disease  
PDB: Protein Data Bank  
PEI: Polyethylenimine  
PH: Pleckstrin Homology  
PI3K: phosphatidylinositol-3 Kinase  
PI3P: phosphatidylinositol-3 phosphate  
PINOT: Protein Interaction Network Online Tool  
PJA2: E3 ubiquitin-protein ligase Praja-2  
PKA: Protein Kinase A  
PKD: Protein Kinase D  
PP2a: Protein phosphatase 2  
PPI: Protein-protein interaction  
PPUK: Protein Production UK  
PTM: Post translational modification  
PTP: Protein-Tyrosine Phosphatase  
RAC: Ras-related C3 botulinum toxin substrate  
RAF: proto-oncogene serine/threonine-protein kinase  
Ras GEF: Ras Guanosine Exchange Factor  
RGS: Regulator of G protein Signalling  
RHO: Ras homologous  
RIP: Receptor Interacting Protein  
RMS: Root Mean Square  
RNA: Ribonucleic acid  
ROC: Ras Of Complex protein  
ROR1: Inactive tyrosine-protein kinase transmembrane receptor ROR1  
RSK: Ribosomal S6 kinase  
RT: room temperature  
SCC: Squamous Cell Carcinoma  
SDS-PAGE: Sodium dodecyl-sulfate polyacrylamide gel electrophoresis  
SEC: size exclusion chromatography  
SEC-MALLS: size-exclusion chromatography with multi-angle laser light scattering  
SF9: Spodoptera Frugiperda  
siRNA: small interfering RNA  
SN: supernatant  
SNP: Single-Nucleotide Polymorphism  
STK25: Serine/threonine-protein kinase 25  
Strep: Streptavidin  
SV: synaptic vesicles



TBS: Tris-Buffered Saline

TIAM1: T-cell lymphoma invasion and metastasis inducing factor 1

TKL: Tyrosine Kinase Like

TLR: Toll like receptor

TPM: Transcripts Per Kilobase Million

TRIM22: E3 ubiquitin-protein ligase Tripartite Motif 22

UPS: ubiquitin proteasome system

V: Volts

Vps34: vacuolar protein sorting 34

Wnt: Wingless/INT

WT: Wild Type

# Chapter 1: Introduction

## 1.1 Aims

This thesis focuses on understanding the structure and function of the human ROCO protein, Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1).

MFHAS1 is the smallest protein of the ROCO family which is constituted of four proteins: Leucine-Rich Repeat Kinase 2 (LRRK2), Leucine-Rich Repeat Kinase 1 (LRRK1), Death Associated Protein Kinase 1 (DAPK1) and MFHAS1. All ROCO proteins have been involved with human pathology. They also all have a similar core composed of two domains: Ras Of Complex (ROC) and C terminal Of Roc (COR).

MFHAS1 is the least studied out of the four ROCO proteins (Dihanich, 2012). The research carried out to date shows that MFHAS1 gene amplification, overexpression or deletion is associated with cancers (Sakabe *et al.*, 1999; Chen *et al.*, 2016). Its Roc domain has GTPase activity, and the protein is most likely located in the cytoplasm (Dihanich *et al.*, 2014). No structure has yet been elucidated.

This study aims to understand further what mechanisms or activities are involved around MFHAS1. To analyse the structure of MFHAS1 protein, expression, and purification of MFHAS1 protein was completed at Diamond Light Source. Then bioinformatics was used to examine the post-translational modification of MFHAS1, potential interaction with other proteins, and predicted structure. Finally, the signalling pathways around MFHAS1 were investigated at the University of Reading.

The goal of this research is a better understanding of the role of the MFHAS1 protein and, by extension, the ROCO proteins.

## 1.2 General presentation of ROCO proteins

The ROCO proteins are a family of proteins found in most living organisms. They were first described in the amoeba *Dictyostelium discoideum* (*D. discoideum*) and since then have been identified across eukaryotic and prokaryotic organisms (Bosgraaf and Van Haastert, 2003). The ROCO proteins derive their name from the characteristic ROC-COR supradomain present in all members of the family (Bosgraaf and Van Haastert, 2003). The Ras Of Complex protein (ROC) domain was initially described based upon close homology to the small GTPase Ras family. In contrast, the C terminal Of Roc (COR) domain is unique to the ROCO proteins and has an unknown function. The human ROCO protein family consists of four members; Leucine Rich Repeat Kinase 2 (LRRK2); Leucine Rich Repeat Kinase 1 (LRRK1); Death Associated Protein Kinase 1 (DAPK1); and Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1), also called MFH-Amplified Sequences with Leucine-rich tandem repeats 1 (MASL1).

The human ROCO proteins are of interest for several reasons. First, with the notable exception of MFHAS1, they possess a kinase and GTPase domains, making them unique within the human proteome. Secondly, all four proteins are implicated in human disease. The most widely known is LRRK2 for its association with familial Parkinson's Disease (PD) (Funayama *et al.*, 2002; Sosero and Gan-Or, 2023). MFHAS1 and DAPK1 are linked with cancer (Sakabe *et al.*, 1999; Calmon *et al.*, 2007), while LRRK1 is involved in osteopetrosis (Xing *et al.*, 2013). Understanding those proteins' origin and evolutionary pathway might explain the activity difference between the ROCO proteins.

### 1.3 Phylogenetic origins of the ROCO proteins

ROCO proteins are found across plants, prokaryotes, metazoa and mammals. In an attempt to classify ROCO proteins across different species, two main approaches have been developed. Bosgraaf and Van Haastert ranked the ROCO proteins in three groups according to their domain architecture (Bosgraaf and Van Haastert, 2003). In contrast, the second approach compared the gene sequence of the ROCO protein and classified them by phylogenetic analyses in seven branches: the eleven genes from *Dictyostelium*, LRRK, MFHAS1, MFHAS1-like, Plant, DAPK1 and Prokaryotic ROCOs (Marin, van Egmond and van Haastert, 2008).

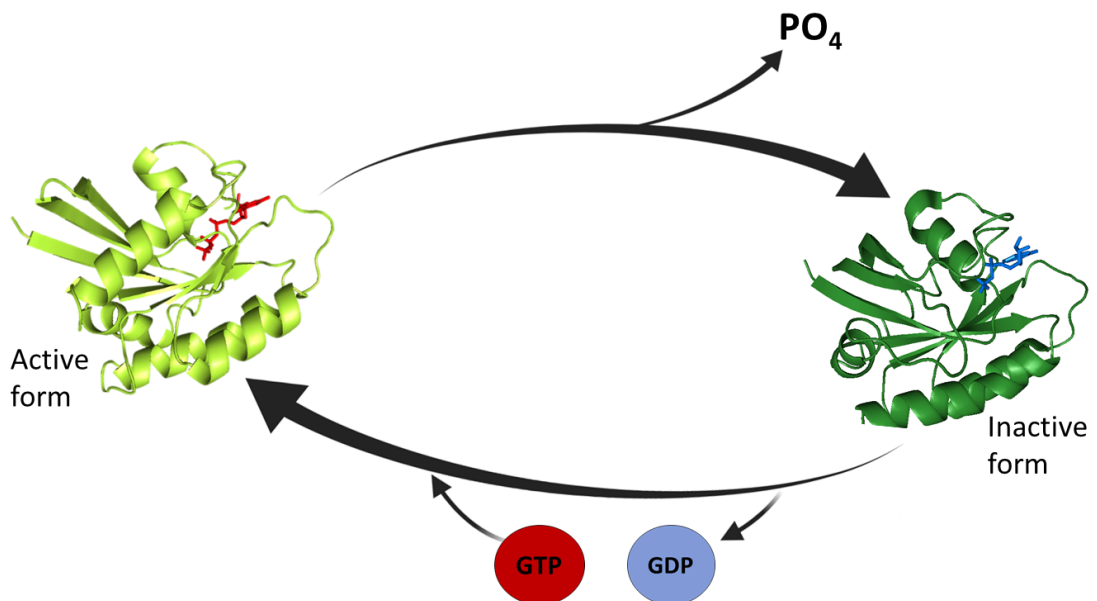
In the 2006 analysis, Marin categorised ROCO proteins' ROC and the COR domain sequence. Analysis of the ROC domains highlight the evolutionary distance between MFHAS1 and LRRK1/2, revealing a closer relationship between MFHAS1 and DAPK1, plants and prokaryotic ROCO genes. Analysis of the COR domains shows MFHAS1 to be closer to DAPK1 and LRRK1/2 than prokaryotic ROCO genes (Marín, 2006). Marin proposed two hypotheses that could explain why the ROCO genes are present in prokaryote and eukaryote organisms. The first one is that ROCO proteins developed prior to the establishment of the eukaryotes and would therefore have evolved with new organisms. The second hypothesis is that the ROCO protein emerged in an early eukaryote and would have been horizontally transmitted to prokaryotic organisms.

## 1.4 Ras proteins

By definition, the human ROCO proteins contain a ROC domain, described as a Ras of complex proteins.

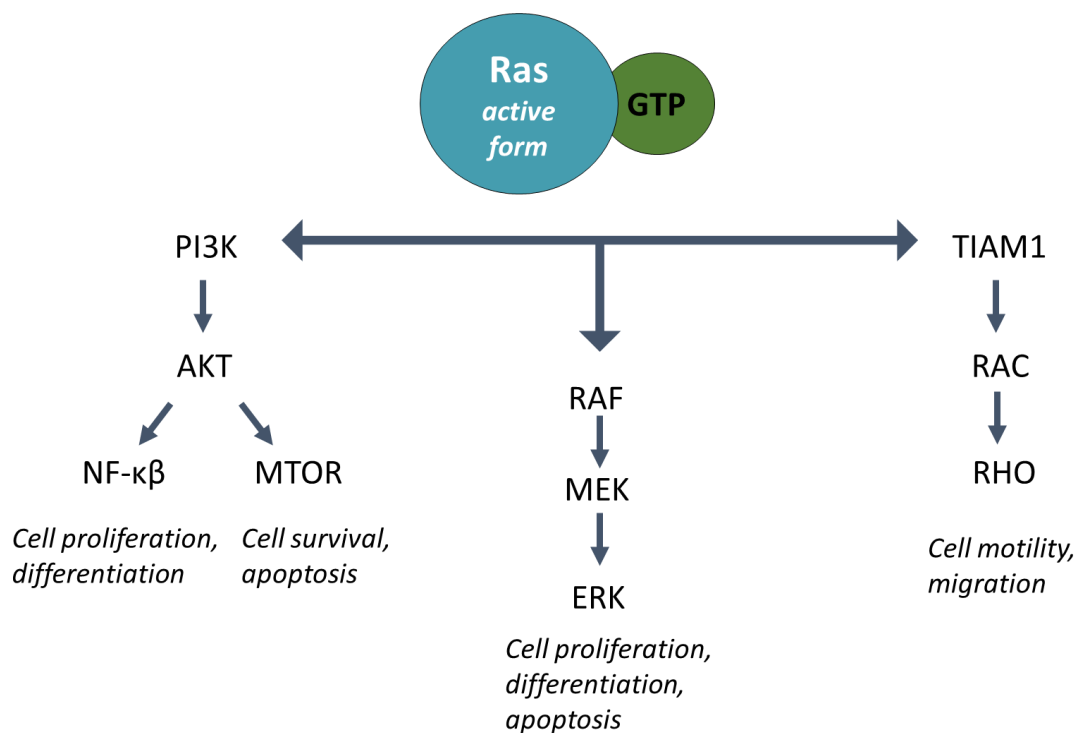
The Ras proteins are prototypic small GTPases, and have been studied for over four decades (Fernández-Medarde and Santos, 2011). They are part of the superfamily of small GTP-binding proteins (GTPases) (Repasky *et al.*, 2009). The three human *RAS* genes encode for four RAS proteins; H-RAS, N-RAS, K-RAS4A and K-RAS4B. MFHAS1 ROC domain shares 19.6%, 17.4% and 21.2% identity with K-RAS, N-RAS and H-RAS, respectively.

RAS proteins act as intracellular molecular switches. They can be in an active or inactive state when bound to Guanosine Triphosphate (GTP) or Guanosine Diphosphate (GDP), respectively (Xiang *et al.*, 2017) (**Figure 1.1**). The guanine nucleotide exchange factor (GEF) activates RAS by releasing GDP and promoting GTP binding. In contrast, GTPase-activating proteins (GAP) catalyse the hydrolysis of GTP into GDP, inactivating the protein.



**Figure 1.1:** Schematic representation of the GTP/GDP switch of the RAS protein. RAS under the inactive form (dark green) is bound to guanosine diphosphate (blue) PDB:4q21. RAS under its active form (light green) is bound to guanosine triphosphate (red) PDB: 5p21.

Under normal conditions, the Ras proteins are responsible for regulating many signalling pathways, including proto-oncogene serine/threonine-protein kinase (RAF)/ Mitogen-activated protein kinase kinase (MEK)/ Extracellular signal-regulated kinase (ERK), phosphatidylinositol-3 Kinase (PI3K)/AKT; and T-cell lymphoma invasion and metastasis inducing factor 1 (TIAM1)/RAC/ Ras homologous (RHO). Those pathways are involved in cell proliferation, differentiation, cell survival, apoptosis, cell mobility and migration (**Figure 1.2**) (Gurung and Bhattacharjee, 2015; Gimple and Wang, 2019; Dillon *et al.*, 2021; Soriano *et al.*, 2021).



**Figure 1.2:** Selection of three downstream pathway under active RAS protein binding to guanosine triphosphate. Pathways includes phosphatidylinositol-3 kinase (PI3K)/AKT; proto-oncogene serine/threonine-protein kinase (RAF)/ Mitogen-activated protein kinase kinase (MEK)/ Extracellular signal-regulated kinase (ERK) and T-cell lymphoma invasion and metastasis inducing factor 1 (TIAM1)/RAC/ Ras homologous (RHO).

Ras mutations are responsible for 30% of cancers, with K-RAS being the most frequent protein involved (Karnoub and Weinberg, 2008; Prior, Lewis and Mattos, 2012). Ras mutations can be somatic or de novo (Dunnett-Kane *et al.*, 2020).

## 1.5 Model organism ROCO

As noted above, ROCO proteins are found across all domains of life, including prokaryotes and eukaryotes (Bosgraaf and Van Haastert, 2003). The association of several of the human ROCO proteins with disease (see below, section 1.6) has generated interest in understanding the cellular mechanisms and the structure of ROCO proteins. However, the large size and toxicity of these proteins when overexpressed make their study challenging. Researchers focused their work on other ROCO proteins from model organisms and single-cell organisms.

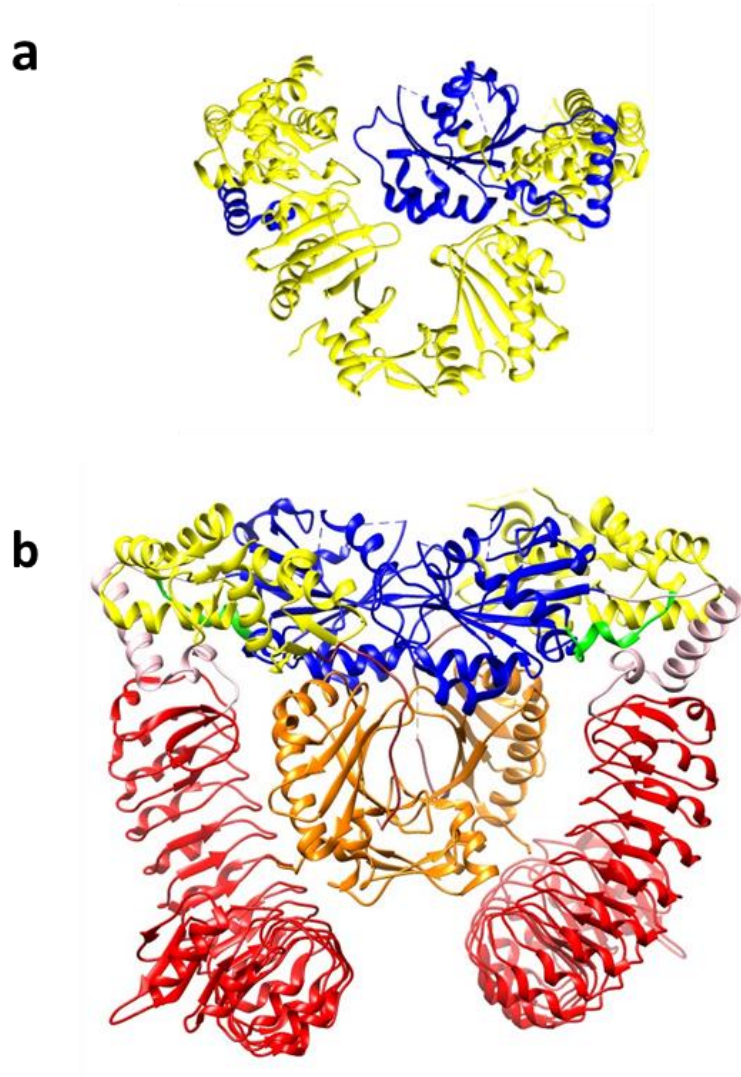
### 1.5.1 *Chlorobium tepidum*

The bacteria *C. tepidum* expresses one ROCO protein, CtROCO. The structure of the ROC-COR supradomain of CtROCO was analysed by truncating the C terminal of the COR domain (Gotthardt *et al.*, 2008). They recorded no GTPase activity and no dimerisation of the protein when truncated. Therefore, they concluded that the COR domain would induce dimerisation. The Roc-COR dimer structure solved by Gotthardt does not display ROC dimerisation (**Figure 1.3**).

Deyaert's team published a crystal structure of LRR-ROC-COR at 3.29 Å resolution in a nucleotide-free state, forming a dimer (**Figure 1.3**). They found that the main interactions are not between the two COR domains but induced by the ROC domain instead (Deyaert *et al.*, 2019). Deyaert's earlier research shows that the structure changed according to the presence or lack of nucleotides to bind the GTPase. Their results show that the CtROCO protein would be in the form of a dimer when no GTP or GDP are bound to it but change its conformation from a dimer to a monomer when binding to a GTP, and GDP release or presence would change the structure back to a dimer (Deyaert *et al.*, 2017).

CtROCO has the same domain architecture as the human ROCO protein MFHAS1 (**Figure 1.4**). They share 34.1% sequence similarity and 23.6% identity.

These data also revealed the importance of the R543 residue described as an arginine finger. R543 is situated in the ROC domain and has been described as essential for stabilising RAS proteins. Substitution of R543 leads to an inhibition of the GTPase activity (Gotthardt et al. 2008). LRRK2 does not contain an arginine finger close to its GTPase activity site.



**Figure 1.3:** Available PDB structures of ROCO protein from *C. tepidum*. **(a)** PDB 3dpu, blue Ras of Complex protein (ROC), yellow: C-terminal of ROC (COR) (Gotthardt et al. 2008). **(b)** PDB 6hlu; red: Leucin rich repeat (LRR), pink: linker, blue: ROC, green: linker, yellow: N-COR, orange: C-COR (Deyaert et al. 2019). (Structure assembly using Chimera).



### 1.5.2 *Dictyostelium discoideum*

The slime mould amoeba, *D. discoideum*, is the organism in which the ROCO proteins were first identified (Bosgraaf and Van Haastert, 2003). The *D. discoideum* genome encodes 11 ROCO proteins: GbpC, Pats1, Roco4, Roco7, Roco6, Roco5, Roco9, Roco10, Roco8, Roco2/QkgA, and Roco11 (**Figure 1.4**). All contain one LRR, the ROC-COR and Mitogen-Activated Protein Kinase Kinase Kinases (MAPKKK) domains but differ at their N or C terminal. GbpC comprises the Ras Guanosine Exchange Factor (GEF) domain and two cyclic Guanosine Monophosphate (cGMP) binding domains in addition to the LRR, ROC-COR and MAPKKK. It has been implicated in chemotaxis by regulating myosin II (Bosgraaf *et al.*, 2002). Another study shows that all cyclic nucleotides interact with the cyclic nucleotide-binding domain, which activates the GEF domain. GEF activation enhances GDP/GTP exchange in the ROC domain and activates the kinase domain (Van Egmond *et al.*, 2008). The ROCO proteins show importance during cell division and late differentiation (van Egmond and van Haastert, 2010).

### 1.5.3 *Caenorhabditis elegans*

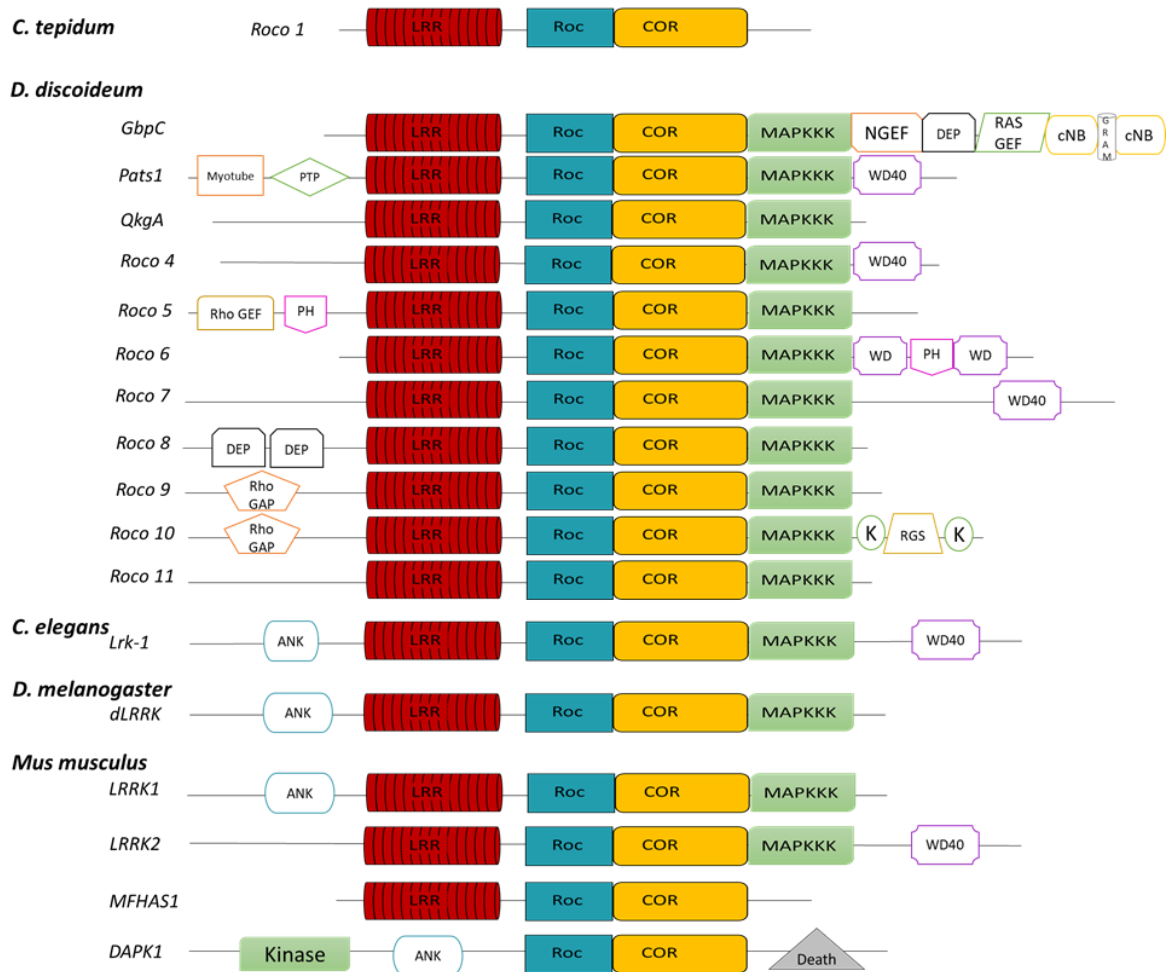
The nematode worm *C. elegans* has two ROCO proteins; Leucine-rich repeat serine/threonine-protein kinase 1 (LRK-1) and Death-associated protein kinase (DAPK-1) (**Figure 1.4**). *lrk-1* has been first described as a homologous gene of LRRK2 (Sakaguchi-Nakashima *et al.*, 2007). The transparency of *C. elegans* makes it practical to study the localisation of proteins *in vivo*. They find that LRK-1 localises to the Golgi. In the *C. elegans* *lrk-1* knockout line, they observe a difference in synaptic vesicles (SV) localisation in the dendrite. It is suggested that LRK-1 protein excludes SV from the dendrite, which was verified using LRK-1 deletion (Fukuzono *et al.*, 2016). In another study where the *lrk-1* gene was deleted, mitochondrial dysfunction was observed in *C. elegans* (Saha *et al.*, 2009). Another approach studied transgenic *C. elegans* with overexpressed wild-type (WT) LRRK2 and LRRK2 mutants, R1441C and G2019S (Yao *et al.*, 2010). The overexpression of WT LRRK2 induced age-dependent symptoms such as neurodegeneration of dopaminergic neurones and locomotor dysfunction. R1441C and G2019S LRRK2 mutants cause more severe symptoms.

#### **1.5.4 *Drosophila melanogaster***

The fruit fly *D. melanogaster* has a ROCO protein named dLRRK and has a domain architecture close to LRRK1 (**Figure 1.4**). The evolutionary studies of ROCO protein by Marin suggest that *D. melanogaster* is not a suitable model to study LRRK2 ROCO protein because it is not a true orthologue of LRRK2 (Marín, 2006, 2008). Despite this recommendation, studies were done on dLRRK to understand LRRK2 and show no impact of dLRRK protein in the early development and the dopaminergic neurones viability (Yue, 2009).

#### **1.5.5 *Mus musculus***

Rodents, similar to all vertebrates, have four ROCO proteins, Malignant fibrous histiocytoma-amplified sequence 1 homolog, Death-associated protein kinase 1, Leucine-rich repeat serine/threonine-protein kinase 1 (Lrrk1), and Leucine-rich repeat serine/threonine-protein kinase 2 (Lrrk2), represented in **Figure 1.4**. Bacterial Artificial Chromosome (BAC)-mediated LRRK2 transgenic mice were produced and used to study LRRK2 function (Yanping Li *et al.*, 2009). LRRK2 mutant G2019S and Y1699C were also made. LRRK2 knockout mice were studied, showing that LRRK2 is not needed for the development of adult mice (Yue, 2009). LRRK2 phosphorylate the Rabs proteins (Steger *et al.*, 2017) LRRK1 in mice was used extensively as a model to understand human LRRK1. Knockout (KO) of LRRK1 in mice induces severe osteopetrosis (Xing *et al.*, 2013). LRRK1 KO mice have also been used to study humoral immune responses (Morimoto *et al.*, 2016).

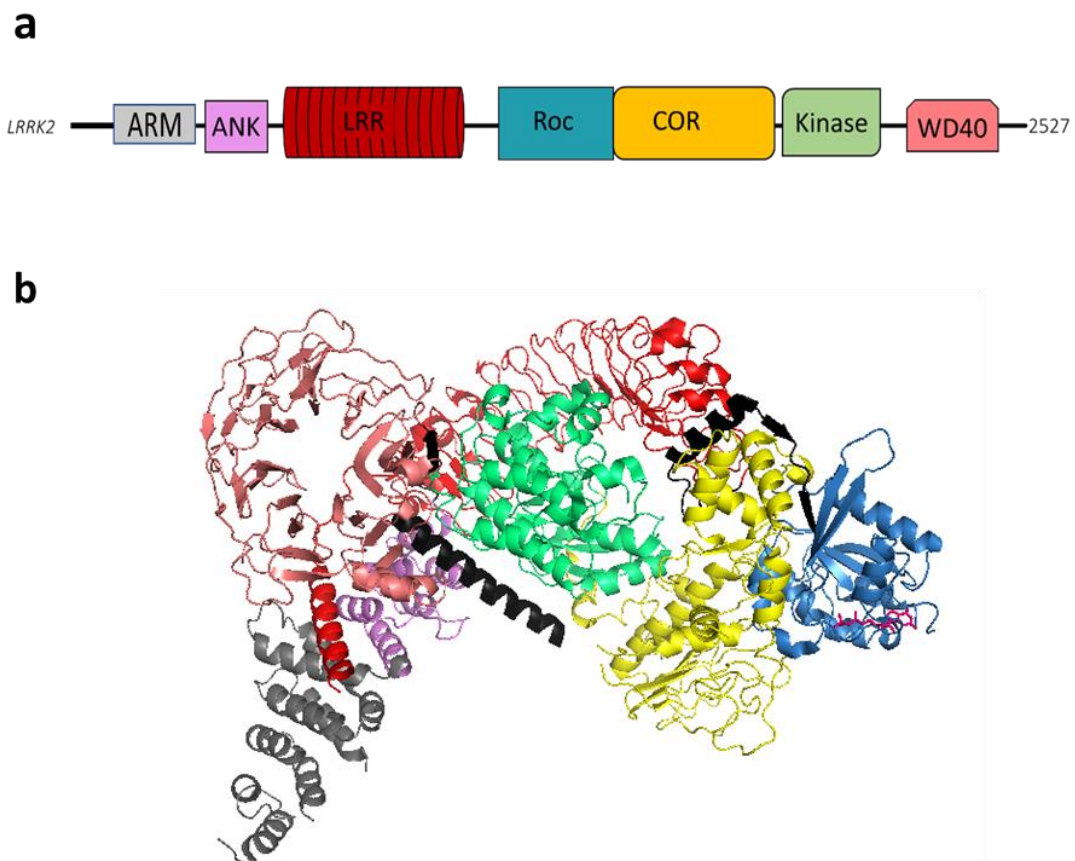


**Figure 1.4:** Schematic representation of the ROCO proteins from *Chlorobium tepidum*, *Dictyostelium discoideum*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Mus musculus*. The main domains present are Leucine Rich Repeat (LRR), Ras Of Complex protein (ROC), C terminal Of Roc (COR), Mitogen-Activated protein Kinase Kinase Kinase (MAPKKK). Other domains present on the proteins are Ankyrin (ANK), WD40, neuronal guanine nucleotide exchange factor (NGEF), Dishevelled/EGL-10/Pleckstrin (DEP), Guanine nucleotide exchange factor for Ras (RAS GEF), cyclic Nucleotide-Binding (cNB), Glucosyltransferases, Rab-like GTPase Activators and Myotubularins (GRAM), Protein-Tyrosine Phosphatase (PTP), Rho Guanine nucleotide Exchange Factors (Rho GEF), Pleckstrin Homology (PH), Rho GTPase-activating protein (Rho GAP) and Regulator of G protein Signalling (RGS).

## 1.6 Human ROCO proteins

### 1.6.1 LRRK2

Leucine-Rich Repeat Kinase 2 (LRRK2) is the largest human ROCO protein, with 2527 amino-acid and seven domains (**Figure 1.5**). The armadillo (ARM) domain found in human LRRK2 is not present in animal. Over ten years, separate the first structure of the Roc domain (Deng *et al.*, 2008) and the full-length structure of LRRK2 (Myasnikov *et al.*, 2021). In 2017, a 16 Angstrom (Å) resolution of full-length LRRK2 was imaged by cryo-Electron Microscopy (cryo-EM) (Sejwal *et al.*, 2017). In 2020, electron-microscopy and tomography technics described a 14 Å structure of full-length LRRK2 and its bound to the microtubule (Watanabe *et al.*, 2020). LRRK2 full-length structure was solved in 2021 in monomeric and dimeric states with a resolution of 3.7 Å and 3.5 Å, respectively (Myasnikov *et al.*, 2021) (**Figure 1.5**).



**Figure 1.5:** (a) Schematic representation of the domains of the ROCO proteins from human leucine-rich repeat kinase 2 (LRRK2). (b) Cryo-EM structure of LRRK2 full-length monomer with Guanosine diphosphate (GDP) bound to the Roc domain. PDB: 7LHW (Myasnikov *et al.*, 2021). Domains colours are matching the schematic. Armadillo (ARM) grey, Ankyrin (ANK) purple, Leucine Rich Repeat (LRR) red, Ras Of Complex protein (ROC) blue, C terminal Of Roc (COR) yellow, kinase green and WD40 pink. Linkers are black.

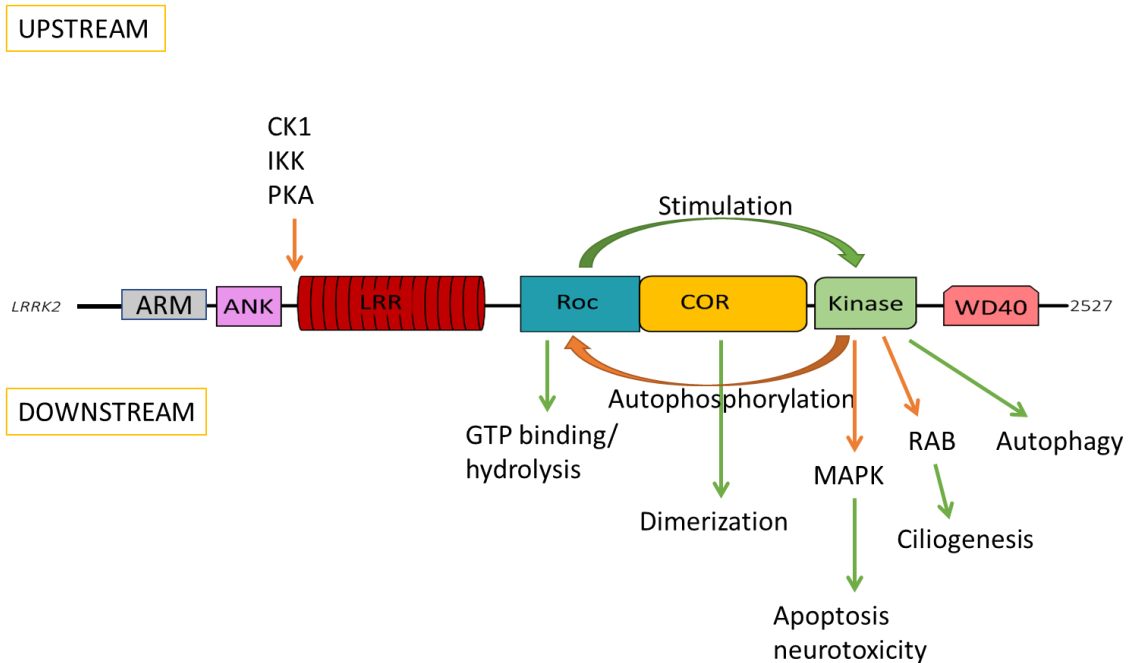
### **1.6.1.1 Pathology**

In 2004 mutations in *LRRK2* were linked to Parkinson's Disease (PD) (Paisan-Rui *et al.*, 2004; Zimprich *et al.*, 2004). Since then, this protein has become the most studied ROCO protein. Mutations in the *LRRK2* gene are the most common cause of sporadic and familial PD (Kumari and Tan, 2009). PD is the second most prevalent neurodegenerative disorder after Alzheimer's disease (AD). It affects the central nervous system, a specific area in the *Substantia nigra*, the *pars compacta*, resulting in the loss of predominantly dopaminergic neurons (Bloem, Okun and Klein, 2021).

LRRK2 is also associated with cancer (Saunders-Pullman *et al.*, 2010; Agalliu *et al.*, 2015; Warø and Aasly, 2018; J. Wang *et al.*, 2020), leprosy (Fu-Ren *et al.*, 2009); Crohn's disease (Barrett *et al.*, 2008) and tuberculosis (Härtlova *et al.*, 2018; Z. Wang *et al.*, 2018).

### **1.6.1.2 Cellular Biology**

The LRRK2 kinase domain can perform autophosphorylation on its ROC and LRR domains (West *et al.*, 2005; Greggio *et al.*, 2009; Vancaenenbroeck *et al.*, 2012) and phosphorylation of other proteins. Ras of Complex (ROC) domain contains a GTPase often described as a Ras-like domain (Mills *et al.*, 2014). It encodes a GTP binding/GTPase activity and is thought to act as an intramolecular switch regulating LRRK2 protein activity and interaction with binding partners (Lewis *et al.*, 2007; Gotthardt *et al.*, 2008; Carlessi *et al.*, 2011). R1441C/G/H mutations in the ROC domain have been reported to decrease GTPase activity which is pathogenic for PD (Lewis *et al.*, 2007; Li *et al.*, 2007). G2019S mutation in the kinase domain is the most common mutation in LRRK2 protein. G2019S increases kinase activity, autophosphorylation and autophagy (West *et al.*, 2005; Greggio *et al.*, 2006). The main activities around LRRK2 are summarised in **Figure 1.6**. Upstream kinases like casein kinase1  $\alpha$  (CK1 $\alpha$ ) (Chia *et al.* 2014), IkappaB kinases (IKKs) (Dzamko *et al.* 2012), and protein kinase A (PKA) (Muda *et al.* 2014) phosphorylate LRRK2. LRRK2 kinase domain phosphorylates a subgroup of 14 RAB GTPase proteins (Alessi and Sammler 2018), including Rab8A, Rab10 and Rab 12 involved in ciliogenesis (Steger *et al.* 2017). LRRK2 has also been linked to the MAPKKK cascade by phosphorylating MAPKK which led to regulation of apoptosis, neurotoxicity, and oxidative stress (Gloeckner *et al.*, 2009). LRRK2 regulates many pathways involved with autophagy (Kania and Parys 2019) and is also involved in the Wntless/INT (Wnt) signaling pathway as a scaffold (Berwick and Harvey, 2012).



**Figure 1.6:** Representation of the main upstream and downstream pathways of Leucine Rich Repeat Kinase 2 (LRRK2) protein containing the following domains Armadillo (ARM), Ankyrin (ANK), Leucine Rich Repeat (LRR), Ras Of Complex protein (ROC), C terminal Of Roc (COR), Kinase and WD40. Green arrows represent activation, and orange arrows represent phosphorylation.

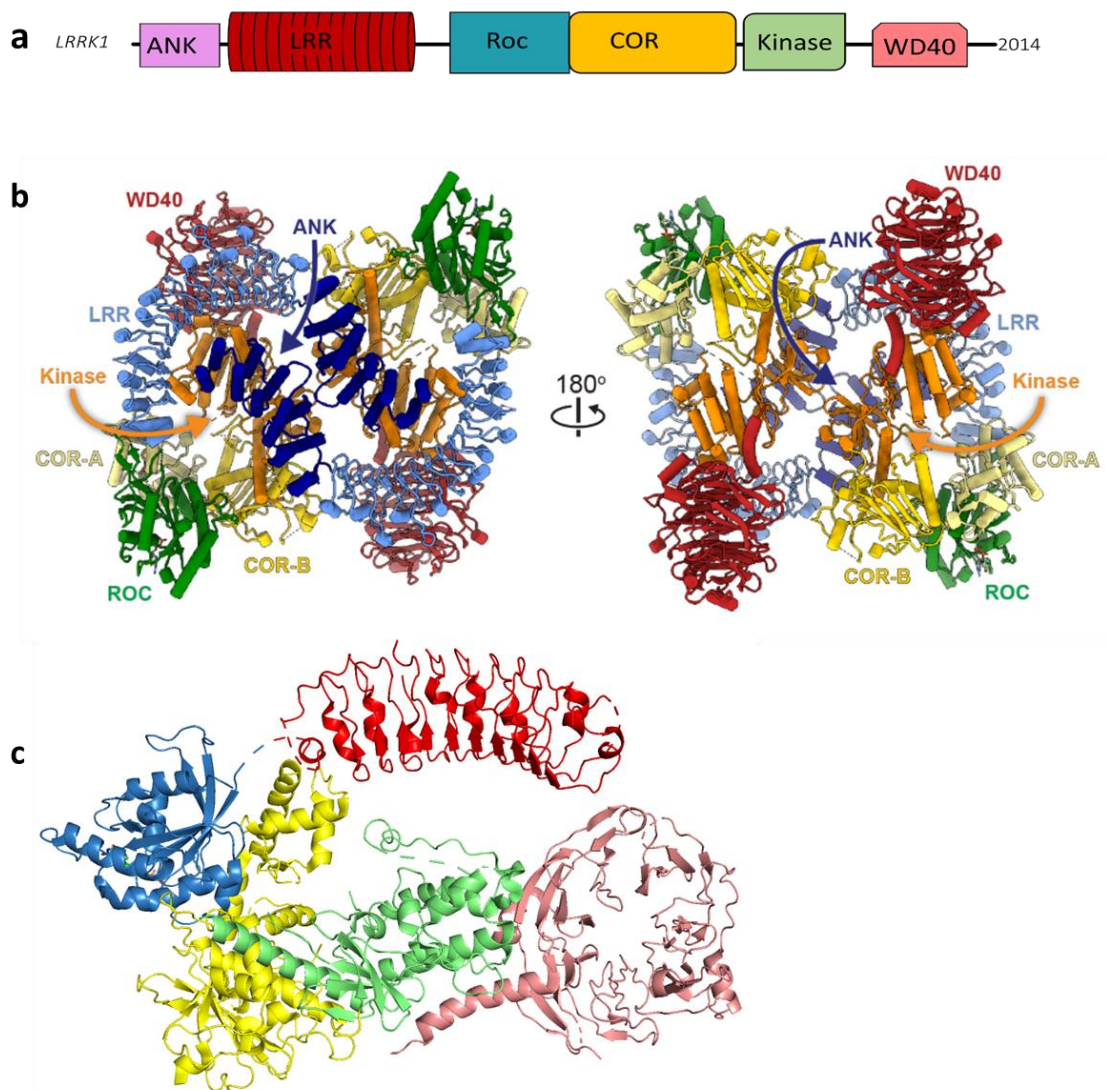
### 1.6.1.3 Expression and localisation in the cells and the body

According to the Genotype-Tissue Expression project (GTEx), *LRRK2* gene expression is the highest in the lung (ENSG00000188906.14), it is assumed that LRRK2 protein is primarily found in the lungs.

LRRK2 protein localisation has been studied using overexpressed LRRK2 and immunofluorescence. It was found that LRRK2 is primarily located in the cytoplasm and the membrane of cellular components like endoplasmic reticulum (ER), mitochondria and Golgi. LRRK2 protein is also found in the microtubular cytoskeleton (Gloeckner *et al.*, 2006). LRRK2, in its dimeric form, is mainly located at the membrane and in the cytosol in its monomeric form (Berger, Smith and Lavoie, 2010).

### 1.6.2 LRRK1

LRRK1 is a 2014 amino-acid human protein with similar domain architecture to LRRK2. It contains six domains; ankyrin repeat (ANK), leucine-rich repeat (LRR), ROC described as a small GTPase (Korr *et al.*, 2006), COR, serine/threonine kinase and WD40 domain (**Figure 1.7**). LRRK1 structure was recently solved by two teams with cryo-EM (Reimer *et al.*, 2022; Metcalfe *et al.*, 2023). The dimeric structure is solved with 4.6 Å, on this date, the paper is on bioRxiv (Reimer *et al.*, 2022). The monomeric structure of LRRK1 is solved with 3.9 Å and the dimeric with 6.4 Å (Metcalfe *et al.*, 2023) (**Figure 1.7**).



**Figure 1.7:** (a) Schematic representation of the domains of the ROCO proteins from human leucine-rich repeat kinase 1 (LRRK1) containing the following domains: Ankyrin (ANK), Leucine Rich Repeat (LRR), Ras Of Complex protein (ROC), C terminal Of Roc (COR), kinase and WD40. (b) Cryo-EM structure of dimeric LRRK1 (image taken from bioRxiv paper (Reimer *et al.*, 2022)). (c) Cryo-EM structure of monomeric LRRK1 PDB: 8FAC (Metcalfe *et al.*, 2023). Domains colours are matching the schematic. Leucine Rich Repeat (LRR) red, Ras Of Complex protein (ROC) blue, C terminal Of Roc (COR) yellow, kinase green and WD40 pink.

### **1.6.2.1 Pathology**

Osteopetrosis is a bone disorder occurring when bone production is more significant than bone resorption and leads to abnormally dense bones, increasing the risk of fracture. A balanced bone mineral density (BMD) is essential for healthy bones. Osteoblast synthesises bones while the osteoclast degrades them. A high throughput screen on a large population found LRRK1 involved in the regulation of osteoclast (Xing *et al.*, 2017). A study of LRRK1 knockout mice shows osteoclast dysfunction leading to severe osteopetrosis (Xing *et al.*, 2013).

Osteosclerotic metaphyseal dysplasia (OSMD) is a rare condition which causes developmental delay and localised increase of bone density called osteosclerosis. It was first described in a Japanese family in 1993 as a new sclerosing bone disease (Nishimura *et al.* 1993). *LRRK1* was linked to this disease in 2016 after whole genome sequencing was studied on three patients (Iida *et al.*, 2018). Since then, few cases have been linked with *LRRK1* mutations (Howaldt *et al.*, 2020; Miryounesi *et al.*, 2020).

### **1.6.2.2 Cellular Biology**

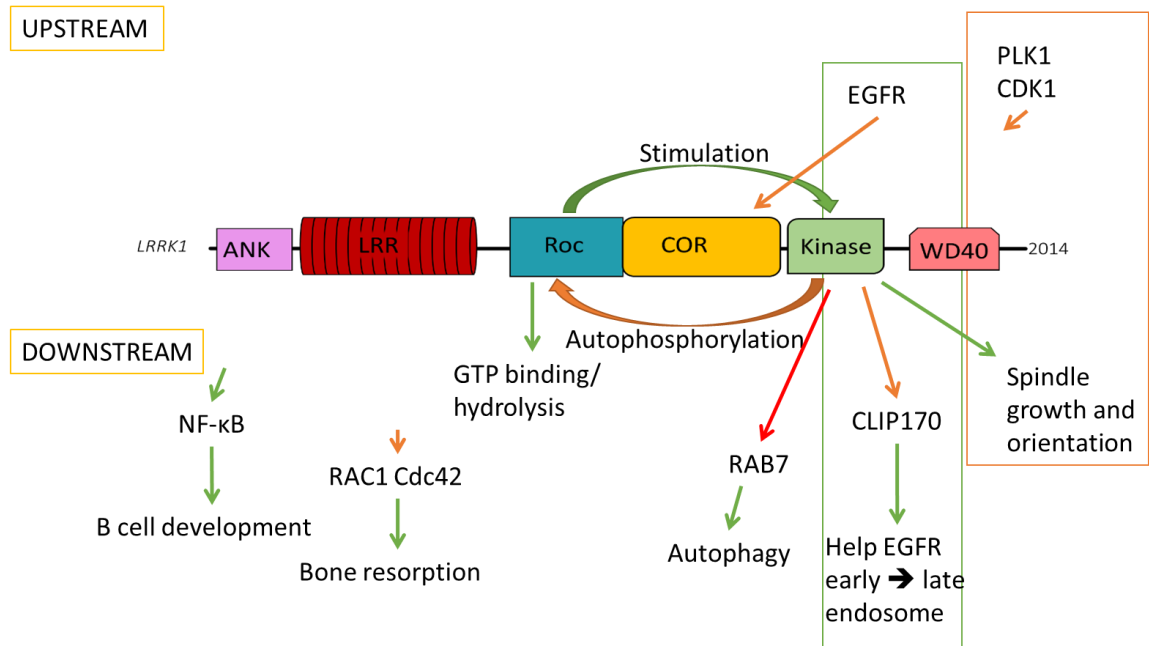
LRRK1 is involved in immunity. LRRK1 regulates the Nuclear factor kappa-light-chain-enhancer of activated B cells (NF- $\kappa$  B) signalling pathway, regulating B cell proliferation and survival (Morimoto *et al.*, 2016).

In 2005, a study identified LRRK1 as a mammalian growth-regulatory factor (GRF) (Harada *et al.*, 2005). In 2011, LRRK1 was reportedly involved in regulating the epidermal growth factor receptor (EGFR) (Hanafusa *et al.*, 2011). EGFR initiates many cellular mechanisms like endocytosis, endosomal trafficking, and MAPK cascade. LRRK1 also regulates bone resorption (Xing *et al.*, 2013).

The main upstream and downstream mechanisms of LRRK1 are represented in **Figure 1.8**. EGFR phosphorylates LRRK1 at T944, which regulates LRRK1 kinase activity (Ishikawa *et al.*, 2012). Downstream to EGFR, LRRK1 phosphorylates Cytoplasmic linker protein 170 (CLIP-170) which facilitates the migration of EGFR from early to late endosomes (Kedashiro *et al.*, 2015). The PLK1/CDK1–LRRK1– signalling pathway regulates spindle formation (Hanafusa and Matsumoto, 2015). LRRK1 activates RAC1/ Cell division control protein 42 (Cdc42) by phosphorylating them, inducing bone resorption (Zeng *et al.*, 2016). LRRK1 regulates



autophagy by binding to TBC1D2 in the lysosome, which inactivates Rab7 (Toyofuku *et al.*, 2015). Parkin- mediated mitophagy is regulated by LRRK1 via S72 phosphorylation of Rab7 (Hanafusa *et al.*, 2019; Fujita *et al.*, 2022). GTP binding to ROC GTPase stimulates the kinase domain. The kinase domain is capable of autophosphorylation and phosphorylation of downstream proteins. (Korr *et al.*, 2006).



**Figure 1.8:** Representation of the upstream and downstream mechanisms of Leucine Rich Repeat Kinase 1 containing the following domains: Ankyrin (ANK), Leucine Rich Repeat (LRR), Ras Of Complex protein (ROC), C terminal Of Roc (COR), Kinase and WD40. Green arrows represent activation, orange arrows represent phosphorylation, and red arrow represents negative regulation. Upstream and downstream pathways that are connected are framed.

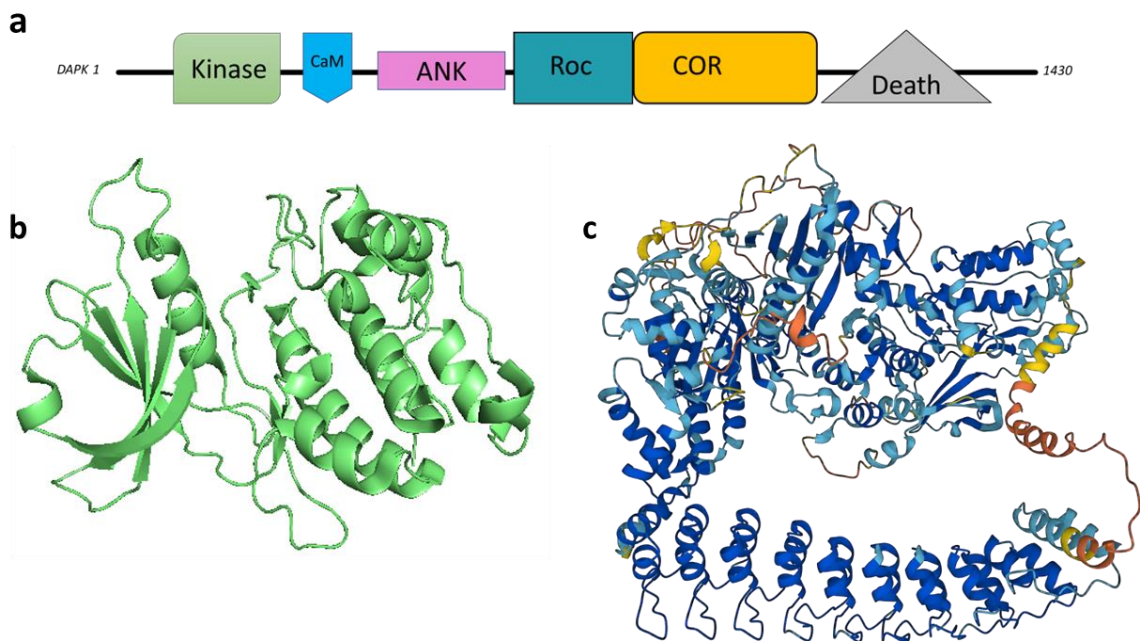
### 1.6.2.3 Expression and localisation in the cells and the body

LRRK1 messenger Ribonucleic acid (mRNA) is present in the adult rat brain, adrenal gland, liver, lung, spleen, and kidney. LRRK1 protein is expressed in the adult human cortex cerebri and the hippocampus (Westerlund *et al.*, 2008). According to the Genotype-Tissue Expression (GTEx) Project, the *LRRK1* gene level is likely to be the highest in EBV-transformed lymphocytes cells (ENSG00000154237.12).

LRRK1 is localised in the cytoplasm (Korr *et al.*, 2006; Schulte *et al.*, 2014). LRRK1 interacts with the epidermal growth factor (EGF) through its kinase domain. LRRK1 and EGF are shown to co-localise in early endosomes (Hanafusa *et al.*, 2011).

### 1.6.3 DAPK1

Death Associated Protein Kinase 1 (DAPK1) is a 1430 amino acid protein containing a serine/threonine kinase, a calmodulin (CaM) domain which regulates the kinase, eight Ankyrin repeats (ANK), a ROC-COR supradomain and a death domain (**Figure 1.9**). DAPK1 is part of the ROCO protein family due to the presence of a ROC-COR domain, but it is also part of the death-associated protein kinase (DAPK) family, which consists of five proteins with a common kinase domain. DAPK1 was first described in 1995; it was found to be a positive mediator of interferon- $\gamma$  (IFN- $\gamma$ ) induced programmed cell death (Deiss *et al.*, 1995). The kinase domain of DAPK1 is distinct from the LRRK proteins in terms of its orientation in the protein (N-terminal instead of C-terminal) and its place on a different branch of the Kinome tree (**S.Figure 1**). The structure of the kinase domain was solved in 2001 at 1.5 Å (Tereshko *et al.*, 2001) (**Figure 1.9**). The full predicted structure is available on Alphafold2 (**Figure 1.9**).



**Figure 1.9:** (a) Schematic representation of the domains of the ROCO proteins from human Death Associated protein Kinase 1 (DAPK1) containing the following domains: Kinase, Calmodulin (CaM), Ankyrin (ANK), Ras Of Complex protein (ROC), C terminal Of Roc (COR), and Death domain. (b) Crystal structure of the kinase domain of DAPK1 which comprise the amino acid 1 – 285 (PDB: 1JKS). (c): The full predicted structure of DAPK1 from Alphafold2.

### **1.6.3.1 Pathology**

DAPK1 acts as a tumour suppressor and is involved in different cancers, including solid tumours. Deoxyribonucleic acid (DNA) methylation of the DAPK1 gene promoter can lead to the silencing of the gene, leading to a decrease in the tumour suppressor action and an increase in tumour development. DAPK1 methylation is involved in many cancers (Calmon *et al.*, 2007; Missaoui *et al.*, 2010; Zhu *et al.*, 2017).

DAPK1 overexpression has been found in Alzheimer's disease (AD) brain patients. Tau proteins are usually present in the brain, but when hyperphosphorylated, they become insoluble and are involved in AD (Mi and Johnson, 2006). DAPK1 has been shown to hyperphosphorylate tau and accentuate tau aggregation, making the neurones more resistant to cell death (Duan *et al.*, 2013; Kim *et al.*, 2014) (**Figure1.10**). DAPK1 could be a target for AD treatment (Xu, Li and Jia, 2019).

### **1.6.3.2 Cellular Biology**

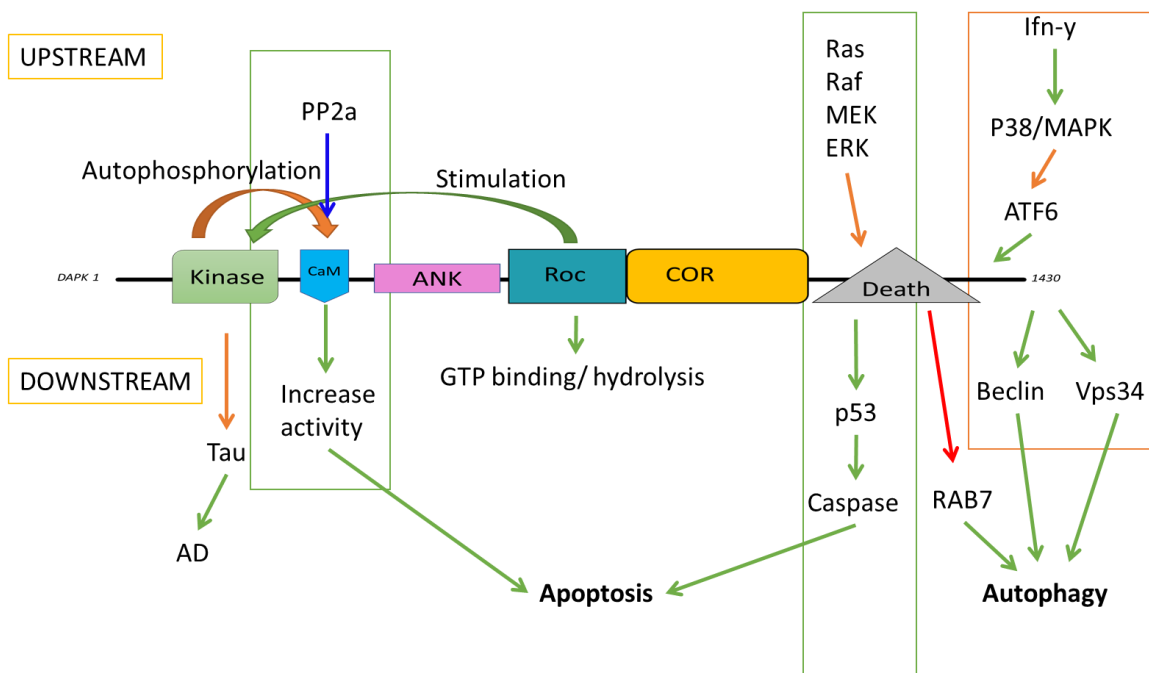
DAPK1 has a GTPase activity deriving from its ROC domain (Jebelli *et al.*, 2012), which may regulates the protein. GTP binding on the P-loop of the ROC domain enhances autoinhibition by autophosphorylation of serine 308 in the CaM domain blocking the kinase domain (Carlessi *et al.*, 2011). The DAPK1 kinase domain is part of the calcium/calmodulin-dependant protein kinase (Manning *et al.*, 2002). In contrast to LRRK2, DAPK1 autophosphorylation acts to repress protein activity.

DAPK1 has several identified upstream regulators, shown in **Figure 1.10**. Protein phosphatase 2 (PP2a) dephosphorylates DAPK1 at S308, increasing its activity and inducing apoptosis (Guenebeaud *et al.*, 2010). Under aberrant growth conditions, extracellular signal-Regulated Kinase (ERK) from the Ras/Raf/MEK/ERK pathway phosphorylate DAPK1 at Ser 735 in the death domain (DD) and activates it, leading to apoptosis (Chen *et al.*, 2005). In normal growth conditions, the pathway Ras/Raf/MEK/ERK is followed by p90 ribosomal S6 kinases (RSK), which inhibit DAPK1 (Anjum *et al.*, 2005).

Activated DAPK1 induces apoptosis by phosphorylating the oncoprotein p53, which provokes caspase-dependent death (Bialik and Kimchi, 2006; Pei *et al.*, 2014) (**Figure1.10**). During a stroke, phosphorylation of the N-methyl-D-aspartate (NMDA) receptor by DAPK1 increases calcium ion (Ca<sup>2+</sup>) influx which leads to neuron death (Tu *et al.*, 2010).

The upstream regulation of DAPK1 in autophagy starts with IFN- $\gamma$  (Gade *et al.*, 2014). IFN- $\gamma$  activates the ASK1-MKK3/MKK6-p38 MAPK pathway. P38 and MAPK phosphorylate ATF6, which migrates from the Golgi to the nucleus, promoting DAPK1 expression. In the downstream regulation of autophagy, DAPK1 phosphorylates Beclin 1. (Zalckvar, Berissi, Eisenstein, *et al.*, 2009). When phosphorylated, Beclin 1 cannot bind to the anti-apoptotic Bcl2. Instead, several activators bind to Beclin 1, inducing autophagy (**Figure 1.10**) (Zalckvar, Berissi, Mizrachy, *et al.*, 2009).

DAPK1 induces autophagosome formation through a phosphorylation cascade (Eisenberg-Lerner and Kimchi, 2012). DAPK1 phosphorylates protein kinase D (PKD), which phosphorylates and activates Vacuolar protein sorting 34 (Vps34). Activated Vps34 then drives production of phosphatidylinositol-3-phosphate (PI(3)P) and formation of the autophagosome (**Figure 1.10**).



**Figure 1.10:** Representation of the upstream and downstream mechanism of Death associated protein kinase 1 (DAPK1) containing the following domains: Kinase, Calmodulin (CaM), Ankyrin (ANK), Ras Of Complex protein (ROC), C terminal Of Roc (COR), and Death domain. Green arrows represent activation, orange arrows represent phosphorylation, red arrows represent negative regulation, and blue arrows represent dephosphorylation. Upstream and downstream pathways that are connected are framed.

### **1.6.3.3 Expression and localisation in the cells and the body**

DAPK1 has been found to suppress migration and regulate cell polarity (Kuo *et al.*, 2006). A 1997 study shows that DAPK kinase is localised at the cytoskeleton and associated with the microtubule (Cohen, Feinstein and Kimchi, 1997). Another study shows that by removing the cytoskeletal interaction domain of DAPK1, the protein is diffused in the cytoplasm (Bialik, Bresnick and Kimchi, 2004).

Genotype-Tissue Expression (GTEx) Project, present that the *DAPK1* gene level is potentially the highest in ovary cells (ENSG00000196730.12).

### 1.6.4 MFHAS1

Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1), also called MFH-Amplified Sequences with Leucine-rich tandem repeats 1 (MASL1), was first described as an oncogene in 1999 in a study on malignant fibrous histiocytomas (MFHs), giving it its name (Sakabe *et al.*, 1999). MFHAS1 has 1052 amino-acid and only three domains: LRR, ROC and COR, making it the only human ROCO protein lacking a kinase domain. (**Figure 1.11**). The domain structure organisation is similar to the *C. tepidum* CtROCO protein. An atomic resolution structure for MFHAS1 has not yet been described, but a study using size exclusion chromatography (SEC) reveals that MFHAS1 is present in two complexes, resulting in two sizes of the protein 600kDa and 200kDa (Dihanich *et al.*, 2014), suggesting the formation of a dimer.



**Figure 1.11:** Schematic representation of Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1) protein containing three domains; Leucine Rich Repeat (LRR), Ras of Complex protein (COR), C terminal of ROC (COR).

#### 1.6.4.1 Pathologies

##### 1.6.4.1.1 Cancer

The first study on the MFHAS1 gene found its location at amplicon 8p23.1. Overexpression of the MFHAS1 gene causes tumours in malignant fibrous histiocytomas (MFH) (Sakabe *et al.*, 1999). MFH is present in soft tissues and bones. Since its discovery in 1999, MFHAS1 gene overexpression, amplification or loss of gene was found to be involved in many cancers (summarised in **Table 1.1**). MFHAS1 seems to have a complex role in cancer as both amplification and gene loss are implicated in cancers.

A study on nude mice shows that MFHAS1 and chimeric MFHAS1 hold tumorigenic activity (Tagawa *et al.*, 2004). Chimeric *MFHAS1* cDNA consists of exon1 at 8p23 (3002 nucleotides) followed by 14q21 element which add 57 nucleotides.

Cancer type	Gene expression	Reference
Malignant Fibrous Histiocytomas (MFH) and gastric cancer	Amplification	(Sakabe <i>et al.</i> , 1999)
Diffuse Large B-Cell Lymphoma (DLBCL)	Amplification	(Mareschal <i>et al.</i> , 2016) (Dubois <i>et al.</i> , 2016)
Squamous cell carcinoma of the vulva	Loss of gene	(Stephen <i>et al.</i> , 2009)
Lymph nodes metastasis (Larynx, pharynx)	Loss of gene	(Alonso Guervós <i>et al.</i> , 2009)
Head and Neck Squamous Cell Carcinoma (HNSCC)	Loss of gene	(Bockmühl <i>et al.</i> , 2001)
Gastric cancer	Overexpression	(Yang <i>et al.</i> , 2007)
Squamous Cell Carcinoma (SCC) of the lung	Loss of gene	(Kang, 2015)
Colorectal Cancer (CRC)	Overexpression	(Chen <i>et al.</i> , 2016)
Immunoblastic B-cell lymphoma	Chimeric	(Tagawa <i>et al.</i> , 2004)
Abdominal tumour	Amplification	(Weng <i>et al.</i> , 2004)

**Table 1.1:** Summary of cancers involved with MFHAS1 gene overexpression, loss, amplification or chimeric.

#### 1.6.4.1.2 Sepsis

A study on proteins containing a Leucine-Rich Repeat domain reported that MFHAS1 could regulate Toll-Like Receptor (TLR) dependant signalling (Ng *et al.*, 2011). They also reported that the MFHAS1 gene was expressed in macrophages infected with *M. tuberculosis* and *S.typhi*. The study also shows that MFHAS1 small interfering RNA (siRNA) knockdown in macrophages increases the production level of Interleukin-6 (IL-6). Suggesting MFHAS1 may have a role in the immune response. Other papers show the implication of MFHAS1 in sepsis through the TLR2/NF- $\kappa$ B pathway (Zhong *et al.*, 2015) and macrophage polarisation (Zhong *et al.*, 2017). Studies in European individuals of African ancestries shows that the MFHAS1 gene is involved in sepsis. Furthermore, the peripheral blood of septic patients shows a significant overexpression of MFHAS1 protein (Hernandez-Beeftink *et al.*, 2022).

#### 1.6.4.1.3 Diabetes mellitus

Diabetes mellitus (DM) is a disorder where hyperglycaemia is present for an extended time (H. Wang *et al.*, 2018). The role of MFHAS1 in hyperglycaemia and inflammation has been investigated in human umbilical vein endothelial cells (HUVEC). MFHAS1 expression can reduce inflammation caused by a high level of glucose. (H. Wang *et al.*, 2018).

Diabetic nephropathy (DN) is a complication of DM resulting in chronic kidney disease (Samsu, 2021). Experiments carried out *in vivo* in diabetic mice show a higher expression of MFHAS1 in diabetic tissues than in wild-type tissues, making MFHAS1 a potential therapeutic target for DN (Lian, Cheng and Wu, 2019).

MFHAS1 gene was associated with type II diabetes, obesity and lipid levels in a study focusing on the Brazilian admixed population (Secolin *et al.*, 2021).

#### 1.6.4.1.4 Others

Two genome-wide association studies found MFHAS1 to be involved in diverse conditions. A neuroticism-associated loci analysis discovers the MFHAS1 gene expressing during foetal life in non-neural genes. (Ohi *et al.*, 2017). MFHAS1 gene is potentially associated with Dietary Approaches to Stop Hypertension (DASH) Diet (Mompeo *et al.*, 2022), suggesting that genes may influence dietary choice.

p.Val106Gly and p.Arg556Cys variants of MFHAS1 are associated with the risk of non-syndromic type III biliary atresia (BA) (Tamaoka *et al.*, 2023). In a group of 15 Japanese children with BA, four present with p.Val106Gly variant, and one has the p.Arg556Cys variant.

#### **1.6.4.2 Cellular function**

The ROC domain of MFHAS1 can bind GTP and is the only active enzymatic domain of MFHAS1. LRRK2 homologous mutation, K422A, present in the MFHAS1 ROC domain, stops GTP binding (Dihanich *et al.*, 2014).

The interactome of MFHAS1 has been investigated using various approaches combined with bioinformatic analysis (Tomkins *et al.*, 2018). Protein microarray screens found 87 positive hits, including six kinase interactors to MFHAS1: CLK1, LIMK1, MAP3K4, NEK11,



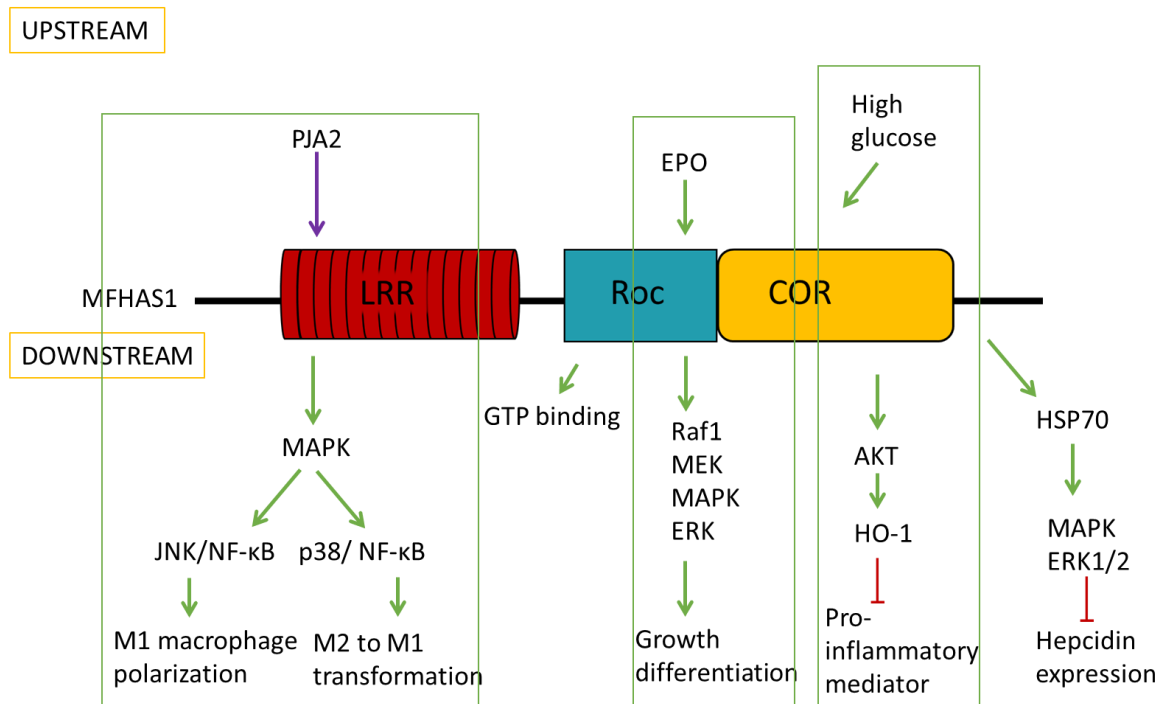
ROR1, and STK25. A biochemical analysis of MFHAS1 using co-immunoprecipitation shows that MFHAS1 interacts with the chaperones Heat shock protein 60 (HSP60) and Heat shock protein 70 (HSP70) (Dihanich *et al.*, 2014). Those results were recently replicated in a hepcidin regulation study (Kumkhaek *et al.*, 2019). Also, using co-immunoprecipitation, HSP60 and HSP70 were found to interact with MFHAS1 at the endogenous level.

MFHAS1 has been implicated in the Toll-like receptor (TLR) cascade. In sepsis, for the first six hours after infection, MFHAS1 inhibits the TLR2 response. Twenty-four hours after infection, MFHAS1 stimulates the JNK/NF- $\kappa$ B/AP-1 signalling pathway (Zhong *et al.*, 2015). Activated TLR2 interacts with MAPK, which then interacts with JNK and p38 leading to IL-6 production. MFHAS1 interacts with MAPK when ubiquitinated by the Ubiquitin-protein ligase, Praja2 (PJA2) (Zhong *et al.*, 2018). MFHAS1 interaction with MAPK would promote activity and increase inflammation through the c-Jun N-terminal Kinase (JNK), p38 and Nuclear Factor Kappa B (NF- $\kappa$ B) pathways. The JNK pathway induces the polarisation of M1 macrophage, while the p38 pathway induces the transformation of M2 into M1 (Zhong *et al.*, 2018) (**Figure 1.12**).

Increased expression of MFHAS1 can be caused by the presence of LPS and stops the inflammation by inhibiting cytokines expression (Shi *et al.*, 2017). Overexpression of MFHAS1 by high glucose in endothelial cells activates the AKT/HO-1 pathway by phosphorylating AKT, which increases the expression of HO-1, an inhibitor of the inflammatory response (H. Wang *et al.*, 2018) (**Figure.1.12**). Overexpression of MFHAS1 *in vitro* and *in vivo* was shown to suppress TLR4 activation (Lian, Cheng and Wu, 2019).

MFHAS1 has also been implicated in the differentiation of erythroid cells into reticulocytes, leading to the formation of red blood cells (Kumkhaek *et al.*, 2013). The erythropoietin (EPO) receptor can positively regulate MFHAS1 through its ROC domain. MFHAS1 interacts with Raf1, which induces the Raf/MEK/ERK cascade (**Figure.1.12**). The MAPK/ERK pathway can induce Elk1 expression, which then induces growth and differentiation of early erythroid cells into reticulocytes. Knockdown of MFHAS1 interrupt the Raf/MEK/ERK pathway and decreases erythroid differentiation (Kumkhaek *et al.*, 2013).

Hepcidin protein regulates the entry of iron into blood circulation. MFHAS1 could bind and activate Heat shock protein 70 (HSP70), which will induce the MAPK/ERK1/2 pathway and reduce hepcidin expression (Kumkhaek *et al.*, 2019).



**Figure 1.12:** Schematic representation of Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1) and its unresolved pathways. MFHAS1 protein contains the following domains: Leucine Rich Repeat (LRR), Ras Of Complex protein (ROC) and C terminal Of Roc (COR). Green arrows represent activation, purple arrows represent ubiquitination, and red arrows represent negative regulation. Upstream and downstream pathways that are connected are framed. Erythropoietin (EPO), E3 ubiquitin-protein ligase Praja-2 (PJA2) and Lipopolysaccharides (LPS) are MFHAS1 upstream activators. Downstream pathways include c-Jun N-terminal Kinase (JNK), p38, Nuclear Factor Kappa B (NF-κB), guanosine triphosphate (GTP), proto-oncogene serine/threonine-protein kinase (RAF1), Mitogen-activated protein kinase kinase (MEK or MAPK), Extracellular signal-regulated kinase (ERK), Heme oxygenase 1 (HO-1) and Heat Shock Protein 70 (HSP70).

#### 1.6.4.3 Expression and localisation in the cells and the body

MFHAS1 overexpression has a cytotoxic effect on the cells, increasing when there is a loss of GTP binding (MFHAS1 mutant). MFHAS1 is found in the cytoplasm, and colocalisation using lysosomal, mitochondrial and cytoskeletal markers shows that MFHAS1 is present in lysosomes but not consistently (Dihanich *et al.*, 2014). Localisation of MFHAS1 protein in the cytoplasm was confirmed by immunofluorescence (Shi *et al.*, 2017). Little is known about MFHAS1 post-translational modification and gene expression which will be discussed in Chapter 3, Bioinformatics.

## **1.7 Thesis structure, aims and objectives**

The main aim of this thesis is to understand the structure and function of MFHAS1. To achieve this, three approaches were used to study different aspects of the biology of MFHAS1.

First, in Chapter 2, expression, and purification of MFHAS1 was attempted, with the goal of facilitating structural studies.

Second, in Chapter 3, bioinformatics approaches are used to understand the post-translational modification of MFHAS1, the predicted proteins interacting with MFHAS1 and its predicted structure.

Third, in chapter 4 the impact of putative post translational modification of MFHAS1 upon downstream signalling pathways was tested using a variety of cellular and biochemical approaches.

Finally, in Chapter 5, the results are reviewed, and future experiments are discussed.

## Chapter 2: Expression and purification of MFHAS1

### 2.1 Introduction

As noted in Chapter 1, Malignant Fibrous Histiocytoma-Amplified Sequence 1 (MFHAS1) protein has a 1052 amino-acid open reading frame and three domains: Leucine-Rich Repeat (LRR), Ras Of Complex (ROC) and C terminal Of ROC (COR). The presence of the ROC and COR domains defines MFHAS1s membership of the human ROCO proteins, including LRRK2, LRRK1 and DAPK1.

Structurally, MFHAS1 shares close homology with the CtROCO protein from *Chlorobaculum tepidum* (*C. tepidum*), which contains the same domain structure (Gotthardt *et al.*, 2008), sharing 34.1% sequence similarity and 27% identity. The structure of the LRR-ROC-COR domains of CtROCO have been determined to a resolution of 3.3 Å (Deyaert, Kortholt and Versées, 2017). At the time of the experiments described in this Chapter, the LRRK2 human protein is the only other protein for which a crystal structure is available that shares some homology to MFHAS1. Determination of the LRRK2 protein structure proved challenging for over a decade before the protein was purified successfully, and the full-length dimer structure has recently been determined (Myasnikov *et al.*, 2021). The structure of the LRRK2 C terminal domain associated with microtubule has also been solved (Deniston *et al.*, 2020). There is no experimentally derived structure of MFHAS1.

The MFHAS1 gene is involved in cancer (Sakabe *et al.*, 1999), but MFHAS1 protein function has not been studied extensively. MFHAS1 has GTPase activity in the ROC domain (Dihanich *et al.*, 2014) and is the only human ROCO protein without a kinase domain. Structural characterisation of MFHAS1 would contribute to the understanding of its function and its role in cancer.

This chapter summarises the methods used to overexpress and purify MFHAS1 in sufficient quantities and purity for structural analyses.

Three steps need to be successful to obtain a sample ready for structural analysis. First, MFHAS1 DNA is cloned in a vector and transformed using a range of constructs and DNA fragments. Secondly, the plasmids are transfected in mammalian or insect cells to express MFHAS1 protein. Finally, the expressed protein is extracted using detergents to obtain a soluble protein which is then purified.

## 2.2 Material and methods

### 2.2.1 Domain delineation

Two methods were used to determine the domain boundaries of MFHAS1.

For the first method, a structure-based approach was taken by using the software Phyre2 with the resultant model structure used as a basis for construct design (Kelley *et al.*, 2015). Phyre2 software has four steps; first homologous sequences are gathered. Secondly, a database of known structures is scanned and aligned with the query (the MFHAS1 sequence), thirdly loops are placed to fill the gap(s) between the template structures, fourthly side chains are added to determine a model of the structure. For the second method, a comparison of the amino acid sequence of human MFHAS1 with other ROCO protein sequences taken together with available structural data was performed. In this case, EMBOSS needle Pairwise Sequence Alignment software and Clustal Omega were used (Madeira *et al.*, 2022). In particular the MFHAS1 protein sequence was compared with Ras proteins to determine the boundaries of the ROC domain. For the COR domain, the sequence was compared with LRRK2 (Mills *et al.*, 2018) and the ROCO protein of *C.tepidum* CtROCO (Gotthardt *et al.*, 2008) protein sequences.

At the time of the experiment, Alphafold2 was not available. In Chapter 3 (Chapter 3.3.1), analysis of MFHAS1 predicted structure from Alphafold2 was showing a similar domain boundary to the one found here.

### 2.2.2 Cloning

A plasmid containing the MFHAS1 open reading frame (ORF) was obtained from DNASU plasmid repository (HsCD00295107). The ORF of MFHAS1 and domains were cloned into pOPINF, pOPINEneo, pOPIN3C-EGFP and pOPINE\_Neo\_GFP\_2Strep\_His vectors designed by Protein Production UK (PPUK) (Berrow *et al.*, 2007). Maps for each vector are displayed in **S. Figure 2**. Vectors were picked for their tags and their capacity to express in mammalian and insect cells. Primers were designed using the web-based application Optic developed at the Oxford Protein Production Facility UK (OPPF-UK) (**Table 2.1**).

<b>pOPINF, pOPINeNeo, pOPIN3C-EGFP</b>	
<b>Forward</b>	<b>Primer</b>
MASL1_AA2Fneo	aggagataaccatgGCTGGGATGGACAGTGGC
MASL1_AA64Fneo	aggagataaccatgGACATTGAGGCACTGAACC
MASL1_AA366Fneo	aggagataaccatgCGGGTGGGTTTGTGGA
MASL1_AA381Fneo	aggagataaccatgCCCTACGAGGTCTGCATGA
MASL1_AA407Fneo	aggagataaccatgCAGCCCCGGCTCAAGC
MASL1_AA621Fneo	aggagataaccatgCTCTCCCCGTGTTGCC
MASL1_AA2Ff	aagttctgtttcaggcccgGCTGGGATGGACAGTGGC
MASL1_AA64Ff	aagttctgtttcaggcccgGACATTGAGGCACTGAACC
MASL1_AA366Ff	aagttctgtttcaggcccgCGGGTGGGTTTGTGGA
MASL1_AA381Ff	aagttctgtttcaggcccgCCCTACGAGGTCTGCATGA
MASL1_AA407Ff	aagttctgtttcaggcccgCAGCCCCGGCTCAAGC
MASL1_AA621Ff	aagttctgtttcaggcccgCTCTCCCCGTGTTGCC
<b>Reverse</b>	<b>Primer</b>
MASL1_AA366Rvneo	gtgaggtgtagtttCCGGGAGAGCTGGCC
MASL1_AA381Rvneo	gtgaggtgtagtttGGGGGGCTGGATCAG
MASL1_AA407Rvneo	gtgaggtgtagtttCTGCACCGCCGGC
MASL1_AA621Rvneo	gtgaggtgtagtttGAGGATCTGCAGCCGG
MASL1_AA1041Rvneo	gtgaggtgtagtttTTGGAACAGGGGCTGATCACAGTCGGCTGGGTGGAACAGGGGCTGATCAC
MASL1_AA366Rvf	atggtctagaaagctttaCCGGGAGAGCTGGCC
MASL1_AA381Rvf	atggtctagaaagctttaGGGGGGCTGGATCAG
MASL1_AA407Rvf	atggtctagaaagctttaCTGCACCGCCGGC
MASL1_AA621Rvf	atggtctagaaagctttaGAGGATCTGCAGCCGG
MASL1_AA1041Rvf	atggtctagaaagctttaTTGGAACAGGGGCTGATCACAGTCGGCTGGGTGGAACAGGGGCTGATCAC
MASL1_AA1041Rv3ceGFP	cagaactccagtttTTGGAACAGGGGCTGATCACAGTCGGCTGGGTGGAACAGGGGCTGATCAC
<b>pOPIN3C-EGFP</b>	
<b>Forward</b>	<b>Primer</b>
OPPF21786FW	aggagataaccatgGACATTGAGGCACTGAACCTG
OPPF21787FW	aggagataaccatgCAGCCCCGGCTCAAGCT
OPPF21788FW	aggagataaccatgCCCTACGAGGTCTGCATGAA
OPPF21790FW	aggagataaccatgCTCTCCCCGTGTTGCC
OPPF21792FW	aggagataaccatgCGGGTGGGTTTGTGGAAGA
MASL1_AA1052Rv3CeGFP	CTGGTTCTGTGCTTTTACCAACATCTCTC <b>TTGGAACAGGGGCTGATCAC</b>
<b>Reverse</b>	<b>Primer</b>
OPPF21783Rev	cagaactccagtttCTGCACCGCCGGCT
OPPF21784Rev	cagaactccagtttCCGGGAGAGCTGGCCAA
OPPF21785Rev	cagaactccagtttGAGGATCTGCACCGGTG
OPPF21793Rev	cagaactccagtttGGGGGGCTGGATCAGTG
OPPF21796Rev	cagaactccagtttCTGGTTTCTGTGCTTTTACCA
<b>pOPINE_Neo_GFP_2Strep_His</b>	
<b>Forward</b>	<b>Primer</b>
OPPF21625FW	aggagataaccatgGCTGGGATGGACAGTGGC
OPPF21633FW	aggagataaccatgGACATTGAGGCACTGAACC
OPPF21635FW	aggagataaccatgCAGCCCCGGCTCAAGC
OPPF21639FW	aggagataaccatgCAGCCCCGGCTCAAGC
OPPF21641FW	aggagataaccatgCTCTCCCCGTGTTGCC
<b>Reverse</b>	<b>Primer</b>
OPPF21657RVshort	cagaactccagtttGGAACAGGGGCTGATCAC
OPPF21784Rev	cagaactccagtttCCGGGAGAGCTGGCCAA
OPPF21657RVshort	cagaactccagtttGGAACAGGGGCTGATCAC
OPPF21785Rev	cagaactccagtttGAGGATCTGCACCGGTG
OPPF21657RVshort	cagaactccagtttGGAACAGGGGCTGATCAC

**Table 2.1:** Forward and reverse primers designed for cloning MFHAS1 protein into pOPINF, pOPINeNeo, pOPIN3C-EGFP and pOPINE\_Neo\_GFP\_2Strep\_His vectors.

The Protein Production UK (PPUK) protocol was adapted for cloning procedures (Bird *et al.*, 2014), with all reactions carried out on ice.

Polymerase Chain Reaction (PCR) was performed with 2X phusion flash master mix from Thermo Fisher Scientific (F-548L). PCR products were incubated for 1 hour at 37 °C with Dpn1 enzyme (New England Biolabs) to remove the methylated DNA. PCR products were analysed using a 1.7% agarose gel (100 mL of Tris/Borate/EDTA (TBE) (1X), 1.7 g of Agarose, Invitrogen SYBR Safe DNA gel stain). DNA loading buffer (0.25 % w/v Bromophenol Blue in 30 %v/v Glycerol) and Hyperladder™ 1kb (BioLine/Meridian Bioscience) were used to run the gel.

The PCR product was purified using AMPure XP Magnetic Bead Purification (Bechman coulter) to separate small PCR products and PCR products larger than 100bp.

POPIN vectors were linearised before being used. 20µg of the vector was incubated for 3 hours at 37°C with NEB buffer 4 (10X New England Biolab), BSA (100X), restriction enzymes KpnI; HindIII (New England Biolabs) and water. After incubation, the solution was run through an agarose gel and then purified using a Macherey-Nagel PCR Purification column.

Purified PCR products were inserted into linearised pOPIN vectors using ClonExpress II One-step cloning kit. An aliquot of Stellar™ competent cells (Takara) was thawed on ice for 5-10 minutes. 1 µL of PCR product was introduced to the competent cells and incubated on ice for 30 minutes. The cells were heat shocked for 30 seconds at 42 °C then placed on ice for 2 minutes. 225 µL of Lysogeny Broth (LB) was then added to the cells. Cells were incubated at 37 °C for 1 to 1.5 hours. Transformed cells were plated on Lysogeny Broth Agar (LBA) plates containing carbenicillin 1:1000, X-gal 1:1000 and Isopropyl β-D-1-thiogalactopyranoside (IPTG) 1:1000 (Sigma). Two replicates were made for each construct, and plates were left overnight at 37 °C.

Successfully transformed cells grew as white colonies. Three white colonies were picked for each experiment and placed into a 96 deep-wells plate with 1.2 mL of LB and carbenicillin 1:1000. They were then incubated overnight at 37 °C using a 200 rpm shaker. The plate was then centrifuged for 30 minutes at 2500 g. DNA purification was performed on the pellet using a miniprep kit (Qiagen). Constructs were verified by PCR reactions using pOPIN forward primers (gac cga aat taa tac gac tca cta tag gg) and the reverse primers corresponding to the construct (see **Table 2.1**). PCR products were analysed with 1.7%

agarose gel to confirm their size and sent to Source Biosciences for Sanger sequencing to confirm the length and fidelity of the cloned insert.

### **2.2.3 Cell culture and protein expression**

Two cell lines were used to express the proteins: Mammalian cells Expi293 (García-Fruitós, 2015) and Insect cells *Spodoptera Frugiperda* (SF9) (Invitrogen) (Vaughn *et al.*, 1977).

Expi293 were maintained at 37 °C with 5 % CO<sub>2</sub> and 120 rpm agitation and passaged twice a week at 4.5x10<sup>5</sup> cell/mL in Expi293 media (Gibco) with 1 % v/v supplement of Penicillin/Streptomycin and 10 % v/v supplement of foetal bovine serum (FBS) (Gibco). One day before transfection, cells were split at 1x10<sup>6</sup> cells/mL and plated in 12 wells plate with 1 mL per well. The plate was incubated overnight at 37 °C with 5 % CO<sub>2</sub> and 120 rpm agitation. 1 µg of DNA was incubated with 100 µL of optiMEM and 5.5 µL of Polyethylenimine (PEI) Max 40 kDa (Polysciences) for 10 minutes at room temperature (RT). The mix was then added to each well of the plate prepared the previous day. The plate was left overnight at 30 °C with 5 % CO<sub>2</sub> and 120 rpm agitation. The following day, enhancers (Sigma) were added to the cells. 16.5 µL valproic acid, 6.5 µL sodium propionate and 18.5 µL glucose were added per well. The plate was incubated for 48 hours at 30 °C with 5 % CO<sub>2</sub> and 120 rpm agitation.

As part of one experiment, cells treatment using Cytochalasin-D and Nocodazole (Thermo Fisher Scientific) was carried out in the cells. Cytochalasin-D and Nocodazole were added to the media 90 minutes prior to collecting the cells. Nocodazole and Cytochalasin-D were added to disrupt the microtubules and the actin filaments respectively.

SF9 cells were maintained at 27 °C at 200 rpm agitation and split every three days at 0.5 x 10<sup>6</sup> cell/mL in SF9 III media (Gibco) with 1 % v/v supplement of Penicillin/Streptomycin.

For transfection, SF9 cells were split at 1 x 10<sup>6</sup> cells/mL and 500 µL was added to each well of a 24 well plate, then left at RT for 1 hour for the cells to attach. The transfection mix was prepared with 2.5 µL of Bacmid, 1 µg of plasmid, 50 µL of SF9 III media and 1.5 µL of FuGeneHD (Promega), then left 10 minutes at RT before being added to the cells. To maintain humidity, the transfection plate was placed in a box with wet paper underneath and place in the incubator at 27°C for 7 days.



After 7 days, the supernatant (virus P0) was collected, and another plate was prepared as previously explained. 5  $\mu\text{L}$  of P0 was added to each well. The plate was left in the incubator at 27°C for 7 days. After 7 days, supernatant (virus P1) was collected. Cells were plated at  $1 \times 10^6$  cells/mL in a 24 deep-wells plate, 3  $\mu\text{L}$  or 30  $\mu\text{L}$  of P1 were added to the cells. Plates were left 3 days at 27 °C with 250 rpm shaking.

Tali® Image-Based Cytometer (T10796) was used to count the percentage of fluorescent cells after transfection.

#### **2.2.4 Protein extraction**

Expi293 and SF9 cells were collected in tubes and centrifuged 15 minutes at 6000 g at 4 °C or 30 minutes at 3000 g at 4 °C as part of the optimisation method. Supernatant was discarded, and pellets were stored at -80 °C or left 30 minutes at -80 °C before being thawed on ice. Pellets were resuspended in 400  $\mu\text{L}$  lysis buffer (for 10 mL: 9.1 mL H<sub>2</sub>O; 11.8 mg CHAPS; 500  $\mu\text{L}$  Hepes 1M; 300  $\mu\text{L}$  NaCl 5 M; 30  $\mu\text{L}$  Dithiothreitol (DTT) 1 M; 20  $\mu\text{L}$  Protease inhibitors; 1  $\mu\text{L}$  Benzonase) and incubated 1 hour at 4 °C with rotation before centrifugation at 3000 g for 30 minutes at 4 °C. Supernatants were collected, and pellets were resuspended in lysis buffers for analysis.

Proteins were analysed by Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Invitrogen™). Precision Plus Protein Dual Color Standards ladder (BIO-RAD) and BenchMark™ Fluorescent Protein Standard ladder (Invitrogen™) were used to calibrate gels. Loading buffer Laemmli was made by adding 10 mL of 1 M Tris HCl, 4 g SDS, 0.2 g Bromophenol blue and 20 mL glycerol. Water was added to obtain a final volume of 100 mL which was aliquoted in 9 mL aliquots. 1 mL of  $\beta$ -mercaptoethanol was added to the aliquot before use. Samples were not heat denatured to keep the GFP stable and visualise the GFP bands.

Gels were run 40 minutes at 200 V. The gel was either put in InstantBlue® Coomassie protein stain (ab119211) or imaged directly using a ChemiDoc™ imaging system BIORAD for the green fluorescent protein (GFP) fluorescence. For immunoblots, SDS-PAGE were transferred onto nitrocellulose membrane using iBlot™2 (Thermo Fisher Scientific). The membranes were blocked for 30 minutes in phosphate-buffered saline tween (1X phosphate-buffered saline (PBS) + 1 % Tween 20) + 5 % milk. Membranes were then

washed twice for 10 minutes in phosphate-buffered saline with Tween 20 (PBST) before incubation for 1 hour in 1:1000 anti-His antibody horseradish peroxidase (HRP) conjugate. The membranes were rewashed twice for 10 minutes in PBST. Pierce™ enhanced chemiluminescence (ECL) (Thermo Fisher Scientific) was prepared by mixing 1 mL of both reagents and was added to the membranes for 3 to 5 minutes before images were taken using ChemiDoc™.

### **2.2.5 Protein solubilisation and detergents screen**

A range of detergents (Sigma) were used for protein extraction and solubilisation. Detergents are described with abbreviations, full name, class, chemical formula, and structure (**Table 2.2**). Dithiothreitol (DTT) was also used at 0.5 mM and 3mM. A range of lysis buffers were tried Tris-Buffered Saline (TBS) (Sigma), 1 M Hepes, 5 M NaCl, protein inhibitor and benzonase. 2 mM guanosine diphosphate (GDP) or 2 mM guanosine triphosphate (GTP) (Sigma) were added to the lysis buffer. Benzonase (Sigma) was used to degrade DNA and RNA of the cells. Urea (Sigma) was used to test the solubility of the proteins.

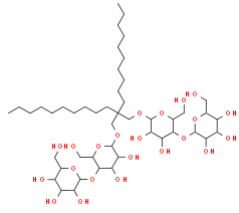
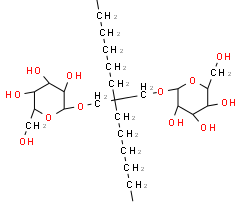
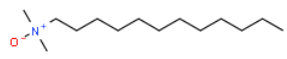
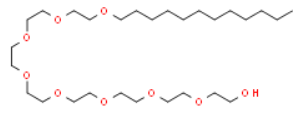
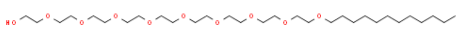
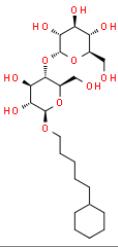
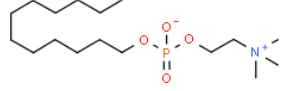
A high salt experiment was performed on a protein sample. Expi293 cells were collected in 50 mL falcon tubes and centrifuged 30 minutes at 3000 g at 4°C. Supernatants were removed, and pellets were placed 30 minutes at -80 °C and thawed before the experiment resumed. Cell pellets were resuspended in 26 mL of lysis buffer (20mM HEPES; 100mM NaCl, 10% v/v glycerol) and sonicated for 2 minutes (5 seconds on/10 seconds off) with 60 % amplitude on ice. Samples were centrifuged two times at 500 g for 15 minutes at 4 °C. Supernatants were kept. Cell lysis was divided into six tubes and centrifuged at 100000 g 15 minutes at 4 °C. Pellets were kept and resuspended in six buffers or detergents with protease inhibitors: Tris (20 mM Tris-Cl pH 7), NaCl (20 mM Tris-Cl pH 7; 140 mM NaCl), Tris NaCl (20 mM Tris-Cl pH 7; 1 M NaCl), Na<sub>2</sub>CO<sub>3</sub> (100 mM Na<sub>2</sub>CO<sub>3</sub> pH 11), DDM 0.1 % and UREA 8 M. Resuspended pellets were left 2 hours with rotation at 4 °C and were centrifuged at 100000 g 15 minutes at 4 °C. Supernatants and pellets were kept and analysed by SDS-PAGE.

The detergent screen protocol was developed at the Membrane Protein Laboratory (MPL) and used here. Stored cell pellets at -80 °C were thawed and used. Lysis buffer (20 mM

HEPES pH 8; 500 mM NaCl; 10 % glycerol. Protease inhibitor added just before use) was added onto the pellets and then sonicated 2 minutes (5 seconds on/ 10 seconds off) with 60 % amplitude on ice. Detergents were added to a final concentration of 0.1 %, 0.5 % or 1 % v/v. Twelve detergent conditions were used in the membrane protein screen; DDM, DDM + CHS, DM, DM + CHS, OG, LMNG, OGNG + CHS, LDAO, C12E8, C12E9, Cymal-5, Fos-choline-12 (detailed in **Table 2.2**). The tubes were incubated 1 hour at 4 °C with 450 rpm agitation before centrifugation at 3000 g for 30 minutes at 4 °C. Supernatants were collected for analysis and purification.

Fluorescent-Size exclusion chromatography (F-SEC) was used to analyse the quality of the extracted protein. F-SEC was performed on protein samples after the detergents screen using the Shimadzu HPLC. Samples were run with filtered buffer (20 mM Tris-HCl; 150 mM NaCl).

Abbreviations	Full name	Class	Structure
<b>ASB-14</b>	Amidosulfobetaine-14	zwitterionic	
<b>Igepal</b>	Polyethylene glycol octylphenol ether	non-ionic	
<b>CHAPS</b>	3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate	zwitterionic	
<b>Triton 100X</b>	Polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether	non-ionic	
<b>NP-40 (Tergitol)</b>	sodium tetradecyl sulfate	non-ionic	
<b>DDM</b>	n-Dodecyl β-d-maltoside	non-ionic	
<b>CHS</b>	Cholesteryl hemisuccinate	anionic	
<b>DM</b>	n-Decyl-β-d-Maltoside	non-ionic	
<b>OG</b>	n-Octyl-β-D-Glucopyranoside	non-ionic	

<b>LMNG</b>	Lauryl Maltose Neopentyl Glycol	amphiphilic	
<b>OGNG</b>	Octyl Glucose Neopentyl Glycol		
<b>LDAO</b>	Lauryldimethylamine oxide	zwitterionic	
<b>C12E8</b>	Dodecyl octaethylene glycol ether		
<b>C12E9</b>	Dodecyl nonaethylene glycol ether		
<b>Cymal-5</b>	Cyclohexyl-pentyl- maltoside		
<b>Fos choline-12</b>	Dodecylphosphocholine		

**Table 2.2:** Summary of detergents used for protein extraction and detergents screen. Detergents are described with abbreviations, full name, class, and structure. Detergents were purchased from Sigma. Images were taken from PubChem and MERCK websites.

## 2.2.6 Protein purification

Four methods of purification have been used. GFP tag purification was carried out with two distinct approaches; strep-tag Magstrep with antibody anti-GFP and Dynabeads™ magnetic beads with nanobodies anti-GFP were used. Then, Histidine (His) tag purification was done using Nickel<sup>2+</sup> Nitrilotriacetic Acid (Ni-NTA) beads. Finally, purification of construct with modified streptavidin binding tag (Strep2) using Strep-Tactin XT Superflow beads (iba) was also used. The protocols used are summarised in following sections.

### 2.2.6.1 Purification using a GFP MFHAS1 construct

Two protocols were used to purify MFHAS1 protein via a GFP tag. The first used strep-tag Magstrep beads, and the second used Dynabeads™ and anti-GFP nanobodies.

For the GFP purification using Streptag Magstrep beads, 20 µL of Streptag Magstrep beads (iba) were placed on a magnetic separator. The beads were resuspended with 200 µL of buffer W (100 mM Tris, 150 mM NaCl, 1 mM EDTA; pH 7.5) two times. 3.5 µL of biotinylated anti-GFP antibodies at 4.3 mg/mL resuspended in distilled water for a final volume of 15 µL were added to the beads and incubated for 1 hour at 4 °C. Beads were then washed two times with 100 µL of buffer W.

200 µL of clear protein lysate previously extracted was added to the beads and placed at 4 °C for 1 hour under agitation. Supernatant was removed by placing the plate on a magnetic separator. Beads were washed two times with 100 µL of buffer W. GFP tags were cleaved using 1 µL or 5 µL of Human Rhinovirus 3C protease (Libby *et al.*, 1988) in 100 µL of buffer W. Beads were left overnight at 4 °C under agitation. The plate was then put on a magnetic separator, and the eluate was collected.

An anti-GFP nanobody was used for the purification and cleavage of GFP tag using Dynabeads™. Anti-GFP nanobodies were made using pRARE2 BirA competent cells (provided by MPL) and following their protocol. An aliquot of pRARE2 BirA competent cells was thawed on ice for 5-10 minutes. 1 µL of GFP nanobody plasmid was mixed with the cell and incubated on ice for 30 minutes. The cells were heat-shocked for 30 seconds at 42 °C then placed on ice for 2 minutes. 225 µL of LB medium was added to the cells. Cells were then incubated at 37 °C for 1 to 1.5 hours. Transformed cells were spread on an agar plate

with 50 µg/mL of carbenicillin (CB); 34 µg/mL of chloramphenicol (Cm) and 50 µg/mL of streptomycin. The plate was then incubated overnight at 37 °C. A colony was picked and grown in LB with 10 µg/mL of CB; 10 µg/mL of Cm and 10 µg/mL of streptomycin overnight at 37 °C with 250 rpm shaking. Terrific Broth, 10 mL of glycerol, 20 mL biotin, CB, streptomycin, and the overnight culture were added and incubated at 37 °C with 250 rpm shaking. The optical density (OD) was measured every 30 minutes until it reached between 1.2 and 1.5. Once the OD was reached, the temperature was dropped to 18 °C and left to grow overnight. The next day, the culture was centrifuged 10 minutes at 4 °C at 6000 g. Supernatant was discarded, and the pellet was resuspended in 15 mL of TES then left in the cold room for 1 hour with stirring. 30 mL/L of ¼ TES was added and left 45 minutes. The solution was centrifuged at 20000 g at 4 °C for 10 minutes. The supernatant (SN) was retained, 6 mL of NaCl and 4 mL of glycerol were added. The solution was filtered using vacuum and purified with the automated Äktaexpress (GE Healthcare) using a His tag column. The fractions containing the nanobodies were concentrated using PD-10 desalting column (Sigma) and then stored at 4 °C.

For the second method of GFP tag purification, 25 µL of Dynabeads™ were transferred into a tube and placed onto a magnetic tube holder to separate the beads to the solution. The solution was discarded. 2.5 µL of the nanobody previously made was diluted in 100 µL PBST and added to the beads. The beads - nanobody solution was incubated 10 minutes at room temperature, then the tube was placed on a magnetic tube holder and SN was discarded. Magnetic beads were resuspended in 200 µL of PBST, SN was discarded. 1 to 2 mL of protein sample previously prepared was added to the beads and incubated 45 minutes at 30 °C with 100 rpm agitation. SN was removed and kept for further analysis. Beads were washed with 10 mM Tris; 150 mM NaCl. 2 µL 3C protease in 200 µL 10 mM Tris; 150 mM NaCl was added to the beads and left overnight at 4 °C. Supernatant was collected and NaOH was added to the beads to detach the nanobody.

#### **2.2.6.2 Purification using nickel affinity chromatography**

Nickel- Nitriloacetic acid (Ni-NTA) coupled to beads has been used to purified 6xHis tagged protein. 6xHis tagged protein were bound to the Ni-NTA column protein and then eluted using imidazole solution.

1 mL of Ni-NTA beads in solution were pipetted in a 15 mL falcon tube. Beads were washed 3 times with water then, 3 times with washing buffer by centrifuging 1 minutes at 500 g. Protein lysate was added to the beads then incubated 1 hour at 4 °C under rotation. Proteins and beads were then transferred to a gravity flow column. Beads were washed 2 times with 14 mL of washing buffer (100 mM Tris HCl; 150 mM NaCl; 20 mM imidazole). Proteins were eluted by adding 200 µL of elution buffer (100 mM Tris HCl; 150 mM NaCl; 300 mM imidazole). At each step, samples were taken and then analysed by SDS-PAGE as described above. Gels were imaged for GFP fluorescence or incubated in Coomassie blue protein stain for 30 minutes before imaging.

### ***2.2.6.3 Purification using Strep-tag II MFHAS1 construct***

The Strep-tag II was used for its high binding selectivity to Strep-Tactin. For the Strep-tag II purification, Strept-Tactin XT purification manufacturer protocol was followed. Beads were added to a chromatography column and equilibrated by adding two column bed volume (CV) of buffer W (100 mM Tris-HCl pH 8; 150 mM NaCl). Previously made clarified cell lysate were added to the column. Flow-through were collected for each fraction. Column was washed five times with one CV of buffer W. Proteins were eluted by adding 3 CV of buffer BXT (100 mM Tris-HCl pH 8; 150 mM NaCl; 50 mM biotin). At each step, the fractions were taken and then analysed by SDS-PAGE. Gels were imaged for GFP fluorescence or left in Coomassie blue for 30 minutes before imaging.



## 2.3 Results

### 2.3.1 Boundary delineation

The domain boundaries of MFHAS1 are defined using two methods, the secondary structure prediction from Phyre2 (Kelley *et al.*, 2015) and sequence alignment software's Clustal Omega and EMBOSS Needle (Madeira *et al.*, 2022).

There are 13 predicted Leucine Rich Repeat (LRR) motifs in the LRR domain of MFHAS1 shown on the structure prediction Phyre2 (**S.Figure 3**). Each are 20 to 29 amino acid (AA) long and comprise the conserved motif LxxLxLxxN/CxL (Kobe and Kajava, 2001). LRR domains have been described by A. Ng (Ng *et al.*, 2011). Based on the predicted LRR motifs of the Leucine Rich Repeat domain, the AA D64 and R366 were chosen for the start and end of the domain.

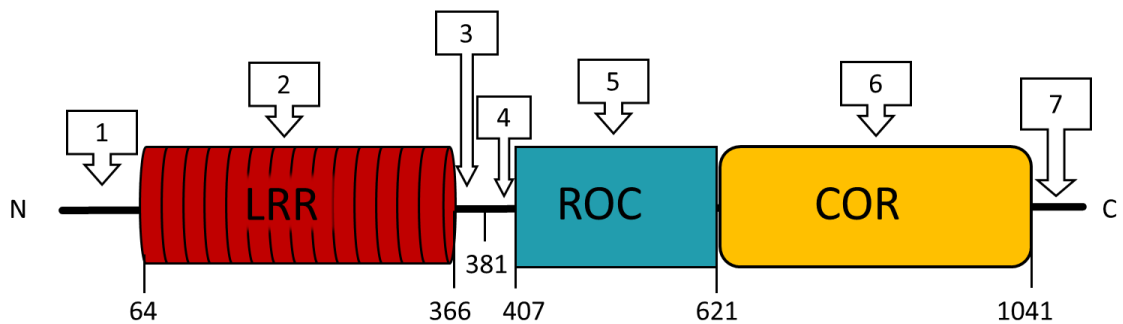
The MFHAS1 protein coding sequence was aligned against CtROCO from *C. tepidum* (Gotthardt *et al.*, 2008), LRRK2 protein (Deng *et al.*, 2008; Mills *et al.*, 2018) and three human Ras proteins (Human KRas, Uniprot P01116; Human HRas, Uniprot P01112 and Human NRas P01111) (**Figure 2.1**) (Bosgraaf and Van Haastert, 2003) using Clustal Omega and EMBOSS Needle.

After analysing the sequence alignment, two positions were determined for the start of the Roc domain: P381 and Q407. For the end of the Roc domain and the start of the COR domain L621 was identified based upon the Phyre2 analysis revealing a disordered region to the N terminus of Q619 that could have implication for the stable production of the domain protein fragment. For the end of the COR domain, it was decided to cut the protein sequence at Q1052. MFHAS1 DNA sequence purchased for this project ends at S1041 resulting in an eleven AA truncation at the C terminus of the protein.

Final boundaries can be found in **Figure 2.2**.

MFHAS1	KGIP----YIAAYQKELAH----SQPAVQPRKLLLMGHKAAGKTLLRHCLTEERVEG	436
LRRK2	KHIGCKAKDIIRFL-----QQRLLKAVPYNRMKLMIVGNTGSGKTTLLQQLMKTKS-	1360
CtROCO	--IVKQKQKAVRQYFQSIIEEARSKEALVHLQEIKVHLIGDMAGKTSLLKQLIGETFQ-	478
HRas	-----MTEYKLVVVGAGGVGKSALTIQLIQNHFV-	29
KRas	-----MTEYKLVVVGAGGVGKSALTIQLIQNHFV-	29
NRas	-----MTEYKLVVVGAGGVGKSALTIQLIQNHFV-	29
	. * : : * ** : * *	
MFHAS1	CPGGGDKKCYPPSPVSKGIEVTSWTA-----DASRGLRFIVYDLAGDESEYVIQP	489
LRRK2	-----DLGMQSATVGDVWDPI---QIRDKRKRDLVNVWDFAGREEFYSTHP	1406
CtROCO	-----PKESQTHGLNVTKQAPNIKLENDDELKECLFHFWDFGGQIMHASHQ	527
HRas	-----DEYDPTIEDSY--R-----KQVWIDGETCLLDILDITAGQEEYSAMRD	69
KRas	-----DEYDPTIEDSY--R-----KQVWIDGETCLLDILDITAGQEEYSAMRD	69
NRas	-----DEYDPTIEDSY--R-----KQVWIDGETCLLDILDITAGQEEYSAMRD	69
	. : . * * * :	
MFHAS1	FFLSPGALYVLVNLATYEPHFPTTVGSFLHRVGARV--PHAVVICVGTADLCGEREL	547
LRRK2	HFMTQRALYLAVYDLSKQQA--EVDAMKPWLFNIKARA--SSSPVILVGTDLVDSDEKQR	1462
CtROCO	FFMTRSSVYMLLLDSR---T--D-SNKHYWLRHIEKYG--GKSPVIVMINKIDENPSYNI	579
HRas	QYMRTEGFLCVFAINNTKS--F-EDIHQYREQIKRVKDSDDVPMVLVGNKCDL-AARTV	125
KRas	QYMRTEGFLCVFAINNTKS--F-EDIHQYREQIKRVKDSDDVPMVLVGNKCDL-PSRTV	125
NRas	QYMRTEGFLCVFAINNTKS--F-ADINLYREQIKRVKDSDDVPMVLVGNKCDL-PTRTV	125
	:: : : : : : : : : : * * * :	
MFHAS1	EKCLDIHRQIALQEKHDAEGLSRLAKVDEALARDFELRSASPHAAYYGVSDKNLRRRK	607
LRRK2	KACMSKIT---KELLNK-----RGFPAIRDYHFN	1489
CtROCO	---EQKKI---NERFPA-----IENRFHR---I-----	598
HRas	---ESRQA---QDLARS-----YGIPYIE---T-----	144
KRas	---DTKQA---QDLARS-----YGIPYIE---T-----	144
NRas	---DTKQA---HELAKS-----YGIPYIE---T-----	144
	:: : .	
MFHAS1	AHFQYLLNHRLQILSPVLPVSCRDPRLRRLRDKLLSVAEHREIFP--NLHRVLPRSWQV	665
LRRK2	-----ATEESDALAKLRKTIINESLNFKIRDQLVVGQLIPDCYVE	1529
CtROCO	-----SCKNGDGVESIAKLSKSAVL---HPDSIYGTPLAPSWIK	634
HRas	-----SAKTRQGVEDAFYTLVRE-----	162
KRas	-----SAKTRQGVEDAFYTLVRE-----	162
NRas	-----SAKTRQGVEDAFYTLVRE-----	162
	: . : :	
MFHAS1	LEELHFQPPQAQRLWLSWMSARLG----LQAGLTEDRLQSALSYLHESGKLLYFEDSP	720
LRRK2	LEKIILSERKNVPIEFVVIDRKRLLQLVRENQLDENELPHAVHFNESGVLVLFHQDPA	1589
CtROCO	VKEKLVEAT---TAQRYLNRTEVEKICN-DSGITDPGERKTLGLYLNMLGIVLYFE--A	687
HRas	-----IRQHLRKLN--PPDESGPGCMSC-----K-CVLS----	189
KRas	-----IRQYRLKKIS--KEE-KTPGCVKI-----KKCIIM-----	189
NRas	-----IRQYRMKKN--SSDDGTQGCMLG-----P-CVVM-----	189
	:: : . : :	

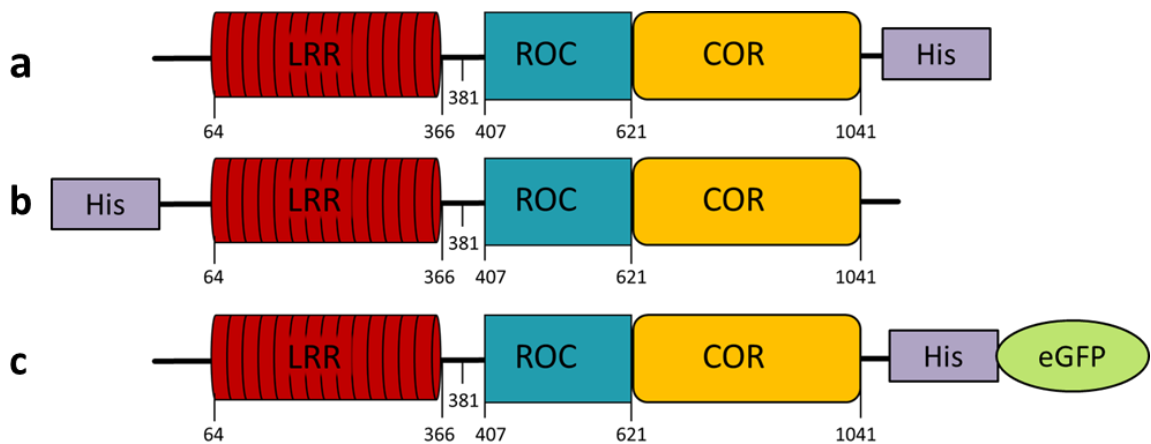
**Figure 2.1:** Clustal Omega of MFHAS1, LRRK2, CtROCO, KRas, NRas and HRas proteins (Uniprot Q5S007, Q8KC98, P01116; P01111; P01112;)



**Figure 2.2:** Shema of Malignant Fibrous Histiocytomas Amplified Sequence 1 protein domains and delineation of each subpart. Leucine Rich Repeat (LRR) domain boundaries are 64 to 366, Ras of Complex (ROC) domain boundaries are 407 to 621 and C terminal of ROC (COR) domain boundaries are 621 to 1041. Numbers on the top of the schema represent the different truncation used for protein expression.

### 2.3.2 Protein expression using vectors containing His and GFP tags

To investigate the expression of the full open reading frame and domain fragments of MFHAS1, a series of 35 constructs were designed and assembled using pOPINF, pOPINeNeo, and pOPINE-3C\_eGFP vectors (**Figure 2.3**). Fifteen constructs were made using pOPINeNeo, seventeen using pOPINF vector, and two using pOPINE-3C\_eGFP vector (**Table 2.3**). Fragments and full-length are produced to compare the expression and toxicity of the domain in the cell. The premise was also to ensure identification of a well performing construct. All three vectors have a His Tag sequence to facilitate purification using nickel columns and to target the His tag using an anti-His antibody. The pOPINE-3CeGFP vector has both a His and GFP tag. All the constructs with the C terminal AA at 1041 were designed using a long primer, which comprises the missing 11AA at the C terminal of MFHAS1. Constructs 33, 34 and 35 were designed with a short primer, meaning that there is 11 AA missing on the MFHAS1 protein. Out of those 35 constructs, 12 were successfully constructed and validated by sequencing (12, 13, 15, 19, 20, 22, 25, 26, 28, 29, 32, 35), highlighted in green in **Table 2.3**.



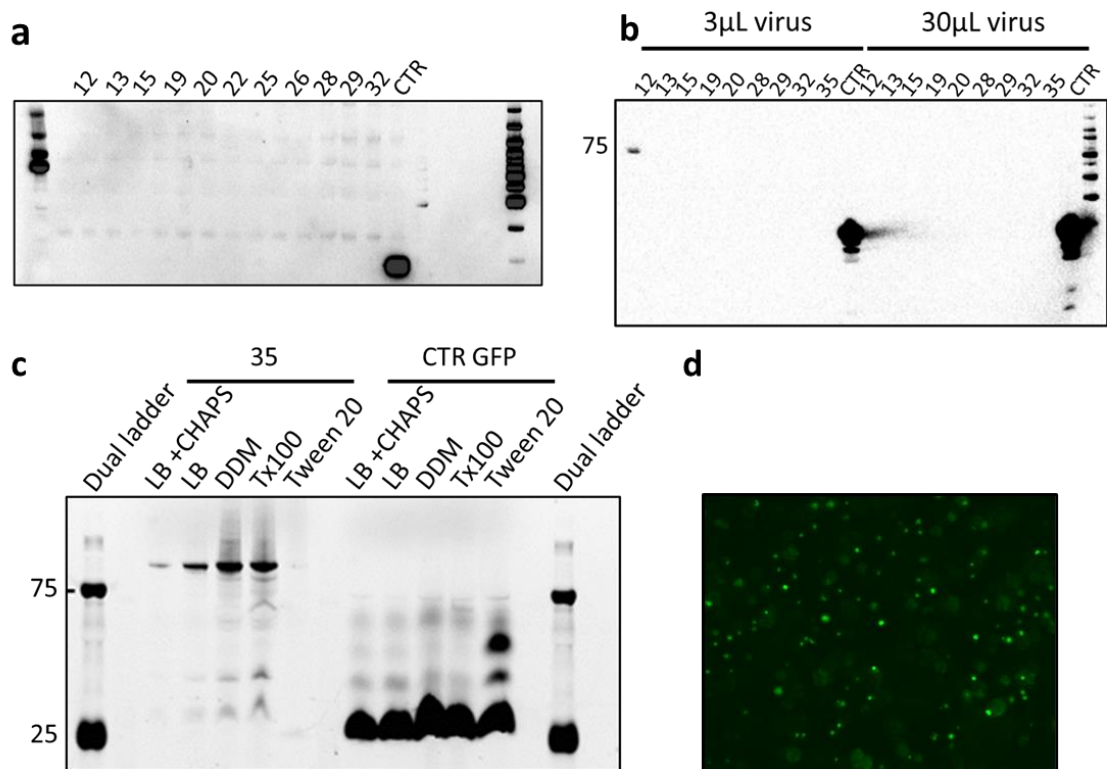
**Figure 2.3:** Schematics of MFHAS1 protein with tags. (a) MFHAS1 expressed with pOPINeNeo vector present one His tag on the C terminal (b) MFHAS1 expressed with pOPINF vector present one His tag on the N terminal (c) MFHAS1 expressed with pOPINE-3C\_eGFP vector present one His tag and one GFP tag on the C terminal

	Gene name	aa_N	aa_C	Vector	Tag	Domains	Molecular weight (kDa)
1	MFHAS1	2	1041	popinEneo	His	FL	115.52
2	MFHAS1	2	407	popinEneo	His	1_4	44.2
3	MFHAS1	2	366	popinEneo	His	1+2	39.6
4	MFHAS1	2	621	popinEneo	His	1_5	68
5	MFHAS1	64	366	popinEneo	His	LRR	33.2
6	MFHAS1	407	1041	popinEneo	His	Roc-COR	71.48
7	MFHAS1	381	1041	popinEneo	His	4_6	74.37
8	MFHAS1	407	621	popinEneo	His	Roc	24.02
9	MFHAS1	621	1041	popinEneo	His	COR	47.59
10	MFHAS1	64	1041	popinEneo	His	2_6	109.18
11	MFHAS1	366	1041	popinEneo	His	3_6	76.13
12	MFHAS1	2	381	popinEneo	His	1_3	41.27
13	MFHAS1	381	621	popinEneo	His	4_5	27
14	MFHAS1	64	621	popinEneo	His	2_5	61.72
15	MFHAS1	366	621	popinEneo	His	3_5	28.67
16	MFHAS1	2	1041	popine 3C eGFP	His GFP	FL	115.52
17	MFHAS1	2	1041	popinF	His	FL	115.52
18	MFHAS1	2	407	popinF	His	1_4	44.19
19	MFHAS1	2	366	popinF	His	1+2	39.57
20	MFHAS1	2	621	popinF	His	1_5	68
21	MFHAS1	64	366	popinF	His	LRR	33.2
22	MFHAS1	407	1041	popinF	His	Roc-COR	71.48
23	MFHAS1	381	1041	popinF	His	4_6	74.37
24	MFHAS1	407	621	popinF	His	Roc	24.02
25	MFHAS1	621	1041	popinF	His	COR	47.59
26	MFHAS1	64	1041	popinF	His	2_6	109.18
27	MFHAS1	366	1041	popinF	His	3_6	76.13
28	MFHAS1	2	381	popinF	His	1_3	41.27
29	MFHAS1	381	621	popinF	His	4_5	27
30	MFHAS1	64	621	popinF	His	2_5	61.72
31	MFHAS1	366	621	popinF	His	3_5	28.67
32	MFHAS1	2	1041	popinF	His	FL	115.52
33	MFHAS1	2	1041	popinF	His	FL	115.52
34	MFHAS1	2	1041	popinEneo	His	FL	115.52
35	MFHAS1	2	1041	popinE 3C eGFP	His GFP	FL	115.52

**Table 2.3:** Summary of all the constructs made using vectors popinEneo, popinF and popinE3CeGFP. Constructs in red failed at the cloning step, constructs in green were cloned successfully and validated by sequencing.

All the verified constructs were transfected into mammalian and insect cells. MFHAS1 protein was extracted using the following lysis buffer: CHAPS; Hepes; NaCl; DTT; protease inhibitors, and benzonase. After immunoblot analysis probing the membrane with an anti-his antibody, no proteins were detected (**Figure 2.4.a and b**).

It was decided to test whether the inability to detect the His tag by immunoblot was due to low protein expression or limited protein solubility. A multiplex protein extraction protocol was applied using a range of lysis and protein extraction buffers. Full-length MFHAS1, construct 35, containing His and GFP tags, was used for further expression analysis. The GFP tag allows for verification of the transfection in the cell using a microscope (**Figure 2.4.d**) and in-gel fluorescence. GFP fluorescence was observed in the transfected cells two days after transfection. Based on previous protocols developed by the Oxford protein purification laboratory, CHAPS, DDM, Triton 100 and Tween-20 were initially used to extract the protein. The GFP fluorescence is visible in SDS-gel (**Figure 2.4.c**). As shown in panel (c), using DDM and Triton 100 detergents for extraction resulted in the maximum yield for MFHAS1 protein. His tag was not detected.



**Figure 2.4:** (a) and (b): Immunoblot analysis of transfected cells. Lysis buffer used to extract the protein: CHAPS; Hepes; NaCl; DTT; Protease inhibitors; Benzonase. Transferred on nitrocellulose membrane, detected with mouse anti-His antibody 1:10000. (a) Expi293 cells were transfected with constructs 12, 13, 15, 19, 20, 22, 25, 26, 28, 29, 32 and GFP control. No bands are present on the membrane for the MFHAS1 constructs, only GFP control is working (b) SF9 insect cells transfected with virus containing the constructs 12, 13, 15, 19, 20, 28, 29, 32, 35 and GFP control. 3 $\mu$ L or 30 $\mu$ L virus were added to the cells. No bands are present on the membrane for the MFHAS1 constructs apart from the GFP control. (c): SDS-page in gel GFP fluorescence looking at the extraction of construct 35 (full length MFHAS1) and GFP control using the following conditions: Lysis buffer with CHAPS; lysis buffer without CHAPS; DDM; Triton 100; Tween-20. Lysis buffer (LB): CHAPS; Hepes; NaCl; DTT; Protease inhibitors; Benzonase. (d): GFP fluorescence of Expi293 cells in suspension two days after transfection with MFHAS1 protein construct.

In this first part, vectors were transfected into both Insect and Mammalian cells to investigate whether MFHAS1 could be expressed at high levels in different cell systems. MFHAS1 constructs containing His tag were not detected (**Figure 2.4**). Sequencing of the construct shows that the His tag and the protein are present but cannot be detected by immunoblot. As the GFP fluorescence could be detected but not the His tag, all the constructs have been redesigned in pOPINE-3CeGFP vector, which contains a GFP and a His tag.

Although mammalian and insect cells are widely used for recombinant protein expression, MFHAS1 protein is found in humans. It was decided to focus on Expi293T cells for future experiments to benefits for correct folding and potential cofactors.

### 2.3.3 Protein expression using vectors containing GFP and His tags.

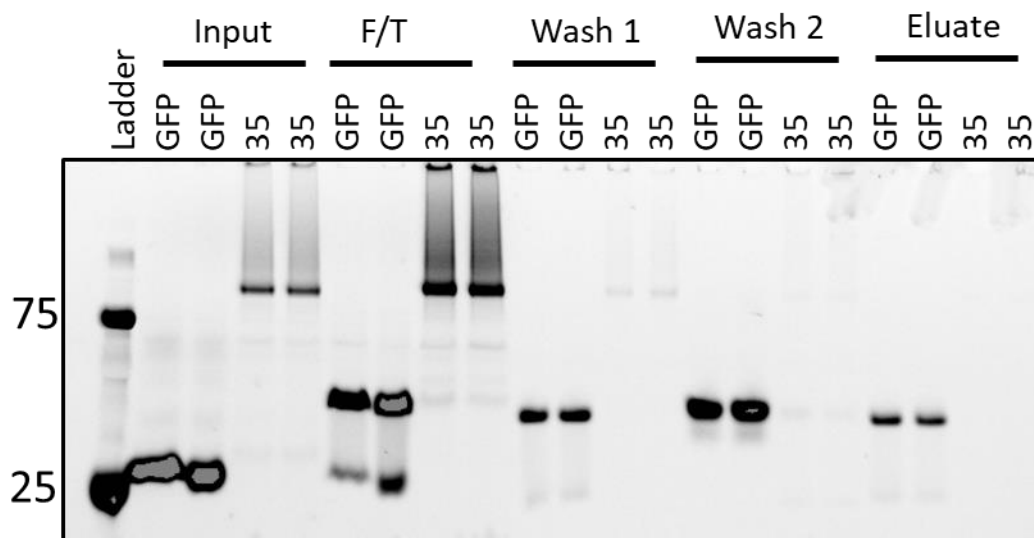
Based upon the previously described plasmid generation, a series of 20 constructs using the pOPINE\_3C\_eGFP vector with His and GFP tags were assembled. Twelve of these constructs were validated by direct sequencing (highlighted in green in **Table 2.4**). Constructs 49 to 54 were designed with an extended primer, adding the missing 11 AA to the C terminal of MFHAS1. These constructs were then transfected into mammalian cells (Expi293) to examine protein expression by in-gel fluorescence following lysis and extraction.

Number	MW protein	Gene name	aa_N	aa_C	Vector	Tag	Domain
35	115.52	MFHAS1	2	1041	pOPINE-3C-eGFP	GFP	FL short
36	44.2	MFHAS1	2	407	pOPINE-3C-eGFP	GFP	1_4
37	39.6	MFHAS1	2	366	pOPINE-3C-eGFP	GFP	1+2
38	68	MFHAS1	2	621	pOPINE-3C-eGFP	GFP	1_5
39	33.2	MFHAS1	64	366	pOPINE-3C-eGFP	GFP	LRR
40	71.48	MFHAS1	407	1041	pOPINE-3C-eGFP	GFP	Roc-COR
41	74.37	MFHAS1	381	1041	pOPINE-3C-eGFP	GFP	4_6
42	24.02	MFHAS1	407	621	pOPINE-3C-eGFP	GFP	Roc
43	47.59	MFHAS1	621	1041	pOPINE-3C-eGFP	GFP	COR
44	109.18	MFHAS1	64	1041	pOPINE-3C-eGFP	GFP	2_6
45	76.13	MFHAS1	366	1041	pOPINE-3C-eGFP	GFP	3_6
46	41.27	MFHAS1	2	381	pOPINE-3C-eGFP	GFP	1_3
47	27	MFHAS1	381	621	pOPINE-3C-eGFP	GFP	4_5
48	61.72	MFHAS1	64	621	pOPINE-3C-eGFP	GFP	2_5
49	116.97	MFHAS1	2	1052	pOPINE-3C-eGFP	GFP	FL*
50	72.8	MFHAS1	407	1052	pOPINE-3C-eGFP	GFP	5_7
51	75.69	MFHAS1	381	1052	pOPINE-3C-eGFP	GFP	4_7
52	48.91	MFHAS1	621	1052	pOPINE-3C-eGFP	GFP	6+7
53	110.5	MFHAS1	64	1052	pOPINE-3C-eGFP	GFP	2_7
54	77.45	MFHAS1	366	1052	pOPINE-3C-eGFP	GFP	3_7

**Table 2.4:** Summary MFHAS1 constructs made using vector popinE3CeGFP. Construct in red were not successfully cloned, construct in green were cloned and used for the experiments.

Cells transfected with MFHAS1 protein were lysed with previously used lysis buffer (DDM 1% v/v; HEPES; NaCl; DTT; Protease inhibitors; Benzonase). Purification of MFHAS1 protein using strep tag Magstrep and biotinylated anti-GFP antibody was performed. The protein

did not attach to the anti-GFP antibody, and the protein was collected in the flow through instead of the eluate (**Figure 2.5**). The control, (GFP protein) also loses interaction with the GFP antibody during each step.

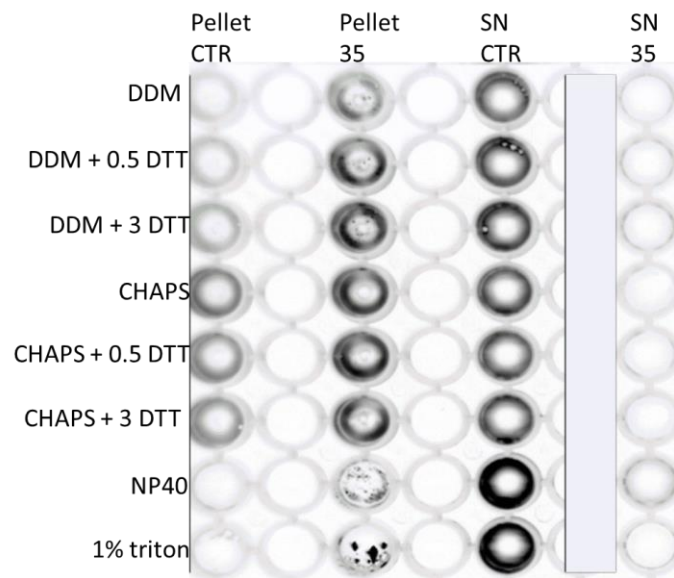


**Figure 2.5:** Purification of MFHAS1 protein construct 35 (full-length) using STREP TAG Magstrep. Cells transfected with GFP were used for the control. SDS-PAGE imaged for GFP fluorescence using ChemiDoc™.

To increase the protein-antibody interaction, lysis buffer and detergents were changed to a milder solution that would not break the bond between antibody and protein.

A series of experiments were carried out to identify the optimum protein extraction conditions using a range of conditions/detergents. To compare the solubilisation of the protein in different detergents, the supernatant (SN) and resuspended pellet of the same samples have been placed in parallel in a 96 well plates (**Figure 2.6**). The cells were resuspended in lysis buffer (1M HEPES, 5M NaCl, protein inhibitor and benzonase) and a combination of reagents and detergents at different concentrations. Dithiothreitol (DTT) at 0.5 mM and 3mM, CHAPS, NP-40 1% (also called Tergitol) and n-Dodecyl  $\beta$ -D-maltoside (DDM) 10% were used. All detergents are detailed in **Table 2.2**. **Figure 2.6** shows that for MFHAS1 full length construct (35), the majority of the fluorescence is present in the pellet while for the control (GFP protein), the fluorescence is in the SN. NP-40 1% displayed the most efficient solubilisation.

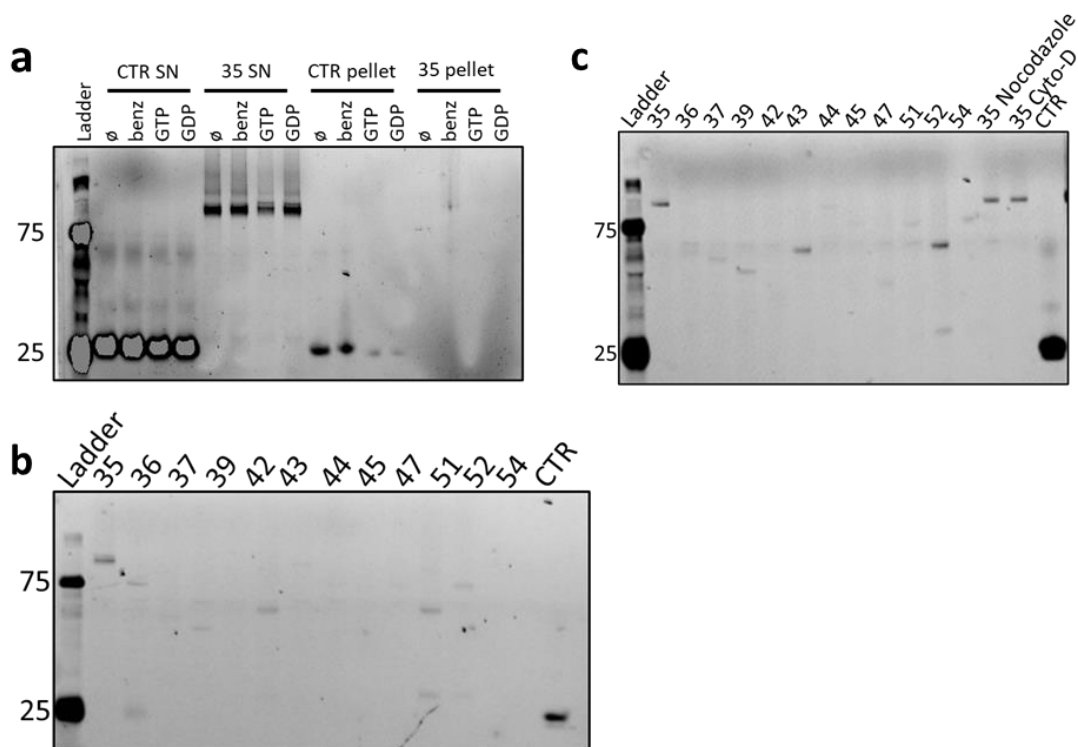




**Figure 2.6:** Resuspended pellet and supernatant (SN) of full length MFHAS1 (construct 35) and green fluorescent protein (GFP) control (CTR), display together on a 96 well plate. Picture taken for the GFP fluorescence. The reagents used are written on the columns on the left. For every condition, construct 35 shows more fluorescence in the pellet while the control GFP has more fluorescence in the SN. Grey rectangle was added to the figure after the image was taken.

Based upon data highlighting the importance of guanine nucleotide binding in ROCO protein function (Deng *et al.*, 2008; Dihanich *et al.*, 2014), guanosine diphosphate (GDP) and guanosine triphosphate (GTP) were added to the lysis buffer to test whether an excess of these nucleotides could improve the efficiency of the combination TBS/Tergitol (**Figure 2.7.a**). GTP and GDP do not improve the solubility of the protein. The twelve constructs successfully cloned (35, 36, 37, 39, 42, 43, 44, 45, 47, 51, 52 and 54) were used to test the following solubilisation methods. Mechanical action was used to break the cellular membrane by sonication and resuspension of the cell pellet with a needle (**Figure 2.7.b**). However, neither of these methods increased soluble protein yield. Cells were treated with Cytochalasin-D and Nocodazole. Cytochalasin-D inhibits the elongation of the actin filament, while Nocodazole stop the polymerisation of the microtubule. If MFHAS1 protein interacted with the cell's cytoskeleton as previously reported for LRRK2 (Meixner *et al.*, 2011), these treatments could increase protein availability for solubilisation. The expression of full-length MFHAS1 (35) after addition of Cytochalasin-D and Nocodazole is shown in **Figure 2.7.c**. There is no visible difference after treatment with Cytochalasin-D and Nocodazole. Lysis buffer containing ASB-14 was used for the construct 35 to 54 (**Figure 2.7.c**).

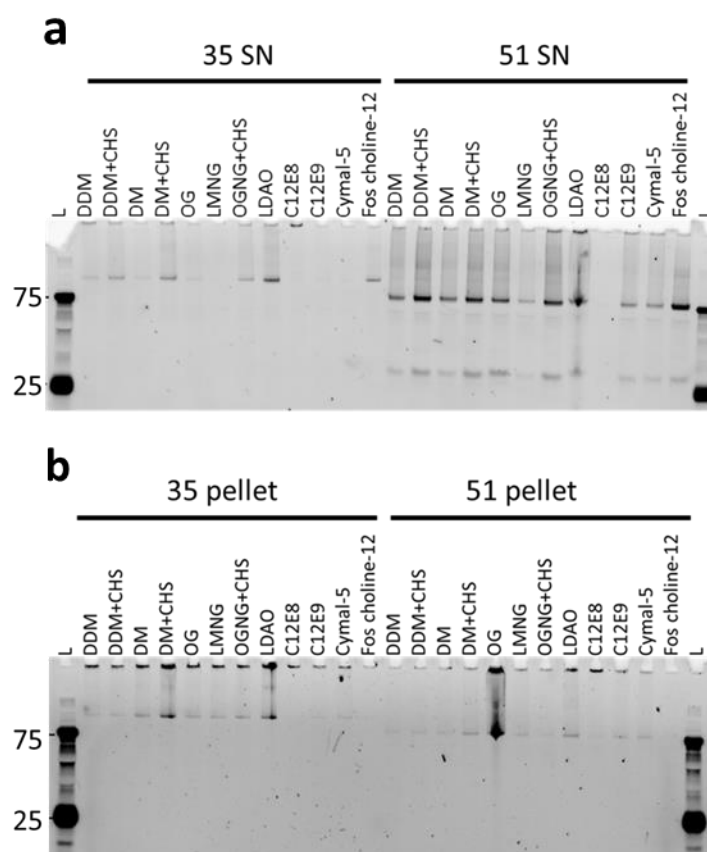
Different transfections were used for each condition, resulting in protein level variation.



**Figure 2.7:** GFP constructs of MFHAS1 transfected in Expi293 cells. In-gel fluorescence of GFP after an SDS-page run. **(a)** Lysis buffer containing TBS, 1 % NP-40. Benzonase, 2 mM guanosine diphosphate (GDP) or 2 mM guanosine triphosphate (GTP) were added to different protein sample. Full-length MFHAS1 (construct35) and control (GFP protein) were compared **(b)** Lysis buffer containing 1 % NP-40/TBS. The samples were passed through a needle when mixed with the Lysis buffer **(c)** Lysis buffer containing ASB-14 for the construct 35 to 54. Cytochalasin-D and Nocodazole were added to the cells 90 minutes before harvesting, the cells were then lysed with ABS-14.

MFHAS1 protein extraction is low, more detergents were tried for a better yield.

As studies show that LRRK2 associates with membrane and membranous area of the cells (Biskup *et al.*, 2006; Berger, Smith and Lavoie, 2010), a hypothesis that MFHAS1 could interact with the membrane was made. From this hypothesis, it was decided to work with the membrane protein laboratory (MPL) and use their protocol. Twelves detergents: DDM, DDM + CHS, DM, DM + CHS, OG, LMNG, OGNG + CHS, LDAO, C12E8, C12E9, Cymal-5, Foscholine-12 (described in **Table 2.2**) were used on twelves constructs (35, 36, 37, 39, 42, 43, 44, 45, 47, 51, 52 and 54). The experiment was repeated three times to assess the reproducibility of the results. SN and pellets band intensity were compared using in-gel GFP fluorescence (**Figure 2.8**).

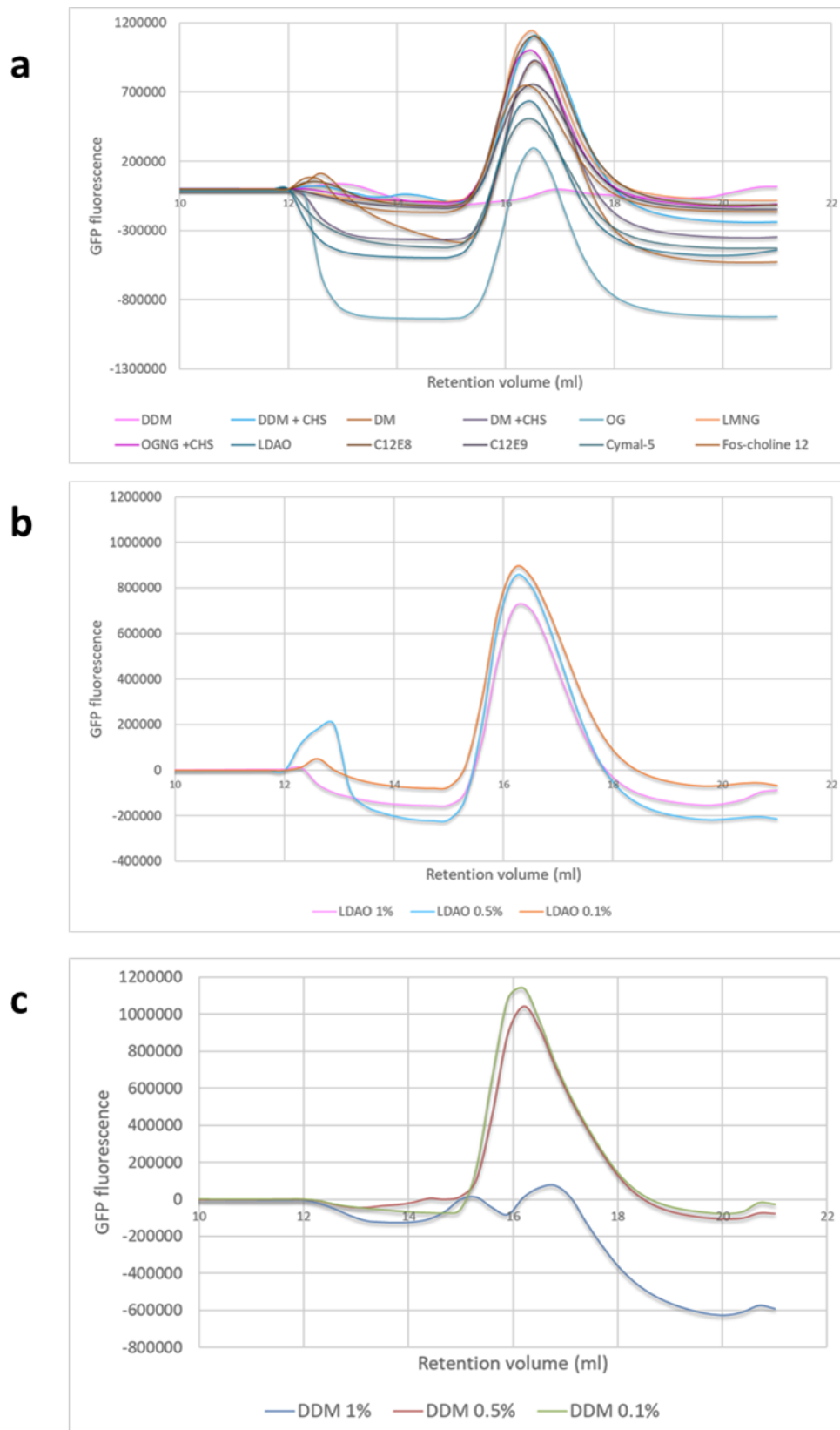


**Figure 2.8:** Detergent screen on full-length MFHAS1 (constructs 35) and the Roc-COR domain (construct 51). Both gels are imaged for in gel GFP fluorescence. Detergents used are DDM, DDM + CHS, DM, DM + CHS, OG, LMNG, OGNG + CHS, LDAO, C12E8, C12E9, Cymal-5 and Fos-choline-12. **(a):** SDS page gel of the supernatant (SN) of the proteins. **(b):** SDS page gel of the protein pellet.

A comparison of the protein in the pellet versus the protein in the SN shows that the best level of expression and extraction is observed when using DDM 0.1% and LDAO 0.1%.

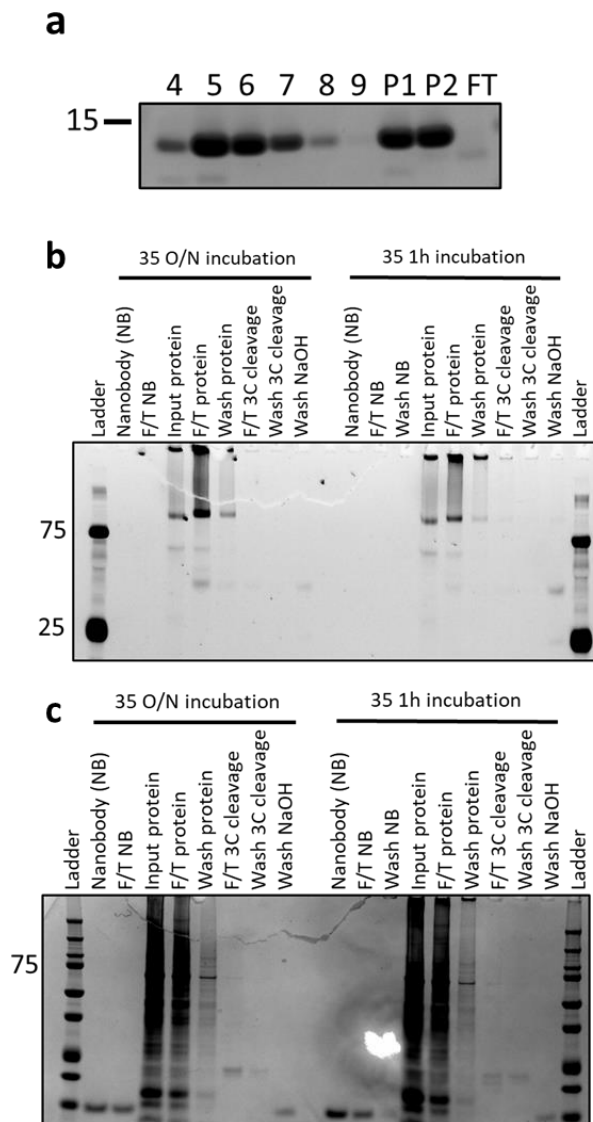
Fluorescent-Size exclusion chromatography (F-SEC) was performed on full-length MFHAS1 (construct 35) lysed with twelve different detergents as part of the detergent screen (**Figure 2.9**). F-SEC provides a size-dependent profile for extracted protein. If the protein is not degraded or aggregated, the profile will manifest as a curve with a distinct peak. In contrast, if the protein is degraded or aggregated, the SEC profile will present as an irregular profile. Thus, the protocol allows us to find the detergent which has the best profile and which concentration of the detergent works the best with regard to maintaining

the integrity of the extracted protein. Three concentrations of LDAO (**Figure 2.9.b**) and DDM (**Figure 2.9.c**) were assessed. LDAO 0.1 % has the highest peak curve out of the three concentrations of LDAO; however, none of the profiles show any major indication of protein degradation. DDM 0.1% shows the best profile, DDM 1% indicates a possible degradation or aggregation of the protein.



**Figure 2.9:** Fluorescent-Size exclusion chromatography (F-SEC) results on full-length MFHAS1. **(a)** Detergents used in the detergent screen. DDM; DDM + CHS; DM; DM +CHS; OG; LMNG; OGNG +CHS; LDAO; C12E8; C12E9; Cymal-5 and Fos-choline **(b)** LDAO detergent at 1%, 0.5% and 0.1% concentration. **(c)** Detergent DDM at 1%, 0.5% and 0.1%.

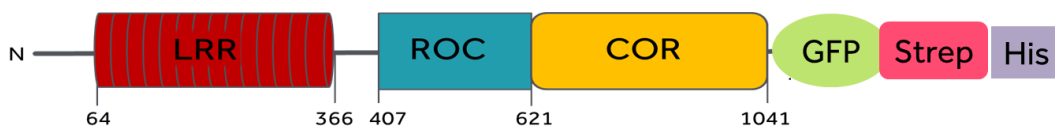
Purification of MFHAS1 protein using the GFP tag was attempted a second time with a different method. A nanobody anti-GFP was made (**Figure 2.10.a**) and conjugated with Dynabeads. Then the protein of interest was incubated with the nanobody-Dynabeads complex. Two incubation times were used, 1 hour or overnight. The majority of the protein is in the flow-through (**Figure 2.10.b and c**). This suggests that there is a low affinity interaction between the protein and the nanobody. The protein did not display a strong interaction with the nanobody. The link between the protein and the nanobody was not strong enough.



**Figure 2.10:** (a) Anti-GFP Nanobodies production. 4,5,6,7,8,9: purification fraction. P1 and P2: final nanobody after PD-10. FT: Fraction total. (b) and (c) Purification of MFHAS1 protein construct 35 (full-length) with Dynabeads and nanobody anti-GFP using 3C cleavage experiment with LDAO detergent. Comparing overnight (O/N) incubation versus 1 hour incubation. (b) SDS page gel imaged for GFP fluorescence (c) Coomassie blue gel. F/T: flow-through

### 2.3.4 Protein expression using vectors containing GFP, 2Strep and His tags.

As the previous approaches to detect His tag and purify MFHAS1 protein using GFP tag failed, it was decided to change vector and use pOPINeNeo\_GFP\_2StrepHis containing His, GFP and 2Strep tags (**Figure 2.11**). The GFP tag is helpful to see if the transfection works in the cells and protein detection in SDS-page gel. The 2strep and His tags give two chances to purify the protein. Either using Ni-NTA for the His tag or Strept-Tactin XT purification protocol with the 2Strep tag. The His tag is placed after the GFP and 2Strep tags. As the His tag is placed further away from the protein compared to the previous vector used, the His tag has a higher probability of being detectable.



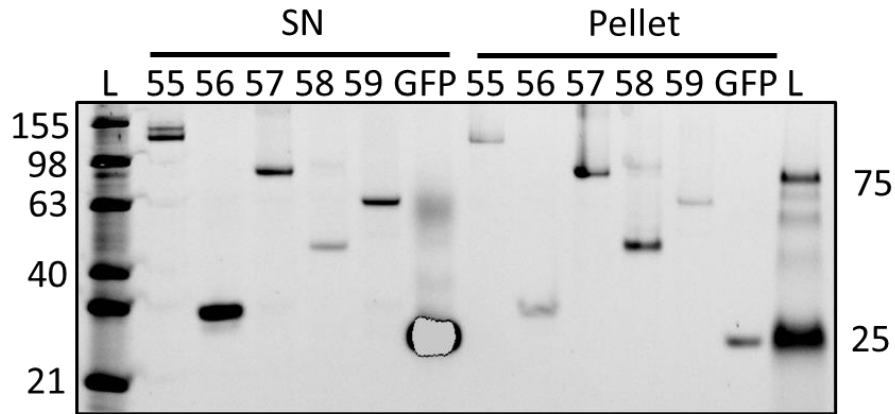
**Figure 2.11:** Schema of MFHAS1 protein and tags. Domain delineation is based on previous experiment. GFP tag is 28kDa, 2Strep tag is 10kDa and His tag is 2.5kDa.

Five constructs were made using the pOPINeNeo\_GFP\_2StrepHis vector (**Table 2.5**): a full length, the LRR domain, Roc-COR domain, Roc domain and COR domain. Constructs were successfully cloned and expressed in mammalian cells apart from construct 56 (LRR domain), which expressed at the wrong molecular weight. It should express around 60 kDa to account for the GFP tag.

Number	MW protein	Gene name	aa_N	aa_C	Vector	Tag	Domain
55	115.52	MFHAS1	2	1041	popinEneo-GFP-2strep-His	GFP-2Strep His	FL short
56	33	MFHAS1	64	366	popinEneo-GFP-2strep-His	GFP-2Strep His	LRR
57	71	MFHAS1	407	1041	popinEneo-GFP-2strep-His	GFP-2Strep His	Roc-COR
58	24	MFHAS1	407	621	popinEneo-GFP-2strep-His	GFP-2Strep His	Roc
59	47	MFHAS1	621	1041	popinEneo-GFP-2strep-His	GFP-2Strep His	COR

**Table 2.5:** Summary of the 5 constructs made with vector popinEneo\_GFP\_2StrepHis. FL: Full length; LRR: Leucine Rich Repeat; Roc: Ras Of Complex protein COR: C terminal Of Roc.

Proteins were extracted using lysis buffer (20mM HEPES pH8; 500mM NaCl; 10% glycerol, protease inhibitor added just before use) and sonicated for 2 minutes (5sec ON – 10sec OFF) 60% amplitude on ice and addition of 0.1% DDM (**Figure 2.12**). A comparison of the SN and pellets bands of the SDS-PAGE revealed a darker band for the SN, indicating higher protein solubility.



**Figure 2.12:** Picture of SDS-page gel looking at GFP fluorescence. Expi293 cells transfected with MFHAS1 pOPIN<sub>neo</sub>\_GFP\_2StrepHis construct. Construct 55, 56, 57, 58, 59 and GFP images of supernatant and pellet. GFP: control GFP; L: Benchmark fluorescent ladder. Lysis buffer: 20mM HEPES pH8; 500mM NaCl; 10% glycerol. Protease inhibitor added just before use. Sonication for 2 minutes (5sec ON – 10sec OFF) and addition of 0.1%DDM.

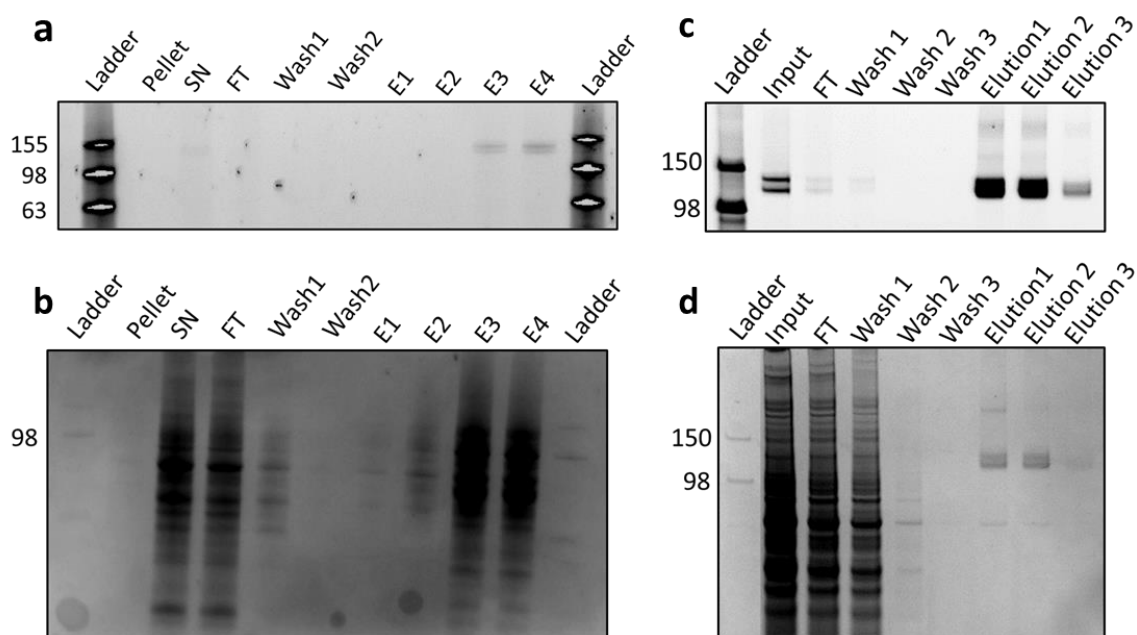
To purify MFHAS1 protein, a vector containing a modified streptavidin binding tag (Strep2) and a Histidine tag was chosen for this experiment. Streptavidin and Histidine tags are part of the main affinity tags used for protein purification (Kosobokova, Skrypnik and Kosorukov, 2016). Purifications were first attempted on the full-length MFHAS1 protein (construct 55).

Strep2 attached to Strep- Tactin matrix and elution is done with desthiobiotin. Purification using the His tag uses a matrix containing NiNTA and is eluted with imidazole (Kosobokova, Skrypnik and Kosorukov, 2016).

The purified protein is collected in the eluate. His purification using NiNTA beads was performed on full-length MFHAS1. In-gel fluorescence of the GFP tag revealed that the input protein is weak, but the protein is present in the eluate (**Figure 2.13.a**). The same gel was stained with Coomassie blue, revealing many unspecific bands in the eluate, which means that the purification was unsuccessful (**Figure 2.13.b**).

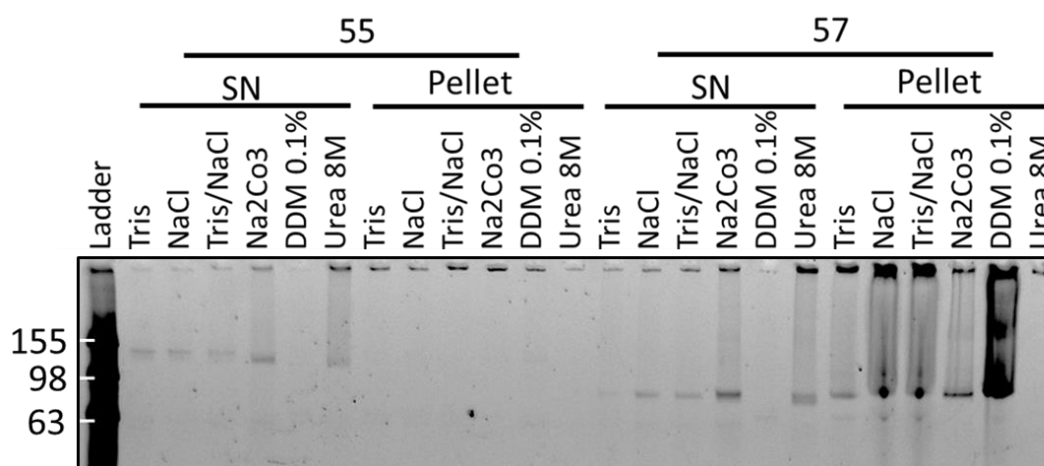


With the Streptavidin purification, the protein input is a clear double band. A small amount of protein goes in the flow-through, but most of the protein is collected in the eluate (**Figure 2.13.c**). The fluorescent band on the eluate is brighter than the input because the volume is smaller, so the protein is more concentrated. On the Coomassie gel of this experiment, the eluate shows one clear band of the size of MFHAS1 protein and two other faint bands (**Figure 2.13.d**). Strep-tag purification works well, and MFHAS1 full-length protein was purified in small quantities.



**Figure 2.13:** SDS-page gel of construct 55 (full-length MFHAS1) purifications. L: benchmark fluorescent ladder. F/T: flow through SN: Supernatant. **(a)** In gel GFP fluorescence of NiNTA (his tag) purification. **(b)** Coomassie blue gel of NiNTA (his tag). **(c)** In gel GFP fluorescence of Strep tactin purification. **(d)** Coomassie blue gel of Strep tactin purification.

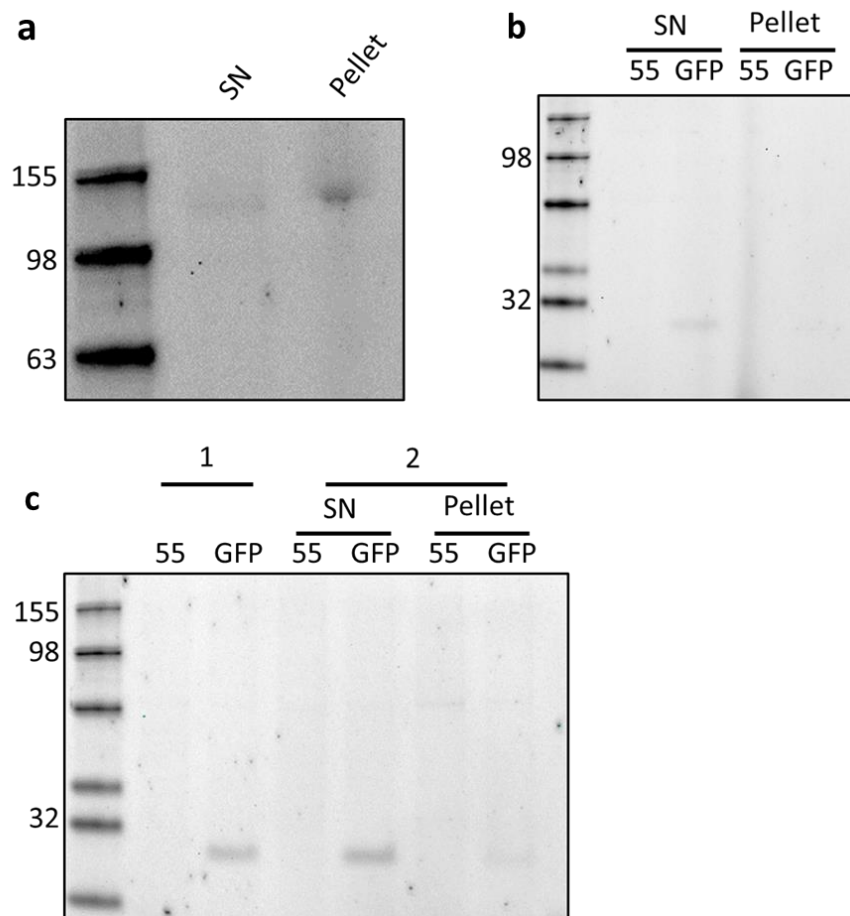
To improve protein purification, a high salt experiment has been performed on the full-length and the Roc-COR domains (55, 57 constructs) (**Figure 2.14**). The following detergents and buffers were used: Tris; NaCl; Tris/NaCl; Na<sub>2</sub>CO<sub>3</sub>; DDM 0.1 % and UREA 8 M. Urea, a strong protein denaturant (Nozaki and Tanford, 1963), was used to examine whether MFHAS1 protein would be denatured and extracted under these conditions. High salt can affect protein solubility (Sinha and Khare, 2014).



**Figure 2.14:** In gel GFP fluorescence of high salt experiment done on full length and Roc-COR domain of MFHAS1 (construct 55, 57). Buffers used: Tris; NaCl; Tris/NaCl; Na<sub>2</sub>CO<sub>3</sub>; DDM 0.1% and UREA 8M SN and pellets are shown for each condition.

At this point, experiments stopped for six months due to COVID. Upon return, the protein pellets left at -80 °C were lysed and run on an SDS-PAGE (**Figure 2.15.a**). The protein is primarily found in the pellet; this could be due to protein degradation.

Transfection in Expi293T cells was repeated in a different environment and growth system than previously used due to the unavailability of consumables. The in-gel fluorescence shows no protein expression for sample 55 and very low expression for the GFP control (**Figure 2.15.b**). This suggests a very low transfection percentage of the cells. Conditions were changed, and new flasks were tried for growth and protein expression, with the percentage of transfected cells (fluorescent green) read using Tali® Image-Based Cytometer. The Tali counts the total and fluorescent cells, then gives a percentage of fluorescent cells (transfected). Only transfections rate of over 70 % were kept for protein extraction. Different lysis conditions were also tried, filtering the lysate instead of centrifugation and using DMSO 0.5 % instead of DDM (**Figure 2.15.c**). None of the conditions used resulted in the expression and purification of soluble protein as previously demonstrated.



**Figure 2.15:** In gel GFP fluorescence of full-length MFHAS1 (construct 55) and control GFP (CTR). **(a)** Cells left in -80 °C for 6 months. SN and pellet of construct 55 cells lysed with 20 mM HEPES pH 8; 500 mM NaCl; 10 % glycerol, protease inhibitor added just before use **(b)** 55 and GFP constructs recently transfected. Cells lysed with 20 mM HEPES pH 8; 500 mM NaCl; 10 % glycerol, protease inhibitor added just before use) SN and pellet show almost no band. **(c)** 1:no centrifugation, filter 0.2 μm 2: DMSO 0.5 % instead of DDM.

## 2.4 Discussion

The work described in this chapter aimed to express and purify MFHAS1 protein in quantities and purity for structural studies via macromolecular crystallography and/or Cryo-Electron Microscopy. The expression and purification of soluble protein was ultimately successful, but due to time restraints caused by COVID restriction, structural studies were not initiated.

In this chapter, the domain delineation of MFHAS1 protein was determined using Phyre2 and sequence alignment (Kelley *et al.*, 2015; Madeira *et al.*, 2022). Determining the correct domain boundaries is essential and can impact protein expression (Nettleship, Rada and Owens, 2019). Expressing the whole protein and each domain allows us to compare the expression level of each domain. When the experiments were devised, Alphafold2 did not exist, and the LRRK2 protein structure had not been determined. MFHAS1 structure prediction was based on the *C. tepidum* ROCO protein. The domain boundaries can be reviewed now using Alphafold2 software (Jumper *et al.*, 2021; Varadi *et al.*, 2021). New boundaries will be discussed further in the Bioinformatic chapter.

The MFHAS1 DNA open reading frame sequence was cloned in four vectors with different tags. A high throughput cloning method was applied (Nettleship, Rada and Owens, 2019), with 59 constructs cloned in total. This high number of clones maximised the chances of success. It was essential to have at least one full-length MFHAS1 clone working.

Vectors were designed to allow expression in both insect and mammalian cells (Berrow *et al.*, 2007). Insect and mammalian expression systems are similar in their post-translational mechanisms like phosphorylation and glycosylation. Working with mammalian cells has the advantage of speed; it takes one week to express proteins in mammalian cells instead of one month in insect cells (McKenzie and Abbott, 2018).

MFHAS1 constructs were transfected in insect and mammalian cells. In the first constructs cloned in vectors containing His tag only or His and GFP tag, the His tag could not be

detected using anti-His antibody in immunoblot. Sequencing of the construct shows that the His tag and the protein are present. The His tag might not be detectable because MFHAS1 may form a dimer, as observed for other ROCO proteins such as LRRK2, LRRK1, DAPK1 and *Chlorobium tepidum* CtROCO (Gotthardt *et al.*, 2008; Klein *et al.*, 2009; Sejwal *et al.*, 2017; Mills *et al.*, 2018). The potential dimerisation may act to mask the His tag. It could also be due to the protein misfolding around the His epitope. To overcome this, vectors containing at least two tags were selected. Vectors containing GFP and His tags were used so if the His tag could not be detected, the purification could be done with the GFP tag.

An 11 AA truncation was present on most of the construct containing the C-terminal of MFHAS1. This truncation would give a slightly shorter structure and may impact the dimerisation of the MFHAS1 protein (Gotthardt *et al.*, 2008). This truncation may not be impactful for structural studies but may have further implications in functional studies.

The extraction of soluble protein was the most challenging step of this project. Many detergents and chemicals were tried before finding an optimal combination.

Several mechanisms can trigger protein insolubility. Proteins can aggregate and form aggresomes (García-Mata *et al.*, 1999). Leibly *et al.* hypothesised that the proteins are soluble inside the cells and could aggregate during cell lysis. They describe and analyse several detergents and additives that act to solubilise bacterially expressed proteins. The detergents were added to the lysis buffer to help the protein to keep its conformation and not to precipitate (Leibly *et al.*, 2012).

There are three main classes of detergents: ionic, non-ionic and Zwitterionic (García-Fruitós, 2015). Ionic detergents such as SDS act by partly denaturing secondary and tertiary protein structures and were not used in the experiments. Zwitterionic detergents have both a positive and negative charge, making them neutral compounds able to break the bounds between proteins without denaturing them. Non-ionic detergent can break lipid-protein interactions and keep an intact structure of the protein but cannot break protein-protein interaction. The action of these detergents is made possible by the presence of a hydrophilic head and a hydrophobic tail, allowing them to form a micelle around proteins, thereby maintaining solubility.

Overexpression of MFHAS1 protein was described to be cytotoxic in a previous study (Dihanich *et al.*, 2014). In this research, cells overexpressed with MFHAS1 were found to have a higher cell death. This data is not shown as it was not quantified. In comparison, overexpression of LRRK2 is also cytotoxic (Berwick and Harvey, 2011). The cytotoxicity of overexpressed MFHAS1 impacted the experiment; a higher volume of cells needed to be transfected in order to produce enough protein to work with.

MFHAS1 protein shares 24.5% identity with LRRK2 and around 20% with Ras proteins. LRRK2 was found to be associated with the membrane and to have an increased kinase activity when associated with the membrane (Berger, Smith and Lavoie, 2010).

LRRK2 is found to bind to microtubules, tubulin and filaments (Law *et al.*, 2014), unlike LRRK1 which does not bind to microtubule (Snead *et al.*, 2022). LRRK2 is also associated with membrane-bound organelles in human and rat brains (Biskup *et al.*, 2006). The Ras proteins have also been found to attach to cellular membranes (Gysin *et al.*, 2011). MFHAS1 ROC domain is described as a Ras protein with GTPase activity. The hypothesis that MFHAS1 is a membrane-associated protein can be drawn from these studies.

From this hypothesis, a protocol developed at the Membrane Protein Laboratory was used to find a detergent and lysis buffer to extract MFHAS1 protein. Performing a membrane protein detergent screen helped determine a suitable detergent for MFHAS1 protein solubilisation.

Unfortunately, after coming back from lockdown, the experiments were shown to be not reproducible. The environment and materials used were changed, resulting in decreased transfection efficiency and low to null protein solubility. After the transfection stopped working, transfection in insect cells could have been tried, as few papers on the structure of LRRK2 use insect cells for protein expression (Mills *et al.*, 2018; Deniston *et al.*, 2020). Cofactors like Heat shock protein 60 (Hsp60) or Hsp70 could have been used to help solubilise the protein (Dihanich *et al.*, 2014; Mokry, Abrahão and Ramos, 2015). However, time did not allow for extensive re-optimisation of techniques.

LRRK2 gene and protein have been studied for two decades before a structure was successfully determined. The LRRK2 gene was first found to be involved with familial Parkinson's disease in 2004 (Paisan-Rui *et al.*, 2004; Zimprich *et al.*, 2004). The first structural characterisation of the LRR domain of LRRK2 was achieved in 2012

(Vancraenenbroeck *et al.*, 2012). The structure of the C terminal of LRRK2 (domains ROC-COR-Kinase-WD40) was solved in 2020 (Deniston *et al.*, 2020). The full-length LRRK2 structure was elucidated by Cryo-EM in 2021 (Myasnikov *et al.*, 2021). These studies highlight the challenge of producing and extracting the ROCO proteins.

Myasnikov *et al.*, produced LRRK2 protein by transfecting HEK293F with recombinant baculoviruses of LRRK2; cells were lysed with 20 mM Tris pH 8.0, 200 mM NaCl, 10% glycerol, 2 mM DTT and protease inhibitors.

To date, there is no available experimentally determined structure for MFHAS1 protein. To take a different approach, the predicted structure of MFHAS1 by AlphaFold2 has been investigated using bioinformatics (Chapter 3). MFHAS1 protein interaction network and post-translational modifications will also be studied. MFHAS1 signalling pathways are analysed in Chapter 4 to understand its biological function better.

## Chapter 3: Bioinformatic analysis of MFHAS1

### 3.1 Introduction

Bioinformatic approaches applied at genomic and proteomic scale provide powerful tools to understanding the structure and function of genes and proteins. Large scale dataset from high-throughput (HTP) experiments are available in online publicly available repositories covering DNA, RNA and protein data. At protein level, data are mostly collected by tandem mass spectrometry (Keerthikumar and Mathivanan, 2017). Web based proteomic software allows access to processed data or make them more accessible for scientists to interpret and use. This huge quantity of data provides valuable information, especially for protein such as MFHAS1 which are not well studied. A cellular context can be provided by nomenclature like Gene Ontology (GO).

Bioinformatic approaches have also been developed to predict protein structures (Jisna and Jayaraj, 2021). Understanding three-dimensional structure is important to understand the function of specific proteins and develop drugs to modulate this function. Solving protein structure via crystallography or Electron microscopy can, however, be very technically challenging and time consuming, leading to a big gap between the protein sequence referenced in Uniprot and the solved structure (Bertoline *et al.*, 2023). A way to bypass this is to model protein structure using machine learning based algorithms such as Alphafold2 (AF2) and Intfold7 (Jumper *et al.*, 2021; McGuffin *et al.*, 2023). Protein modelling software compare all solved protein structures to a given protein sequence to model a structure that can be close to experiment data if similar protein sequences are available. Those predicted structures do not replace experimentally derived data but yield an insight into the proteins structure. Following Chapter 2 where MFHAS1 protein purification and hence derivation was unsuccessful, predicted protein structures from Alphafold2 and IntFold7 for MFHAS1 were used and analysed.

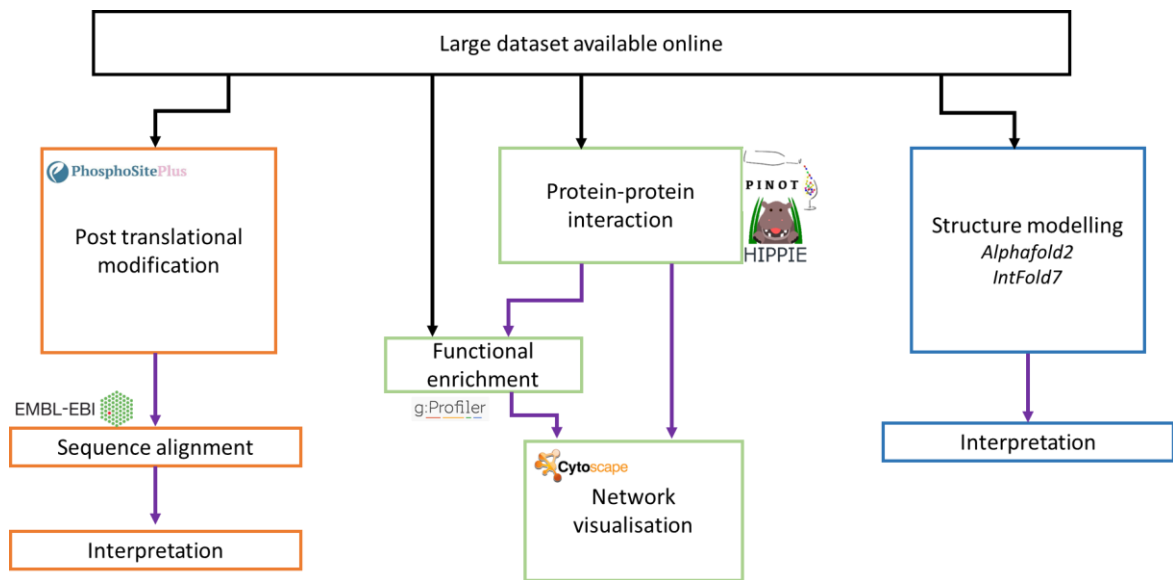
Post translational modifications (PTM) are essential for the regulation of protein function. Phosphorylation by kinases can activate and deactivate proteins. This regulation can lead to diseases like cancer (Nishi, Shaytan and Panchenko, 2014; Ardito *et al.*, 2017). To date the PTM of MFHAS1 are incompletely characterised. Only the ubiquitination of MFHAS1 by



PJA2 has been partly described (Zhong *et al.*, 2018). The PhosphoSitePlus® website summarises all reported PTM from the literature, including large scale high throughput studies. For this study, the focus was brought to the phosphorylation and ubiquitination sites of MFHAS1.

Protein interacting with MFHAS1 are of interest in order to understand which cellular pathways MFHAS1 is involved in. MFHAS1 interactions with the following protein have been reported: Heat Shock protein 60 (HSP 60) (Dihanich *et al.*, 2014), Heat Shock protein 70 (HSP 70) (Kumkhaek *et al.*, 2019), Raf1 (Kumkhaek *et al.*, 2013) and E3 ubiquitin-protein ligase Praja-2 (PJA2) (Zhong *et al.*, 2018). Protein-protein interaction software are used to find other potential proteins interacting with MFHAS1.

In this chapter, bioinformatic tools and resources to gain insights into the function and biology of MFHAS1 are used. The aims of this chapter are to model and analyse the structure of MFHAS1 protein using Alphafold2 and Intfold7. To understand the post translational modification (PTM) reported in PhosphoSitePlus® (Hornbeck *et al.*, 2015), and try to understand the impact of phosphorylation and ubiquitination on MFHAS1. From the reported phosphorylation sites, predicted interacting kinases were listed using kinases consensus sequences websites. In order to define MFHAS1 protein interactions using protein-protein interaction software; PINOT and HIPPIE and functional enrichment software g:Profiler have been applied. A pipeline of the bioinformatic work is summarise in **Figure 3.1**.



**Figure 3.1:** Pipeline of the bioinformatic analysis presented in this chapter. Datasets available online are analysed through post translational modification, protein-protein interaction, and structure modelling.

## 3.2 Materials and methods

### 3.2.1 Protein modelling structure

Modelling of MFHAS1 protein structure prediction was carried out using two platforms; Alphafold2 (Jumper *et al.*, 2021; Varadi *et al.*, 2022) and IntFOLD (McGuffin *et al.*, 2017; McGuffin *et al.*, 2019, 2023).

Alphafold2 was created by DeepMind and the EMBL-European Bioinformatics Institute (EMBL-EBI) as an artificial intelligence (AI) system. Alphafold2 software was using multiple sequence alignments (MSAs) and pairwise features to create the predicted structure. Uniprot ID “Q9Y4C4” was entered in the Alphafold2 search bar. Predicted 3D structure with confidence score and predicted aligned error were downloaded and analysed.

IntFOLD (Version 7.0) contains three subparts, IntFOLD7 which predicts protein structure by fragment assembly and contact prediction methods. It also focuses on the domain prediction, protein-ligand binding prediction. ModFOLD dock gives data about the quality of the model and finally MultiFOLD allows multimeric Assemblies. IntFOLD7 was used to predict tertiary structure of MFHAS1 and contains DISOclust and FunFOLD. Disorder prediction of the protein structure from DISOclust was analysed and used to predict the boundaries of MFHAS1. FunFOLD was used to predict ligand binding site residues.

The MFHAS1 protein sequence was pasted into the IntFOLD prediction page ([https://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD7\\_form.html](https://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD7_form.html)). Results were downloaded and analysed.

The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC was used to visualise and compare protein structures.

### 3.2.2 Post translational modification.

PhosphoSitePlus® (Hornbeck *et al.*, 2015) was used to study the post translational modifications (PTM). PhosphoSitePlus® is an online repository which groups and catalogs PTM across the human proteome such as phosphorylation, ubiquitination, acetylation, methylation, and O-glycosylation. The references were either high throughput (HTP) or low throughput (LTP). HTP PTM were data collected using exclusively mass spectrometry. LTP PTM data were collected using other methods than mass spectrometry such as pull-down.

To predict the kinases carrying out phosphorylation on the PTM sites, two Kinase interaction Prediction software were used to identify consensus sequence (Bradley *et al.*, 2021). NetPhos - 3.1 was used to predict the kinases interacting on the phosphorylation site (Blom, Gammeltoft and Brunak, 1999; Blom *et al.*, 2004). This software was making prediction for 17 kinases: ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38 MAPK. Prediction scores were display between 0 and 1. Scores higher than 0.5 were indicating a positive prediction. The second kinase prediction software that was also used was available through the PhosphoSitePlus® portal (Johnson *et al.*, 2023). 303 protein serine and threonine kinases were profiled against phosphorylation sites, with synthetic peptide libraries. PhosphoSitePlus® predicted kinases score superior to zero were favorable.

Conservation of the phosphorylation and ubiquitination motifs across the ROCO proteins was tested by protein sequences alignment. Alignment of more than two protein sequences together, Clustal omega (Madeira *et al.*, 2022) was used. Clustal omega sequences alignment software was performed by seeded guide trees and hidden Markov models (HMM) techniques.

Alignment of two sequences was performed by EMBOSS Needle alignment software, which use the Needleman-Wunsch alignment algorithm (Needleman and Wunsch, 1970).

### 3.2.3 Protein network analysis

In order to characterise the network interactome for MFHAS1, two approaches were applied. Protein-protein interaction (PPI) was performed with two software: Protein Interaction Network Online Tool (PINOT) (Tomkins *et al.*, 2020) ([https://www.reading.ac.uk/bioinf/PINOT/PINOT\\_form.html](https://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html)) and Human Integrated Protein-Protein Interaction rEference (HIPPIE) (Alanis-Lobato, Andrade-Navarro and Schaefer, 2017).

PINOT was used to generate a list of proteins interacting with MFHAS1 on the 08/04/2023. After submission of input query using Uniprot IDs, the PPI were retrieved live from seven databases (BioGrid, bhf-ucl, IntAct, MINT, UniProt, MBIInfo, InnateDB) via the Proteomics Standard Initiative Common Query Interface (PSICQUIC). Confidence scoring is created for each protein interaction by adding the number of distinct methods used to access the interaction and the number of publication in which the interaction is reported (Tomkins *et al.*, 2020). The second PPI tool used was HIPPIE on the 08/04/2023. Their PPI repository was updated from PSICQUIC interface using seven databases (IntAct, MINT, BioGRID, HPRD, DIP, BIND and MIPS) (Alanis-Lobato, Andrade-Navarro and Schaefer, 2017).

First layer network was generated by submitting MFHAS1 (UniprotID: Q9Y4C4) into PINOT and HIPPIE then merging the two datasets. Second layer interactome network was created by submitting the first layer data into PINOT. A filter was added to only keep proteins with a confidence scoring of 4 and more. To visualise PPI network, Cytoscape software was used (Shannon *et al.*, 2003). Lists of PPI were downloaded into Cytoscape, and networks were established to allow visualisation and clustering analysis.

In order to understand the function of the proteins interacting with MFHAS1, functional enrichment of the protein present in the PPI list was carried out using g: Profiler (Raudvere *et al.*, 2019), and visualised using cytoscape.

To group proteins of the second layer network into functional cluster, the Cytoscape tool AutoAnnotate was used to automatically create the clusters (Kucera *et al.*, 2016) based upon shared relationships among the nodes.

### 3.3 Results

#### 3.3.1 Protein modelling structure

Protein structure modelling using machine learning approaches has become more accessible recently, especially since the launch of AlphaFold2 (AF2). AF2 has generated high-accuracy models for proteins on an unprecedented scale (Bertoline *et al.*, 2023). Structural modelling provides a powerful alternative and complementary approach to experimentally derived data, but their use is limited. In particular, prediction software may have limitations in modelling intrinsically disordered regions or novel structures (Bertoline *et al.*, 2023). As efforts to derive a structure for MFHAS1 were unsuccessful via crystallography or cryo-EM, it was decided to model its structure using AF2 and IntFOLD7.

DeepMind developed AF2 and uses deep learning methods and multiple sequence alignment to predict protein structure accurately. Multiple sequence alignment methods predict the distance between pairs of amino acids. Structure modelling algorithms are tested every two years through Critical Assessment of techniques for protein Structure Prediction (CASP), where many modelling software compete to predict a protein structure with the most accuracy. At the CASP14 competition, AF2 was ranked first (Jumper *et al.*, 2021).

The IntFOLD software package was developed at the University of Reading (Mcguffin *et al.*, 2023). IntFOLD contains ModFOLD, MultiFOLD and IntFOLD7. At the CASP15, ModFOLD dock was the leading software, IntFOLD7 ranked competitively, and MultiFOLD was among the top ten software (Mcguffin *et al.*, 2023).

IntFOLD and AF2 are powerful tools for structure prediction focusing on slightly different aspects of protein modelling. While they both predict protein structure with high confidence, AF2 uses artificial intelligence and presents a database of protein modelling structures, while IntFOLD operates with a machine learning algorithm and also yields information about the potential ligands' interaction and domain boundaries.

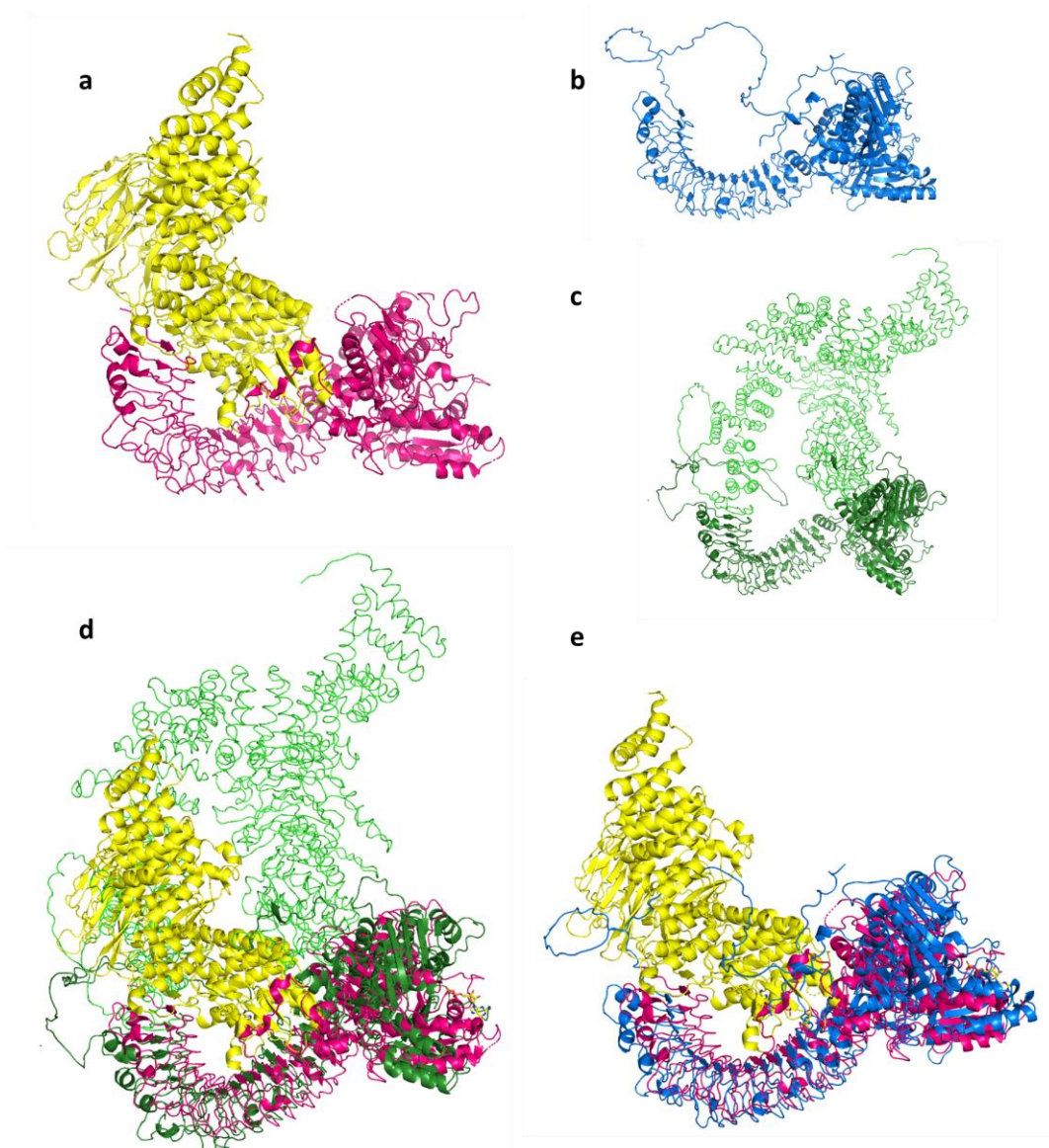
### **3.3.1.1 Assessing the quality of the protein model**

#### 3.3.1.1.1 Comparison of LRRK2 crystal structure with predicted models

To understand the accuracy of the predicted structure from AF2 and IntFOLD in the context of the ROCO proteins, a comparison of LRRK2 monomeric crystal structure 7LHW (Myasnikov *et al.*, 2021) and modelled structures of LRRK2 is performed.

The structures are aligned in PyMOL, and give a Root Mean Square Deviation (RMSD) score. RMSD is calculated by the average of the absolute values of the differences between two structures. The closer the RMSD is to 1, the better the alignment is (Walker, 2009).

Only the LRR-ROC-COR domains of LRRK2 are compared. This choice was made to match MFHAS1 domains. Also, IntFOLD has a limitation of 1000 amino acid length of the protein sequence that can be submitted. Amino acid sequence 880 to 1879 of LRRK2 was submitted to IntFOLD. The same sequence was highlighted in dark green in the AF2 model (**Figure 3.2**). Alignment of LRRK2 with AF2 shows an RMSD of 8.452 Å. LRRK2 sequence alignment with IntFOLD presents an RMSD of 6.548 Å. According to those results, the model structure from IntFOLD is closer to the solved structure.

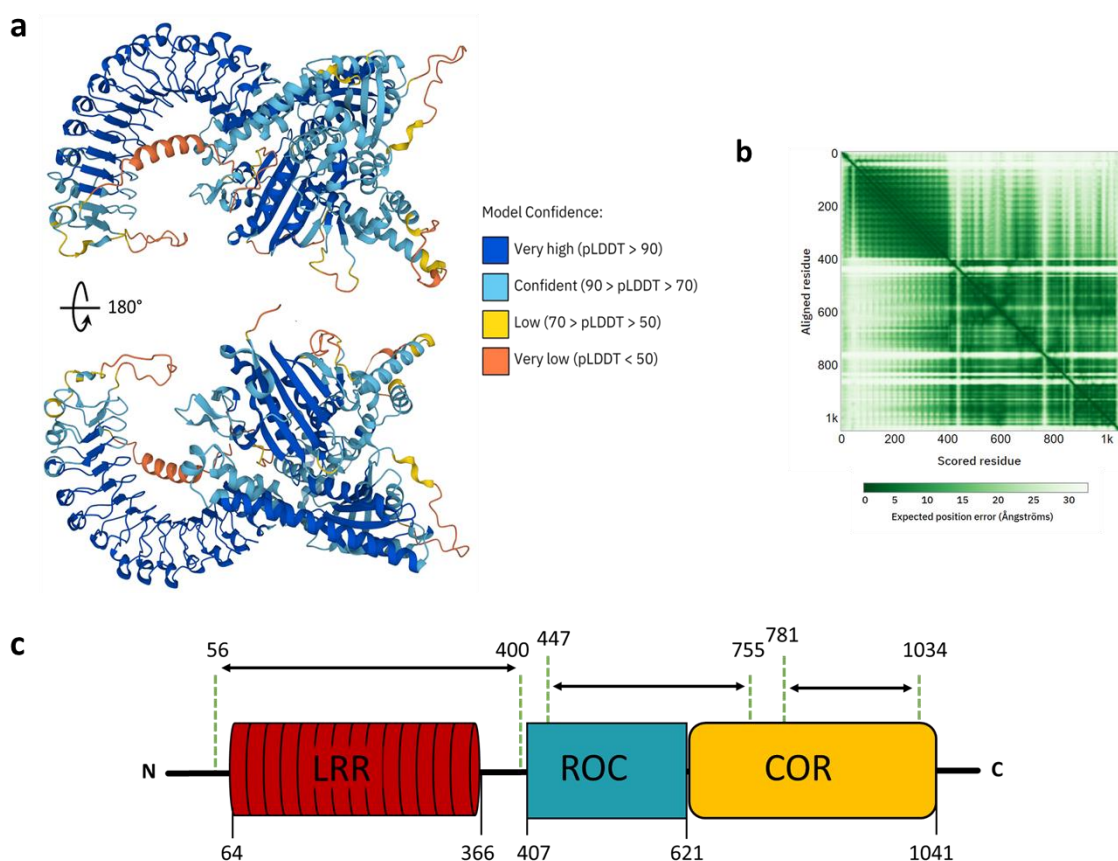


**Figure 3.2:** Comparison of LRRK2 protein solved structure with modelled structure from AlphaFold2 and IntFOLD. **(a):** LRRK2 monomeric structure 7LHW is represented in pink for the residues 880 to 1879 and yellow for the other amino acids (AA). **(b):** IntFOLD model structure of LRRK2 **(c):** AlphaFold2 model structure of LRRK2 **(d):** Comparison of alphaFold2 (AF2) structure with the solved structure. AF2 model is shown in green, AA sequence 880 to 1879 is in dark green. **(e):** Comparison with IntFOLD model structure (blue) AA sequence 880 to 1879.



### 3.3.1.1.2 MFHAS1 predicted structure from AlphaFold2

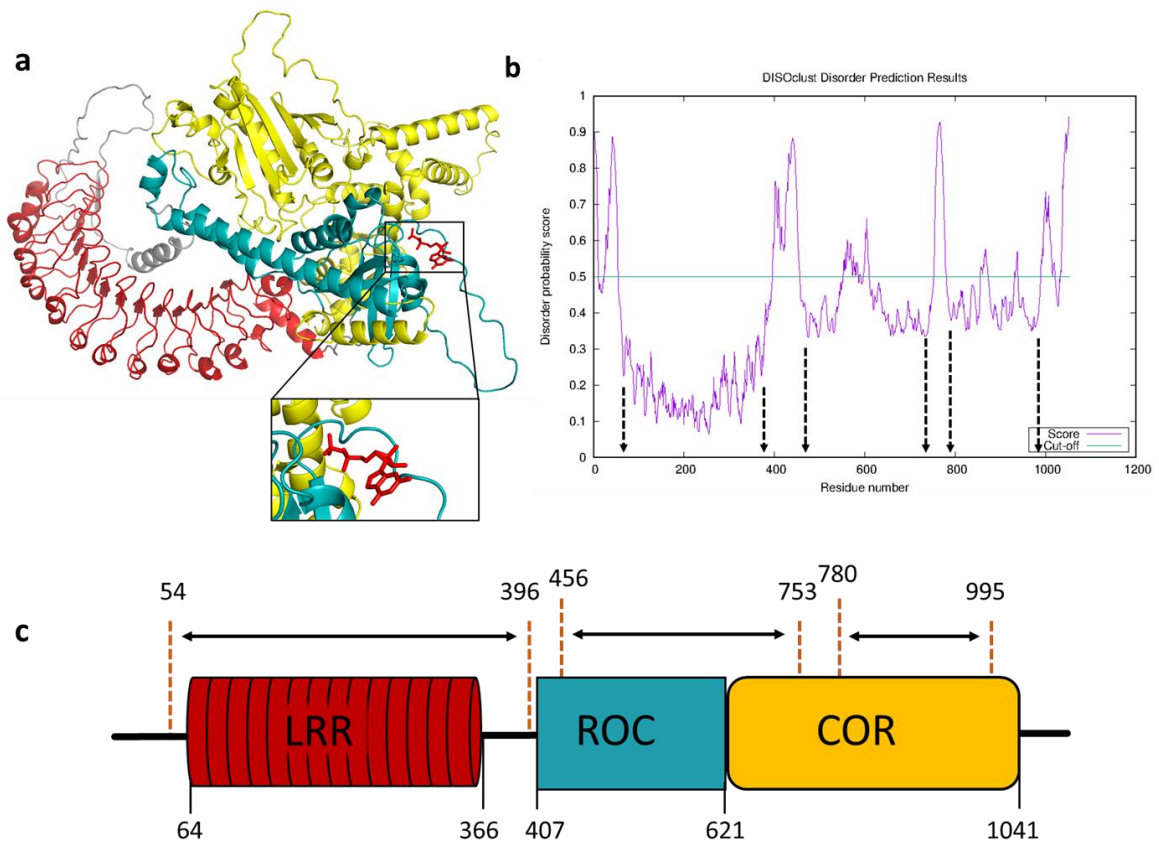
The modelling structure of MFHAS1 predicted by AF2 shows the confidence score of the sequence (**Figure 3.3**). The predicted Local Distance Difference Test score (pLDDT) corresponds to the confidence score per residue. Different colours represent it: dark blue shows a very high confidence score, light blue shows a confident score, yellow shows a low, and orange shows a very low confidence score. The LRR domain of MFHAS1 has a very high confidence score. Most of the ROC and COR domains are predicted with high or confident confidence scores, apart from a few disordered loops with low to very low scores. The N-terminal of MFHAS1 also has a very low score. Predicted aligned error data is also available for the structure; it is helpful to predict inter-domain accuracy and consequently provide information about domain boundaries. MFHAS1 domains could be defined as 56 aa to 400 aa (LRR), 447 to 755 aa (ROC), and 781 to 1034 aa (COR) (**Figure 3.3**).



**Figure 3.3:** AlphaFold2 prediction for MFHAS1 protein. **(a):** MFHAS1 predicted structure with confidence score. Very high predicted Local Distance Difference Test score (pLDDT) is dark blue, confident pLDDT is light blue, low pLDDT is yellow and very low pLDDT is orange. **(b):** Graphical representation of the predicted aligned error of MFHAS1 structure. With the expected position error in Ångströms. Dark green low error, light green high error. **(c):** Transposition of the domain boundaries from AF2 (above) into the MFHAS1 protein ideogram.

### 3.3.1.1.3 MFHAS1 predicted structure from IntFOLD

IntFOLD software comprises IntFOLD7 structure prediction but also DISOclust, which predicts the disordered region of the predicted protein (Mcguffin, 2008) and FunFOLD, which predicts the ligand binding residues (Roche, Tetchner and McGuffin, 2011). MFHAS1 protein model presents a monomer structure. Guanosine diphosphate (GDP) is found to be a potential ligand to MFHAS1 within the ROC domain (**Figure 3.4**). Disorder prediction shows high disorder at the N-terminal, around the residue 450 and around 750 of MFHAS1 protein (**Figure 3.4**).

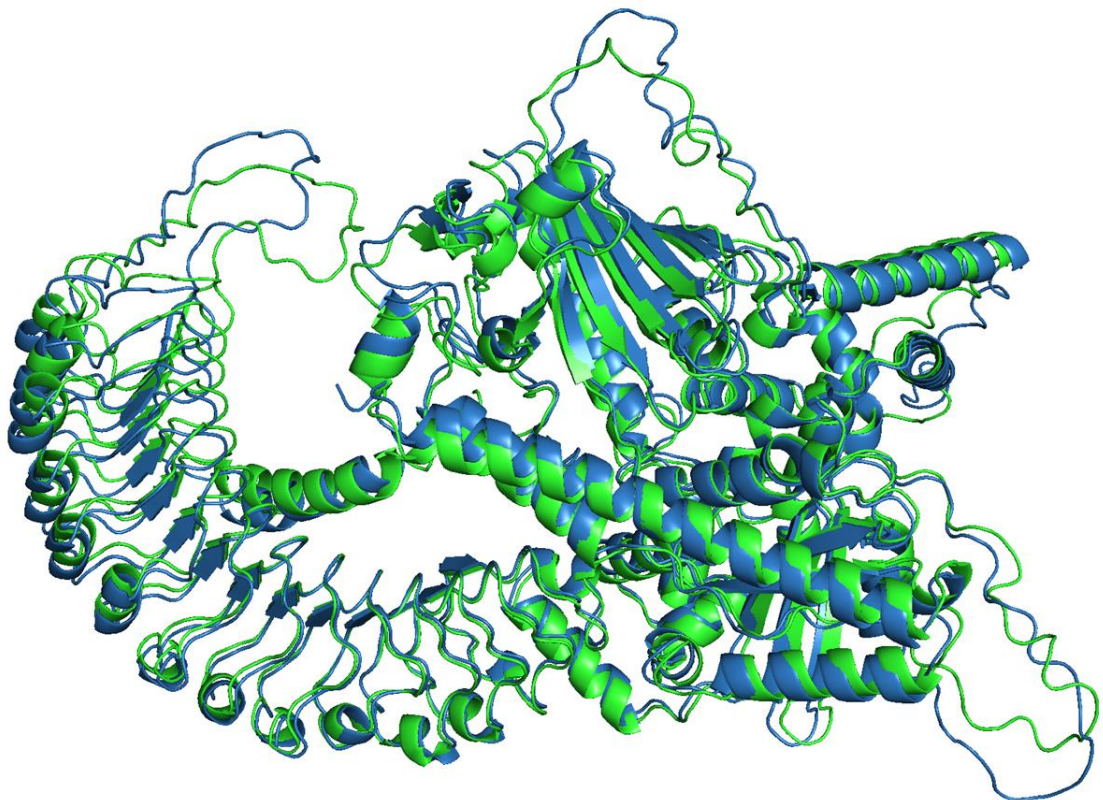


**Figure 3.4:** MFHAS1 prediction structure from IntFold. **(a):** IntFOLD7 and FunFOLD predicted a monomer structure of MFHAS1 with guanosine diphosphate ligand. **(b):** Graphic of the disorder prediction from DISOclust. Disorder score above 0.5 is considered disordered. Black arrows were added to highlight the domain boundaries **(c):** Transposition of the main disordered regions onto MFHAS1 ideogram. Boundaries below the ideogram were found in Chapter 2. Arrows above the ideogram are the ordered regions.

#### 3.3.1.1.4 Comparison predicted model of MFHAS1 from AF2 and IntFOLD

A comparison of both MFHAS1 predictions is performed to confirm the model. AF2 and IntFOLD7 predicted structures are aligned on PyMOL (**Figure 3.5**). The alignment's Root Mean Square Deviation (RMSD) is 1.182 Å. Both prediction structures are very close and would endorse the quality of the prediction.

Furthermore, the disorder region found by IntFOLD matches the predicted aligned error from AF2.

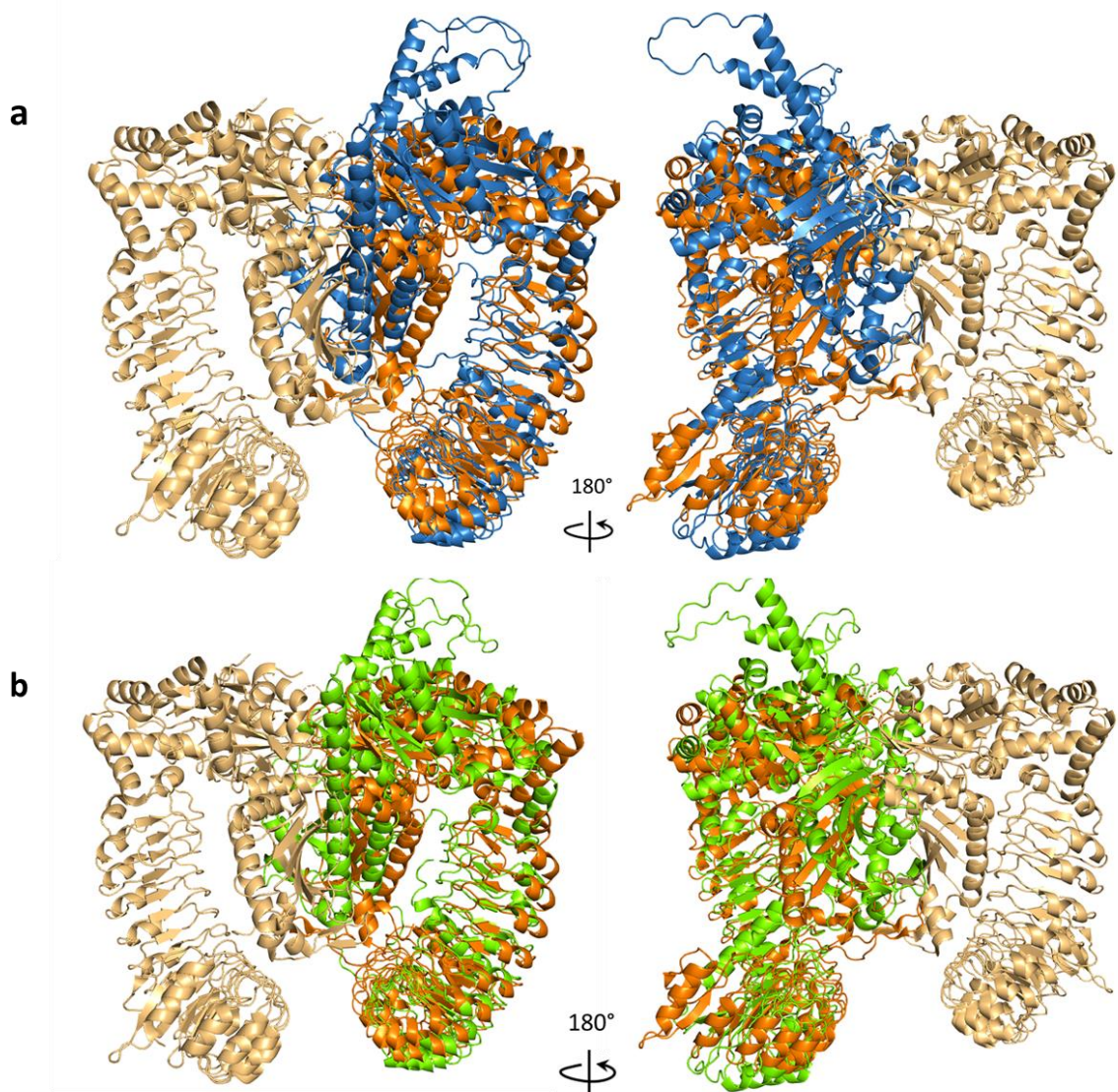


**Figure 3.5:** Model structure comparison IntFOLD7 (blue) and AlphaFold2 (green) of the MFHAS1 protein. Alignment performed on PyMOL.



### 3.3.1.1.5 Comparison of MFHAS1 predicted structures with CtROCO protein

The *Chlorobaculum tepidum* ROCO protein (CtROCO) has the same domain structure as MFHAS1 but only shares 23.6% sequence identity. CtROCO crystal structure was solved in 2019 PDB: 6HLU (Deyaert *et al.*, 2019). They found CtROCO protein in the form of a dimer. Comparison of MFHAS1 model structures with the CtROCO crystal structure shows the differences and similarities between those two proteins (**Figure 3.6**). Alignment of the IntFOLD structure shows an RMSD of 8.556 Å. Alignment of the AF2 structure shows an RMSD of 7.984 Å. According to these data, the AF2 predicted structure is closer to the CtROCO structure than the IntFOLD model.



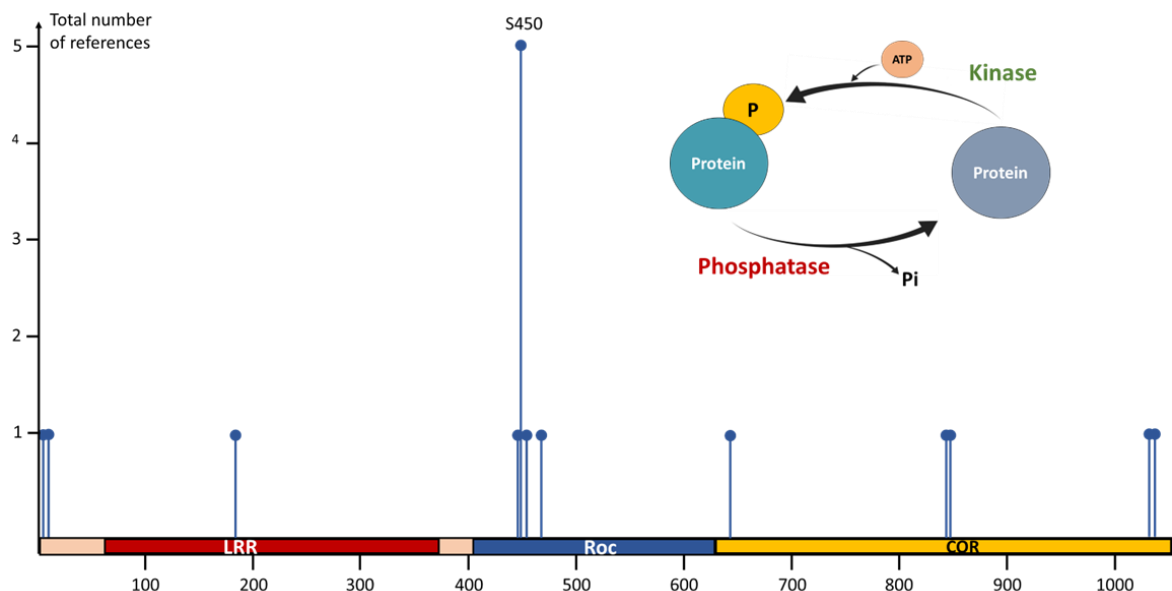
**Figure 3.6:** Comparison of the MFHAS1 model with *Chlorobaculum tepidum* ROCO (CtROCO) protein structure (PDB: 6HLU). **(a)** Alignment of the IntFOLD7 predicted structure (blue) and CtROCO (orange). **(b):** Alignment of the AlphaFold2 predicted structure (green) and the CtROCO (orange).

### 3.3.2 Post-translational modification

Post-translational modification (PTM), such as phosphorylation and ubiquitination, can be defined as a reversible covalent modification of the protein by adding a functional group (Bradley, 2022). The modification impacts the protein by increasing its range of function (Khoury, Baliban and Floudas, 2011). In this study, the focus was placed on phosphorylation and ubiquitination sites as crucial modulators of signal transduction.

#### 3.3.2.1 Phosphorylation site

Phosphorylation is one of the most studied PTM (Ardito *et al.*, 2017). It consists of the covalent attachment of a phosphate group by a kinase to the protein (**Figure 3.7**). Canonical phosphorylation occurs on three amino acids, serine, tyrosine and threonine, within a consensus site (Blom, Gammeltoft and Brunak, 1999). PTM of MFHAS1 were investigated using PhosphoSitePlus® (Hornbeck *et al.*, 2015). PhosphoSitePlus® groups all the PTM found via experiments and classifies them by HTP (PTM found via mass spectrometry) or LTP (PTM found by any experiments other than mass spectrometry). Each data is identified, curated, and edited to provide accurate information. Eleven phosphorylation sites are referenced for MFHAS1. Apart from S450, referenced five times, all the other sites are referenced once (**Figure 3.7**).



**Figure 3.7:** Schematic representation of the reported phosphorylation sites of MFHAS1. Adapted from PhosphoSitePlus®. And schematic representation of the phosphorylation mechanism.

### 3.3.2.1.1 Kinase prediction

To find which kinase could interact with and modify MFHAS1 on these putative phosphorylation sites, all the consensus sequences of the phosphorylation sites found in MFHAS1 were analysed by kinase prediction software. Two prediction tools were used to compare the data: NetPhos and the kinase prediction portal of PhosphoSitePlus®. The predicted kinases and their confidence scores are shown in **Table 3.1**. These tools do not allow for tyrosine phosphorylation prediction. NetPhos uses a scoring system from 0 to 1, with results from 0.5 and above being positive predictions. PhosphoSitePlus® uses log2 (score) with a score superior to 0 being positive.

MFHAS1 (human)	PhosphoSitePlus® Kinase library	Score PhosphoSitePlus®	NetPhos	Score NetPhos
S6-p	CDK16	2.554	GSK3	0.456
T11-p	ERK7	5.613	CaM-II	0.438
	ERK5	4.505	GSK3	0.435
	IRAK4	3.400	p38MAPK	0.422
T185-p	SBK	4.156	unsp	0.771
	CAMK2A	3.380		
Y447-p				
S450-p	KIS	9.022	GSK3	0.532
	P38D	8.491	cdk5	0.528
	CDK19	7.366		
S455-p	GSK3B	2.741		
S468-p	PERK	3.101		
	HRI	3.011		
S644-p	PAK6	4.883	unsp	0.994
	RSK4	3.507	PKC	0.676
	TTBK1	3.465	CKII	0.574
S844			cdc2	0.516
	NIM1	3.835	PKC	0.655
	NEK6	2.465		
	HUNK	1.819		
Y848-p				
T1033-p	P38D	6.815	p38 MAPK	0.546
	P38B	6.330	cdk5	0.517
	P38A	5.712	GSK3	0.509
S1038-p	ERK2	3.734	unsp	0.902
	ERK1	3.544	cdk5	0.623
	JNK2	3.379	p38MAPK	0.527

**Table 3.1:** Kinase's prediction of the consensus sequence of the phosphorylation site referred in MFHAS1. PhosphoSitePlus® and NetPhos tools are used. The top predicted kinases are display in the table along with their score.

The S450 phosphorylation site was focused upon for better investigation because it was reported more time than the other sites and because the kinase prediction score is the highest from PhosphoSitePlus® for this site.

The predicted kinases for S450 are Serine/threonine-protein kinase Kist (KIS), Mitogen-activated protein kinase 13 (p38d) and Cyclin-dependent kinase 19 (CDK19) from PhosphoSitePlus®. Glycogen synthase kinase-3 beta (GSK3) and Cyclin-dependent kinase 5 (CDK5) are found in NetPhos.

It is interesting to note that each software predicts a Cyclin-dependent kinase. CDK19 variants are associated with Epileptic encephalopathy (Chung *et al.*, 2020), while CDK5 phosphorylates many proteins, including small GTPases (Walkup *et al.*, 2015), histone H1, p53 and dysregulation of CDK5 is reported in neurodegenerative diseases (Gong *et al.*, 2003).

p38d, also called MK13, is a serine/threonine kinase and one of the p38 MAPK isoforms (Jiang *et al.*, 1997). KIS is also a serine/threonine kinase; its expression was found in acute lymphoblastic leukaemia cases (Nakamura *et al.*, 2008).

GSK3 is involved in numerous pathways and phosphorylates over 100 substrates (Beurel, Grieco and Jope, 2015).

### 3.3.2.1.2 Sequence alignment

To test if the consensus sequences of MFHAS1 phosphorylation sites are conserved in other ROCO proteins, sequence alignment was performed using Clustal Omega (Madeira *et al.*, 2022) and Needle tools (Needleman and Wunsch, 1970). Clustal omega was used to have an overview of the human ROCO proteins: LRRK2, LRRK1, DAPK1 and *C.tepidum* ROCO protein (CtROCO) (**S.Figure.4**). EMBOSS Needle was also used to compare two sequences simultaneously to provide a more accurate alignment to the MFHAS1 sequence. Most alignments were found using EMBOSS Needle only.

Alignment between MFHAS1 (S450) and DAPK1 (S734) has a four amino acid conservation forming a PPSP sequence (**Figure 3.8**). The PPSP motif, highly constrained by the presence of multiple proline residues, has previously been highlighted as essential for signalling in LRP6 (Brennan *et al.*, 2004). DAPK1 (S734) is a phosphorylation site referenced nine times on PhosphoSitePlus® (seven HTP and two LTP). Serine 734 of DAPK1 is phosphorylated by ERK, leading to the activation of p53 (Chen *et al.*, 2005; Kwon *et al.*, 2016). Position of DAPK1 S734 is shown on **Figure 3.8**.

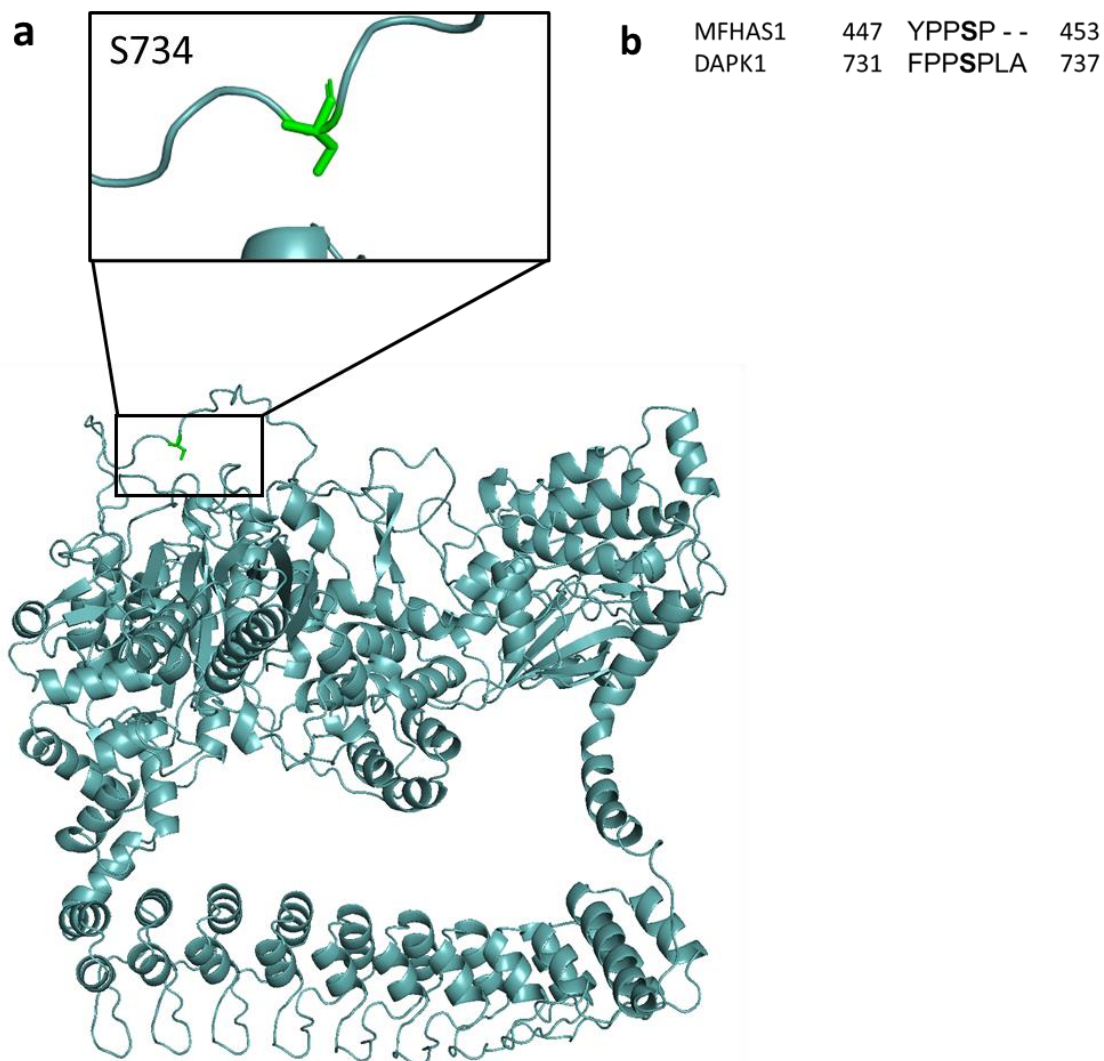
MFHAS1 (S455) aligns with LRRK2 (T1368), CtROCO (T484) and DAPK1 (S743) sites. LRRK2 (T1368) is a phosphorylation site reported four times in PhosphoSitePlus®, including one LTP (Sheng *et al.*, 2012).

MFHAS1 (S644) aligns with LRRK2 (S1536), CtROCO (S615) and DAPK1 (S971) sites. The LRRK2 (S1536) phosphorylation site was reported once on PhosphoSitePlus® via HTP methods.

MFHAS1 (Y848) aligns with LRRK1 (Y989) using Clustal omega and LRRK1 (Y1063) when looking through the EMBOSS Needle. LRRK1 (Y1063) phosphorylation site is reported seven times via HTP methods.

The following phosphorylation sites align with other ROCO proteins without conserved sites. MFHAS1 (S6) with CtROCO (S86), MFHAS1 (T11) with LRRK1 (T43), MFHAS1 (Y447) with LRRK2 (S1360), MFHAS1 (T1033) and (S1038) with CtROCO (S930) and (935) respectively.





**Figure 3.8:** Phosphorylation site S734 of DAPK1. **(a):** Predicted structure of DAPK1 from Alphafold2 and zoom in on the phosphorylation site S734. **(b):** Alignment of MFHAS1 and DAPK1 PPSP motif and phosphorylation site MFHAS1 (S450) and DAPK1 (S734) using EMBOSS Needle.

Cross-species comparison of the phosphorylation sites of MFHAS1 was performed by aligning the protein sequences of human, mouse, rat, African elephant, taurus, whale, chimpanzee, gorilla, rabbit, and guinea pig MFHAS1.

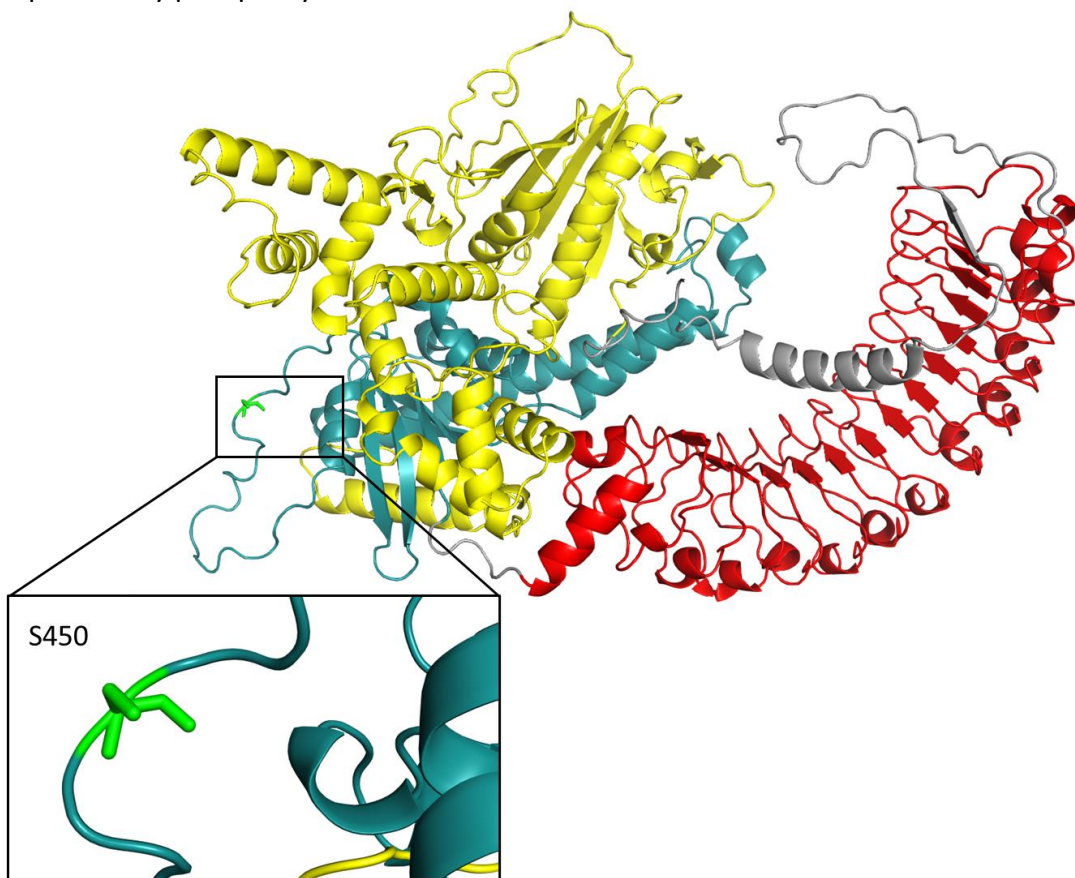
Apart from Y447 and S450, all the phosphorylation sites of MFHAS1 are conserved across all species. The PPSP motif around S450 is conserved in the African elephant, taurus, whale, chimpanzee, and gorilla MFHAS1 but not in the rabbit, mouse, and rat. The Y447 phosphorylation site is not conserved in the whale and rabbit. Overall, the COR domain of MFHAS1 is well-conserved across species.

### 3.3.2.1.3 Phosphorylation sites in the context of MFHAS1 structure

Phosphorylation can act to reversibly alter protein structure and function (Johnson and Barford, 1993). Understanding where the phosphorylation site resides can yield information about how the phosphorylation could change the protein's conformation and, therefore, function (Raju, 2019).

Due to the multiple reports of phosphorylation at residue S450, this site was chosen for further investigation. Using the AF2-modelled structure of MFHAS1, the S450 residue was identified and highlighted (**Figure 3.9**). Domains of MFHAS1 are colour-coded: LRR is red, ROC domain is blue, COR is yellow, and the linkers are grey. The S450 site is zoomed in and highlighted in green.

The S450 phosphorylation site is present on a disordered loop in the middle of the Roc domain. The site's location makes it accessible for other proteins to interact with MFHAS1 and potentially phosphorylate it.



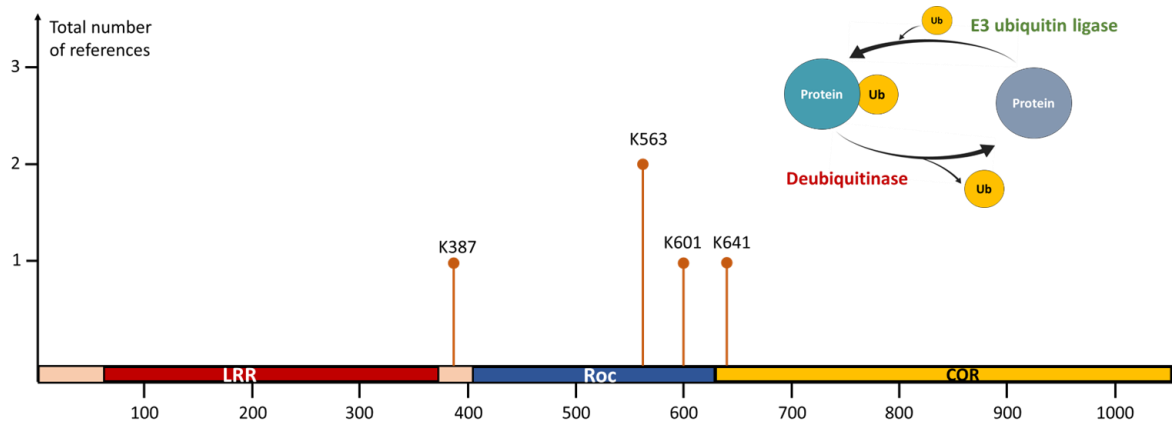
**Figure 3.9:** Phosphorylation site S450 of MFHAS1. MFHAS1 protein prediction from Alphafold2. S450 site is represented in green. MFHAS1 domains are coloured: Red: LRR, Blue: ROC, Yellow: COR and grey: linker regions.

### 3.3.2.2 Ubiquitination site

Ubiquitination is a post-translational modification of proteins observed in eukaryotic cells, which consists of the covalent attachment of the ubiquitin protein to another protein (**Figure 3.10**) (Damgaard, 2021). The ubiquitin protein is 76 amino acids long and can be found as a free form as well as bound to a protein. There are three forms of ubiquitination: mono-ubiquitination, which consists of attaching one ubiquitin to a protein; multiple mono-ubiquitination, which is the attachment of a single ubiquitin molecule at multiple sites onto the protein; and polyubiquitin, which is the addition of many ubiquitin proteins at one site of the protein forming a chain of ubiquitin (Damgaard, 2021).

Ubiquitination of proteins can have a wide range of consequences on the protein function and stability. It can tag a protein for degradation via the ubiquitin-proteasome system (UPS) (Deshaies and Joazeiro, 2009) and can also play a role in regulating protein-protein interactions (Piper and Stringer, 2011), protein localisation, protein activity, proteasomal degradation, autophagy, endosomal cycling and cell death (Dikic and Schulman, 2022). A conserved set of enzymes performs ubiquitination: ubiquitin is first attached to the E1 ubiquitin-activating enzyme, then it is transferred to the E2 ubiquitin-conjugating enzyme, and finally, the ubiquitin is attached to a lysine residue on the target protein by the E3 ubiquitin ligase (Callis, 2014). This study focuses on the E3 ubiquitin ligase that has been shown to interact with MFHAS1 (Zhong *et al.*, 2018).

Ubiquitination sites of MFHAS1 are summarised in PhosphoSitePlus® (**Figure 3.10 and Table 3.2**). Four ubiquitination sites are found in MFHAS1: K563 is referenced twice, K387, K601 and K641 are referenced once. The K387 ubiquitination site is of interest for its potential connection with PJA2. PJA2 E3 ubiquitin ligase was previously reported to bind to MFHAS1 via the LRR domain (Zhong *et al.*, 2018).



**Figure 3.10:** Schematic representation of the reported ubiquitination sites of MFHAS1. Adapted from PhosphoSitePlus®. And schematic representation of the ubiquitination mechanism.

MFHAS1	Sequence	HTP
K387-ub	PPYEVCmKGIPYIAA	1
K563-ub	RQIALQEkHDAEGLS	2
K601-ub	AYYGVSdKnlRRRKA	1
K641-ub	HLRRLRDkLLsVAEH	1

**Table 3.2:** Summary table of the ubiquitination sites found in MFHAS1 protein from PhosphoSitePlus®. The sequences and the number of references is shown for each site. HTP: High throughput

### 3.3.2.2.1 Sequence alignment

Alignment of MFHAS1 sequence with LRRK1/2, DAPK1 and CtROCO was performed with Clustal Omega (Madeira *et al.*, 2022), and MFHAS1 ubiquitination sites were analysed for conservation across these proteins. MFHAS1 (K387) is aligned with LRRK2 (K1314), which is not a reported ubiquitination site for LRRK2. MFHAS1 (K563), with two reports of ubiquitination, is aligned with CtROCO (K582). MFHAS1 (K601) residue is aligned with

CtROCO (K611), DAPK1 (K939) and LRRK2 (K1502) (**Table 3.3**). DAPK1 (K939) is a known ubiquitination site, reported four times in PhosphoSitePlus®. CtROCO (K611) and LRRK2 (K1502) are not reported ubiquitination sites. The alignment shows a poorly conserved sequence between the proteins. DAPK1 (K968) is aligned with MFHAS1 (K641) but is not reported as a ubiquitination site in the phosphosite data set.

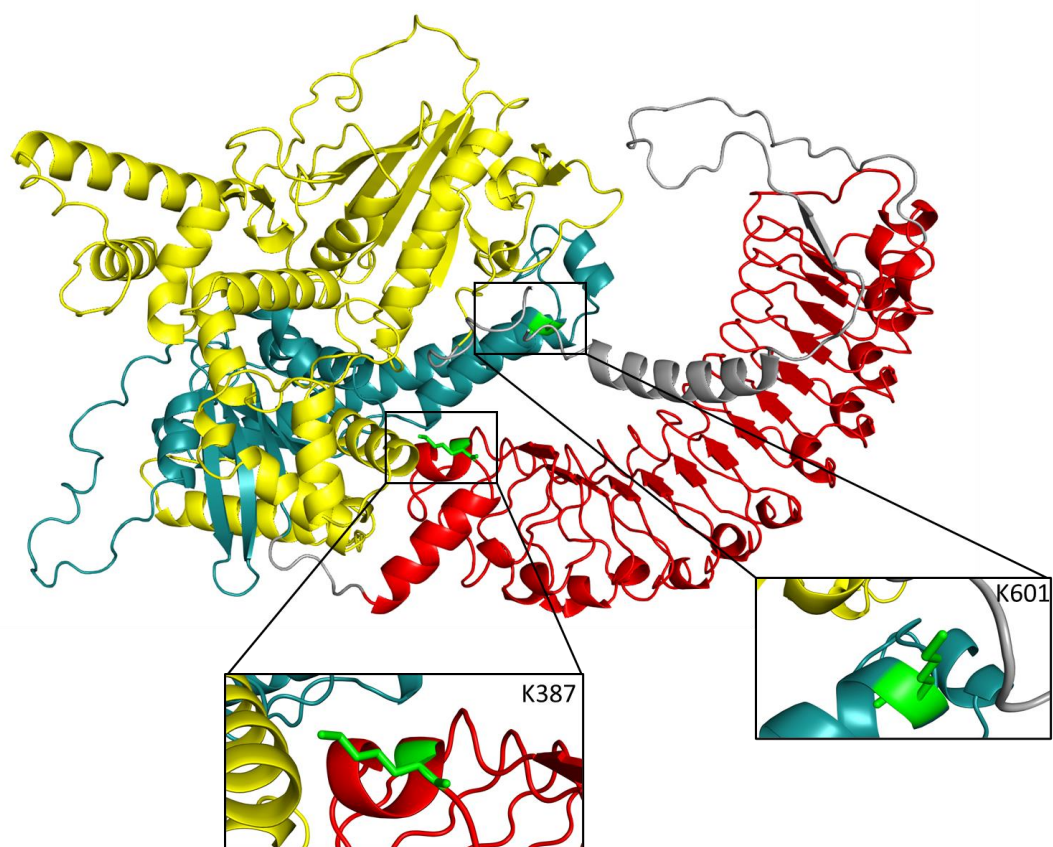
<b>MFHAS1 ubiquitination sites</b>	<b>Proteins aligned with MFHAS1</b>	<b>Alignment sequence</b>			
<b>K601</b>	LRRK2 (K1502)	MFHAS1	597	GVSD <b>K</b> NLRR	605
	DAPK1 (K939)	LRRK2	1498	AKLR <b>K</b> T I I N	1506
		LRRK1	810	- EGQEGLRQ	817
		DAPK1	935	ASG <b>S</b> KDMKV	943
		CtROCO	608	- S I A <b>K</b> S LKS	615
CtROCO (K611)					

**Table 3.3:** Summary of the MFHAS1 ubiquitination sites aligned with the other human ROCO protein and CtROCO protein.

### 3.3.2.2.2 Ubiquitination sites in the context of MFHAS1 structure

Ubiquitination sites are studied in MFHAS1 predicted structure. The most interesting sites for this study are K601 for its conservation with other ROCO proteins and K387 for its potential interaction with PJA2 (Zhong *et al.*, 2018).

MFHAS1 predicted structure from AF2 highlights the ubiquitination sites (**Figure 3.11**). Each functional protein domain of MFHAS1 is coloured to define them. Ubiquitination sites are shown in green. K387 is located at the end of the Leucine Rich Repeat (LRR) domain of MFHAS1 which is known to be a protein-protein interaction domain (Kobe and Kajava, 2001). This ubiquitination site is studied in more detail in the next part of the chapter and in chapter 4. K601 is present in the ROC domain at the end of an alpha helix.



**Figure 3.11:** Ubiquitination site K387 and K601 of MFHAS1. MFHAS1 protein prediction from AlphaFold2 zoomed in on K601 and K387 ubiquitination sites. MFHAS1 domains are coloured: Red: LRR, Blue: ROC, Yellow: COR and grey: linker regions Green represent the ubiquitination sites. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. LRR: Leucine-Rich Repeat. ROC: Ras Of Complex protein. COR: C terminal Of Roc.

### 3.3.3 Protein-protein interaction network analysis

#### 3.3.3.1 MFHAS1 protein potential interactors

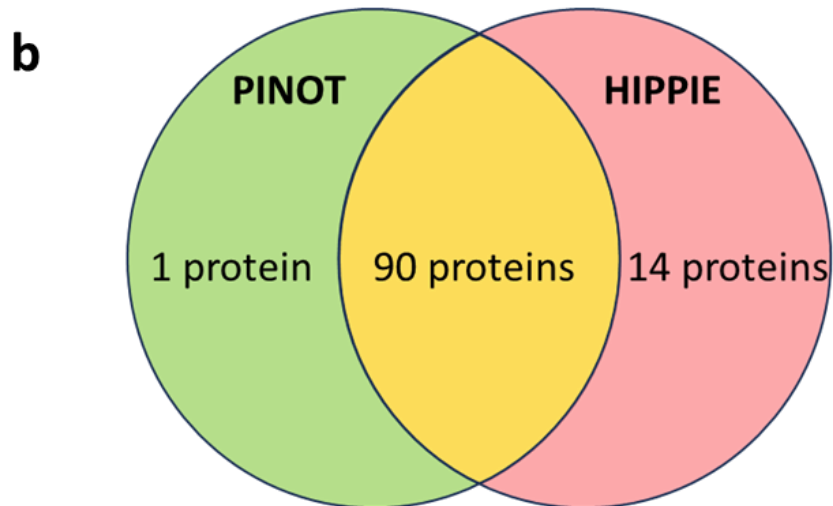
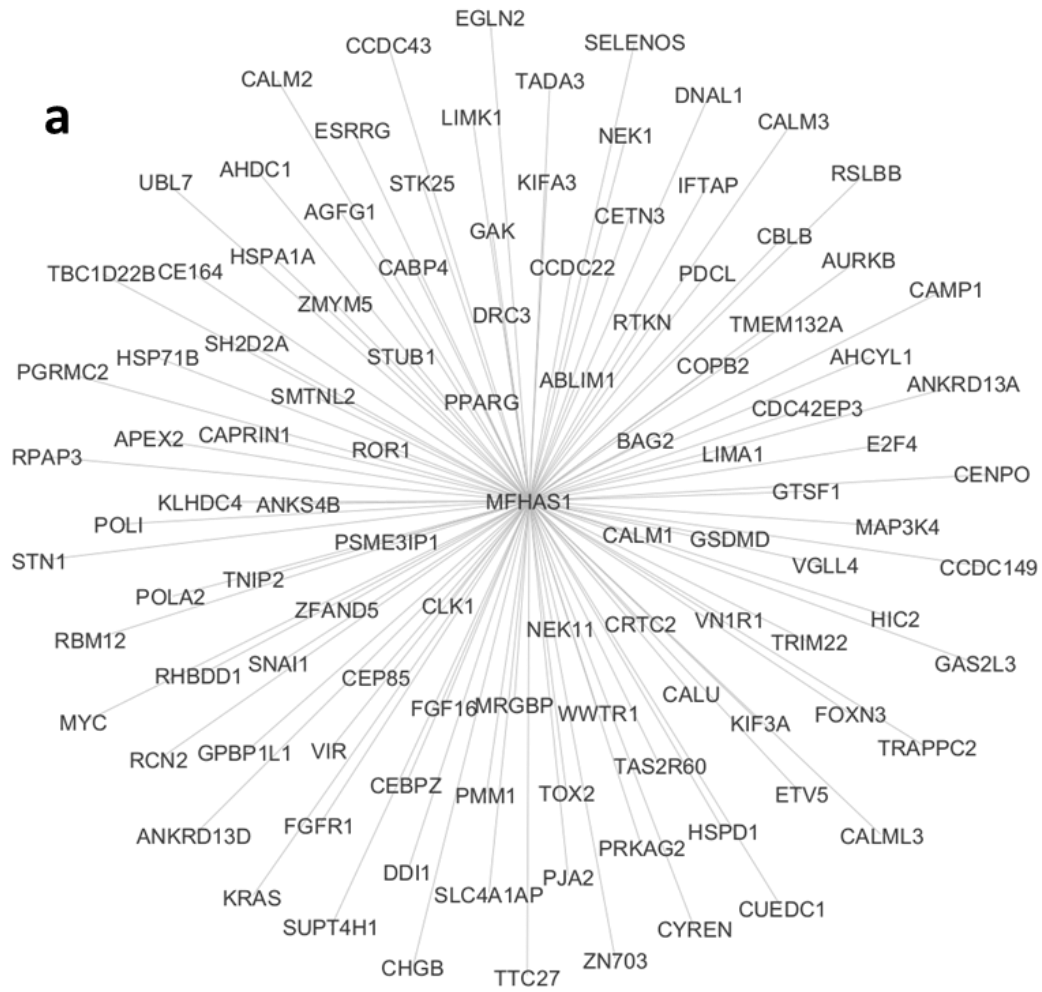
Studying the protein interacting with MFHAS1 and the function of these close interactors will help gain insight into MFHAS1 function (Manzoni *et al.*, 2018; Tomkins *et al.*, 2018). Protein-protein interaction (PPI) analysis is a powerful approach to understanding protein function and biological processes (Tomkins and Manzoni, 2021; Durham *et al.*, 2023).

As MFHAS1 possesses a GTPase activity in the ROC domain but does not include a kinase domain like the other ROCO proteins, it can be hypothesised that a kinase protein may interact with MFHAS1 to act as an effector.

To identify proteins interacting with MFHAS1, two web-based tools were used: first, Protein Interaction Network Online Tool (PINOT) from the University of Reading (Tomkins *et al.*, 2020), and second Human Integrated Protein-Protein Interaction rEference (HIPPIE) (Alanis-Lobato, Andrade-Navarro and Schaefer, 2017). PINOT identified 91 proteins directly interacting with MFHAS1, while HIPPIE recovered 104 proteins. Data from both software were merged and duplicates removed to obtain a master list of first-layer protein interactome (**Table 3.4**). Most of the proteins from both lists are in common apart from the RASK protein, which is only found in PINOT, and 14 proteins only found with HIPPIE (**Figure 3.12**). From this data, a network can be drawn and analysed (**Figure 3.12**).

PINOT scoring system adds the number of distinct methods used for detection and the number of publications reporting an interaction. HIPPIE, on the other hand, gives a score based on the type of methodology used, the functional association of the protein interacting and the number of sources. These scoring differences explain why HIPPIE and PINOT obtained slightly different results.





**Figure 3.12:** First layer interactome of MFHAS1 **(a)**: Network of merged data from PINOT and HIPPIE. Network visualised with Cytoscape. **(b)** Venn diagram of the protein found with PINOT and HIPPIE. It showcases the protein found by both tools and the protein only found with one tool.



First layer interactome from PINOT and HIPPIE				
ABLIM1	CE164	FOXN3	PDCL	STK25
AGFG1	CEBPZ	GAK	PGRMC2	STN1
AHCYL1	CENPO	GAS2L3	PJA2	STUB1
AHDC1	CEP85	GPBP1L1	PMM1	SUPT4H1
ANKRD13A	CETN3	GSDMD	POLA2	TADA3
ANKRD13D	CHGB	GTSF1	POLI	TAS2R60
ANKS4B	CLK1	HIC2	PPARG	TBC1D22B
APEX2	COPB2	HSP71	PRKAG2	TMEM132A
AURKB	CRTC2	HSPD1	PSME3IP1	TNIP2
BAG2	CT135	IFTAP	RBM12	TOX2
CABP4	CUEDC1	KIF3A	RCN2	TRAPPC2
CALM	CYREN	KIFA3	RHBDD1	TRIM22
CALML3	DDI1	KLHDC4	ROR1	TTC27
CALU	DNAL1	KRAS	RPAP3	UBL7
CAMP1	DRC3	LIMA1	RSLBB	VGLL4
CAPRIN1	E2F4	LIMK1	RTKN	VIR
CBLB	EGLN2	MAP3K4	SELENOS	VN1R1
CCDC149	ESRRG	MRGBP	SH2D2A	WWTR1
CCDC22	ETV5	MYC	SLC4A1AP	ZFAND5
CCDC43	FGF16	NEK1	SMTNL2	ZMYM5
CDC42EP3	FGFR1	NEK11	SNAI1	ZN703

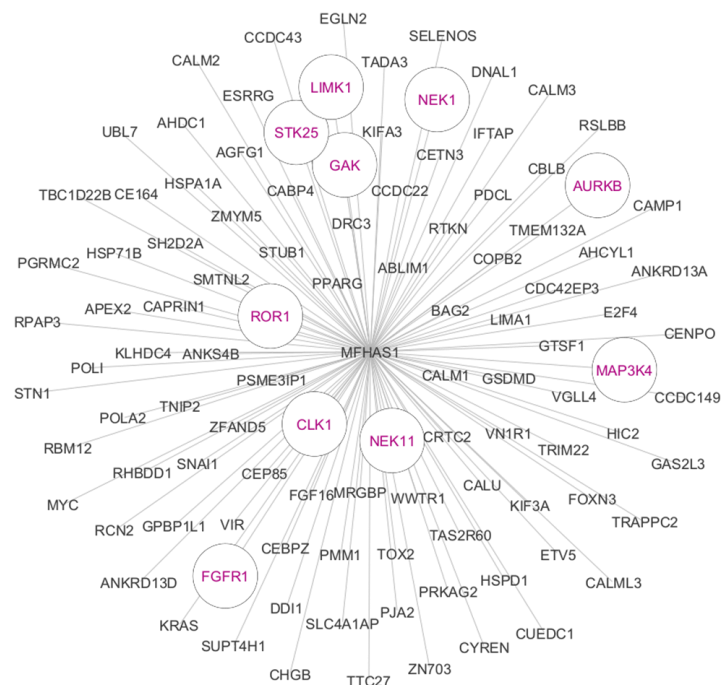
**Table 3.4:** First layer interactome from PINOT and HIPPIE. Lists from HIPPIE and PINOT are merged. Proteins only found from HIPPIE are highlighted in orange. Protein only found in PINOT is highlighted in green. Proteins are ranked by alphabetical order.

### 3.3.3.1.1 MFHAS1 interaction with the phosphorylation machinery

The MFHAS1 protein contains many putative phosphorylation sites, suggesting it might interact with one or more protein kinases. Kinases present in the first interactome of MFHAS1 are filtered by comparing the list of interactome from PINOT and HIPPIE with a list of 537 human kinases from kinhub. The network of the first layer interactome around MFHAS1 is visualised with Cytoscape (**Figure 3.13**).

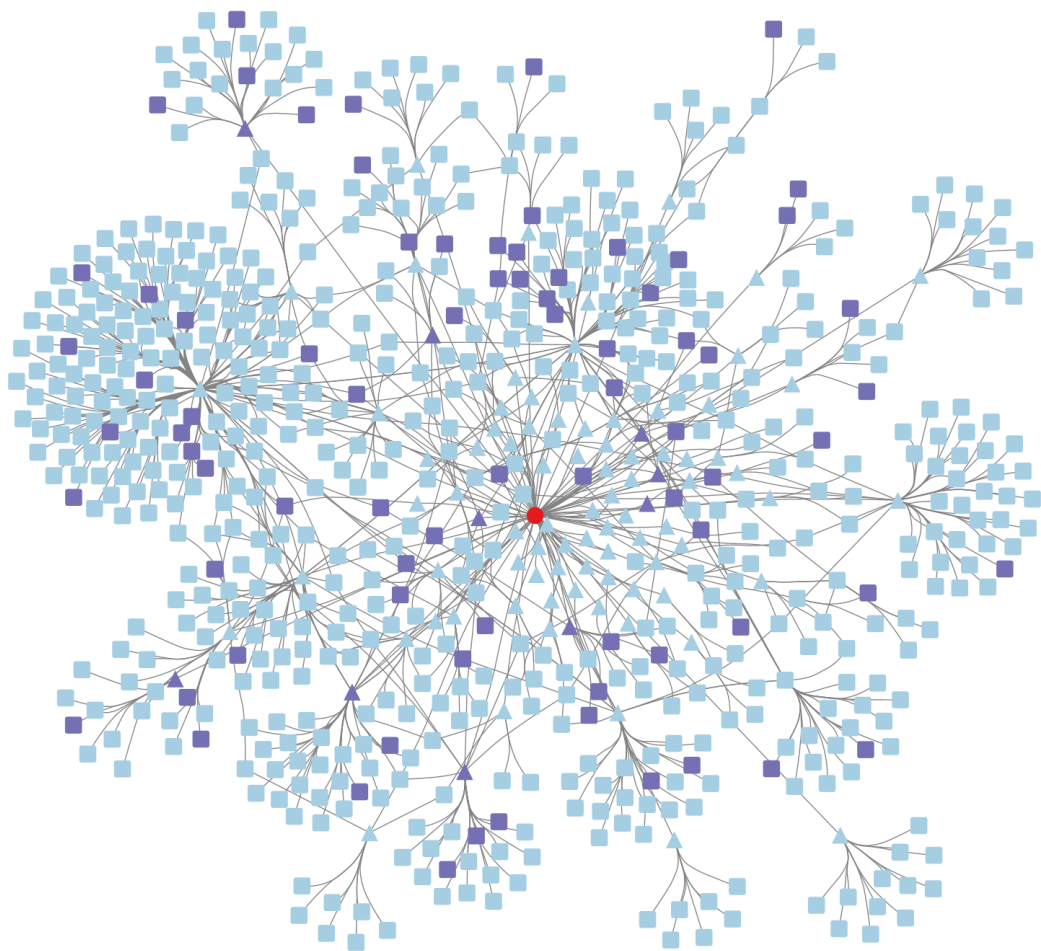
Ten kinases are present in the first layer interactome of MFHAS1, Serine/threonine-protein kinase 25 (STK25), Cyclin-G-associated kinase (GAK), Fibroblast growth factor receptor 1 (FGFR1), Dual specificity protein kinase CLK1 (CLK1), LIM domain kinase 1 (LIMK1), Inactive tyrosine-protein kinase transmembrane receptor ROR1 (ROR1), Serine/threonine-protein kinase Nek11 (NEK11), Aurora kinase B (AURKB), Serine/threonine-protein kinase Nek1 (NEK1) and Mitogen-activated protein kinase kinase kinase 4 (MAP3K4).

STK25, GAK, CLK1, LIMK1, ROR1, NEK11, AURKB, NEK1 and MAP3K4 were identified by protein microarray screen and are described in detail in Tomkins's paper (Tomkins *et al.*, 2018). FGFR1 is a tyrosine-protein kinase involved in small-cell lung cancer (Schultheis *et al.*, 2014).



**Figure 3.13:** First layer interactome of the proteins potentially interacting with MFHAS1. Kinases are in purple and circled. Data were collected from PINOT and HIPPIE tools then merged. Network is visualised with cytoscape software.

A second layer interactome was generated by submitting all the proteins from the first layer in PINOT to further analysis. 9255 proteins are present in the second layer; a filter was added to remove all the proteins, with a final score under four. Final scores are calculated by adding the number of method and publication references. Proteins interacting with MFHAS1 from the first layer all have a score of two, so they were added to the list after filtering. This adjusts the number of proteins in the second layer to 746 unique proteins. The filtering is added only to keep strong interaction. In the unfiltered list of the second layer, 345 kinases are present. The filtered network contains 746 proteins, including 90 kinases (**Figure 3.14**). Two other ROCO proteins are present in the list of kinases: LRRK2 and DAPK1 (**Table 3.5**). They are suggesting a common interactor between the three ROCO proteins. Tomkins and coworkers had previously identified five common interactors between the ROCO proteins using a microarray screen (CDC42EP3, VGLL4, BAG2, STUB1 and CBLB) (Tomkins *et al.*, 2018).

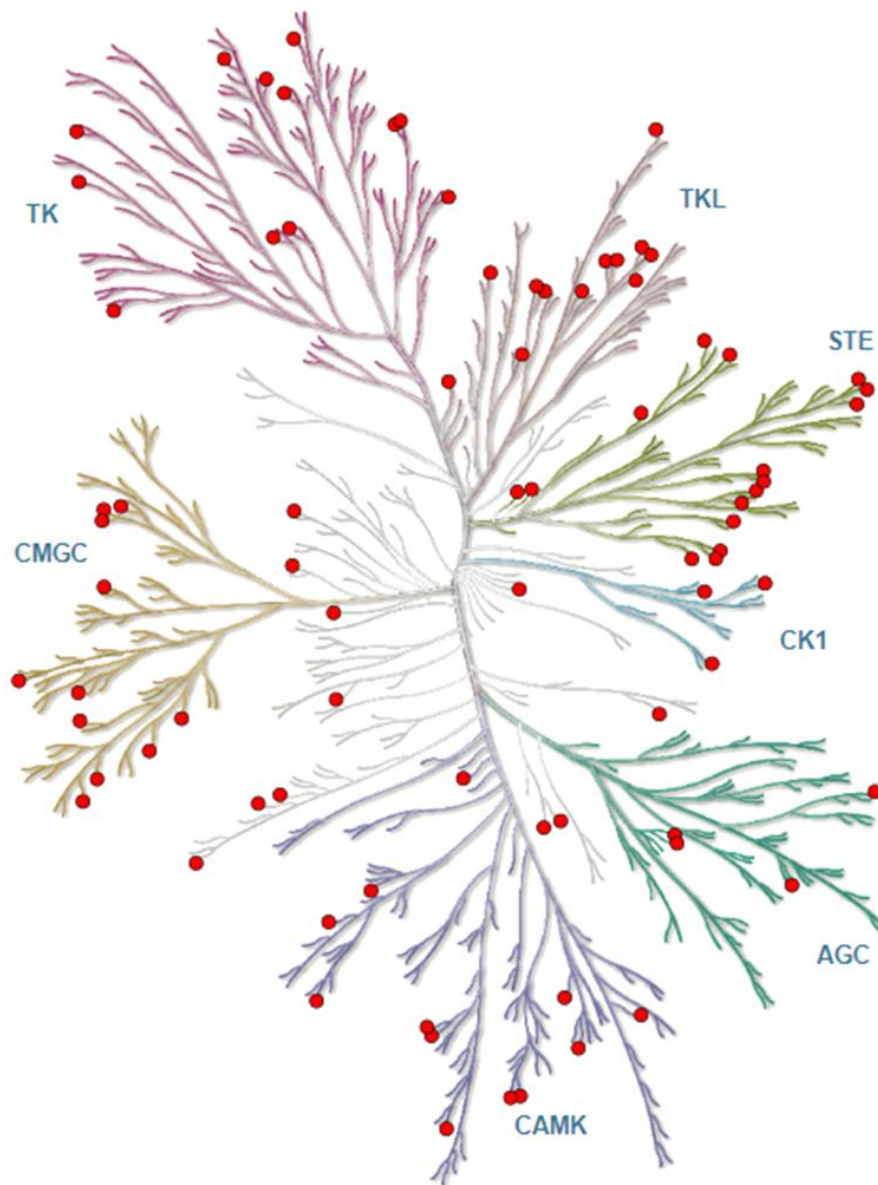


**Figure 3.14:** Network representation of the first and second layer interactome protein around MFHAS1. First layer proteins are represented with a triangle. Second layer proteins are represented with a square. Purple nodes are the kinases. MFHAS1 is indicated by the red round node at the centre of the network. Network built with Cytoscape.

ABL1	CDK4	HASPIN	MAP2K6	PAK1	SPEG
AKT2	CDK5	ILK	MAP3K11	PAK2	SRC
ARAF	CDK9	IRAK1	MAP3K14	PDK3	STK11
ATR	CLK1	JAK3	MAP3K2	PINK1	STK24
AURKB	CLK2	KSR1	MAP3K4	PLK1	STK25
AURKC	CLK3	KSR2	MAP3K5	PRKAA1	STK26
AXL	CSNK1E	LATS1	MAP3K8	PRKAA2	SYK
BCKDK	CSNK2A1	LATS2	MAPK1	PRKCZ	TAF1
BRAF	DAPK1	LCK	MAPK8	PRKDC	TAOK1
BRD4	EGFR	LIMK1	MOS	PTK2	TAOK2
CAMK2D	ERBB2	LIMK2	MYLK	RAF1	TRIM24
CAMKV	ERN1	LRRK2	NEK1	ROR1	TRIM28
CASK	FGFR1	MAP2K1	NEK11	RPS6KA1	TRRAP
CDK1	GAK	MAP2K3	NEK2	RPS6KA3	TTBK2
CDK20	GSK3B	MAP2K4	NTRK1	SIK2	VRK1

**Table 3.5:** Protein kinases present in the second layer interactome of MFHAS1. ROCO proteins LRRK2 and DAPK1 are highlighted in blue. Kinases present in the first layer are highlighted in red.

From the kinases present in the first and second layer of MFHAS1, the MAPK signalling pathway is well represented, with twelve kinases involved in MAP kinase pathways. All the kinases present in the second layer are mapped onto the kinome tree to understand their relationship to each other (Eid *et al.*, 2017). Kinases from the MFHAS1 interactome are present on all seven branches of the kinome tree with no specific group predominant (**Figure 3.15**).



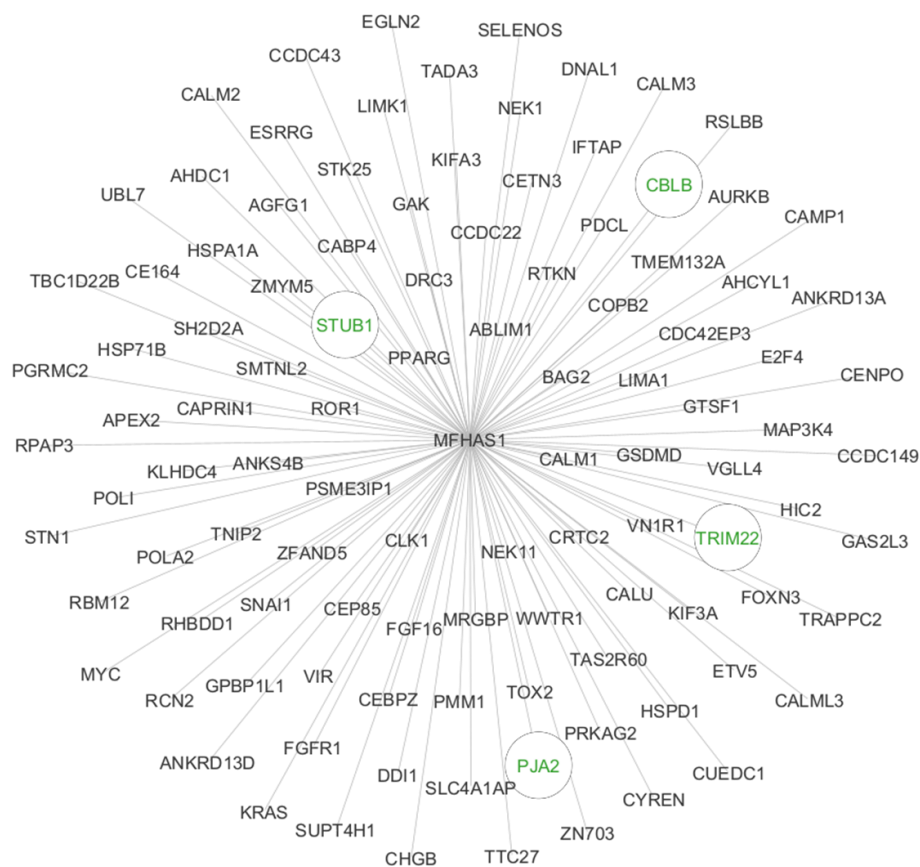
"Illustration reproduced courtesy of Cell Signaling Technology, Inc. ([www.cellsignal.com](http://www.cellsignal.com))"

**Figure 3.15:** Human Kinome tree. The 90 kinases present in the second layer of protein-protein interaction of MFHAS1 are pinpointed in red. Interactive KinMap from Kinhub (Eid *et al.*, 2017)

### 3.3.3.1.2 MFHAS1 interaction with the ubiquitination machinery

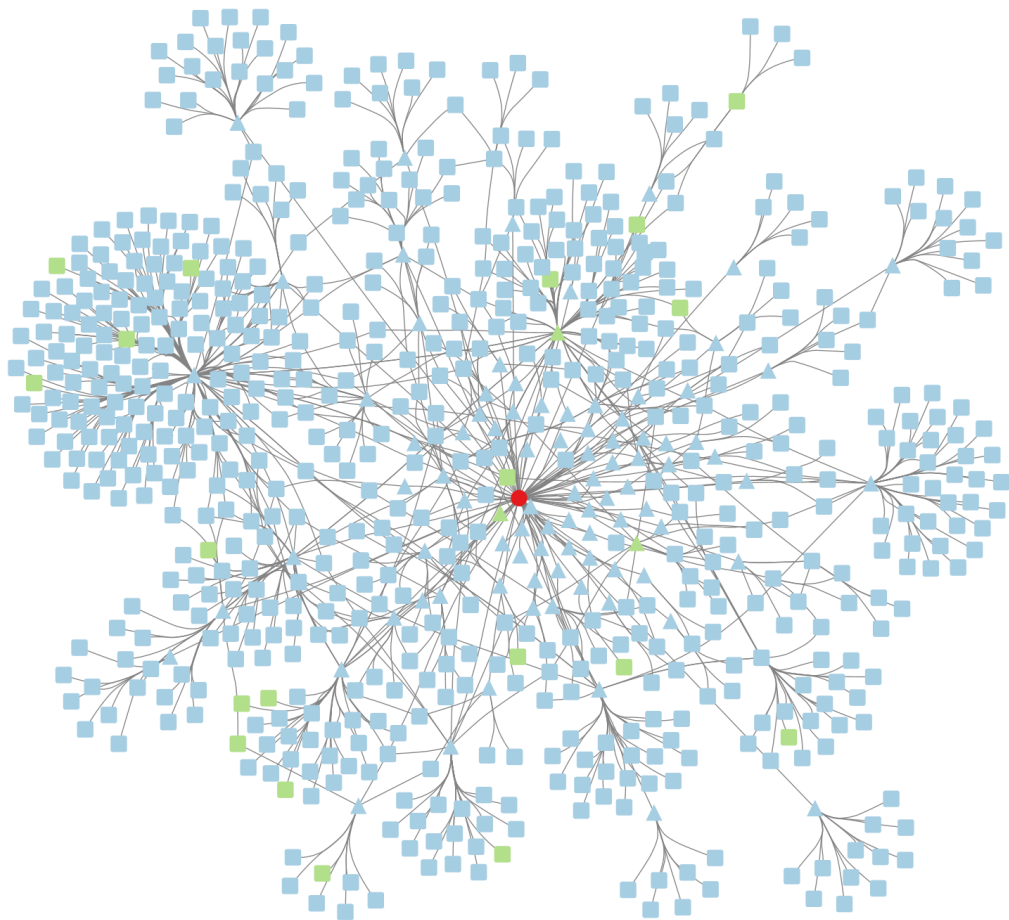
Proteins present in the first and second layer of MFHAS1 network are compared with a list of E3 ubiquitin- protein ligases (Medvar *et al.*, 2016). Interestingly, there were several E3 ligases that have been shown to interact with MFHAS1. E3 ubiquitin ligase act in few steps; first they recognise and bind to a target protein, then they facilitate the binding of a ubiquitin to the target protein. Finally, for the protein target that are poly-ubiquitinated, the E3 ligase elongates the chain of ubiquitin by adding ubiquitin protein (Zheng and Shabek, 2017).

In the first layer four E3 ubiquitin- protein ligases are found: E3 ubiquitin-protein ligase Casitas B-lineage lymphoma (CBLB), E3 ubiquitin-protein ligase Tripartite Motif 22 (TRIM22), E3 ubiquitin-protein ligase Carboxyl terminal of Hsp70-Interacting Protein (CHIP) also called STUB1 and E3 ubiquitin-protein ligase Praja-2 (PJA2) (**Figure 3.16**).



**Figure 3.16:** First layer interactome of the proteins potentially interacting with MFHAS1. Data from PINOT and HIPPIE are merged. The following E3 ubiquitin-protein ligases are found in this network are circled and written in green. Casitas B-lineage lymphoma (CBLB), Tripartite Motif 22 (TRIM22), Carboxyl terminal of Hsp70-Interacting Protein (CHIP) also called STUB1 and Praja-2 (PJA2). Network is visualised with Cytoscape software.

Microarray experiments were conducted to determine CBLB, TRIM22 and STUB1 protein interaction with MFHAS1 (Tomkins *et al.*, 2018), and a pull-down experiment was used to detect PJA2 (Zhong *et al.*, 2018). STUB1 and CBLB are common interactors between all four ROCO proteins (Tomkins *et al.*, 2018). CBLB is involved in innate and adaptive immunity (Liu, Langdon and Zhang, 2014; Tang, Langdon and Zhang, 2019). It also plays a role in Multiple Drug-Resistant Gastric and Breast Cancer Cells by inhibiting cell migration (Xu *et al.*, 2017). TRIM22 is found to have antiviral activities (Pagani, Poli and Vicenzi, 2021) and is involved in the regulation of p53 (Hatakeyama, 2011). STUB1 is involved in immunity (Zhan, Wang and Ge, 2017). It also functions as a central quality control in the cells (Paul and Ghosh, 2014). PJA2 inhibit the Wnt/ $\beta$ -catenin pathway (Song *et al.*, 2018) and is involved in gastric cancer (Zheng *et al.*, 2021). The second and first layer contains 126 E3 ubiquitin ligases in the unfiltered list of protein of the secondary layer. 22 E3 ubiquitin ligase proteins are present after implementing a threshold to keep only the proteins reported four times or more. E3 ubiquitin ligases are shown in green (**Figure 3.17**).



**Figure 3.17:** Network representation of the first and second layer interactome protein around MFHAS1. First layer proteins are represented with a triangle. Second layer proteins are represented with a square. Green nodes are the E3 ubiquitin ligases. MFHAS1 is indicated by the red round node at the centre of the network. Network built with Cytoscape.



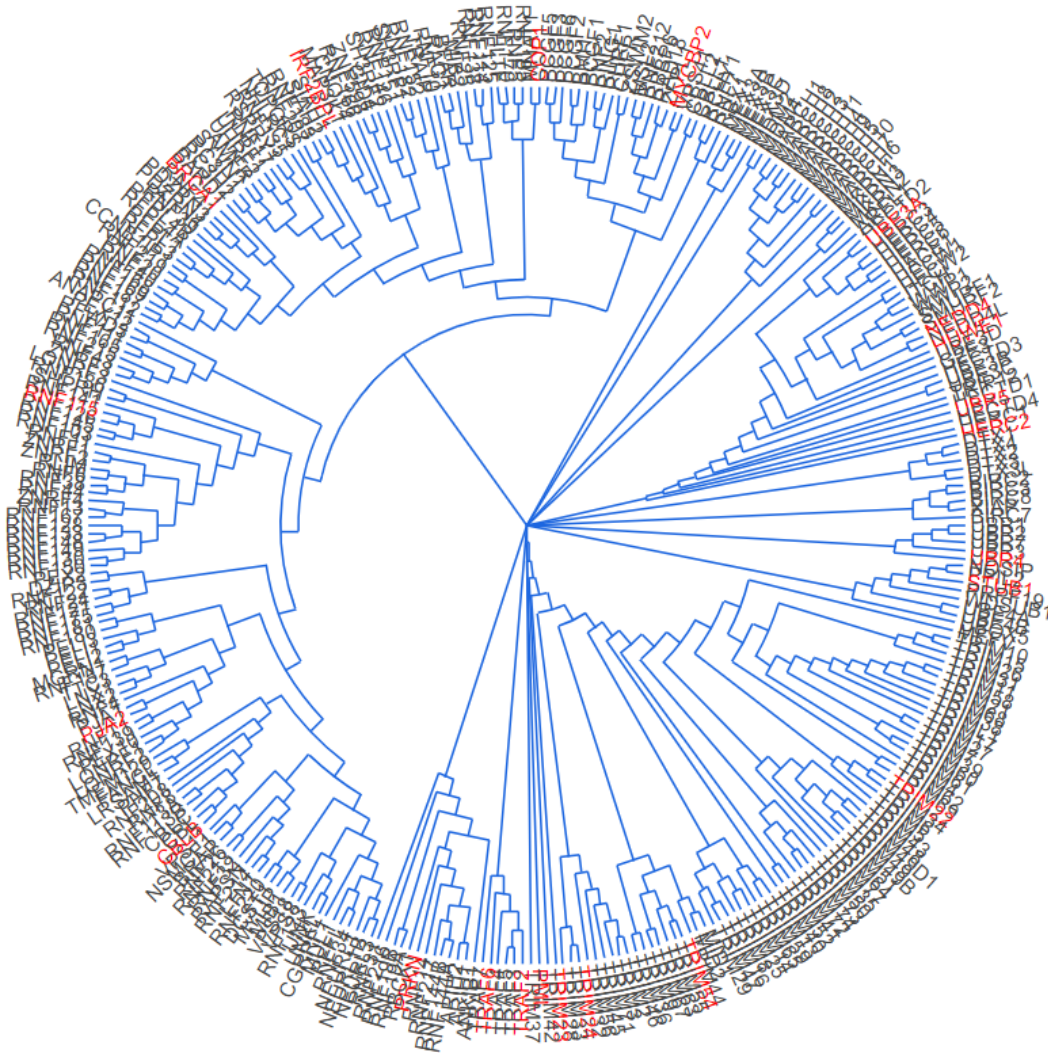
The 22 E3 ubiquitin ligases proteins and their main functions are summarised in **Table 3.6**.

<b>Uniprot</b>	<b>Full name</b>	<b>Main Implications</b>
<b>BRCA1</b>	Breast cancer type 1 susceptibility protein	DNA damage repair
<b>CBLB</b>	E3 ubiquitin-protein ligase CBL-B	negatively regulates T-cell receptor and B-cell receptor
<b>COP1</b>	E3 ubiquitin-protein ligase COP1	Jun, p53
<b>HERC2</b>	E3 ubiquitin-protein ligase HERC2	DNA damage
<b>HUWE1</b>	E3 ubiquitin-protein ligase HUWE1	MCL1, p53/TP53, H1,H2A/B, H3, H4
<b>IRF2BP2</b>	Interferon regulatory factor 2-binding protein 2	immature B cell differentiation
<b>MYCBP2</b>	E3 ubiquitin-protein ligase MYCBP2	Work with HUWE1 to regulate circadian clock gene expression
<b>NEDD4</b>	E3 ubiquitin-protein ligase NEDD4	proteasomal degradation
<b>PJA2</b>	E3 ubiquitin-protein ligase Praja-2	proteasomal degradation, activation of p38, JNK, MAP pathways, ciliogenesis
<b>PML</b>	Protein PML	protein sumoylation, cobalt ion binding
<b>PRKN</b>	E3 ubiquitin-protein ligase parkin	HSP70 protein binding, downstream of PINK1
<b>RNF115</b>	E3 ubiquitin-protein ligase RNF115	negative regulation of epidermal growth factor receptor signaling pathway
<b>STUB1</b>	E3 ubiquitin-protein ligase CHIP	enzyme activator
<b>TRAF2</b>	TNF receptor-associated factor 2	cell survival and apoptosis, ubiquitinat BIRC3, RIPK1, TICAM1
<b>TRAF6</b>	TNF receptor-associated factor 6	activation of NF-kappa-B and JUN
<b>TRIM22</b>	E3 ubiquitin-protein ligase TRIM22	immunity
<b>TRIM23</b>	E3 ubiquitin-protein ligase TRIM23	autophagy during viral infection
<b>TRIM24</b>	Transcription intermediary factor 1-alpha	chromatin binding, ubiquitinat for degradation
<b>TRIM54</b>	Tripartite motif-containing protein 54	microtubule binding
<b>UBE3A</b>	Ubiquitin-protein ligase E3A	brain development
<b>UBR4</b>	E3 ubiquitin-protein ligase UBR4	calmodulin binding
<b>UBR5</b>	E3 ubiquitin-protein ligase UBR5	DNA repair

Table 3.6: E3 ubiquitin ligase present in the second layer interactome of MFHAS1.



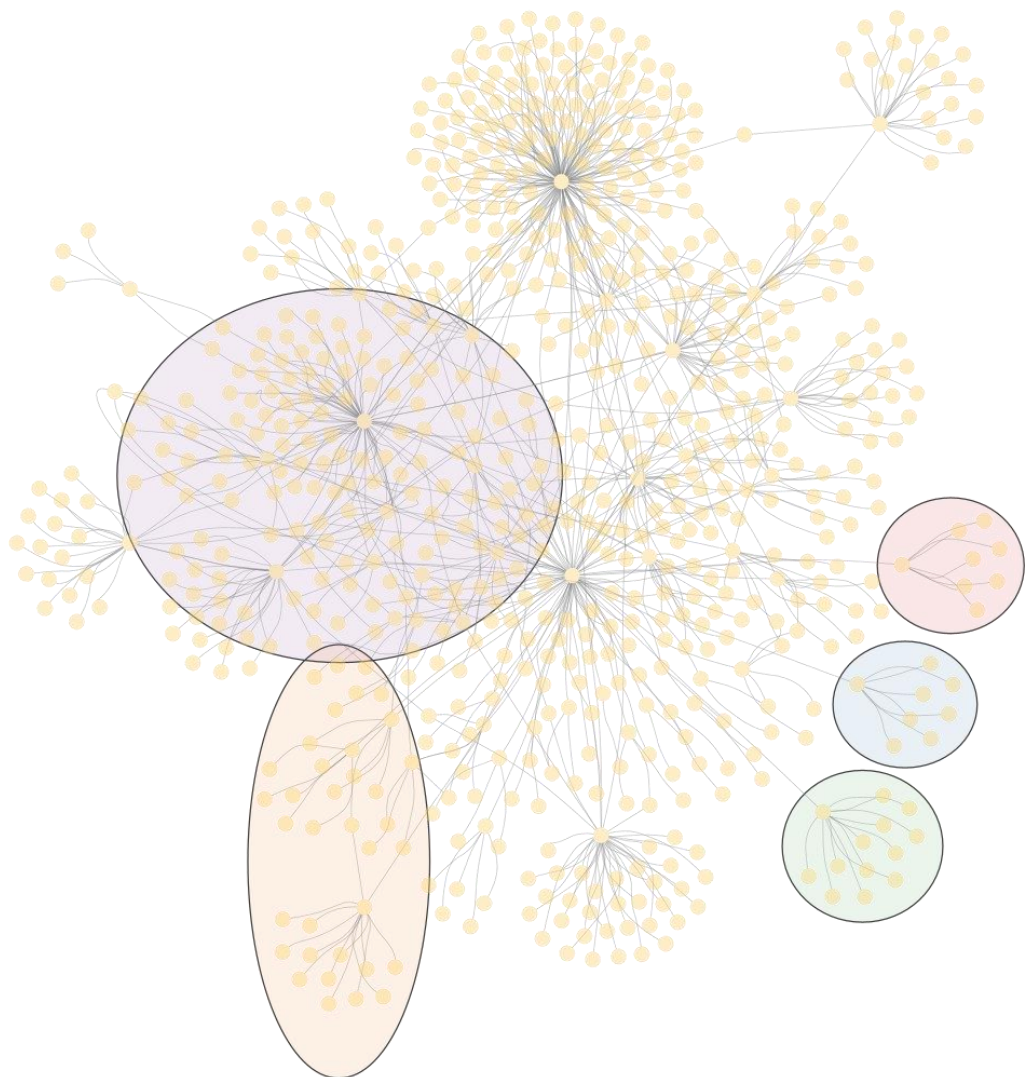
Phylogenetic tree of all the E3 ubiquitin ligases protein was used to visualised where the 22 E3 ligases present in the MFHAS1 PPI seats (Liu *et al.*, 2019). E3 ubiquitin are highlighted in red (**Figure 3.18**). The E3 ligase proteins are present in different parts of the tree with no particular branch predominant.



**Figure 3.18:** Phylogenetic tree of the E3 ubiquitin ligase based on the alignment of the full-length proteins. Proteins present in the second layer interactome of MFHAS1 are highlighted in red. Phylogenetic tree made by ubihub (Liu *et al.*, 2019).

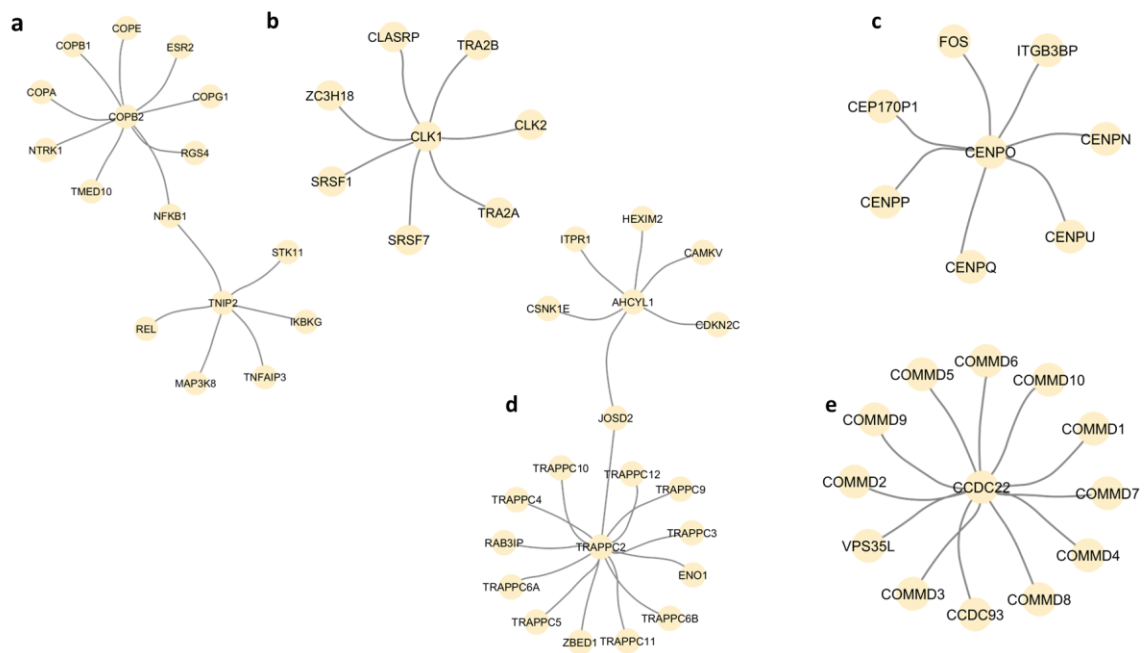
### 3.3.3.1.3 Clusters analysis

To further analyse the second layer interactome of MFHAS1, a cluster analysis was performed with the Auto annotate tool in Cytoscape (**Figure 3.19**). This analysis groups proteins from the interactome which interact with each other (showing a high degree of connectedness) into clusters. Clusters present in the second layer of MFHAS1 disclose information about which proteins around MFHAS1 are forming a complex and which function they are involve in. It is interesting to perform to understand which protein complex and pathways are connected to MFHAS1.



**Figure 3.19:** Cluster analysis of the second layer protein-protein interaction of MFHAS1. Five clusters were found shown in purple, orange, red, blue, and green.

Each cluster is isolated and analysed individually in **Figure 3.20**.



**Figure 3.20:** Isolation and detail of clusters present in MFHAS1 second layer of protein-protein interaction.

The 16 nodes cluster contains the proteins, COPB2, COPB1, COPE, ESR2, COPG1, RGS4, TMED10, NTRK1, COPA, Nuclear Factor kappa-light-chain-enhancer of activated B cells 1 (NF- $\kappa$ B1), TNIP2, STK11, IKKKG, TNFAIP3, MAP3K8 and REL (**Figure 3.20.a**). Using affinity purification mass spectrometry, TNP2 is found to up and down regulates NF-B. TNP2 is also involve with cellular transport machinery and RNA transcript processing (Banks *et al.*, 2016). A functional map of TNF/ NF- $\kappa$ B1 identified 221 interactors including some proteins present in this cluster (Bouwmeester *et al.*, 2004). TNIP2, MAP3K8 and NF- $\kappa$ B1 p105 form a ternary complex called Tpl2 (Banks *et al.*, 2016). Bioinformatic tools demonstrate correlation between MAP3K8 overexpression and NF- $\kappa$ B inflammation in cancer (Hao *et al.*, 2021).

COPA, COPB1, COPB2, COPE, and COPG1 are part of the coat protein complex I (COPI) (Futatsumori *et al.*, 2000). COPI complex contains seven subunits, and its formation is initiated by the small GTPase Arf1. The COPI complex is part of the retrograde transport, transporting lipids and proteins from the Golgi to the endoplasmic reticulum (Arakel and Schwappach, 2018). The structure of the COPI complex was described within the cell (Bykov

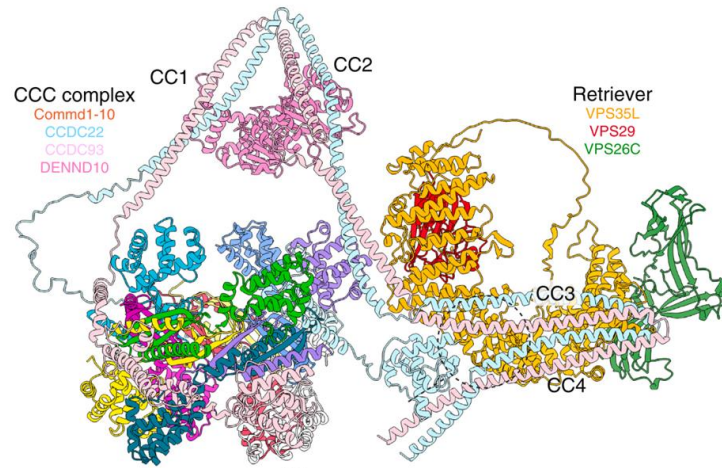
*et al.*, 2017) and a 9 Å structure was defined by cryo-electron tomography (Dodonova *et al.*, 2017).

CLK1, CLK2, TRA2A, TRA2B, SRSF7, SRSF1, ZC3H18 and CLASRP proteins are part of an eight proteins cluster. All are involve in cell cycle and splicing regulation of mRNA (Dominguez *et al.*, 2016; Winczura *et al.*, 2018). (**Figure 3.20.b**). CDC2-like kinases (CLK) 1 phosphorylates serine/arginine rich protein (SR) like SRSF1 and SRSF7 present in the cluster (Das and Krainer, 2014).

The eight proteins cluster contains CENPO, FOS, ITGB3BP, CENPN, CENPU, CENPQ, CENPP and CEP170P1 (**Figure 3.20.c**). Many proteins of this cluster are centromere protein (Su *et al.*, 2022). Centromere protein (CENP) N, U,O, P and Q are involve in centromere kinetochore complex (Liu and Liu, 2022).

Twenty nodes AHCYL1, CSNK1E, ITPR1, HEXIM2, CAMKV, CDKN2C, JOSD2, TRAPPC 2,3,4,5,6A,6B,9,10,11,12, ENO1, ZBED1 and RAB3IP (**Figure 3.20.d**). Transport protein particle (TRAPP) proteins can indirectly activate Rab1 GTPase (Brunet and Sacher, 2014). TRAPPC 2,3,4,5,6A,6B,9,10 are part of the TRAPP II complex which play a role with COPI coat complex in the pre-Golgi compartment (Barrowman *et al.*, 2010).

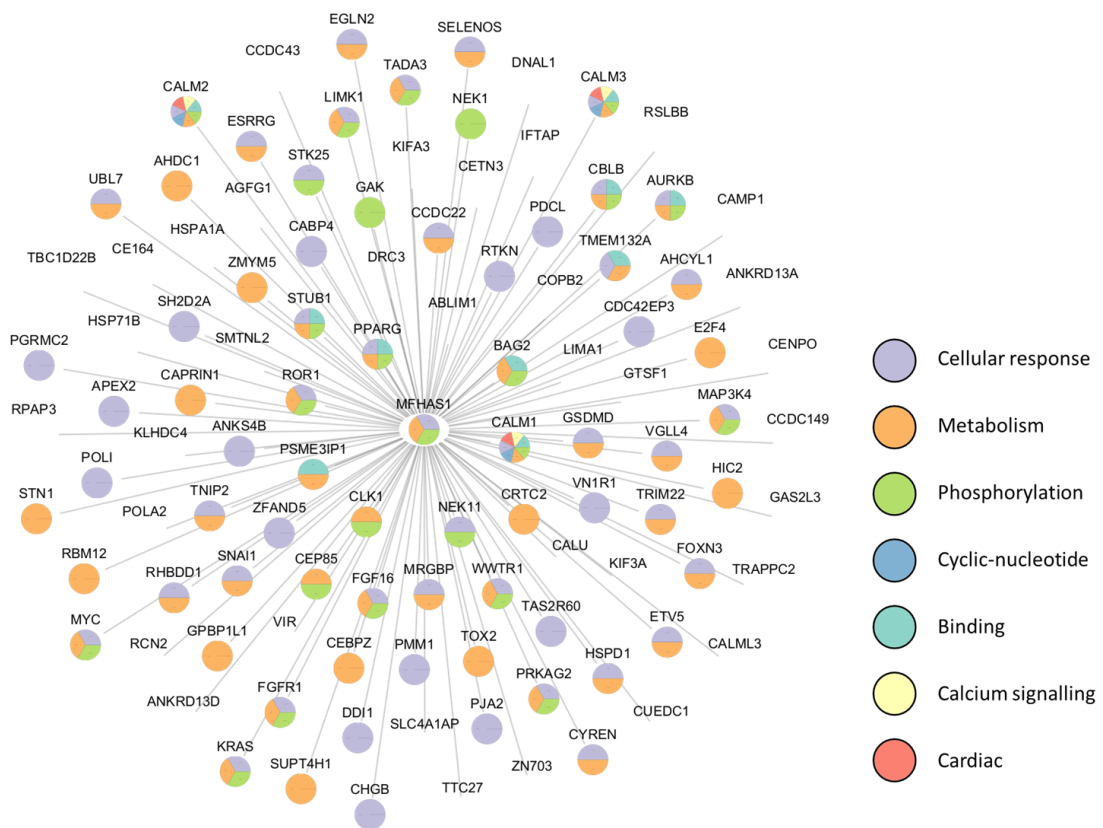
A 13 nodes cluster contains the proteins, CCDC22, COMMD1, COMMD2, COMMD3, COMMD4, COMMD5, COMMD6, COMMD7, COMMD8, COMMD9, COMMD10, CCDC93, and VPS35L (**Figure 3.20.e**). All those protein form a complex called the commander complex (Laulumaa and Varjosalo, 2021). Commander complex is involved in endosomal recycling and linked to Ritscher-Schinzel syndrome. A structure of the commander complex is shown in **Figure 3.21** (Healy *et al.*, 2023).



**Figure 3.21:** Structure of the commander complex from (Healy *et al.*, 2023). Cryo-Electro microscopy, crystallography and AlphaFold2 were used to assemble this structure.

### 3.3.3.2 Functional categorisation

The list of 105 proteins interacting with MFHAS1 is analysed for functional enrichment using g:profiler web software. Data from g:profiler groups the protein by their common function (term). Terms with similar functions were grouped into more generic terms. Seven terms were selected: metabolism, cellular response, cyclic nucleotide, binding, calcium signalling, phosphorylation and cardiac. A pie chart is created for each protein with one or many functions (Figure 3. 22).



**Figure 3.22:** Functional enrichment of MFHAS1 first layer interaction using g:profiler. Seven common functions are found in the network and are represented in a pie chart for each protein involved. Network was constructed using Cytoscape.

Metabolism and binding are generic terms. They do not give any biological insight into specific functions of the network. CALM1, CALM2, and CALM3 are involved in cardiac, calcium signalling and cyclic nucleotide, which they do not share with other proteins. These three functions only present in the CALM proteins are not representative of the functions present in the network.

Functional enrichment data demonstrates that MFHAS1 is involved in metabolism, phosphorylation, and cellular response. Phosphorylation also groups protein modifications. The cellular response comprises a response to stress and stimulus.

MAP3K4 functions are also metabolism, phosphorylation, and cellular response. PJA2 is only present in cellular response.

Cellular response and phosphorylation are the two main functions present in the network.

Functional enrichment analysis also yields information about the pathways present around MFHAS1 using the Kyoto Encyclopedia of Genes and Genomes pathways (KEGG). Cancer-related pathways are present in the first layer interactome of MFHAS1, including “Glioma” and “Melanoma”. Ras and Rap1 signalling pathways are also reported. A complete list of the pathways can be found in **Supplementary Figure 5**.



### 3.3.3.3 Gene expression

From the list of kinases and E3 ubiquitin ligases that interact with MFHAS1, MAP3K4 and PJA2 were selected to perform further research. PJA2 is selected for its potential interaction with the MFHAS1 K387 ubiquitination site (Zhong *et al.*, 2018). The same paper hypothesises an activation of the MAPK signalling by MFHAS1.

In order to assess potential co-expression of these proteins, implying a shared function, the gene expression of *PJA2* and *MAP3K4* were assessed and compared with *MFHAS1*. Data were collected from the GTEX portal and are measured with Transcripts Per Kilobase Million reads (TPM). Proteins expressed in the same tissue increase the probability of the proteins being implicated in a similar biological processes and pathways.

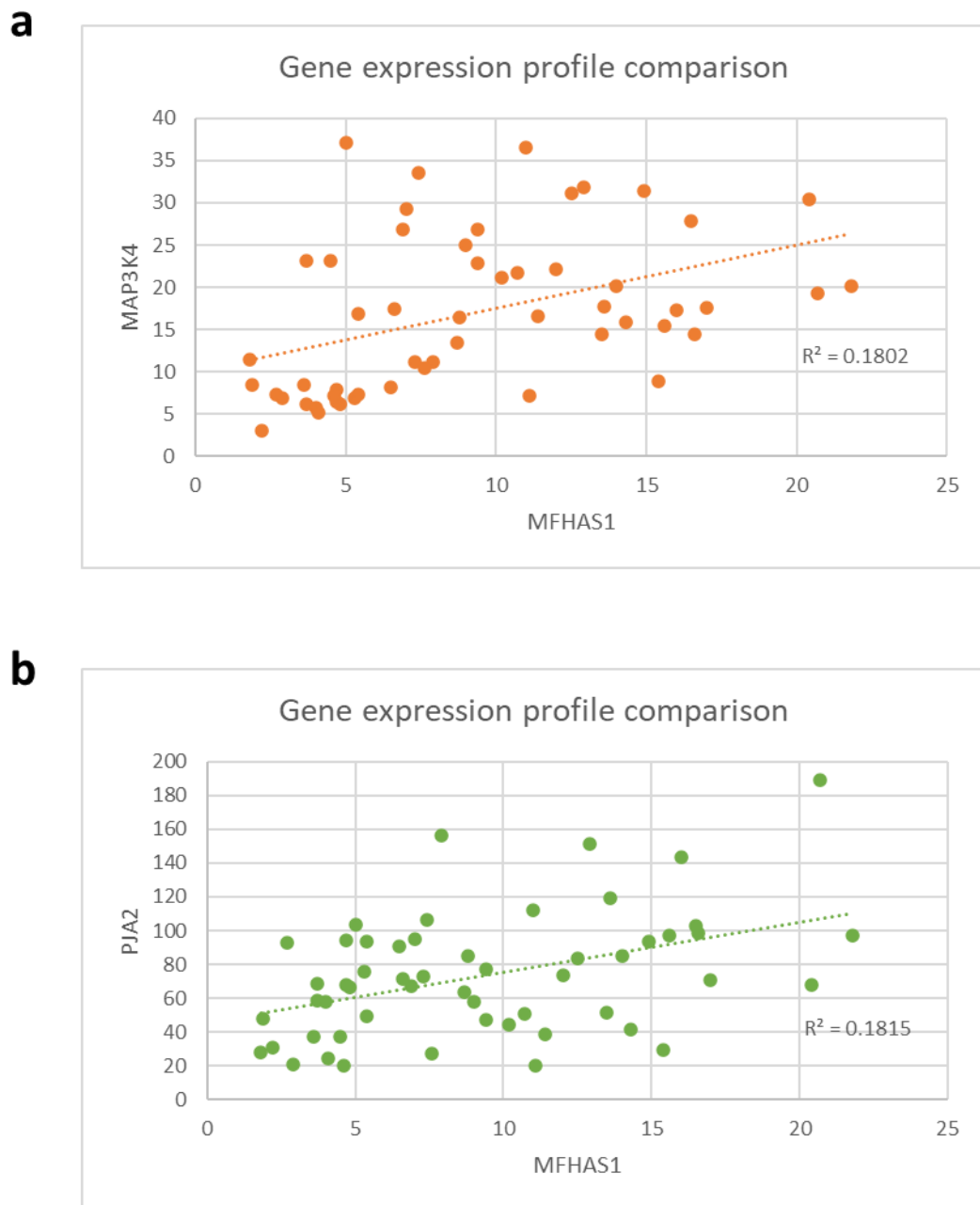
MAP3K4 is mainly expressed in the nerve–tibial (Median TPM: 37.12), uterus (Median TPM:36.64) and cervix–ectocervix (Median TPM: 33.45). PJA2 expression is the highest in the artery–tibial (Median TPM: 189.1), the brain–frontal cortex (Median TPM: 156.7) and the cerebellar hemisphere (Median TPM: 151.4). MFHAS1 gene is expressed principally in the colon – sigmoid (Median TPM: 21.77), the artery – tibial (Median TPM: 20.72) and cells – EBV transformed lymphocytes (Median TPM: 20.41) (**Figure 3. 23**). It is worth noting that mRNA and protein expression do not always correlate.

MFHAS1 gene expression is overall lower than MAP3K4 and PJA2. MFHAS1 and MAP3K4's highest expression tissues do not match, but the expression levels of MAP3K4 in the colon and artery are 20 and 19, respectively, which is close to the MFHAS1 data. PJA2 and MFHAS1 share the artery tissues as the highest expression tissue for PJA2 and second for MFHAS1. Correlation analysis between MAP3K4 and MFHAS1 data median across all tissues indicates a correlation of 0.4244. The correlation between PJA2 and MFHAS1 presents a r value of 0.4260. A number above 0 and the closest to 1 indicates a positive correlation between the two data sets.



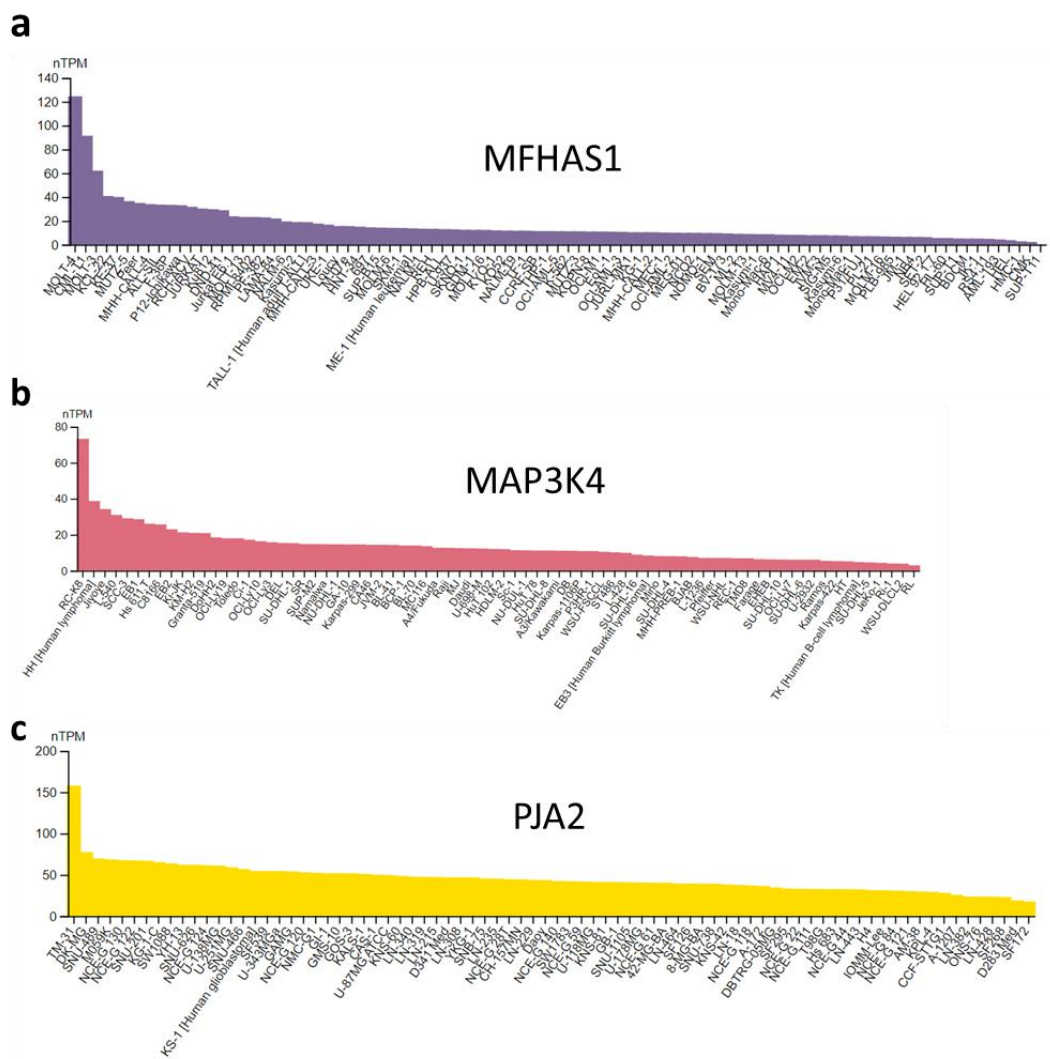


Scattered plots are designed with the data from the gene expression profile to show the correlation of gene expression between different tissues (**Figure 3.24**). The data indicates that the gene expression of MFHAS1 and MAP3K4 are not correlated with a  $r^2$  value of 0.1802. MFHAS1 and PJA2 gene expression are also not correlated with a  $r^2$  value of 0.1815.



**Figure 3.24:** Scattered plot of gene expression profile from the GTEX portal. **(a)** comparison of MFHAS1 and MAP3K4. **(b):** comparison of MFHAS1 with PJA2.

Dataset from the human protein Atlas presents the cell lines in which the protein has the highest expression. High expression in the same cell could mean a higher probability for the proteins to interact and provides information relevant for future cellular modelling (see chapter 4). MFHAS1 RNA is expressed the highest in the MOLT-4 cell line with a nTPM of 124.5. This cell line is derived from leukaemia. MAP3K4 is predominantly expressed in RC-K8 established from lymphoma with 73.2 nTPM. PJA2 is highly expressed in TM-31, a cell line derived from brain cancer; its nTPM is 157.7 (**Figure 3.25**).



**Figure 3.25:** RNA expression data in cell lines. (a): MFHAS1 (b): MAP3K4 (c): PJA2. Data extracted from the human protein atlas.

### 3.4 Discussion

In this chapter, bioinformatic approaches were applied in order to gain a deeper understanding of the function and cellular role of MFHAS1. Structure modelling software AF2 and IntFOLD were used to model the structure of the MFHAS1 protein. PTMs were studied by comparison with other ROCO proteins and kinases prediction to understand the role of phosphorylation and ubiquitination on MFHAS1. The first and second layers of PPI around MFHAS1 were analysed for functional enrichment and to understand the protein interacting with MFHAS1.

Modelling structure provides an insight into how MFHAS1 protein would look like in three dimensions. Two model structures of LRRK2 from Alfold2 and IntFOLD were compared and accessed with a known protein structure of LRRK2. The range of the RMSD between the model and the laboratory structure was high (6 Å and 8 Å), meaning that the modelling structure differed from the cryo-EM structure. A RMSD value of zero implies that the structures are identical. The higher the RMSD value increases, the more difference there is between the two structures, typically a RMSD below 3 Å is considered to be a good alignment (Kukol, 2015).

The two MFHAS1 predicted structures are not very similar to the CtROCO protein structure, presenting an RMSD of 8.5 Å and 7.9 Å. Predicted structures are built using multiple sequence alignments from proteins with similar sequences and deep learning approaches (Jisna and Jayaraj, 2021). In the case of MFHAS1, CtROCO has a similar domain architecture and is used to model MFHAS1. Using CtROCO for modelling structure could explain the overlaps in conformation between the two proteins.

The modelled structures of MFHAS1 from AF2 and IntFOLD7 present an RMSD of 1.182 Å, meaning that the two models are very similar, reinforcing the probability of the prediction being close to a physiologically relevant structure.

However, there is a limitation to those models; predictions only present a monomeric structure. CtROCO and LRRK2 experimentally derived structures are found under monomeric and dimeric forms. Having a dimeric structure for MFHAS1 would be interesting to understand its conformation and where PTM are sitting. Another limitation of the

predicted structure is the disordered regions, which cannot be predicted. To define the conformation of these regions, cryo-EM needs to be used.

Overall, predicted structures are a powerful alternative to experimentally derived structures, and in the context of these investigations, have allowed a greater understanding of where PTM sit within the protein. Focusing on the phosphorylation site S450, the predicted structures of MFHAS1 shows that the phosphorylation site is present on a disordered region of the ROC domain. The region is situated on the outside of the protein, allowing space for potential interactions, and possibly alterations in conformation. Ubiquitination sites K387 and K601 are both present on an ordered region of MFHAS1. Interestingly, this follows a pattern of phosphorylation sites present in disordered regions and ubiquitination sites in ordered regions that has been previously reported (Bludau *et al.*, 2022).

Post-translational modification and protein-protein interaction results can be assessed in combination.

PPI analysis found 105 proteins in the first layer of MFHAS1 using PINOT and HIPPIE, which means that 105 proteins are potentially directly interacting with MFHAS1. The proteins interacting with the first layer form the second layer interactome. The second layer contains 746 proteins.

Phosphorylation is a crucial PTM that is part of important physiological processes such as regulating metabolism, cell proliferation and apoptosis. The phosphorylation site S450 is the leading reported PTM site of MFHAS1, with five references. The serine is flanked by three prolines, forming the motif proline-proline-serine-proline. The PPSP motif is reported as an essential motif for signalling in the low-density lipoprotein receptor-related protein (LRP6) (Brennan *et al.*, 2004).

S450 aligns with DAPK1 (S734), which also contains the PPSP motif. DAPK1 (S734) is a phosphorylation site referenced nine times on PhosphoSitePlus® (seven HTP and two LTP). Serine 734 of DAPK1 is phosphorylated by ERK, leading to the activation of p53 (Chen *et al.*, 2005; Kwon *et al.*, 2016). MFHAS1 is reportedly implicated in the Raf/MEK/ERK pathway by interacting with Raf1 (Kumkhaek *et al.*, 2013) and HSP70 (Kumkhaek *et al.*, 2019).

The kinase prediction software shows that KIS is potentially phosphorylating MFHAS1 at S450. KIS kinase is present in leukaemia cells, including the MOLT4 cell line. A High KIS protein level stimulates the cell cycle progression in leukaemia cells (Nakamura *et al.*, 2008). It is interesting to note that MFHAS1 expression is the highest in the MOLT4 leukaemia cell line. This site is further investigated in the cellular biology chapter (Chapter 4).

Ten kinases are present in the first layer and 90 in the second layer interactome. The potential kinases interacting with MFHAS1 are of primary interest and can be verified by immunoprecipitation. The focus is brought to the MAP3K4 kinase present in the first layer interactome. MFHAS1 was reported to potentially interact with a MAPK protein (Zhong *et al.*, 2018). MAPK1 was tested for interaction but was not found to bind to MFHAS1 (Kumkhaek *et al.*, 2019).

The kinases predicted by the kinase prediction software are not found in the PPI network from PINOT. The kinase prediction is based on the amino acid sequence and might not be accurate. The PINOT data are based on experimental data, mostly from microarray or mass spectrometry for the proteins around MFHAS1. Both datasets give an idea of the proteins that may be around MFHAS1, but those data need to be confirmed through experiments such as immunoprecipitation.

Proteins found to interact with MFHAS1 from PINOT and HIPPIE in the first layer have been experimentally described once and might need to be validated with low throughput techniques (co-immunoprecipitation).

KRas4B (S181) is phosphorylated by the protein kinase C (PKC), causing Ras trafficking and cell death (Xiang *et al.*, 2017). Interestingly, the PKC is a predicted kinase for MFHAS1 (S644) and (S844). PKC is part of regulating chromatin in the nucleus (Lim, Sutton and Rao, 2015).

LRRK2 and DAPK1 are present in the second layer interactome of MFHAS1, suggesting that the ROCO proteins may indirectly interact with each other. Common interactors between the ROCO protein may involve them in common pathways (Tomkins *et al.*, 2018).

Ubiquitination is a reversible PTM, involved in a number of cellular processes (Dikic and Schulman, 2022). Four E3 ubiquitin ligases are present in the first layer interaction and 22

in the second layer. Ubiquitination site K601 of MFHAS1 is aligned with DAPK1 (K939), a reported ubiquitination site through HTP methods. K939 has no known function. Overall, the ubiquitination sites are poorly conserved across the ROCO proteins. Few ubiquitination sites are reported for the Ras proteins, which either activate, inhibit or are involved in protein degradation and Ras trafficking and cell death (Xiang *et al.*, 2017).

Examining the literature, PJA2 might be ubiquitinating MFHAS1 in the region 64-364. A comparison of those residues to the reported ubiquitination site of MFHAS1 in Phosphositeplus® deduces it might be the ubiquitination site of MFHAS1 K387 (Zhong *et al.*, 2018). This paper also found that the ubiquitination of MFHAS1 by PJA2 induces the JNK and p38 pathways, which induced M1 macrophage polarisation and M1 to M2 macrophage transformation. PJA2 is also present in the first layer interactome of MFHAS1 from the PPI data.

Interestingly, five common interactors of all four of the ROCO proteins were found by microarray screens (Tomkins *et al.*, 2018): BAG2, CBLB, CDC42EP3, STUB1, and VGLL4. STUB1 (also called CHIP) and CBLB are both E3 ubiquitin ligases present in the first layer interactome of MFHAS1. For LRRK2 and DAPK1, STUB1 forms a complex with HSP90 (Zhang, Nephew and Gallagher, 2007; Ko *et al.*, 2009; Chen and Wu, 2018). It would be interesting to understand if MFHAS1 would form a similar complex and interact with HSP90.

A functional enrichment and clustering analysis were performed to further understand the proteins interacting with MFHAS1.

Functional enrichment of the first layer found seven specific functions. MFHAS1 is involved in metabolism, phosphorylation-protein modification and cellular response. MFHAS1 function of cellular response is to stress or stimulus. It is interesting to note that the MAPK signalling can be stimulated by cellular stress (Alam, Uddin and Brown, 2012).

Overall, the protein's function in the first layer mainly involves protein signalling.

Looking at the KEGG data from g: profiler, which gives an insight into which pathways are present in the network, many pathways are related to cancers, and others are related to the Ras and Rap1 proteins.

Understanding the function of the clusters present in the second layer interactome of MFHAS1 gives an insight into the function of MFHAS1 protein. Six clusters were found in the second layer interactome of MFHAS1.

The COPI complex, whose role is in the retrograde transport from the Golgi to the endoplasmic reticulum, and the TRAPPII complex are present. COPI and TRAPPII complexes are reported to interact and work together to transport liquid droplets (Li *et al.*, 2017). The commander complex is also present and involved in endosomal recycling.

In addition, two clusters are involved with histones and centromere. One cluster contains protein with cell cycle function and mRNA splicing regulation.

From the function of these clusters, MFHAS1 can be hypothesised to be involved in cell division and transport between the reticulum endoplasmic and the Golgi. This is an area of MFHAS1 cellular function that merits further investigation.

Analysis of the bioinformatics data gives a direction to find proteins and pathways potentially implicated with MFHAS1. This bioinformatics chapter is a preamble for the cell biology chapter (Chapter 4). Some proteins found with bioinformatic approaches are selected for further study.



## Chapter 4: MFHAS1 signalling pathways.

### 4.1 Introduction

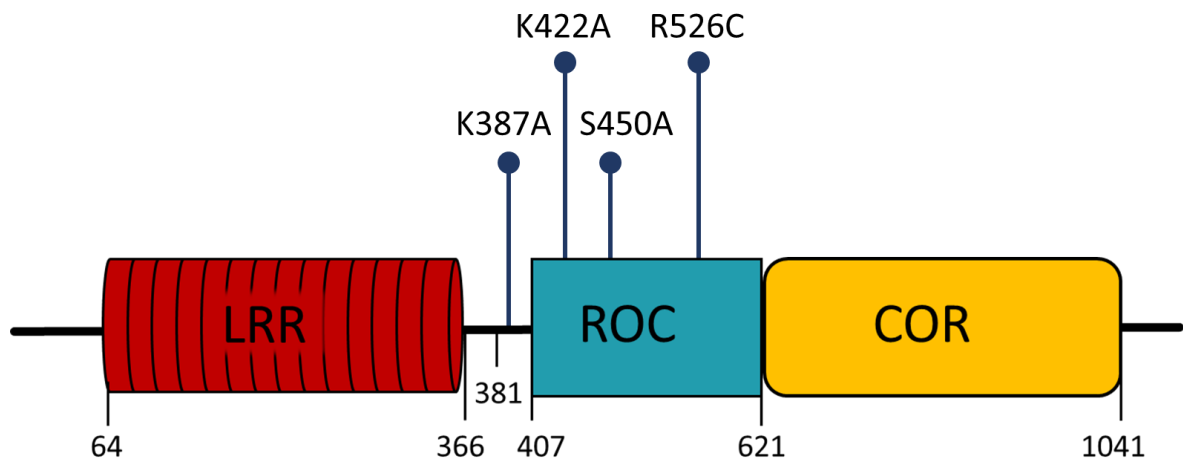
MFHAS1 is the least studied human ROCO protein (Dihanich, 2012). As of September 2023, the PubMed database counts 38 papers for the terms “MFHAS1” or “MASL1” against 3245 for “LRRK2”, 670 for “DAPK1”, and 99 for “LRRK1”. Despite those numbers, MFHAS1 is a noteworthy protein to study predominantly for its link to cancer and for the wider context of its membership of the ROCO family (Bosgraaf and Van Haastert, 2003). The MFHAS1 gene was first identified in 1999 as an oncogene (Sakabe *et al.*, 1999). Since then, a little over ten studies have found the MFHAS1 gene to be involved in various cancers, but the molecular pathways through which MFHAS1 is implicated in the aetiology of cancer are yet to be defined. MFHAS1 has also been implicated in sepsis via activation of the JNK and p38 mitogen-activated protein kinase pathways (Shi *et al.*, 2017; Zhong *et al.*, 2018), both of which are central to main cellular signal transduction pathways. A summary of the pathways in which MFHAS1 is involved can be found in Chapter 1.6.4.2. As is clear from this, a deeper understanding of these pathways will yield a better understanding of this protein’s function.

The investigation of MFHAS1 pathways is based on the literature and the previous findings from the bioinformatics chapter (Chapter 3.2). As a result of this research, the focus was oriented on various residues in the MFHAS1 protein, which have key roles in regulating the signalling function such as kinase function and phosphorylation activity, and possibly the stability or the activity of the protein, for instance, ubiquitination sites.

This part of the study aimed to generate point mutations in the coding sequence of MFHAS1 and examine whether this impacts upon the known functions or biochemical properties of the protein. As established in the earlier chapters, ubiquitination of MFHAS1 and phosphorylation of the protein at specific residues might significantly impact the protein’s stability or downstream signalling pathways.

Mutagenesis of residues potentially crucial for the protein function followed by cellular and biochemical analysis is a well-established method (Cole and Prabakaran, 2020).

A series of protein mutants are designed: K387A, K422A, S450A and R526C (**Figure 4.1**). The mutated protein is expressed and used in immunoblot and immunoprecipitation experiments.



**Figure 4.1:** Schematic of MFHAS1 protein and the four mutations investigated in this chapter; K387A, K422A, S450A and R526A. Domain delineations were defined in Chapter 2. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. LRR: Leucine-Rich Repeat. ROC: Ras Of Complex protein. COR: C terminal Of Roc.

K387 is a ubiquitination site reported on PhosphoSitePlus. This site was chosen based on a paper reporting the interaction of MFHAS1 with the E3 ubiquitin ligase PJA2 (Zhong *et al.*, 2018). They found that PJA2 would ubiquitinate MFHAS1 within the amino acids 64 and 364. As no ubiquitination sites are reported in this sequence, the closest site, K387, was selected as a probable target for ubiquitination. By mutating K387, the interaction between PJA2 and MFHAS1 may decrease or be ablated. Alternatively, mutating this residue (K387A) could show if PJA2 ubiquitinates MFHAS1 at the K387 site and whether this affects the steady-state level of MFHAS1.

MFHAS1 phosphorylation site S450 is reported five times on PhosphoSitePlus (Yu *et al.*, 2007; Olsen *et al.*, 2010; Rigbolt *et al.*, 2011; Beli *et al.*, 2012), which is this protein's most reported PTM site. It is located in the Switch I portion of the ROC domain. Examining the impact of the S450A mutation in mammalian cell models will help to understand the role of phosphorylation on MFHAS1 and its impacts on the protein's expression.

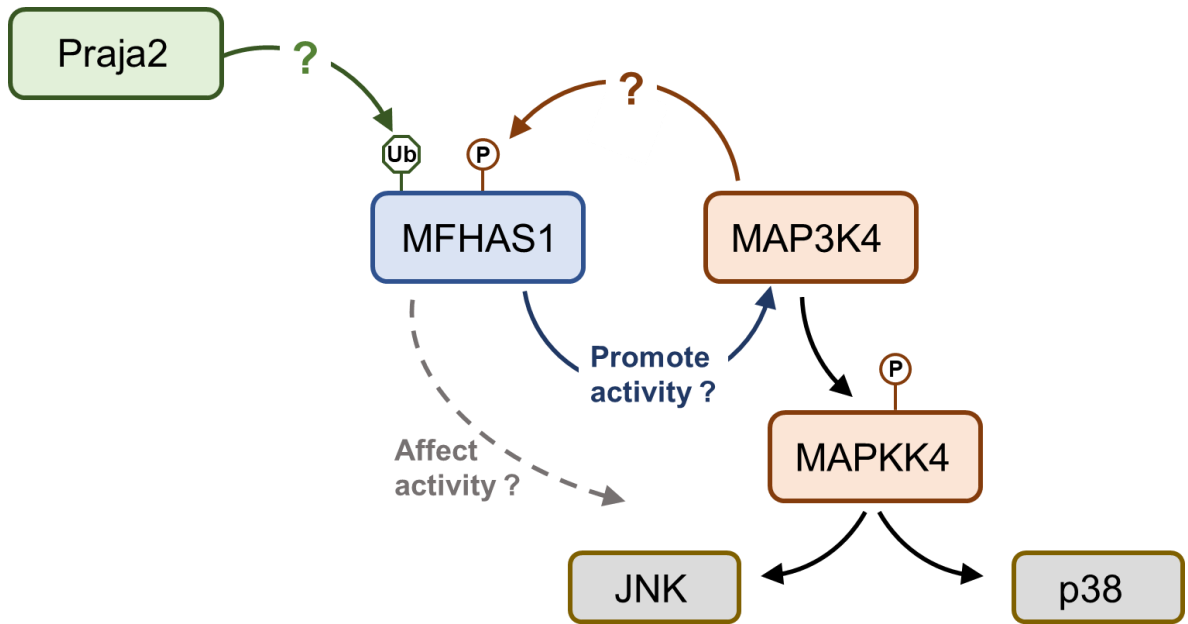
The R526C mutation is homologous to the R1441C mutation in LRRK2 (**S. Figure 4**). The R1441C change in LRRK2, linked to Parkinson's disease (Zimprich *et al.*, 2004) has been

shown to disrupt the GTP hydrolysis of LRRK2 (Lewis *et al.*, 2007) and impact downstream signalling (Price *et al.*, 2018). R526C mutation may also impact the GTPase activity of MFHAS1 or the binding with other proteins.

The GTP-binding mutant MFHAS1 K422A has previously been characterised and provides a defined enzymatic control for these proposed experiments (Dihanich *et al.*, 2014). A repeat of some experiments and further investigation is done to understand the impact of the mutation on protein interaction. K422 is conserved with the RAS proteins (K16) and LRRK2 (K1347), and plays a critical role in coordinating hydrolytic release of the GTP gamma phosphate (Yao *et al.*, 2010).

MFHAS1 protein interactions and cellular pathways are studied in this chapter using the HEK293T cellular model. Previous research on MFHAS1 used HEK293T (Dihanich *et al.*, 2014); therefore, the same cell line was used. The results from the PPI and PTM analysis of the bioinformatic chapter and the literature were merged to build a hypothetical signalling pathway around MFHAS1 (**Figure 4.2**). Data described in Chapter Three reveal that MFHAS1 may interact with MAP3K4 protein, and papers have demonstrated the involvement of MFHAS1 in the JNK/MAPK pathway (Zhong *et al.*, 2018; Kumkhaek *et al.*, 2019). Kumkhaek and team suggest that MFHAS1 and its chaperone, HSP70, are implicated via the MAPK/ERK1/2 signalling pathway in the negative regulation of hepcidin expression (Kumkhaek *et al.*, 2019). PJA2 is shown to form a complex with MFHAS1 on its LRR domain (Zhong *et al.*, 2018).

Based on the proposed model from Zhong's paper, which suggests that ubiquitination of MFHAS1 by PJA2 would activate the TLR2/JNK/p38/NF- $\kappa$ B through MAPK (Zhong *et al.*, 2018), it was decided to look at the p38 and JNK pathways. The proposed hypothesis for this chapter is that MFHAS1 could promote the activity of MAP3K4, which then modulates phosphorylation of MAPKK4, and activates the JNK and p38 pathways (**Figure 4.2**). The hypothesis is also to understand if MAP3K4 would phosphorylate MFHAS1 and test with different methods if PJA2 interacts with MFHAS1.

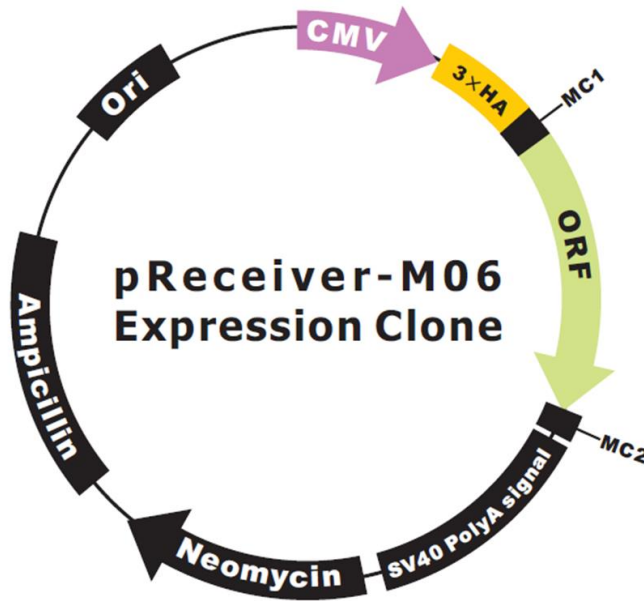


**Figure 4.2:** Proposed pathways surrounding MFHAS1 protein. It shows that PJA2 could ubiquitinate MFHAS1, MAP3K4 could phosphorylate MFHAS1. MFHAS1 could promote the activity of MAP3K4 which would impact the JNK and p38 pathways. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. MAPKK4: Mitogen-activated protein kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. JNK: c-Jun N-Terminal Protein Kinase. P38: p38 Mitogen-activated protein kinase.

## 4.2 Material and method

### 4.2.1 DNA preparation

MFHAS1 constructs were purchased from GeneCopoeia. pReceiver-M06 vector containing MFHAS1 cDNA and a hemagglutinin (HA) tag at the N terminal was used (GeneCopoeia EX-R0032-M06) (**Figure 4.3**).



**Figure 4.3:** Map of the pReceiver-M06 vector used to clone MFHAS1 DNA. Hemagglutinin tag is present.

Four MFHAS1 mutant construct were generated, encoding for the MFHAS1 coding variants; K422A, K387A, S450A and R526C. The MFHAS1 K422A mutant was provided by Patrick Lewis (Dihanich *et al.*, 2014). All other mutations were generated by site-direct mutagenesis using the QuikChange II Site-Directed Mutagenesis Kit from Agilent.

### 4.2.2 Mutagenesis using Agilent kit

The three mutant constructs of MFHAS1 (K387A, S450A and R526C) were generated with QuikChange II Site-Directed Mutagenesis Kit (Agilent). Primers were designed using the QuikChange Primer Design Program (**Table 4.1**).

<b>K387A</b>	
Forward	5'-ctacgaggtctgcat <b>ggc</b> gggatcccctacatc
Reverse	5'-gatgtaggggatcccc <b>cc</b> atgcagacctcgtag
<b>S450A</b>	
Forward	5'- aagtgctaccaccc <b>ggc</b> acctccccctgtgagc
Reverse	5'- gctcacagggggaggt <b>g</b> ccggtgggtagcactt
<b>R526C</b>	
Forward	5'-gggtcggggcg <b>tg</b> cggtccccacgc
Reverse	5'-gcgtggggcac <b>g</b> cacgccccgacct

**Table 4.1:** Forward and reverse primers designed for MFHAS1 site-directed mutagenesis using the QuickChange Primer Design Program. The mutated bases are shown in bold.

The mutant strand synthesis reaction using thermal cycling was performed under the following conditions. The control reaction was prepared with 5  $\mu$ L of 10X reaction buffer, 2  $\mu$ L of pWhitescript 4.5 kb control plasmid (5 ng/  $\mu$ L), 1.25  $\mu$ L of nucleotide control primer #1, 1.25  $\mu$ L of nucleotide control primer #2, 1  $\mu$ L of dNTP mix (provided by kit) and 38.5  $\mu$ L of ddH<sub>2</sub>O. The individual mutagenesis reactions were each prepared with 5  $\mu$ L of 10X reaction buffer, 1  $\mu$ L of plasmid with the wild type MFHAS1 mRNA (50 ng/  $\mu$ L), 1.2  $\mu$ L of nucleotide forward primers (10  $\mu$ M), 1.2  $\mu$ L of nucleotide reverse primer (10  $\mu$ M), 1  $\mu$ L of dNTP mix and 40.6  $\mu$ L of ddH<sub>2</sub>O. 1  $\mu$ L of *PfuUltra* HF DNA polymerase (2.5 U/  $\mu$ L) was added to all the samples and control. The following cycling parameters were used: Step 1: 95 °C for 30 seconds, one cycle, Step 2: 95 °C for 30 seconds, 55 °C for 1 minute, 68 °C for 10 minutes, 16 cycles.

The amplification products were digested by adding 1  $\mu$ L of the *Dpn* I restriction enzyme (10 U/  $\mu$ L) to the samples. Samples were mixed by pipetting up and down and then incubated at 37 °C for 1 hour. Digestion was performed to remove the parental, non-mutated double stranded DNA.

Transformation of the digested constructs into XL1-Blue Supercompetent cells (Agilent) was performed on ice. An aliquot of 50  $\mu$ L of XL1-Blue Supercompetent cells was thawed

on ice for each sample. 1  $\mu$ L of the digested sample was added to the competent cells. Reactions were incubated for 30 minutes on ice, then heat shocked for 45 seconds at 42 °C and placed on ice for 2 minutes. 0.5 mL of preheated SOC was added to the samples before incubation at 37 °C for 1 hour with 225 rpm shaking. Samples were centrifuged, and most supernatant was removed. Pellet was resuspended with the remaining supernatant. Luria-Bertani (LB) Broth (Difco 244610) and agar plates were made with ampicillin. Resuspended samples were plated on the plates and incubated overnight at 37 °C.

One colony of each sample were picked and grown in 5 mL of LB and ampicillin overnight at 37 °C with 250 rpm. Cultures were centrifuged for 30 minutes at 3500 g at 4 °C. Miniprep or maxiprep (Sigma) were performed on the pellets. For the miniprep, a 5 mL of the overnight bacteria culture was pelleted and resuspended in 200  $\mu$ L of resuspension solution, then 200  $\mu$ L of lysis solution was added and gently mixed. The resuspended pellet was left for 5 min to allow it to clear. 350  $\mu$ L of neutralisation was added, and the tube was inverted 4 to 6 times. Debris were pelleted by centrifugation for 10 min at full speed. The binding column was prepared according to provider's instructions. 500  $\mu$ L of column preparation solution was added to the column before spinning for 1 min at 12000 g. The flow-through was discarded. The clear lysate was transferred to the column and centrifuged for 1 min at 5000 g. Contaminants were removed by adding 750  $\mu$ L wash solution to the column and centrifuging for 1 min at 5000 g. The flow-through was discarded. The column was transferred to a new collection tube, and 100  $\mu$ L of elution buffer was added. The column was centrifuged for 1 min at 5000 g. Flow-through was kept, and the DNA concentration was measured with a nanodrop (Thermo Scientific™).

Mutated MFHAS1 constructs were sent for sequencing (Eurofins genomics) to confirm the presence of the mutations in the MFHAS1 cDNA sequence.

### **4.2.3 Cell culture**

Human Embryonic Kidney cells (HEK293T) were used for the experiment (CRL-3216™) (Graham *et al.*, 1977). HEK293T were maintained at 37 °C with 5 % CO<sub>2</sub> and passaged three times a week in Dulbecco's Modified Eagle Medium (DMEM) (Gibco™ 11960044) with 10 % v/v supplement of fetal bovine serum (FBS) (Gibco™ 10270106) and 1 % v/v supplement of Penicillin/Streptomycin (Gibco™ 15140122). HBSS (Gibco™ 14175095) was used to wash

the cells, and trypsin-EDTA (0.05%) (Gibco™ 25300062) was used to detach the cells from the flask.

For the transfection, HEK293T cells were plated at  $2.5 \times 10^5$  or  $5 \times 10^5$  cells/mL in 6 wells plates. For immunoprecipitation (IP), transfection was done at  $1 \times 10^6$  or  $3 \times 10^6$  cells/mL in 100 mm Petri dishes. 500  $\mu$ L DMEM, 1500 ng of plasmid DNA and 7  $\mu$ L of transfection reagent turbofect (Thermo Scientific™ R0533) were mixed in an Eppendorf tube and incubated for 15 min at room temperature before being added to the cells. The media was changed 4 hours after transfection. The transfected cells were incubated for 48 hours at 37 °C with 5 % CO<sub>2</sub>. Five replicates were made for each experiment.

Two other cell lysates were used to validate the MFHAS1 antibody (PA5-62347). Human acute lymphoblastic leukaemia (MOLT-4) whole cell lysate (Abcam ab7912 – ATCC CRL-1582) was purchased and used in immunoblot. MFHAS1 RNA expression was the highest in MOLT-4 cell line.

MFHAS1 Reverse siRNA transfection of HEK cells with Knock-Down (KD) MFHAS1 cell lysate was performed by Ben O'Callaghan at the University College of London using the following protocol.

Each well of six wells plates was seeded with HEK cells at  $1 \times 10^6$  cells/mL. 100 nM of siRNA and 0.3% v/v DharmaFECT1 (Dharmacon SMARTpools) were transfected in HEK cells. Cells were collected three days post-transfection. Cells were washed with ice-cold DPBS and collected in 100  $\mu$ L WC lysis buffer. Lysates were centrifuged, and the soluble fraction was kept.

#### **4.2.4 Protein extract sample preparation**

All the cell extract samples were prepared on ice. Media was discarded, and plates and Petri dishes were washed with HBSS. Cells were lysed with 1X RIPA buffer (50 mM Tris-HCl, pH 8.0, with 150 mM sodium chloride, 1 % Igepal CA-630 (NP-40), 0.5 % sodium deoxycholate, and 0.1% sodium dodecyl sulfate (Sigma R0278)) with 1X supplement of Halt protease and phosphatase inhibitor (78430 Thermo scientific 100x). IP lysis buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 % NP-40, 1 mM EDTA, 5% glycerol (Thermo Scientific 11820084)) was used instead of RIPA buffer for the cell prepared for immunoprecipitation.



Cells were detached with a cell scraper, transferred into an Eppendorf tube, and lysed for 30 min in the cold room at 4 °C with rotation. The tubes were centrifuged for 10 min at 4 °C and 1000 g. The supernatants were collected in a fresh tube.

Proteins were quantified with Bicinchoninic acid assay (BCA Pierce™ Thermo scientific) (Smith *et al.*, 1985). 10 µL of standard and samples were added to a 96 wells plate. 200 µL of BCA working reagent (50 part working reagent A and 1 part working reagent B) was added to each well. The plate was incubated 30min at 37°C and read at 562nm with a spectrophotometer. Samples were diluted to the same concentration.

#### **4.2.5 Determining the protein expression level by immunoblot**

SDS Sample Buffer (4x Nupage) + 10 % β-mercaptoethanol was added to each sample. Samples were denatured for 5 minutes at 95 °C and centrifuged for 1 min at 10,000 g. 10 µL of each sample and 4 µL of the protein ladder (26619 Thermo Scientific) were run on SDS- PAGE for 10 min at 200 V and 40 min at 150 V with 1X MES buffers (NP0002 Invitrogen) or 1X MOPS (NP0001-02 Invitrogen).

After migration, the gel was transferred onto a nitrocellulose membrane (IB23001 Invitrogen) using iBlot™ 2 Dry Blotting system (IB21001 Invitrogen).

Membranes were placed in Phosphate-Buffered Saline-Tween 20 v/v 0.5% (PBS-T) and 5 % milk, rotating for 1 hour. Membranes were incubated in primary antibody 5 % milk + PBS for 1.5 hours at RT or O/N at 4 °C. A range of specific primary and secondary antibodies were used (summarised in **Table 4.2**). Membranes were washed with PBS thrice for 10 min, then placed in secondary antibody 1:5000 5 % milk PBS. Membranes were rewashed three times with PBS.

The fluorescence was activated with Enhanced Chemiluminescence (ECL). ECL solution A was prepared as follows, 100 µL of luminol, 44 µL of coumaric acid, 1 mL of Tris-HCL (1M PH 8.5) and water was added up to 10 mL. Solution B was prepared with 6.4 µL of H<sub>2</sub>O<sub>2</sub>, 1 mL of Tris-HCL (1M PH 8.5) and water up to 10 mL. 1 mL of solutions A and B was added on each membrane for 2 min. The membranes were read with GeneSys software, Imager Syngene G:Box and analysed using imageJ.

Target	AB	Reference	Company	MW	Dilution	Host
<b>p38</b>	p38 MAPK	#9212S	Cell signaling	40kDa	1.:1000	Rabbit
<b>phospho p38 (Thr 180/Tyr 182)</b>	phospho p38 MAPK	#4511S	Cell signaling	43kDa	1.:1000	Rabbit
<b>MAP2K3</b>	MKK3b	#9238S	Cell signaling	40kDa	1.:1000	Rabbit
<b>MAP2K3 (Ser 189/ Ser 207)</b>	phospho MKK3	#9231S	Cell signaling	40kDa	1.:1000	Rabbit
<b>MAP2K4</b>	SEK1/ MKK4	#9152S	Cell signaling	44kDa	1.:1000	Rabbit
<b>MAP2K4 (Ser80) inhibit protein</b>	Phospho SEK1/ MKK4	#9155S	Cell signaling	44kDa	1.:1000	Rabbit
<b>MAP2K4 (Ser 257) activate protein</b>	Phospho SEK1/ MKK4	#4514S	Cell signaling	44kDa	1.:1000	Rabbit
<b>MAP3K4</b>	MAP3K4	ab186125	Abcam	182kDa	1.:1000	Rabbit
<b>HA tag</b>	HA tag	26183	Invitrogen	NA	1.:1000	Mouse
<b>MFHAS1</b>	MFHAS1	PA5-62347	Invitrogen	115kDa	1.:1000	Rabbit
<b>PJA2</b>	PJA2	PA5-101571	Invitrogen	80kDa	1.:1000	Rabbit
<b>Alpha-tubulin</b>	Alpha-tubulin	62204	Invitrogen	55kDa	1.:1000	Mouse
<b>Mouse Gamma Immunoglobins Heavy and Light chains</b>	Anti-mouse	62-6520	Invitrogen	NA	1.:5000	Goat
<b>Rabbit Gamma Immunoglobulin Fc region</b>	Anti-rabbit	A16116	Invitrogen	NA	1.:5000	Goat

**Table 4.2:** Primary and secondary antibodies used in immunoblotting. Kilodalton (kDa).

## **4.2.6 Investigating protein-protein interactions using immunoprecipitation.**

### **4.2.6.1 Buffer and beads preparation:**

Tris-Buffer Saline (TBS) 10X was prepared with 2.4 g of Tris, 88 g of NaCl and 100 mL of distilled water. The solution was adjusted to pH 7.6. 1X TBS with 0.05 % Tween 20 (TBS-T) was prepared by diluting 10 mL of 10X TBS in 90 mL of distilled water, and 50  $\mu$ L of Tween 20 was added.

25 $\mu$ L (0.25mg) of Pierce Anti-HA Magnetic Beads (Fisher Scientific 13464229) were placed into a 1.5 mL microcentrifuge tube. 175  $\mu$ L of 1X TBS-T was added to the beads and gently vortexed to mix. The tube was placed into a magnetic stand to collect the beads against the side of the tube. The supernatant was discarded. 1mL of 1X TBS-T was added to the tube. The tube was inverted several times, and the supernatant was discarded.

### **4.2.6.2 Immunoprecipitation**

For each step, 40  $\mu$ L was kept for analysis.

Supernatants of lysed cells were added to the pre-washed magnetic beads and incubated at room temperature for 30 minutes with rotation. The beads were collected with a magnetic stand, and the unbound samples were removed, keeping 40  $\mu$ L of it for further analysis (labelled as flow through in the following experiments). 300  $\mu$ L of 1X TBS-T was added to the tube and gently mixed. The beads were collected, and the supernatant was discarded keeping 40  $\mu$ L of it for further analysis (labelled as 1<sup>st</sup> wash in the following experiments). This wash was repeated two times. 300  $\mu$ L of ultrapure water was added to the tubes and gently mixed. The beads were collected, and the supernatant was discarded. 50  $\mu$ L of SDS-PAGE Sample Buffer was added to the tubes to eluate the protein. The tubes were then gently vortexed to mix and incubated at 95 °C for 5 minutes. The tubes were placed on the magnetic tube holder to separate the beads and collect the supernatant containing the target proteins.

#### **4.2.7 Statistical analysis**

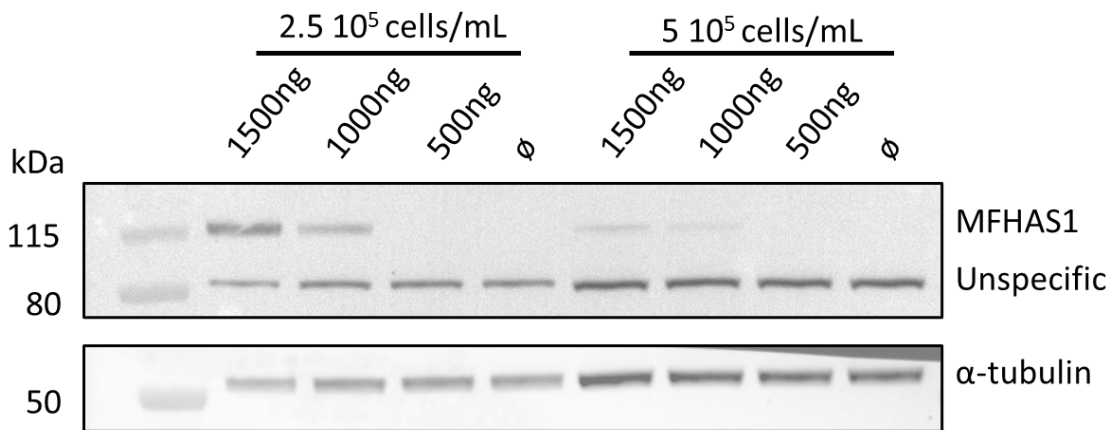
Immunoblot membranes were analysed and quantified with ImageJ. Each band was selected and quantified on ImageJ. Each sample was normalised against the tubulin control.

Statistical analyses were performed on GraphPad Prism 9.5.1. The significance between each sample was calculated using an ordinary one-way ANOVA test. Tukey-Kramer's follow-up test was performed to compare each column's mean with the mean of every other column.

## 4.3 Results

### 4.3.1 Transfection optimisation

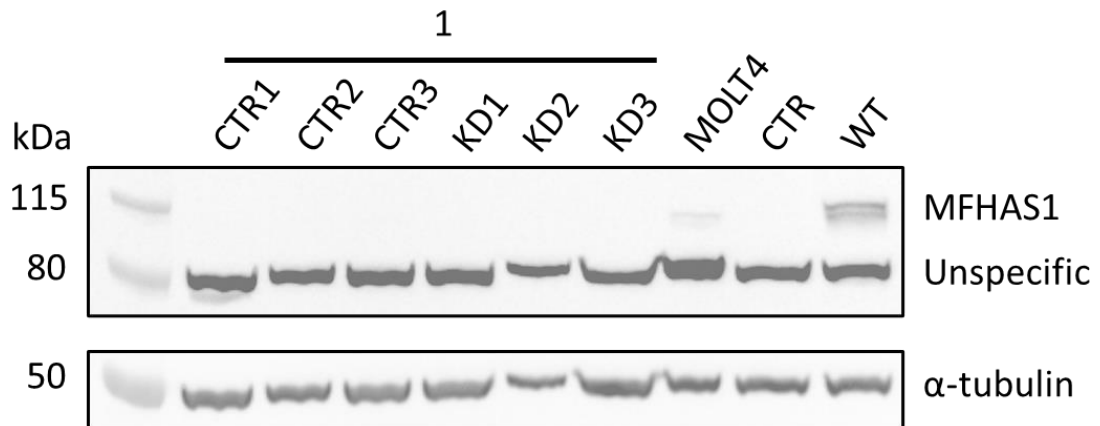
Prior to investigating the biology of the MFHAS1 protein, optimisation of the transfection conditions was carried out. Two cell concentrations,  $2.5 \times 10^5$  and  $5 \times 10^5$  cells/mL, and three DNA concentrations, 1500 ng, 1000 ng, and 500 ng, were compared (**Figure 4.4**).



**Figure 4.4:** Immunoblot analysis of different transfection conditions using HEK293T cells. The membrane is probed with anti-MFHAS1 and anti- $\alpha$ -tubulin. Three DNA concentrations are tried (1500ng, 1000ng and 500ng) with two cell concentrations ( $2.5 \times 10^5$  and  $5 \times 10^5$  cells/mL). MFHAS1 and a non-specific putative band are identified. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence.

The highest intensity band is at 115 Kilodalton (kDa), the expected molecular mass for MFHAS1. The concentration of  $2.5 \times 10^5$  cells/mL and 1500 ng of DNA shows the thickest band on the immunoblot at 115 kDa, making it the best condition to work with. Transfection with 500 ng of MFHAS1 DNA shows no presence of MFHAS1 protein; it may be due to the DNA concentration being too low. There was a possible non-specific protein band appearing at 80 kDa. To investigate whether this non-specific protein band represent an alternative form of MFHAS1 or a degradation product of the full-length protein, an experiment was performed by comparing MFHAS1 expression levels in HEK cells transfected with MFHAS1 targeting siRNA (silencing and sample collection were performed by Ben O'Callaghan at University College London, in HEK293 cells line), MOLT4 cells lysate,

which has a high RNA expression of MFHAS1 (Uhlén *et al.*, 2015), and HEK cells with overexpressed MFHAS1 (**Figure 4.5**).



**Figure 4.5:** Immunoblot to investigate the unspecific band of MFHAS1 AB at 80 kDa. “1” represents the untransfected control (CTR) and Knock-Down (KD) of MFHAS1 in HEK cells performed by Ben O’Callaghan at the University College of London. Human acute lymphoblastic leukemia (MOLT4) whole cell lysate from Abcam is run. Untransfected HEK293T cells (CTR) and MFHAS1 wild-type (WT) construct transfected in HEK293T are from this chapter’s experiments. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence

The endogenous level of MFHAS1 mRNA is very low (nTPM: 7.6) and shows the same result as the siRNA MFHAS1 samples. As shown in **Figure 4.5**, the unspecific band at 80 kDa is present even when MFHAS1 endogenous expression is decreased in the cells. MFHAS1 band at 115 kDa is present in the sample with overexpressed WT MFHAS1 and the MOLT4 cell line. The anti-MFHAS1 AB might be detecting an unspecific protein. MFHAS1 construct used for this chapter contains an HA tag. As it is not possible to only detect MFHAS1 protein at 115 kDa, an anti-HA AB will be used to detect exogenously expressed HA-tagged MFHAS1 in the following experiments. WT MFHAS1 present a double band indicating a possible transient phosphorylation on MFHAS1. Interestingly, the endogenous level of MFHAS1 in MOLT4 is lower than the WT overexpressed MFHAS1 in HEK cells. In this chapter, MFHAS1 is overexpressed at the highest possible level to obtain a high amount of protein. It would be of interest to overexpress MFHAS1 at a lower level, resembling the endogenous level, to lower the impact on the cell.

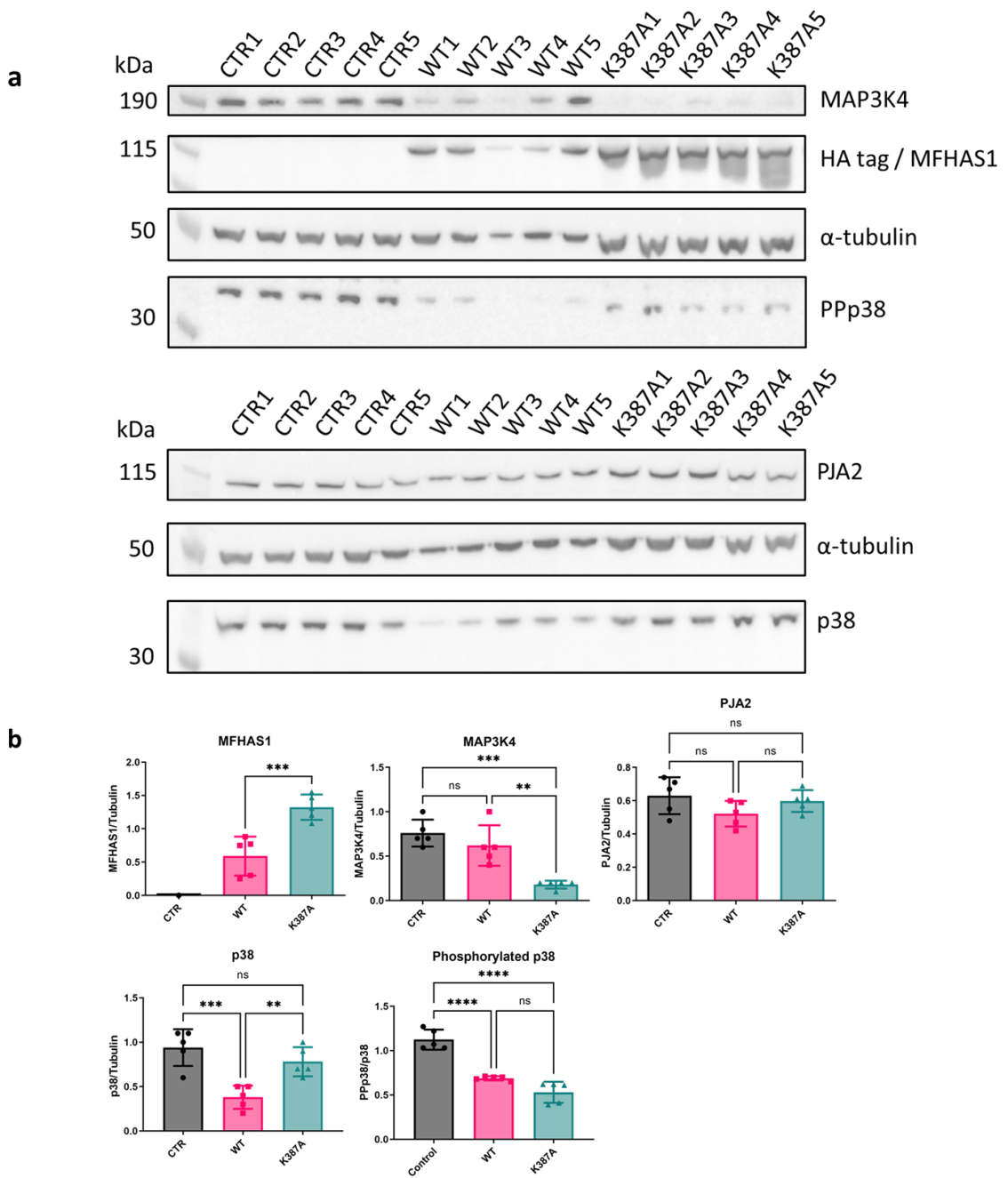
### 4.3.2 Effect of ubiquitination site mutation K387A

The MFHAS1 K387 is a putative ubiquitination site reported on PhosphoSitePlus via the High throughput method, mass spectrometry (Udeshi *et al.*, 2013). This site has been mutated to inhibit protein ubiquitination at the Lysine 387 residue. Mutation of this site could help develop an understanding of what the physiological role of this ubiquitination event is and provide a valuable tool for validating ubiquitination at this residue.

#### 4.3.2.1 Protein expression level

To study the effects of K387A mutation on MFHAS1 protein, Wild-Type (WT) MFHAS1 and mutant K387A MFHAS1 construct expressing the HA epitope-tagged MFHAS1 proteins were transfected into HEK293T cells. Five replicates were collected and separated via SDS-PAGE before performing an immunoblot (**Figure 4.6**). Steady-state levels of MFHAS1, PJA2 (also called Praja2), MAP3K4 and p38 are normalised over  $\alpha$ -tubulin levels, while phosphorylated p38 is normalised over p38 levels and analysed using GraphPad Prism (**Figure 4.6**). MFHAS1 was detected with anti-HA antibodies.

The protein ratio over  $\alpha$ -tubulin was calculated, and one-way ANOVA with Tukey's multiple comparison tests was performed. Results show a significant MFHAS1 protein increase in steady-state protein levels in the presence of the K387A ( $P= 0.0002$ ) (**Figure 4.6**), suggesting that this mutation may have an impact on the stability/turnover of the protein. MAP3K4 protein expression significantly decreases when MFHAS1 is mutated at the K387A site compared to the control ( $P= 0.0003$ ) and the wild-type MFHAS1 transfected cells ( $P= 0.0025$ ). PJA2 protein level stays consistently the same between conditions. Phosphorylated p38 expression decreases significantly when WT ( $P < 0.0001$ ) and K387A ( $P < 0.0001$ ) MFHAS1 protein is overexpressed. The steady-state level of p38 decreases when WT MFHAS1 is overexpressed compared to the control ( $P= 0.0006$ ). p38 under the condition of K387A MFHAS1 has no significant decrease of expression compared to the control but increases significantly compared to the WT condition ( $P= 0.0077$ ).

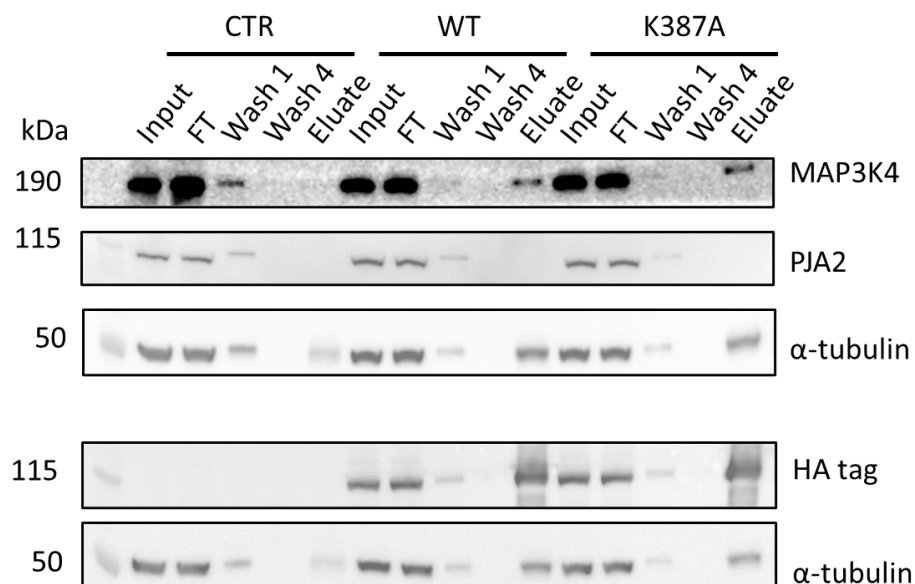


**Figure 4.6:** Impact of K387A mutation on the steady state protein levels of selected proteins. **(a):** Immunoblot images of control (CTR), Wild Type (WT) and K387A MFHAS1 samples (n=5). Membranes are probed with HA, MAP3K4, p38, phosphorylated p38, PJA2 and  $\alpha$ -tubulin antibodies. **(b):** Statistical analysis of proteins expression level of MFHAS1, MAP3K4, phosphorylated p38, p38 and PJA2. Values were normalised over the  $\alpha$ -tubulin. Phosphorylated p38 was normalised over p38. One-way ANOVA with Tukey's multiple comparison test; ns  $P > 0.05$ ; \*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ ; \*\*\*\* $P \leq 0.0001$ . MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. P38: p38 Mitogen-activated protein kinase. HA: hemagglutinin.



#### 4.3.2.2 Immunoprecipitation

To investigate potential protein/protein interaction and the effect of K387A on these interactions, immunoprecipitation (IP) was performed on cell pellets previously transfected with WT or K387A MFHAS1 alongside an untransfected control sample. Transfection was carried out in Petri dishes, and the cells were collected in IP lysis buffer. Three replicates were collected for each sample. MFHAS1 constructs containing an HA tag were pulled down with anti-HA beads. Lysed cells (input) are added to the beads; the supernatant is collected (flow-through), and then the beads are washed four times and eluted with SDS. Immunoblot was performed and probed for MAP3K4, PJA2, HA tag and  $\alpha$ -tubulin (**Figure 4.7**).



**Figure 4.7:** Immunoprecipitation experiment to investigate proteins interacting with MFHAS1. Immunoblot of immunoprecipitation of K387A and wild-type (WT) overexpressed MFHAS1. Membranes are probed with MAP3K4, HA, PJA2 and  $\alpha$ -tubulin. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. HA: hemagglutinin.

Results show a band for MAP3K4 in the eluate of the WT and mutant MFHAS1, suggesting that MAP3K4 is interacting and coprecipitating with MFHAS1. There is no visible difference between the WT and the K387A mutant for the quantity of MAP3K4 protein immunoprecipitated with MFHAS1. This suggests that the mutation has no impact on the binding of MAP3K4, which also correlates well with the previous data (**Figure 4.6**), showing that this mutation has no impact on MAPK downstream signalling of MFHAS1. This result was found on three replicates (**S.Figure.6**). This experiment has shown no interaction

between MFHAS1 and PJA2, as no protein band is seen for PJA2 in control, WT or mutant eluates, suggesting that the interaction might be very short-lived or absent under the conditions used in this experiment.

### **4.3.3 Effect of phosphorylation site mutation S450A**

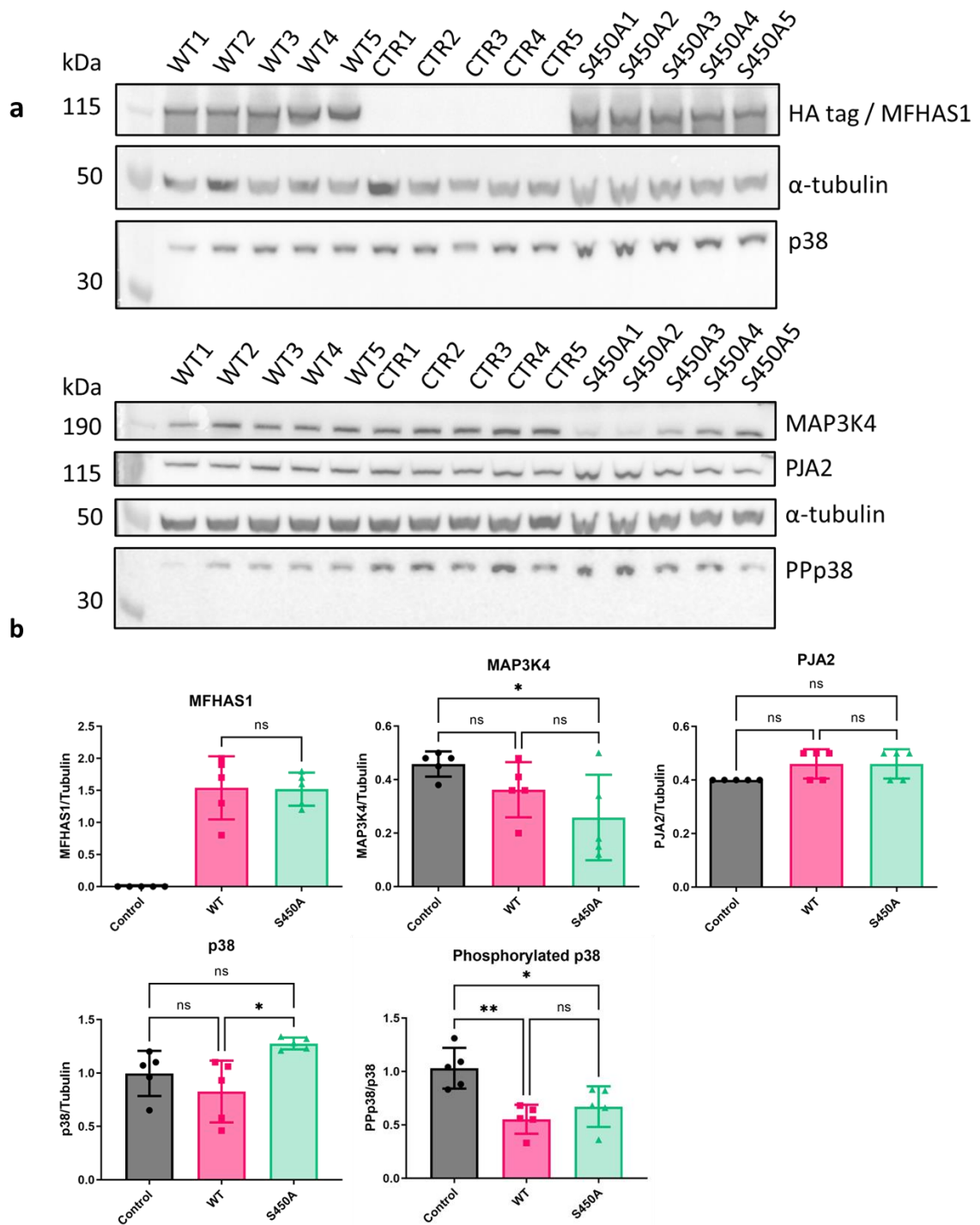
The S450 phosphorylation site has been reported five times on PhosphoSitePlus, making it the most highly reported PTM site of MFHAS1. Mutation of the serine to alanine (preventing phosphorylation) was carried out to test the impact of post translational modification on this residue. The consequences of the mutated phosphorylation site on the downstream pathway of MFHAS1 and protein interaction were then assessed.

#### **4.3.3.1 Protein expression level**

Five control, WT and S450A MFHAS1 replicates were expressed in HEK293T cells, which were then collected and lysed. Samples were run on an SDS-PAGE and transferred to a nitrocellulose membrane. Membranes are probed for MAP3K4, PJA2, HA, p38, phosphorylated p38 and  $\alpha$ -tubulin to assess alterations in steady-state protein levels (**Figure 4.8**).

Protein levels are normalised to  $\alpha$ -tubulin, and one-way ANOVA with Tukey's multiple comparison tests was performed to test for significant alterations. The statistical analysis of the immunoblot shows no change in MFHAS1 steady-state level between WT and S450A (**Figure 4.8**). PJA2 protein expression level does not change between the control, WT and mutant. MAP3K4 protein level decreased significantly between the control and S450A mutant MFHAS1 expressing cells ( $P= 0.0398$ ) but not between the control and WT. p38 protein levels increased for the S450A mutant compared to the expression of the WT protein ( $P= 0.0136$ ). Phosphorylated p38 steady-state level decreased significantly for the WT condition compared to the control ( $P= 0.0049$ ) but there is no significant change between the WT and S450A mutant.

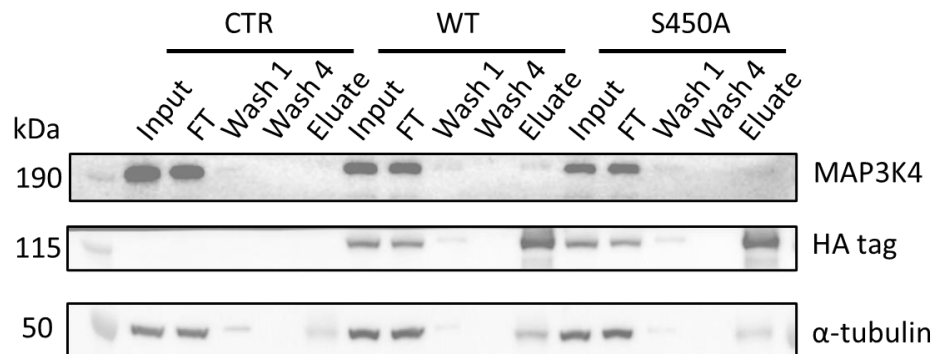
This experiment suggests that S450A mutation does not alter the steady-state levels of MFHAS1 proteins; however, overexpression of MFHAS1 in HEK cells leads to decreased p38 phosphorylation dependent on the serine 450 site.



**Figure 4.8:** Impact of S450A mutation on the expression level of selected proteins. **(a):** Immunoblot images of control (CTR), Wild Type (WT) and S450A MFHAS1 samples (n=5). Membranes are probed with HA, MAP3K4, p38, phosphorylated p38, PJA2 and  $\alpha$ -tubulin antibodies. **(b):** Statistical analysis of proteins expression level of MFHAS1, MAP3K4, phosphorylated p38, p38 and PJA2. Values were normalised over the  $\alpha$ -tubulin. Phosphorylated p38 was normalised over p38. One-way ANOVA with Tukey's multiple comparison test; ns  $P > 0.05$ ; \* $P < 0.05$  \*\* $P \leq 0.01$ . MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. P38: p38 Mitogen-activated protein kinase. HA: hemagglutinin.

#### 4.3.3.2 Immunoprecipitation

S450A and WT MFHAS1 constructs are expressed in HEK cells alongside untransfected cells before performing an IP with anti-HA beads and running an immunoblot (**Figure 4.9**). HA, MAP3K4 and  $\alpha$ -tubulin are probed.



**Figure 4.9:** Immunoblot of control (CTR), Wild Type (WT) and S450 MFHAS1 after immunoprecipitation. Membrane is probed with MAP3K4, HA and  $\alpha$ -tubulin. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase 4. HA: hemagglutinin.

When probed for HA, the immunoblot shows a clear band in the eluate for WT and S450A, indicating that MFHAS1 was successfully immunoprecipitated. A band is visible in the eluate of WT and mutant MFHAS1 when probed with MAP3K4, suggesting an interaction between MFHAS1 and MAP3K4, resulting in co-immunoprecipitation.

#### 4.3.4 Understanding the effects of PD-like mutations in MFHAS1

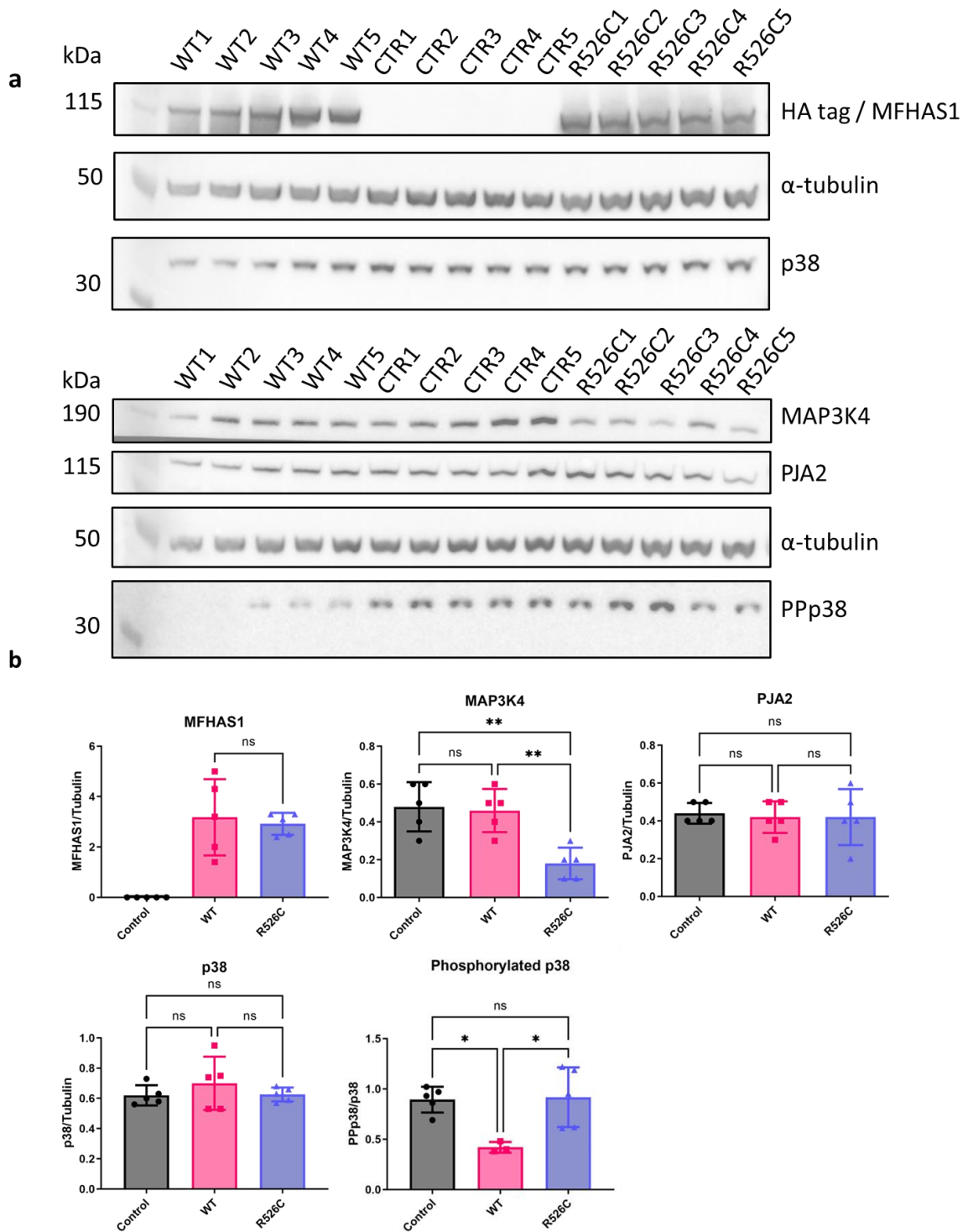
Two mutations that might impact the GTP binding are designed into MFHAS1: K422A and R526A.

R526C is the conserved LRRK2 R1441C (Lewis *et al.*, 2007), a mutation that causes autosomal dominant PD and decreases GTPase activity in LRRK2. The K422A is an artificial mutation, a homologue to the RAS K16 proteins and LRRK2 K1347 (Yao *et al.*, 2010) and has been previously studied (Dihanich *et al.*, 2014). It is designed to prevent GTP binding by ablating coordination of the gamma phosphate of GTP.

##### 4.3.4.1 Protein expression level

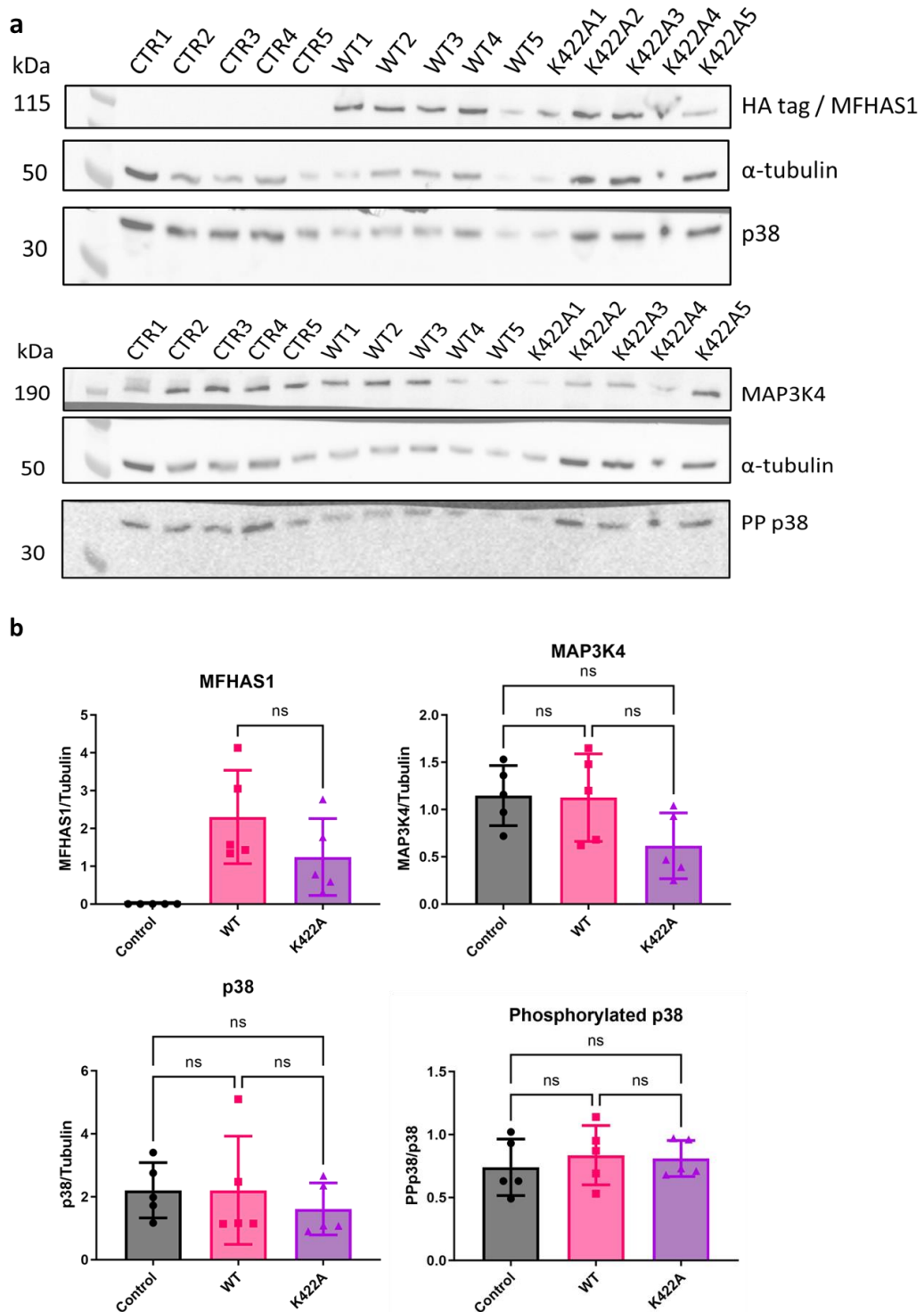
Comparing the expression level will give a better understanding of the consequences of those two mutations. Five controls, WT and mutant replicates were expressed in cells and prepared for immunoblotting. Membranes were probed with MAP3K4, PJA2, HA, p38, phosphorylated p38 and  $\alpha$ -tubulin (**Figures 4.10 and 4.11**).

One-way ANOVA with Tukey's multiple comparison tests was performed on protein levels normalised to  $\alpha$ -tubulin or p38 for phosphorylated p38. MFHAS1 protein expression level is consistent between WT and R526C. The ratio of protein is calculated over the  $\alpha$ -tubulin. Immunoblot shows no difference in the steady-state level of PJA2 across control, WT and mutant. MAP3K4 expression level is significantly lower for R526C mutant than WT ( $P=0.0047$ ) and control ( $P=0.0029$ ). The steady-state expression level of p38 does not change significantly between the different conditions. Phosphorylated p38 steady-state level decreases significantly when WT MFHAS1 is overexpressed compared to the control and R526C mutant.



**Figure 4.10:** Impact of R526C mutation on the expression level of selected proteins. **(a):** Immunoblot images of control (CTR), Wild Type (WT) and R526C MFHAS1 samples (n=5). Membranes are probed with HA, MAP3K4, p38, phosphorylated p38, PJA2 and  $\alpha$ -tubulin antibodies. **(b):** Statistical analysis of proteins expression level of MFHAS1, MAP3K4, phosphorylated p38, p38 and PJA2. Values were normalised over the  $\alpha$ -tubulin. Phosphorylated p38 was normalised over p38. One-way ANOVA with Tukey's multiple comparison test; ns  $P > 0.05$ ; \*\* $P \leq 0.01$ . MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. P38: p38 Mitogen-activated protein kinase. HA:

The Immunoblot of K422A mutation and the statistical analysis of the protein expression level are shown in **Figure 4.11**.



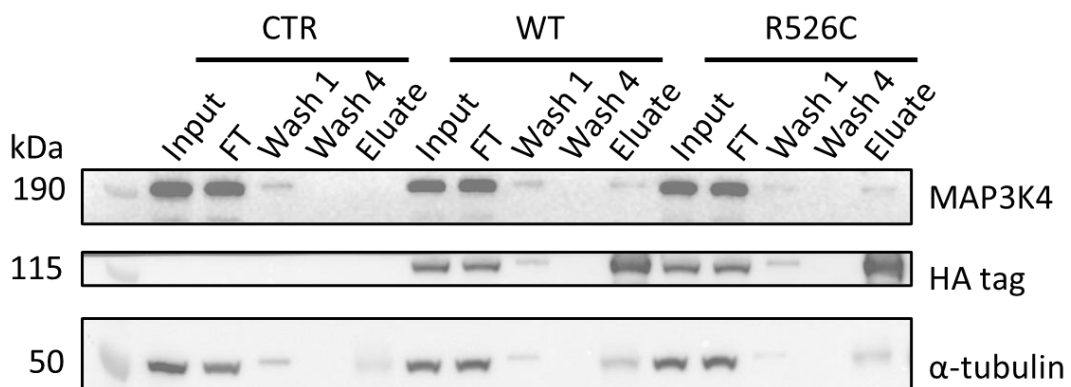
**Figure 4.11:** Impact of K422A mutation on the expression level of selected proteins. **(a):** Immunoblot images of control (CTR), Wild Type (WT) and K422A MFHAS1 samples (n=5). Membranes are probed with HA, MAP3K4, p38, phosphorylated p38 and  $\alpha$ -tubulin antibodies. **(b):** Statistical analysis of proteins expression level of MFHAS1, MAP3K4, phosphorylated p38 and p38. Values were normalised over the  $\alpha$ -tubulin. Phosphorylated p38 was normalised over p38. One-way ANOVA with Tukey's multiple comparison test; ns P>0.05. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. P38: p38 Mitogen-activated protein kinase. HA: hemagglutinin.



Immunoblot of K422A mutation compared to the WT MFHAS1 and untransfected cells K422A mutant shows no significant changes across the proteins probed for this experiment.

#### 4.3.4.2 Immunoprecipitation

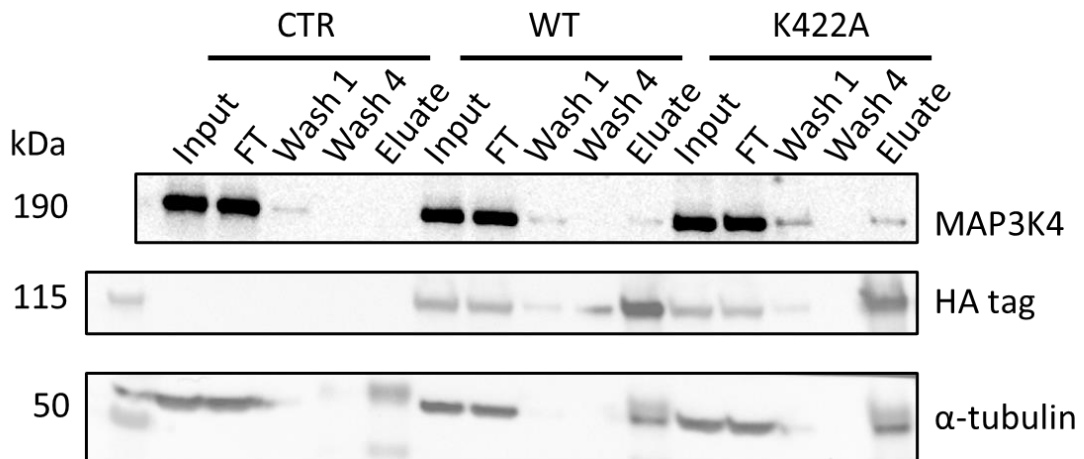
To test for disruption of the interaction between MFHAS1 and MAP3K4 in the presence of the R526C mutation, immunoprecipitation of control, WT and R526C cell lysate samples was performed, and immunoblot analysis was carried out. Membranes are probed with MAP3K4, HA and  $\alpha$ -tubulin (**Figure 4.12**).



**Figure 4.12:** Immunoblot of immunoprecipitation experiment performed on control (CTR), wild type (WT) and R526C MFHAS1 mutant. Membrane is probed for MAP3K4, HA and  $\alpha$ -tubulin. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. HA: hemagglutinin.

The result of the IP experiment shows a band for MAP3K4 after elution of WT and R526C MFHAS1. The same result is seen on three replicates (**S.Figure.6**). There is no apparent change between WT and mutant samples.

Immunoprecipitation was performed on K422A mutant cell lysate and compared with WT and untransfected control to test for the impact of GTP binding upon the relationship between MFHAS1 and MAP3K4 (**Figure 4.13**).



**Figure 4.13:** Immunoblot of immunoprecipitation (IP) experiment performed on control (CTR), wild type (WT) and K422A MFHAS1 mutant. Membrane is probe for MAP3K4, HA and  $\alpha$ -tubulin. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. HA: hemagglutinin.

Results of the IP experiment show a band for MAP3K4 protein for the WT and K422A mutant MFHAS1. A Coomassie blue stain of SDS gel of the IP experiment was also carried out and showed a clear eluate (**S.Figure.7**).

## 4.4 Discussion

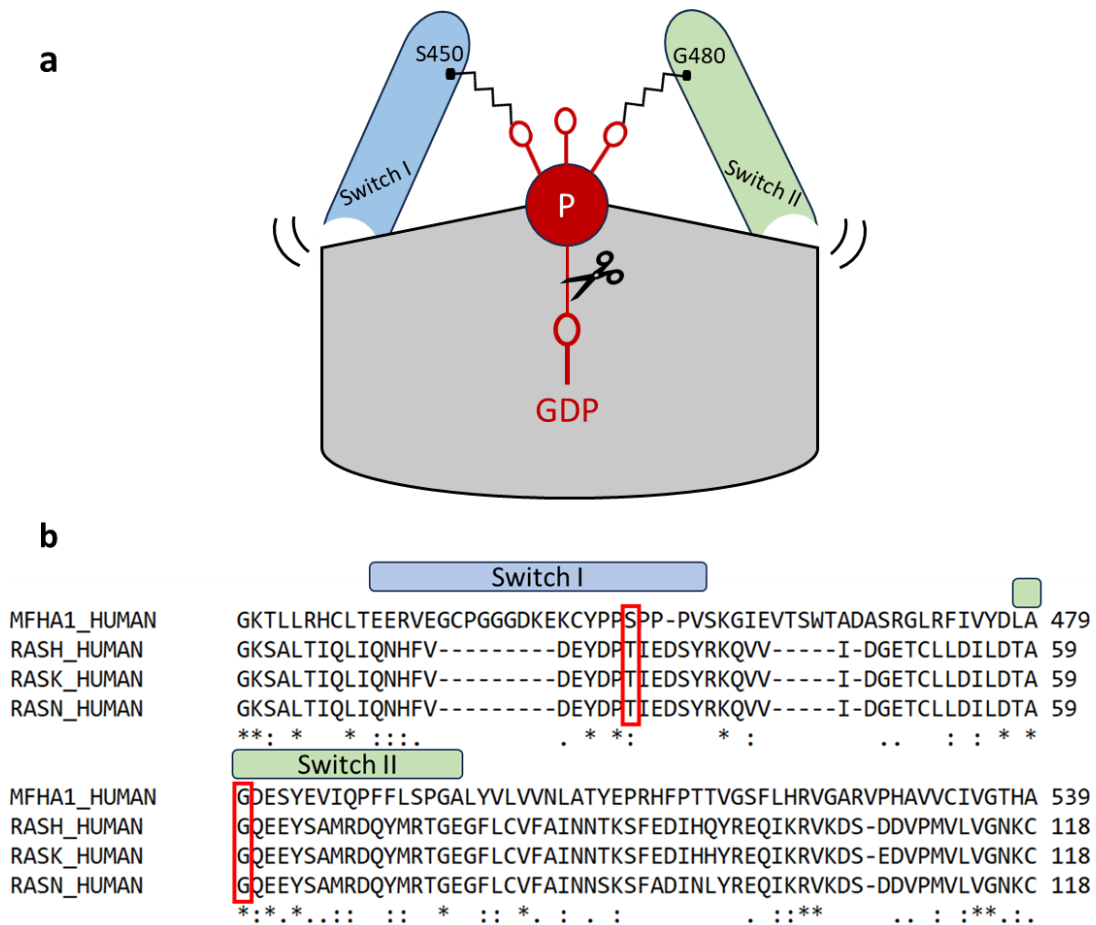
In this chapter, MFHAS1 mutants were designed to investigate the effect of targeted artificial mutations on the biochemistry and physiological functions of MFHAS1. Wild-type (WT) and mutant MFHAS1 were overexpressed in HEK cells. Immunoblot and immunoprecipitation were performed to analyse the samples. Four mutations were investigated: K387A, S450A, R526C and K422A.

K387 is a putative MFHAS1 ubiquitination site; it was selected for functional analysis following the report by Zhong and coworkers (2018) on the interaction between MFHAS1 and PJA2 on the amino acid sequence 64-364 of MFHAS1 (Zhong *et al.*, 2018). The closest reported MFHAS1 ubiquitination site to this sequence is K387. Steady-state levels of PJA2 were compared between the control, WT and K387A MFHAS1. These results show no change in PJA2 expression level, suggesting that MFHAS1 does not impact the expression of PJA2. Zhong's paper shows the interaction between MFHAS1 and PJA2 via mass spectrometry and immunoprecipitation when both proteins are overexpressed. In this chapter, immunoprecipitation of MFHAS1 via its HA tag shows no interaction with endogenous PJA2. The substantial difference in experimental conditions may explain the discrepancy in results. This paper also describes the activation of the TLR2/JNK/p38/NF- $\kappa$ B pathway by ubiquitinated MFHAS1 through MAPK. With regard to this, MAP3K4 was listed as a potential interacting kinase protein in the bioinformatic chapter and was selected to test in the experiment. The results of IP experiments suggest that MAP3K4 does interact with MFHAS1.

The S450 phosphorylation site in MFHAS1 was chosen for investigation because it has the most reports in the PhosphoSitePlus database and is located in the ROC functional domain of MFHAS1. More precisely, S450 is homologous to T35 in the RAS protein, which is an essential amino acid in the switch I of the RAS protein (**Figure 4.14**) (Vetter and Wittinghofer, 2001; Pálffy, Menyhárd and Perczel, 2020). It is worth mentioning that T35 of HRas is a putative phosphorylation site (Kinoshita *et al.*, 1997).

Phosphorylation can cause a change in binding energy (Nishi, Hashimoto and Panchenko, 2011) and can affect protein function (Nishi, Shaytan and Panchenko, 2014). The S450A mutation does not impact MFHAS1 protein steady-state expression level; however,

MAP3K4 expression decreases when S450A MFHAS1 is overexpressed. PJA2 protein levels do not change.



**Figure 4.14:** Comparison of RAS proteins with MFHAS1. **(a)** Schematic representation of the GTPase activity in the ROC domain of MFHAS1. Figure adapted from (Vetter and Wittinghofer, 2001). **(b)** Clustal sequence alignment of MFHAS1, KRAS, NRAS and HRAS proteins. Switch I and switch II are shown on the alignment. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. ROC: Ras Of Complex protein. GDP: Guanosine Diphosphate. GTP: Guanosine Triphosphate.

MFHAS1 K422 site is aligned with K16 of the RAS proteins and LRRK2 K1347 (Yao *et al.*, 2010). K16 is an essential site in the Ras proteins, forming a bond between K16 and GDP or GTP (Poorebrahim *et al.*, 2022; Bao *et al.*, 2023). In 2014, the MFHAS1 K422A mutation was investigated by GTPase assay and GTP pull-down and found a complete loss of GTP binding with K422A mutant (Dihanich *et al.*, 2014). In this chapter, the K422A mutation impacts none of the proteins studied. The protein expression does not significantly change with the mutation, which suggests that this mutation has no apparent impact on the MFHAS1 protein level or pathways through these experiments.

R526C mutation is homologous to the LRRK2 Parkinson's associated R1441C mutation. In LRRK2, R1441C mutation causes familial PD and has been shown to disrupt the GTPase activity. R1441C decrease the stability of the ROC domain of LRRK2 (Yongchao Li *et al.*, 2009). Experiments could show if there is a shared or equivalent function. R526C MFHAS1 constructs show no decrease in MFHAS1 protein expression level, but MAP3K4 steady-state level decreased significantly in the presence of the mutated MFHAS1. Interestingly, when MFHAS1, LRRK2 and the RAS protein sequences are compared in Clustal, R526 aligns with D105 of the RAS proteins. D105 is part of the amino acid essential for binding to external partners (Stouten *et al.*, 1993).

The expression level of MAP3K4 throughout the mutations shows a consistent decrease when MFHAS1 is mutated, except for the K422A mutation. MAP3K4 is found to be interacting with MFHAS1 from all four IP experiments, and none of the mutations impacts the interaction with MAP3K4. The overexpression of the MFHAS1 protein may influence the results. MAP3K4 protein expression seems to be impacted by mutations in MFHAS1, but the same mutations do not impact the interaction between the two proteins. MFHAS1 might be playing a role in MAP3K4 expression. Although LRRK2 Kinase is closer to the RIP Kinase on the Kinome tree, LRRK2 Kinase is sometimes described as a MAPKKK (Gloeckner *et al.*, 2009). Based upon the results reported in this chapter, MAP3K4 might act as the effector kinase for MFHAS1.

In the IP experiments, the  $\alpha$ -tubulin is present in the eluate. This indicates that  $\alpha$ -tubulin might bind to MFHAS1, furthermore, LRRK2 was shown to bind to specific tubulin isoform (Law *et al.*, 2014). Another control would be needed to access this experiment.

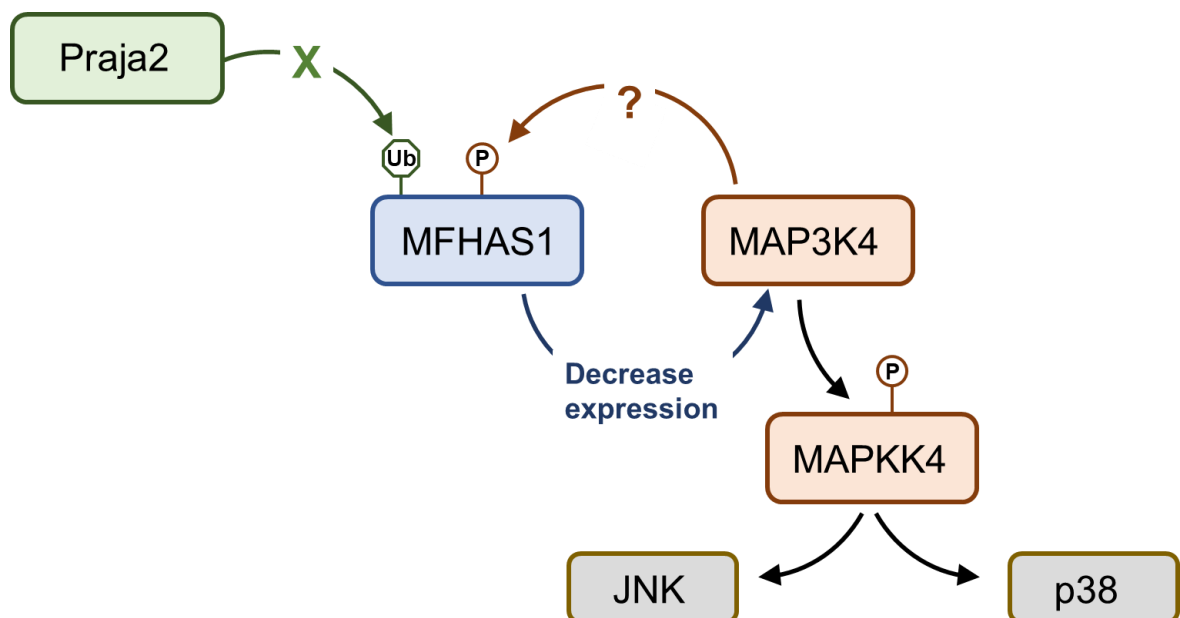
Phosphorylated p38 and p38 are downstream proteins of MAP3K4. Their expression level was compared to understand the impact of MFHAS1 on the MAPKKK/MAPKK/p38 pathway. Phosphorylated p38 expression level decreases significantly when WT MFHAS1 is overexpressed for three experiments out of four. The protein expression level of p38 also decreases when K387A MFHAS1 is overexpressed. The other mutations have no impact on phosphorylated p38. Results would show that overexpression of MFHAS1 impacts the downstream pathway of MAP3K4 and that S450A and R526C mutations would stop MFHAS1 from impacting the phosphorylation of p38. The expression level of the p38 protein is not altered when MFHAS1 is overexpressed.

Phosphorylated p38 is quantified at a 48h timepoint only. It would be interesting to add more timepoint to observe if the level of p38 phosphorylation changes.

The effect of mutated MFHAS1 on the steady state level of other proteins was looked at once with five replicates and should be repeated to confirm these findings. Overexpression of MAP3K4 and PJA2 could be interesting to study and understand the impact on the interaction.

It is important to note that the cytotoxicity of overexpressed MFHAS1 protein was previously reported (Dihanich *et al.*, 2014). It was also observed in this study but not quantified.

The signalling pathway with the results found in this chapter is summarised in **Figure 4.15**.



**Figure 4.15:** Schematic representation of the signalling pathway around MFHAS1 after interpretation of the results. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase kinase 4. MAPKK4: Mitogen-activated protein kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. JNK: c-Jun N-Terminal Protein Kinase. P38: p38 Mitogen-activated protein kinase.

In conclusion, MFHAS1 interacts with MAP3K4, and none of the mutations disrupt this interaction. PJA2's steady state level stays consistent throughout all the conditions, and no interaction is reported with overexpressed MFHAS1.

To further investigate those four mutations, a pull-down using GTP beads could be performed; it would help to understand the impact of the mutations, especially the PD-like mutation present in the ROC domain of MFHAS1.

Other putative PTM referenced for MFHAS1 could be mutated to comprehend their impact on MFHAS1 and its surrounding proteins—mainly the ubiquitination site K601 present in the ROC domain, which aligns with DAPK1 (K939).

Detection of other proteins potentially interacting with MFHAS1 could be done. Proteins found in the first layer of interaction from the protein-protein interaction experiment performed in the bioinformatic chapter could be accessed. In particular, the kinases STK25, GAK, CLK1, LIMK1, ROR1, NEK11, AURKB, NEK1 and FGFR1 and the E3 ubiquitin ligases STUB1, CBLB and TRIM22 could be tested with IP.

Previous studies have supported a cytosolic localisation for MFHAS1 (Dihanich *et al.*, 2014), but the impact of the mutation on the localisation of MFHAS1 could be made by colocalisation in HEK cells. Confocal images with WT and mutant MFHAS1 could be compared to see the impact of the mutation on the localisation of MFHAS1. The localisation of interacting proteins could be determined to determine if MFHAS1 interact with other proteins in specific organelles.

## Chapter 5: Conclusions and discussion

This thesis aimed to understand the structure and function of the MFHAS1 protein. Three approaches were used to gain insight into the biology of MFHAS1.

First, the objective was to solve the protein structure by producing and purifying recombinant MFHAS1 protein.

Second, bioinformatic tools were used to predict the structure of MFHAS1, to determine the proteins interacting with MFHAS1, and to analyse the post-translational modifications and regulation of this protein.

Finally, targeted mutations modulating putative post translational modification sites were inserted into MFHAS1 to test the impact on the biochemistry and cellular role of this protein.

Future work ideas were discussed at the end of each chapter. A more detailed plan is described herein.

The protein production chapter highlighted the substantial challenges of purifying a large, multi-domain protein. A number of purification strategies were used and are detailed in Chapter 2. This chapter concludes that purification using the Streptavidin fusion tag works well. The new domain delineation presented in the bioinformatic chapter can be used to redesign the constructs according to new boundaries.

More recent papers solving the structures of other human ROCO proteins, notably LRRK1 and LRRK2, used insect cells (Mills *et al.*, 2018; Deniston *et al.*, 2020; Reimer *et al.*, 2022; Metcalfe *et al.*, 2023) or baculovirus vector driven expression in HEK293F (Myasnikov *et al.*, 2021). Building on insights from these studies, insect cells could be tried again as an expression system for MFHAS1. Another high-throughput detergent screen can be performed on the expressed protein from insect cells, seeking an optimised purification protocol for MFHAS1.

This chapter focused on purifying the MFHAS1 protein in order to carry out crystallographic or cryo-electron-microscopy characterisation of protein structure. In the event that purified protein can be isolated, other analytical methods could be attempted, like size-exclusion chromatography with multi-angle laser light scattering (SEC-MALLS), which gives



the protein's estimated size and shape. Analytical ultracentrifugation (AUC) examines at the sedimentation velocity or equilibrium and gives information about the molecular weight of a protein and its hydrodynamic shape, the percentage of monomer, multimer and aggregates. Circular Dichroism (CD) measures the absorption difference of the structural asymmetry light. It can estimate the secondary and overall tertiary structure, as well as providing important insights into the stability and folding of the protein.

The bioinformatics chapter provided information on the structure of MFHAS1 with the predicted structures from AlphaFold2 and IntFOLD7. Both predicted structures are monomers, but there is evidence that MFHAS1 is also present as a dimer (Dihanich *et al.*, 2014). Developing a dimer structural model for MFHAS1 would be interesting to understand where the domains, particularly the catalytic ROC domain, sit within the dimeric structure. According to the results in the cellular pathways where MAP3K4 is found to interact with MFHAS1, a structure of the two predicted structures interacting would be interesting to understand in the molecular basis of this interaction. In recent years, a breakthrough happened in structure prediction due to the advance in artificial intelligence opening up the possibility of gaining insights *in silico* prior to wet lab validation (Durham *et al.*, 2023).

In this chapter, the protein-protein interaction (PPI) networks were visualized, and the functional enrichment of the first layer interaction was done. The functional enrichment gives information about the cellular functions present around MFHAS1. An expanded analysis could investigate the co-expression of the proteins and understand the impact of the overexpression of one protein upon the others (Zhao *et al.*, 2023).

Post-translational modification (PTM), including phosphorylation and ubiquitination sites present in MFHAS1, were studied, and compared with members of the ROCO proteins. Following the nomination of putative phosphorylation sites on MFHAS1, kinase prediction was performed for each phosphorylation site.

The final chapter investigates the cellular pathway around MFHAS1. To test the interactions and pathways implicated by data from chapter 3, immunoprecipitation analyses were performed, and two proteins were focused on to test if they interacted with MFHAS1: PJA2

and MAP3K4. Due to constraints of time, a wider validation screen was not possible and so other kinases and E3 ubiquitin ligases proteins found to be interacting with MFHAS1 in the bioinformatic chapter could be investigated by IP as a future continuation of the research described in this thesis.

Endogenous MAP3K4 was found to interact with overexpressed MFHAS1 wild-type and mutant. Further investigation could be done by knocking down, inhibiting or overexpressing MAP3K4 and see if it impacts MFHAS1 protein expression level and downstream pathways, as well as a targeted analysis of domain interactions to test for the molecular level basis of the interaction.

Contrary to previously published work, PJA2 was not found to interact with MFHAS1 (Zhong *et al.*, 2018). This data should be repeated, and PJA2 overexpressed to match the conditions reported in Zhong and co-workers experiments.

At a functional level, IP using Guanosine Triphosphate (GTP) beads could be performed to understand the GTP binding of MFHAS1 and the impact of the mutations on GTP binding. Previously, the ability of MFHAS1 to hydrolyse GTP was assessed for MFHAS1 WT and K422A and no activity was reported (Dihanich *et al.*, 2014). More conditions could be tried as the other mutations studied and the addition of co-factors, and if purified recombinant protein becomes available this would open up the possibility of more sensitive assays.

Mass spectrometry (MS) on IP sample performed by three research found PJA2, HSP60 and HSP70 proteins interacting with MFHAS1 (Dihanich *et al.*, 2014; Zhong *et al.*, 2018; Kumkhaek *et al.*, 2019). MS experiment could be repeated to investigate if more proteins interacting with MFHAS1 would be found. Coomassie blue gel after IP shows a very clear eluate, but few bands are present that could be of interest (**S. Figure 7**).

Whole genome RNA sequencing and MS could also be performed on MFHAS1 KO and WT cell lines to see if the absence of MFHAS1 disrupts other proteins or pathways.

Ubiquitination and phosphorylation sites are reported in MFHAS1 proteins. To investigate further the ubiquitination sites, a few approaches can be used. In Chapter 4, the focus was put on the K387 site, but three other ubiquitination sites are reported. The mutation of those ubiquitination sites and comparing protein expression and interaction could be

interesting. Multiple ubiquitin sites can be mutated on the same construct to understand the impact if MFHAS1 cannot be ubiquitinated.

Ubiquitination sites can be validated by mass spectrometry (MS) (Sun and Zhang, 2022). Biochemistry and imaging can be used to study ubiquitination (van Wijk *et al.*, 2019). Ubiquitin chains can be identified by immunoblotting using anti-ubiquitin antibodies (Emmerich and Cohen, 2015). Different AB can identify if the ubiquitination is monoubiquitination, multi- monoubiquitination or polyubiquitination. Inhibition of deubiquitinases is necessary to keep the ubiquitin chains.

Pharmacological approaches could be tried to inhibit ubiquitination and understand their role (Edelmann, Nicholson and Kessler, 2011). Proteasomal, autophagy and Chaperone-mediated autophagy blockers could be tried to see whether they change the steady state level of MFHAS1 (Myung, Kim and Crews, 2001; Juste and Cuervo, 2019).

Phosphorylation plays a significant role in proteins. It can modify its structure, affect the biological function, and be needed for protein-protein interaction. As for the ubiquitination sites, mutation of all reported phosphorylation sites could showcase the importance of phosphorylation on MFHAS1 protein. Stable isotope labelling by amino acids during cell culture (SILAC) coupled with MS can give information about changes in phosphorylation (Raju, 2019). Developing and then applying phosphospecific antibodies targeting S450 would provide a powerful tool for assaying MFHAS1 regulation in cells, as has been the case for LRRK2.

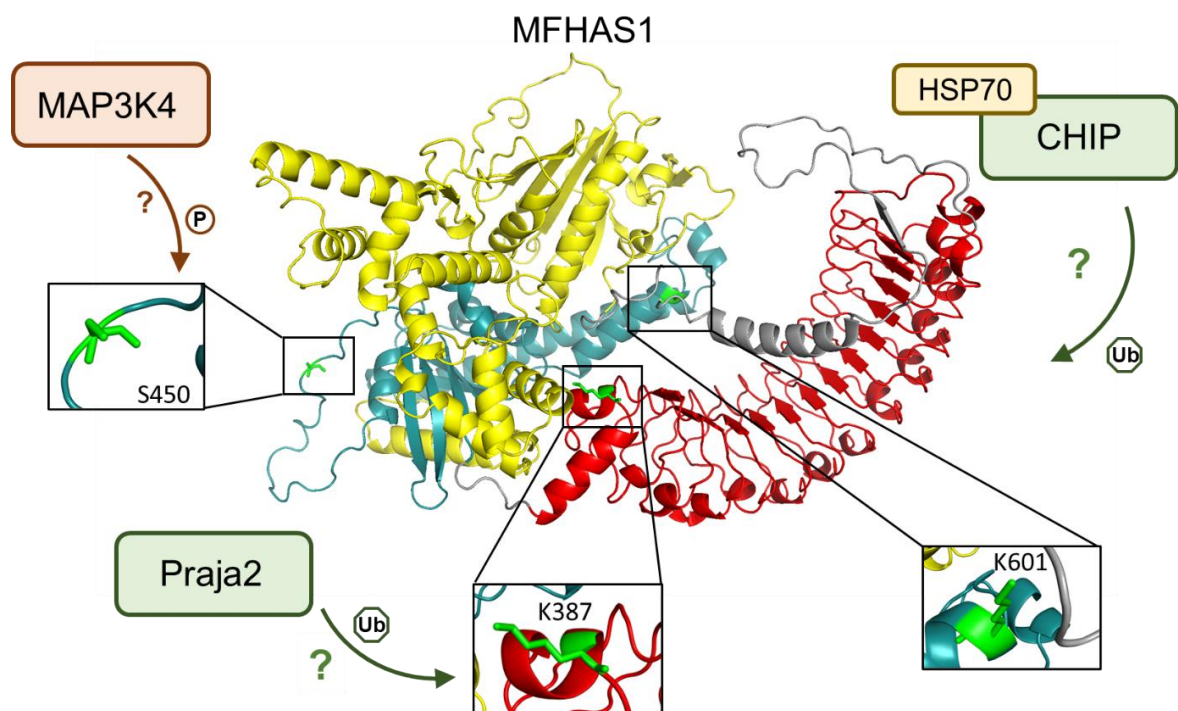
In conclusion, the research performed during this thesis allowed a better understanding of the MFHAS1 interactions with other proteins and possible functions. It also provides an insight into the structure of MFHAS1 and where the PTMs reside.

The main finding is the validation of an interaction of overexpressed MFHAS1 with MAP3K4. MAP3K4 present a kinase domain on its C-terminus; the rest of its structure is not determined. MAP3K4 activates the p38/JNK pathways, regulating inflammation and vascular inflammation (Herlaar and Brown, 1999; Zhou *et al.*, 2021). MAP3K4 is also involved in cancer (Whitmarsh and Davis, 2007) and osteoporosis (Zhang *et al.*, 2018). It is interesting to note that the central pathology linked with LRRK1 is osteoporosis.

CHIP E3 ubiquitin ligase is found to interact with all four ROCO proteins (Tomkins *et al.*, 2018). HSC70, HSP70 and HSP90 are protein chaperones CHIP needs to function correctly (T. Wang *et al.*, 2020). HSP70 interacts with MFHAS1 (Dihanich *et al.*, 2014; Kumkhaek *et al.*, 2019), reinforcing that CHIP might ubiquitinate MFHAS1 (**Figure 5.1**).

The findings in the thesis may have implications beyond MFHAS1. MFHAS1 is part of the ROCO protein family; and it may be that findings across the ROCO protein can be integrated and assessed by analogy. Based on the literature, a hypothetical pathway showcases LRKK2 phosphorylating a MAPKKK or p38 protein, which then regulates autophagy (Obergasteiger *et al.*, 2018).

Finally, as a summation of the analysis in this thesis, we can construct and continue to refine a model where MFHAS1 might take part in the MAPK- p38/JNK cascade. In this context, it would be interesting to directly test whether MAP3K4 can phosphorylate MFHAS1 as well as assessing the direct downstream substrates of MAP3K4 in the presence of potential modulation by MFHAS1 (**Figure 5.1**).



**Figure 5.1:** Schematic representation of MFHAS1 hypothetical pathway based on the finding in this thesis. MAP3K4 is found to be interacting with MFHAS1 but the location is not determined yet. PJA2/Praja2 and CHIP might be ubiquitinating MFHAS1. S450, K387 and K601 have been determined to be the mains post translational modification sites of MFHAS1.

## References

---

Agalliu, I. *et al.* (2015) 'Higher frequency of certain cancers in LRRK2 G2019S mutation carriers with Parkinson disease a pooled analysis', *JAMA Neurology*, 72(1), pp. 58–65. doi: 10.1001/jamaneurol.2014.1973.

Alam, M. A., Uddin, S. J. and Brown, L. (2012) *Mitogen-activated protein kinase and natural phenolic compounds in cardiovascular remodeling*. 1st edn, *Studies in Natural Products Chemistry*. 1st edn. Elsevier B.V. doi: 10.1016/B978-0-444-59530-0.00006-X.

Alanis-Lobato, G., Andrade-Navarro, M. A. and Schaefer, M. H. (2017) 'HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks', *Nucleic Acids Research*, 45(D1), pp. D408–D414. doi: 10.1093/nar/gkw985.

Alonso Guervós, M. *et al.* (2009) 'Genetic differences between primary larynx and pharynx carcinomas and their matched lymph node metastases by multiplex ligation-dependent probe amplification', *Oral Oncology*. Elsevier Ltd, 45(7), pp. 600–604. doi: 10.1016/j.oraloncology.2008.08.005.

Anjum, R. *et al.* (2005) 'The tumor suppressor DAP kinase is a target of RSK-mediated survival signaling', *Current Biology*, 15(19), pp. 1762–1767. doi: 10.1016/j.cub.2005.08.050.

Arakel, E. C. and Schwappach, B. (2018) 'Formation of COPI-coated vesicles at a glance', *Journal of Cell Science*, 131(7). doi: 10.1242/jcs.218347.

Ardito, F. *et al.* (2017) 'The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)', *International Journal of Molecular Medicine*, 40(2), pp. 271–280. doi: 10.3892/ijmm.2017.3036.

Banks, C. A. S. *et al.* (2016) 'TNIP2 is a Hub Protein in the NF- $\kappa$ B Network with Both Protein and RNA Mediated Interactions', *Molecular and Cellular Proteomics*. © 2016 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 15(11), pp. 3435–3449. doi: 10.1074/mcp.M116.060509.

Bao, H. Y. *et al.* (2023) 'Binding modes of GDP, GTP and GNP to NRAS deciphered by using Gaussian accelerated molecular dynamics simulations', *SAR and QSAR in Environmental Research*. Taylor & Francis, 34(1), pp. 65–89. doi: 10.1080/1062936X.2023.2165542.

Barrett, J. C. *et al.* (2008) 'Genome-wide association defines more than 30 distinct susceptibility loci

for Crohn's disease', *Nature Genetics*, 40(8), pp. 955–962. doi: 10.1038/ng.175.

Barrowman, J. *et al.* (2010) 'TRAPP complexes in membrane traffic: Convergence through a common Rab', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 11(11), pp. 759–763. doi: 10.1038/nrm2999.

Beli, P. *et al.* (2012) 'Proteomic Investigations Reveal a Role for RNA Processing Factor THRAP3 in the DNA Damage Response', *Molecular Cell*. Elsevier Inc., 46(2), pp. 212–225. doi: 10.1016/j.molcel.2012.01.026.

Berger, Z., Smith, K. A. and Lavoie, M. J. (2010) 'Membrane localization of LRRK2 is associated with increased formation of the highly active Lrrk2 dimer and changes in its phosphorylation', *Biochemistry*, 49(26), pp. 5511–5523. doi: 10.1021/bi100157u.

Bertoline, L. M. F. *et al.* (2023) 'Before and after AlphaFold2: An overview of protein structure prediction', *Frontiers in Bioinformatics*, 3(February), pp. 1–8. doi: 10.3389/fbinf.2023.1120370.

Berwick, D. C. and Harvey, K. (2011) 'LRRK2 signaling pathways: The key to unlocking neurodegeneration?', *Trends in Cell Biology*. Elsevier Ltd, 21(5), pp. 257–265. doi: 10.1016/j.tcb.2011.01.001.

Berwick, D. C. and Harvey, K. (2012) 'LRRK2 functions as a wnt signaling scaffold, bridging cytosolic proteins and membrane-localized LRP6', *Human Molecular Genetics*, 21(22), pp. 4966–4979. doi: 10.1093/hmg/dd342.

Beurel, E., Grieco, S. F. and Jope, R. S. (2015) 'Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases', *Pharmacology and Therapeutics*. Elsevier Inc., 148, pp. 114–131. doi: 10.1016/j.pharmthera.2014.11.016.

Bialik, S., Bresnick, A. R. and Kimchi, A. (2004) 'DAP-kinase-mediated morphological changes are localization dependent and involve myosin-II phosphorylation', *Cell Death and Differentiation*, 11(6), pp. 631–644. doi: 10.1038/sj.cdd.4401386.

Bialik, S. and Kimchi, A. (2006) 'The Death-Associated Protein Kinases: Structure, Function, and Beyond', *Annual Review of Biochemistry*, 75(1), pp. 189–210. doi: 10.1146/annurev.biochem.75.103004.142615.

Bloem, B. R., Okun, M. S. and Klein, C. (2021) 'Parkinson's disease', *The Lancet*. Elsevier Ltd, 397(10291), pp. 2284–2303. doi: 10.1016/S0140-6736(21)00218-X.

- Blom, N. *et al.* (2004) 'Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence', *Proteomics*, 4(6), pp. 1633–1649. doi: 10.1002/pmic.200300771.
- Blom, N., Gammeltoft, S. and Brunak, S. (1999) 'Sequence and structure-based prediction of eukaryotic protein phosphorylation sites', *Journal of Molecular Biology*, 294(5), pp. 1351–1362. doi: 10.1006/jmbi.1999.3310.
- Bludau, I. *et al.* (2022) 'The structural context of posttranslational modifications at a proteome-wide scale', *PLoS Biology*, 20(5), pp. 1–23. doi: 10.1371/journal.pbio.3001636.
- Bockmühl, U. *et al.* (2001) 'Association of 8p23 deletions with poor survival in head and neck cancer', *Otolaryngology - Head and Neck Surgery*, 124(4), pp. 451–455. doi: 10.1067/mhn.2001.114794.
- Bosgraaf, L. *et al.* (2002) 'A novel cGMP signalling pathway mediating myosin phosphorylation and chemotaxis in Dictyostelium', *EMBO Journal*, 21(17), pp. 4560–4570. doi: 10.1093/emboj/cdf438.
- Bosgraaf, L. and Van Haastert, P. J. M. (2003) 'Roc, a Ras/GTPase domain in complex proteins', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1643(1–3), pp. 5–10. doi: 10.1016/j.bbamcr.2003.08.008.
- Bouwmeester, T. *et al.* (2004) 'A physical and functional map of the human TNF- $\alpha$ /NF- $\kappa$ B signal transduction pathway', *Nature Cell Biology*, 6(2), pp. 97–105. doi: 10.1038/ncb1086.
- Bradley, D. *et al.* (2021) 'Sequence and Structure-Based Analysis of Specificity Determinants in Eukaryotic Protein Kinases', *Cell Reports*, 34(2). doi: 10.1016/j.celrep.2020.108602.
- Bradley, D. (2022) 'The evolution of post-translational modifications', *Current Opinion in Genetics and Development*. Elsevier, 76, p. 101956. doi: 10.1016/j.gde.2022.101956.
- Brennan, K. *et al.* (2004) 'Truncated mutants of the putative Wnt receptor LRP6/Arrow can stabilize  $\beta$ -catenin independently of Frizzled proteins', *Oncogene*, 23(28), pp. 4873–4884. doi: 10.1038/sj.onc.1207642.
- Brunet, S. and Sacher, M. (2014) 'in sickness and in health: The role of TRAPP and associated proteins in disease', *Traffic*, 15(8), pp. 803–818. doi: 10.1111/tra.12183.
- Bykov, Y. S. *et al.* (2017) 'The structure of the COPI coat determined within the cell', *eLife*, 6, pp. 1–18. doi: 10.7554/eLife.32493.
- Callis, J. (2014) 'The Ubiquitination Machinery of the Ubiquitin System', *The Arabidopsis Book*, 12, p. e0174. doi: 10.1199/tab.0174.

- Calmon, M. F. *et al.* (2007) 'Methylation profile of genes CDKN2A (p14 and p16), DAPK1, CDH1, and ADAM23 in head and neck cancer', *Cancer Genetics and Cytogenetics*, 173(1), pp. 31–37. doi: 10.1016/j.cancergencyto.2006.09.008.
- Carlessi, R. *et al.* (2011) 'GTP binding to the ROC domain of DAP-kinase regulates its function through intramolecular signalling', *EMBO Reports*. Nature Publishing Group, 12(9), pp. 917–923. doi: 10.1038/embor.2011.126.
- Chen, C. H. *et al.* (2005) 'Bidirectional signals transduced by DAPK-ERK interaction promote the apoptotic effect of DAPK', *EMBO Journal*, 24(2), pp. 294–304. doi: 10.1038/sj.emboj.7600510.
- Chen, M. L. and Wu, R. M. (2018) 'LRRK 2 gene mutations in the pathophysiology of the ROCO domain and therapeutic targets for Parkinson's disease: A review Julie Y.H. Chan', *Journal of Biomedical Science*. Journal of Biomedical Science, 25(1), pp. 1–11. doi: 10.1186/s12929-018-0454-0.
- Chen, W. *et al.* (2016) 'MFHAS1 promotes colorectal cancer progress by regulating polarization of tumor-associated macrophages via STAT6 signaling pathway', *Oncotarget*, 7(48). doi: 10.18632/oncotarget.12807.
- Chung, H. lok *et al.* (2020) 'De Novo Variants in CDK19 Are Associated with a Syndrome Involving Intellectual Disability and Epileptic Encephalopathy', *American Journal of Human Genetics*. ElsevierCompany., 106(5), pp. 717–725. doi: 10.1016/j.ajhg.2020.04.001.
- Cohen, O., Feinstein, E. and Kimchi, A. (1997) 'DAP-kinase is a Ca<sup>2+</sup>/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity', *EMBO Journal*, 16(5), pp. 998–1008. doi: 10.1093/emboj/16.5.998.
- Cole, S. and Prabakaran, S. (2020) 'PhosphoEffect: Prioritizing Variants On or Adjacent to Phosphorylation Sites through Their Effect on Kinase Recognition Motifs', *iScience*. Elsevier Inc., 23(8), p. 101321. doi: 10.1016/j.isci.2020.101321.
- Damgaard, R. B. (2021) 'The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities', *Cell Death and Differentiation*. Springer US, 28(2), pp. 423–426. doi: 10.1038/s41418-020-00703-w.
- Das, S. and Krainer, A. R. (2014) 'Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer', *Molecular Cancer Research*, 12(9), pp. 1195–1204. doi: 10.1158/1541-7786.MCR-14-0131.



- Deiss, L. P. *et al.* (1995) 'Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the  $\gamma$  interferon-induced cell death', *Genes and Development*, 9(1), pp. 15–30. doi: 10.1101/gad.9.1.15.
- Deng, J. *et al.* (2008) 'Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase', *Proceedings of the National Academy of Sciences*, 105(5), pp. 1499–1504. doi: 10.1073/pnas.0709098105.
- Deniston, C. K. *et al.* (2020) 'Structure of LRRK2 in Parkinson's disease and model for microtubule interaction', *Nature*, 588(7837), pp. 344–349. doi: 10.1038/s41586-020-2673-2.
- Deshai, R. J. and Joazeiro, C. A. P. (2009) 'RING domain E3 ubiquitin ligases', *Annual Review of Biochemistry*, 78, pp. 399–434. doi: 10.1146/annurev.biochem.78.101807.093809.
- Deyaert, E. *et al.* (2017) 'A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover', *Nature Communications*. Springer US, 8(1), pp. 1–12. doi: 10.1038/s41467-017-01103-4.
- Deyaert, E. *et al.* (2019) 'Structure and nucleotide-induced conformational dynamics of the *Chlorobium tepidum* Roco protein', *Biochemical Journal*, 476(1), pp. 51–66.
- Dihanich, S. (2012) 'MASL1: a neglected ROCO protein', *Biochemical Society Transactions*, 40(5), pp. 1090–1094. doi: 10.1042/BST20120127.
- Dihanich, S. *et al.* (2014) 'GTP binding controls complex formation by the human ROCO protein MASL1', *FEBS Journal*, 281(1), pp. 261–274. doi: 10.1111/febs.12593.
- Dikic, I. and Schulman, B. A. (2022) 'An expanded lexicon for the ubiquitin code', *Nature Reviews Molecular Cell Biology*. Springer US, 24(April). doi: 10.1038/s41580-022-00543-1.
- Dillon, M. *et al.* (2021) 'Progress on Ras/MAPK Signaling Research and Targeting in Blood and Solid Cancers', *cancers*, 13(5059), pp. 1–35. Available at: <https://doi.org/10.3390/cancers13205059>.
- Dodonova, S. O. *et al.* (2017) '9Å structure of the COPI coat reveals that the Arf1 GTPase occupies two contrasting molecular environments', *eLife*, 6, pp. 1–29. doi: 10.7554/eLife.26691.
- Dominguez, D. *et al.* (2016) 'An extensive program of periodic alternative splicing linked to cell cycle progression', *eLife*, 5(MARCH2016), pp. 1–19. doi: 10.7554/eLife.10288.
- Duan, D. X. *et al.* (2013) 'Phosphorylation of tau by death-associated protein kinase 1 antagonizes the kinase-induced cell apoptosis', *Journal of Alzheimer's Disease*, 37(4), pp. 795–808. doi:

10.3233/JAD-130377.

Dubois, S. *et al.* (2016) 'Next-generation sequencing in diffuse large B-cell lymphoma highlights molecular divergence and therapeutic opportunities: A LYSA study', *Clinical Cancer Research*, 22(12), pp. 2919–2928. doi: 10.1158/1078-0432.CCR-15-2305.

Dunnett-Kane, V. *et al.* (2020) 'Germline and sporadic cancers driven by the RAS pathway: parallels and contrasts', *Annals of Oncology*, 31(7), pp. 873–883. doi: 10.1016/j.annonc.2020.03.291.

Durham, J. *et al.* (2023) 'Recent advances in predicting and modeling protein–protein interactions', *Trends in Biochemical Sciences*, 48(6), pp. 527–538. doi: 10.1016/j.tibs.2023.03.003.

Edelmann, M. J., Nicholson, B. and Kessler, B. M. (2011) 'Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases.', *Expert reviews in molecular medicine*, 13(November), pp. 1–17. doi: 10.1017/s1462399411002031.

Van Egmond, W. N. *et al.* (2008) 'Intramolecular activation mechanism of the Dictyostelium LRRK2 homolog Roco protein GbpC', *Journal of Biological Chemistry*, 283(44), pp. 30412–30420. doi: 10.1074/jbc.M804265200.

van Egmond, W. N. and van Haastert, P. J. M. (2010) 'Characterization of the Roco protein family in Dictyostelium discoideum', *Eukaryotic Cell*, 9(5), pp. 751–761. doi: 10.1128/EC.00366-09.

Eid, S. *et al.* (2017) 'Kinmap: A web-based tool for interactive navigation through human kinome data', *BMC Bioinformatics*. BMC Bioinformatics, 18(1), pp. 1–6. doi: 10.1186/s12859-016-1433-7.

Eisenberg-Lerner, A. and Kimchi, A. (2012) 'PKD is a kinase of Vps34 that mediates ROS-induced autophagy downstream of DAPK', *Cell Death and Differentiation*. Nature Publishing Group, 19(5), pp. 788–797. doi: 10.1038/cdd.2011.149.

Emmerich, C. H. and Cohen, P. (2015) 'Optimising methods for the preservation, capture and identification of ubiquitin chains and ubiquitylated proteins by immunoblotting', *Biochemical and Biophysical Research Communications*. Elsevier Ltd, 466(1), pp. 1–14. doi: 10.1016/j.bbrc.2015.08.109.

Fernández-Medarde, A. and Santos, E. (2011) 'Ras in cancer and developmental diseases', *Genes and Cancer*, 2(3), pp. 344–358. doi: 10.1177/1947601911411084.

Fu-Ren, Z. *et al.* (2009) 'Genomewide Association Study of Leprosy', *The new england journal of medicine original*, pp. 2609–2618.

- Fujita, K. *et al.* (2022) 'The ULK complex–LRRK1 axis regulates Parkin-mediated mitophagy via Rab7 Ser-72 phosphorylation', *Journal of Cell Science*, 135(23). doi: 10.1242/jcs.260395.
- Fukuzono, T. *et al.* (2016) 'Chaperone complex BAG2-HSC70 regulates localization of *Caenorhabditis elegans* leucine-rich repeat kinase LRK-1 to the Golgi', *Genes to Cells*, 21(4), pp. 311–324. doi: 10.1111/gtc.12338.
- Funayama, M. *et al.* (2002) 'A new locus for Parkinson's Disease (PARK8) maps to chromosome 12p11.2-q13.1', *Annals of Neurology*, 51(3), pp. 296–301. doi: 10.1002/ana.10113.
- Futatsumori, M. *et al.* (2000) 'Identification and characterization of novel isoforms of COP I subunits', *Journal of Biochemistry*, 128(5), pp. 793–801. doi: 10.1093/oxfordjournals.jbchem.a022817.
- Gade, P. *et al.* (2014) 'Regulation of the Death-Associated Protein Kinase 1 Expression and Autophagy via ATF6 Requires Apoptosis Signal-Regulating Kinase 1', *Molecular and Cellular Biology*, 34(21), pp. 4033–4048. doi: 10.1128/mcb.00397-14.
- García-Fruitós, E. (2015) *Insoluble proteins, Methods in Molecular Biology*. doi: 10.1007/978-1-4939-2205-5\_19.
- Gimple, R. C. and Wang, X. (2019) 'RAS: Striking at the Core of the Oncogenic Circuitry', *Frontiers in Oncology*, 9(September), pp. 1–16. doi: 10.3389/fonc.2019.00965.
- Gloeckner, C. J. *et al.* (2006) 'The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity', *Human Molecular Genetics*, 15(2), pp. 223–232. doi: 10.1093/hmg/ddi439.
- Gloeckner, C. J. *et al.* (2009) 'The Parkinson disease-associated protein kinase LRRK2 exhibits MAPKKK activity and phosphorylates MKK3/6 and MKK4/7, in vitro', *Journal of Neurochemistry*, 109(4), pp. 959–968. doi: 10.1111/j.1471-4159.2009.06024.x.
- Gong, X. *et al.* (2003) 'Cdk5-mediated inhibition of the protective effects of transcription factor MEF2 in neurotoxicity-induced apoptosis', *Neuron*, 38(1), pp. 33–46. doi: 10.1016/S0896-6273(03)00191-0.
- Gotthardt, K. *et al.* (2008) 'Structure of the Roc-COR domain tandem of *C. tepidum*, a prokaryotic homologue of the human LRRK2 Parkinson kinase', *EMBO Journal*, 27(16), pp. 2239–2249. doi: 10.1038/emboj.2008.150.

- Graham, F. L. *et al.* (1977) 'Characteristics of a human cell line transformed by DNA from human adenovirus type 5', *Journal of General Virology*, 36(1), pp. 59–72. doi: 10.1099/0022-1317-36-1-59.
- Greggio, E. *et al.* (2006) 'Kinase activity is required for the toxic effects of mutant LRRK2/dardarin', *Neurobiology of Disease*, 23(2), pp. 329–341. doi: 10.1016/j.nbd.2006.04.001.
- Greggio, E. *et al.* (2009) 'The Parkinson's disease kinase LRRK2 autophosphorylates its GTPase domain at multiple sites', *Biochemical and Biophysical Research Communications*. Elsevier Inc., 389(3), pp. 449–454. doi: 10.1016/j.bbrc.2009.08.163.
- Guenebeaud, C. *et al.* (2010) 'The Dependence Receptor UNC5H2/B Triggers Apoptosis via PP2A-Mediated Dephosphorylation of DAP Kinase', *Molecular Cell*, 40(6), pp. 863–876. doi: 10.1016/j.molcel.2010.11.021.
- Gurung, A. B. and Bhattacharjee, A. (2015) 'Significance of Ras Signaling in Cancer and Strategies for its Control', *Oncology & Hematology Review (US)*, 11(02), p. 147. doi: 10.17925/ohr.2015.11.02.147.
- Hanafusa, H. *et al.* (2011) 'Leucine-rich repeat kinase LRRK1 regulates endosomal trafficking of the EGF receptor', *Nature Communications*, 2(1). doi: 10.1038/ncomms1161.
- Hanafusa, H. *et al.* (2019) 'LRRK1 phosphorylation of Rab7 at S72 links trafficking of EGFR-containing endosomes to its effector RILP', *Journal of Cell Science*, 132(11). doi: 10.1242/jcs.228809.
- Hanafusa, H. and Matsumoto, K. (2015) 'LRRK1 regulates spindle orientation by phosphorylating CDK5RAP2', *Cell Cycle*, 14(21), pp. 3349–3350. doi: 10.1080/15384101.2015.1093446.
- Hao, J. *et al.* (2021) 'Effect of MAP3K8 on Prognosis and Tumor-Related Inflammation in Renal Clear Cell Carcinoma', *Frontiers in Genetics*, 12(September), pp. 1–13. doi: 10.3389/fgene.2021.674613.
- Harada, J. N. *et al.* (2005) 'Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis', *Genome Research*, 15(8), pp. 1136–1144. doi: 10.1101/gr.3889305.
- Härtlova, A. *et al.* (2018) 'LRRK2 is a negative regulator of Mycobacterium tuberculosis phagosome maturation in macrophages', *The EMBO Journal*, 37(12), pp. 1–17. doi: 10.15252/emboj.201798694.
- Hatakeyama, S. (2011) 'TRIM proteins and cancer', *Nature Reviews Cancer*. Nature Publishing Group, 11(11), pp. 792–804. doi: 10.1038/nrc3139.
- Healy, M. D. *et al.* (2023) 'Structure of the endosomal Commander complex linked to Ritscher-

- Schinzel syndrome', *Cell*, 186(10), pp. 2219–2237.e29. doi: 10.1016/j.cell.2023.04.003.
- Herlaar, E. and Brown, Z. (1999) 'p38 MAPK signalling cascades in inflammatory disease', *Molecular Medicine Today*, 5(10), pp. 439–447. doi: 10.1016/S1357-4310(99)01544-0.
- Hernandez-Beeftink, T. *et al.* (2022) 'Admixture Mapping of Sepsis in European Individuals With African Ancestries', *Frontiers in Medicine*, 9(March), pp. 1–11. doi: 10.3389/fmed.2022.754440.
- Hornbeck, P. V. *et al.* (2015) 'PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations', *Nucleic Acids Research*, 43(D1), pp. D512–D520. doi: 10.1093/nar/gku1267.
- Howaldt, A. *et al.* (2020) 'Adult osteosclerotic metaphyseal dysplasia with progressive osteonecrosis of the jaws and abnormal bone resorption pattern due to a LRRK1 splice site mutation.', *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, 00(00), pp. 1–11. doi: 10.1002/jbmr.3995.
- Iida, A. *et al.* (2018) 'Identification of biallelic LRRK1 mutations in osteosclerotic metaphyseal dysplasia and evidence for locus heterogeneity', *Journal Medecin Genetic*, 53(8), pp. 568–574. doi: 10.1136/jmedgenet-2016-103756.Identification.
- Ishikawa, K. *et al.* (2012) 'EGFR-dependent phosphorylation of leucine-rich repeat kinase LRRK1 is important for proper endosomal trafficking of EGFR', *Molecular Biology of the Cell*, 23(7), pp. 1294–1306. doi: 10.1091/mbc.E11-09-0780.
- Jebelli, J. D. *et al.* (2012) 'GTP binding and intramolecular regulation by the ROC domain of Death Associated Protein Kinase 1', *Scientific Reports*, 2, pp. 1–7. doi: 10.1038/srep00695.
- Jiang, Y. *et al.* (1997) 'Characterization of the structure and function of the fourth member of p38 group mitogen-activated protein kinases, p38 $\delta$ ', *Journal of Biological Chemistry*. © 1997 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 272(48), pp. 30122–30128. doi: 10.1074/jbc.272.48.30122.
- Jisna, V. A. and Jayaraj, P. B. (2021) 'Protein Structure Prediction: Conventional and Deep Learning Perspectives', *Protein Journal*. Springer US, 40(4), pp. 522–544. doi: 10.1007/s10930-021-10003-y.
- Johnson, J. L. *et al.* (2023) 'An atlas of substrate specificities for the human serine/threonine kinome', *Nature*, 613(7945), pp. 759–766. doi: 10.1038/s41586-022-05575-3.
- Johnson, L. N. and Barford, D. (1993) 'The effects of phosphorylation on the structure and function of proteins', *Annual Review of Biophysics and Biomolecular Structure*, 22, pp. 199–232. doi:

10.1146/annurev.bb.22.060193.001215.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*. Springer US, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Juste, Y. R. and Cuervo, A. M. (2019) 'Analysis of chaperone-mediated autophagy', *Methods in Molecular Biology*, 1880, pp. 703–727. doi: 10.1007/978-1-4939-8873-0\_47.

Kang, J. (2015) 'Genomic alterations on 8p21-p23 are the most frequent genetic events in stage I squamous cell carcinoma of the lung', *Experimental and Therapeutic Medicine*, 9(2), pp. 345–350. doi: 10.3892/etm.2014.2123.

Karnoub, A. E. and Weinberg, R. A. (2008) 'Ras oncogenes: Split personalities', *Nature Reviews Molecular Cell Biology*, 9(7), pp. 517–531. doi: 10.1038/nrm2438.

Kedashiro, S. *et al.* (2015) 'LRRK1-phosphorylated CLIP-170 regulates EGFR trafficking by recruiting p150Glued to microtubule plus ends', *Journal of Cell Science*, 128(4), p. 829. doi: 10.1242/jcs.169102.

Keerthikumar, S. and Mathivanan, S. (2017) *Proteome Bioinformatics, Methods in Molecular Biology*. doi: 10.1007/978-1-4939-6740-7\_4.

Khoury, G. A., Baliban, R. C. and Floudas, C. A. (2011) 'Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database', *Scientific Reports*, 1, pp. 1–5. doi: 10.1038/srep00090.

Kim, B. M. *et al.* (2014) 'Death-associated protein kinase 1 has a critical role in aberrant tau protein regulation and function.', *Cell death & disease*. Nature Publishing Group, 5(5), pp. e1237-12. doi: 10.1038/cddis.2014.216.

Kinoshita, T. *et al.* (1997) 'Raf/MAPK and rapamycin-sensitive pathways mediate the anti-apoptotic function of p21Ras in IL-3-dependent hematopoietic cells', *Oncogene*, 15(6), pp. 619–627. doi: 10.1038/sj.onc.1201234.

Ko, H. S. *et al.* (2009) 'CHIP regulates leucine-rich repeat kinase-2 ubiquitination, degradation, and toxicity', *Proceedings of the National Academy of Sciences of the United States of America*, 106(8), pp. 2897–2902. doi: 10.1073/pnas.0810123106.

Kobe, B. and Kajava, A. V (2001) 'The leucine-rich repeat as a protein recognition motif', *Elsevier Science*, 11, pp. 725–732.

- Korr, D. *et al.* (2006) 'LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain', *Cellular Signalling*, 18(6), pp. 910–920. doi: 10.1016/j.cellsig.2005.08.015.
- Kucera, M. *et al.* (2016) 'AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations', *F1000Research*, 5, pp. 1–12. doi: 10.12688/F1000RESEARCH.9090.1.
- Kukol, A. (2015) *Molecular modeling of proteins, : methods in molecular biology*. second edi. Edited by A. Kukol. Hertfordshire , UK: Humana Press. Available at: <http://www.springer.com/series/7651>.
- Kumari, U. and Tan, E. K. (2009) 'LRRK2 in Parkinson's disease: Genetic and clinical studies from patients', *FEBS Journal*, 276(22), pp. 6455–6463. doi: 10.1111/j.1742-4658.2009.07344.x.
- Kumkhaek, C. *et al.* (2013) 'MASL1 induces erythroid differentiation in human erythropoietin-dependent CD34<sup>+</sup> cells through the Raf / MEK / ERK pathway', *BLOOD*, 121(16), pp. 3216–3228. doi: 10.1182/blood-2011-10-385252.The.
- Kumkhaek, C. *et al.* (2019) 'Role of MFHAS1 in regulating hepcidin expression via the BMP/SMAD and MAPK/ERK1/2 signalling pathways', *British Journal of Haematology*, pp. e108–e112. doi: 10.1111/bjh.15939.
- Kuo, J. C. *et al.* (2006) 'The tumor suppressor DAPK inhibits cell motility by blocking the integrin-mediated polarity pathway', *Journal of Cell Biology*, 172(4), pp. 619–631. doi: 10.1083/jcb.200505138.
- Kwon, T. W. *et al.* (2016) 'DANGER is involved in high glucose-induced radioresistance through inhibiting DAPK-mediated anoikis in non-small cell lung cancer', *Oncotarget*, 7(6), pp. 7193–7206. doi: 10.18632/oncotarget.6887.
- Laulumaa, S. and Varjosalo, M. (2021) 'Commander complex—a multifaceted operator in intracellular signaling and cargo', *Cells*, 10(12). doi: 10.3390/cells10123447.
- Law, B. M. H. *et al.* (2014) 'A direct interaction between leucine-rich repeat kinase 2 and specific  $\beta$ -Tubulin isoforms regulates tubulin acetylation', *Journal of Biological Chemistry*, 289(2), pp. 895–908. doi: 10.1074/jbc.M113.507913.
- Lewis, P. A. *et al.* (2007) 'The R1441C mutation of LRRK2 disrupts GTP hydrolysis', *Biochemical and Biophysical Research Communications*, 357(3), pp. 668–671. doi: 10.1016/j.bbrc.2007.04.006.
- Li, C. *et al.* (2017) 'COPI – TRAPP II activates Rab18 and regulates its lipid droplet association', *The EMBO Journal*, 36(4), pp. 441–457. doi: 10.15252/emj.201694866.

- Li, X. *et al.* (2007) 'Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants', *Journal of Neurochemistry*, 103(1), pp. 238–247. doi: 10.1111/j.1471-4159.2007.04743.x.
- Li, Yanping *et al.* (2009) 'Mutant LRRK2R1441G BAC transgenic mice recapitulate cardinal features of Parkinson's disease', *Nature Neuroscience*. Nature Publishing Group, 12(7), pp. 826–828. doi: 10.1038/nn.2349.
- Li, Yongchao *et al.* (2009) 'The R1441C mutation alters the folding properties of the ROC domain of LRRK2', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier B.V., 1792(12), pp. 1194–1197. doi: 10.1016/j.bbadis.2009.09.010.
- Lian, H., Cheng, Y. and Wu, X. (2019) 'Malignant fibrous histiocytoma amplified sequence 1 alleviates inflammation and renal fibrosis in diabetic nephropathy by inhibiting TLR4', *Bioscience Reports*, 39(11), pp. 1–11. doi: 10.1042/BSR20190617.
- Lim, P. S., Sutton, C. R. and Rao, S. (2015) 'Protein kinase C in the immune system: From signalling to chromatin regulation', *Immunology*, 146(4), pp. 508–522. doi: 10.1111/imm.12510.
- Liu, L. *et al.* (2019) 'UbiHub: A data hub for the explorers of ubiquitination pathways', *Bioinformatics*, 35(16), pp. 2882–2884. doi: 10.1093/bioinformatics/bty1067.
- Liu, Q., Langdon, W. Y. and Zhang, J. (2014) 'E3 ubiquitin ligase Cbl-b in innate and adaptive immunity', *Cell Cycle*, 13(12), pp. 1875–1884. doi: 10.4161/cc.29213.
- Liu, X. and Liu, Y. (2022) 'Comprehensive Analysis of the Expression and Prognostic Significance of the CENP Family in Breast Cancer', *International Journal of General Medicine*, 15(March), pp. 3471–3482. doi: 10.2147/IJGM.S354200.
- Madeira, F. *et al.* (2022) 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic Acids Research*, pp. 1–4.
- Manning, G. *et al.* (2002) 'Materials and Methods for "The protein kinase complement of the human genome"', *Science (New York, N.Y.)*, 298(December). Available at: <http://science.sciencemag.org/>.
- Manzoni, C. *et al.* (2018) 'Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences', *Briefings in Bioinformatics*, 19(2), pp. 286–302. doi: 10.1093/BIB/BBW114.
- Mareschal, S. *et al.* (2016) 'Whole Exome Sequencing of Relapsed/Refractory Patients Expands the



- Repertoire of Somatic Mutations in Diffuse Large B-Cell Lymphoma', *Genes, Chromosomes Cancer*, 396(January), pp. 389–396. doi: 10.1002/gcc.
- Marín, I. (2006) 'The Parkinson disease gene LRRK2: Evolutionary and structural insights', *Molecular Biology and Evolution*, 23(12), pp. 2423–2433. doi: 10.1093/molbev/msl114.
- Marín, I. (2008) 'Ancient origin of the parkinson disease gene LRRK2', *Journal of Molecular Evolution*, 67(1), pp. 41–50. doi: 10.1007/s00239-008-9122-4.
- Marin, I., van Egmond, W. N. and van Haastert, P. J. M. (2008) 'The Roco protein family: a functional perspective', *The FASEB Journal*, 22(9), pp. 3103–3110. doi: 10.1096/fj.08-111310.
- Mcguffin, L. J. (2008) 'Intrinsic disorder prediction from the analysis of multiple protein fold recognition models', *Bioinformatics*, 24(16), pp. 1798–1804. doi: 10.1093/bioinformatics/btn326.
- Mcguffin, L. J. *et al.* (2019) 'IntFOLD: An integrated web resource for high performance protein structure and function prediction', *Nucleic Acids Research*. Oxford University Press, 47(W1), pp. W408–W413. doi: 10.1093/nar/gkz322.
- Mcguffin, L. J. *et al.* (2023) 'Prediction of protein structures , functions and interactions using the IntFOLD7 , MultiFOLD and ModFOLDdock servers', *Nucleic Acids Research*. Oxford University Press, pp. 1–7.
- McGuffin, L. J. *et al.* (2017) 'Accurate template-based modeling in CASP12 using the IntFOLD4-TS, ModFOLD6, and ReFOLD methods', *Proteins: Structure, Function and Bioinformatics*, 86(May 2017), pp. 335–344. doi: 10.1002/prot.25360.
- Medvar, B. *et al.* (2016) 'Comprehensive database of human E3 ubiquitin ligases: Application to aquaporin-2 regulation', *Physiological Genomics*, 48(7), pp. 502–512. doi: 10.1152/physiolgenomics.00031.2016.
- Metcalfe, R. D. *et al.* (2023) 'Structure and regulation of full-length human leucine-rich repeat kinase 1', *Nature Communications*. Springer US, (December 2022), p. 2022.12.21.521433. doi: 10.1038/s41467-023-40532-2.
- Mi, K. and Johnson, G. (2006) 'The Role of Tau Phosphorylation in the Pathogenesis of Alzheimers Disease', *Current Alzheimer Research*, 3(5), pp. 449–463. doi: 10.2174/156720506779025279.
- Mills, R. D. *et al.* (2014) 'Prediction of the Repeat Domain Structures and Impact of Parkinsonism-Associated Variations on Structure and Function of all Functional Domains of Leucine-Rich Repeat

- Kinase 2 (LRRK2)', *Human Mutation*, 35(4), pp. 395–412. doi: 10.1002/humu.22515.
- Mills, R. D. *et al.* (2018) 'The Roc-COR Tandem Domain of Leucine-Rich Repeat Kinase 2 (LRRK2) Forms Dimers and Exhibits Conventional Ras-Like GTPase Properties', *Journal of Neurochemistry*. doi: 10.1111/jnc.14566.
- Miryounesi, M. *et al.* (2020) 'A novel homozygous LRRK1 stop gain mutation in a patient suspected with osteosclerotic metaphyseal dysplasia', *Annals of Human Genetics*, 84(1), pp. 102–106. doi: 10.1111/ahg.12352.
- Missaoui, N. *et al.* (2010) 'Promoter hypermethylation of CDH13, DAPK1 and TWIST1 genes in precancerous and cancerous lesions of the uterine cervix', *Pathology Research and Practice*. Elsevier GmbH., 207(1), pp. 37–42. doi: 10.1016/j.prp.2010.11.001.
- Mompeo, O. *et al.* (2022) 'Genome-Wide Association Analysis of Over 170,000 Individuals from the UK Biobank Identifies Seven Loci Associated with Dietary Approaches to Stop Hypertension (DASH) Diet', *Nutrients*, 14(20). doi: 10.3390/nu14204431.
- Morimoto, K. *et al.* (2016) 'LRRK1 is critical in the regulation of B-cell responses and CARMA1-dependent NF- $\kappa$ B activation', *Scientific Reports*. Nature Publishing Group, 6(May), pp. 1–13. doi: 10.1038/srep25738.
- Myasnikov, A. *et al.* (2021) 'Structural analysis of the full-length human LRRK2', *Cell*. Elsevier Inc., 184(13), pp. 1–9. doi: 10.1016/j.cell.2021.05.004.
- Myung, J., Kim, K. B. and Crews, C. M. (2001) 'The ubiquitin-proteasome pathway and proteasome inhibitors', *Medicinal Research Reviews*, 21(4), pp. 245–273. doi: 10.1002/med.1009.
- Nakamura, S. *et al.* (2008) 'KIS induces proliferation and the cell cycle progression through the phosphorylation of p27Kip1 in leukemia cells', *Leukemia Research*, 32(9), pp. 1358–1365. doi: 10.1016/j.leukres.2008.02.012.
- Needleman, S. B. and Wunsch, C. D. (1970) 'A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins', *Molecular Biology*, 48, pp. 453–463. doi: 10.1016/b978-0-12-131200-8.50031-9.
- Ng, A. C. Y. *et al.* (2011) 'Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity', *Proceedings of the National Academy of Sciences*, 108(Supplement\_1), pp. 4631–4638. doi: 10.2139/ssrn.2796474.

- Nishi, H., Hashimoto, K. and Panchenko, A. R. (2011) 'Phosphorylation in protein-protein binding: Effect on stability and function', *Structure*. Elsevier Ltd, 19(12), pp. 1807–1815. doi: 10.1016/j.str.2011.09.021.
- Nishi, H., Shaytan, A. and Panchenko, A. R. (2014) 'Physicochemical mechanisms of protein regulation by phosphorylation', *Frontiers in Genetics*, 5(AUG), pp. 1–10. doi: 10.3389/fgene.2014.00270.
- Obergasteiger, J. *et al.* (2018) 'A new hypothesis for Parkinson's disease pathogenesis: GTPase-p38 MAPK signaling and autophagy as convergence points of etiology and genomics', *Molecular Neurodegeneration*. *Molecular Neurodegeneration*, 13(1), pp. 1–17. doi: 10.1186/s13024-018-0273-5.
- Ohi, K. *et al.* (2017) 'Spatial and temporal expression patterns of genes around nine neuroticism-associated loci', *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. Elsevier, 77(April), pp. 164–171. doi: 10.1016/j.pnpbp.2017.04.019.
- Olsen, J. V. *et al.* (2010) 'Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis', *Science Signaling*, 3(104). doi: 10.1126/scisignal.2000475.
- Pagani, I., Poli, G. and Vicenzi, E. (2021) 'TRIM22. A multitasking antiviral factor', *Cells*, 10(8), pp. 1–14. doi: 10.3390/cells10081864.
- Paisan-Rui, C. *et al.* (2004) 'Cloning of the Gene Containing Mutations that Cause PARK8 -Linked Parkinson's Disease', *Neuron*, 44, pp. 595–600. doi: 10.1016/j.neuron.2004.10.023.
- Pálffy, G., Menyhárd, D. K. and Perczel, A. (2020) 'Dynamically encoded reactivity of Ras enzymes: opening new frontiers for drug discovery', *Cancer and Metastasis Reviews*. *Cancer and Metastasis Reviews*, 39(4), pp. 1075–1089. doi: 10.1007/s10555-020-09917-3.
- Paul, I. and Ghosh, M. K. (2014) 'The E3 ligase CHIP: Insights into its structure and regulation', *BioMed Research International*, 2014. doi: 10.1155/2014/918183.
- Pei, L. *et al.* (2014) 'DAPK1-p53 interaction converges necrotic and apoptotic pathways of ischemic neuronal death', *Journal of Neuroscience*, 34(19), pp. 6546–6556. doi: 10.1523/JNEUROSCI.5119-13.2014.
- Piper, R. C. and Stringer, D. K. (2011) 'Terminating protein ubiquitination: Hasta la vista, ubiquitin', *Cell Cycle*, 10(18), pp. 3067–3071. doi: 10.4161/cc.10.18.17191.

- Poorebrahim, M. *et al.* (2022) 'Multi-targeting of K-Ras domains and mutations by peptide and small molecule inhibitors', *PLoS Computational Biology*, 18(4), pp. 1–21. doi: 10.1371/journal.pcbi.1009962.
- Price, A. *et al.* (2018) 'The LRRK2 signalling system', *Cell and Tissue Research*. *Cell and Tissue Research*, 373(1), pp. 39–50. doi: 10.1007/s00441-017-2759-9.
- Prior, I. A., Lewis, P. D. and Mattos, C. (2012) 'A comprehensive survey of ras mutations in cancer', *Cancer Research*, 72(10), pp. 2457–2467. doi: 10.1158/0008-5472.CAN-11-2612.
- Raju, T. S. (2019) 'Phosphorylation of Proteins Mechanism of Protein Phosphorylation in Living Cells', *wiley online library*, (Sasai 1965), pp. 163–175.
- Raudvere, U. *et al.* (2019) 'G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update)', *Nucleic Acids Research*. Oxford University Press, 47(W1), pp. W191–W198. doi: 10.1093/nar/gkz369.
- Reimer, J. M. *et al.* (2022) 'Structure of LRRK1 and mechanisms of autoinhibition and activation', *bioRxiv*, 2022, p. 2022.11.22.517582. Available at: <https://www.biorxiv.org/content/10.1101/2022.11.22.517582v1%0Ahttps://www.biorxiv.org/content/10.1101/2022.11.22.517582v1.abstract>.
- Repasky, G. A. *et al.* (2009) 'Role of R-Ras in Cell Growth', in *Handbook of Cell Signaling, Second Edition*. Second Edi. Elsevier Inc., pp. 1753–1762. doi: 10.1016/B978-0-12-374145-5.00214-X.
- Rigbolt, K. T. G. *et al.* (2011) 'System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation', *Science Signaling*, 4(164). doi: 10.1126/scisignal.2001570.
- Roche, D. B., Tetchner, S. J. and McGuffin, L. J. (2011) 'FunFOLD: An improved automated method for the prediction of ligand binding residues using 3D models of proteins', *BMC Bioinformatics*, 12(May). doi: 10.1186/1471-2105-12-160.
- Saha, S. *et al.* (2009) 'LRRK2 modulates vulnerability to mitochondrial dysfunction in *Caenorhabditis elegans*', *Journal of Neuroscience*, 29(29), pp. 9210–9218. doi: 10.1523/JNEUROSCI.2281-09.2009.
- Sakabe, T. *et al.* (1999) 'Identification of a Novel Gene , MASL1 , within an Amplicon at 8p23 . 1 Detected in Malignant Fibrous Histiocytomas by Comparative Genomic Hybridization 1', *Cancer research*, 59, pp. 511–515.

- Sakaguchi-Nakashima, A. *et al.* (2007) 'LRK-1, a *C. elegans* PARK8-Related Kinase, Regulates Axonal-Dendritic Polarity of SV Proteins', *Current Biology*, 17(7), pp. 592–598. doi: 10.1016/j.cub.2007.01.074.
- Samsu, N. (2021) 'Diabetic Nephropathy: Challenges in Pathogenesis, Diagnosis, and Treatment', *BioMed Research International*, 2021. doi: 10.1155/2021/1497449.
- Saunders-Pullman, R. *et al.* (2010) 'LRRK2 G2019S mutations are associated with an increased cancer risk in Parkinson disease', *Movement Disorders*, 25(15), pp. 2536–2541. doi: 10.1002/mds.23314.
- Schulte, E. C. *et al.* (2014) 'Rare variants in LRRK1 and Parkinson's disease', *Neurogenetics*, 15(1), pp. 49–57. doi: 10.1007/s10048-013-0383-8.
- Schultheis, A. M. *et al.* (2014) 'Fibroblast growth factor receptor 1 (FGFR1) amplification is a potential therapeutic target in small-cell lung cancer', *Modern Pathology*, 27(2), pp. 214–221. doi: 10.1038/modpathol.2013.141.
- Secolin, R. *et al.* (2021) 'Exploring a Region on Chromosome 8p23.1 Displaying Positive Selection Signals in Brazilian Admixed Populations: Additional Insights Into Predisposition to Obesity and Related Disorders', *Frontiers in Genetics*, 12(March), pp. 1–10. doi: 10.3389/fgene.2021.636542.
- Sejwal, K. *et al.* (2017) 'Cryo-EM analysis of homodimeric full-length LRRK2 and LRRK1 protein complexes', *Scientific Reports*, 7(1), pp. 1–12. doi: 10.1038/s41598-017-09126-z.
- Shannon, P. *et al.* (2003) 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks', *Genome Research*, 13(22), p. 426. doi: 10.1101/gr.1239303.metabolite.
- Sheng, Z. *et al.* (2012) 'Ser1292 autophosphorylation is an indicator of LRRK2 kinase activity and contributes to the cellular effects of PD mutations.', *Science translational medicine*, 4(164). doi: 10.1126/scitranslmed.3004485.
- Shi, Q. *et al.* (2017) 'MFHAS1 suppresses TLR4 signaling pathway via induction of PP2A C subunit cytoplasm translocation and inhibition of c-Jun dephosphorylation at Thr239', *Molecular Immunology*. Elsevier, 88(March), pp. 79–88. doi: 10.1016/j.molimm.2017.06.017.
- Smith, P. K. *et al.* (1985) 'Measurement of protein using bicinchoninic acid', *Analytical Biochemistry*, 150(1), pp. 76–85. doi: 10.1016/0003-2697(85)90442-7.

- Snead, D. M. *et al.* (2022) *Structural basis for Parkinson's disease-linked LRRK2's binding to microtubules*, *Nature Structural and Molecular Biology*. Springer US. doi: 10.1038/s41594-022-00863-y.
- Song, Y. *et al.* (2018) 'Pja2 inhibits Wnt/ $\beta$ -catenin signaling by reducing the level of TCF/LEF1', *International Journal of Stem Cells*, 11(2), pp. 242–247. doi: 10.15283/ijsc18032.
- Soriano, O. *et al.* (2021) 'The Crossroads between RAS and RHO Signaling Pathways in Cellular Transformation, Motility and Contraction', *Genes*, 12.
- Sosero, Y. L. and Gan-Or, Z. (2023) 'LRRK2 and Parkinson's disease: from genetics to targeted therapy', *Annals of Clinical and Translational Neurology*, 1, pp. 1–15. doi: 10.1002/acn3.51776.
- Steger, M. *et al.* (2017) 'Systematic proteomic analysis of LRRK2-mediated rab GTPase phosphorylation establishes a connection to ciliogenesis', *eLife*, 6(i), pp. 1–22. doi: 10.7554/eLife.31012.
- Stephen, J. K. *et al.* (2009) 'DNA hypermethylation profiles in squamous cell carcinoma of the vulva', *International Journal of Gynecological Pathology*, 28(1), pp. 63–75. doi: 10.1097/PGP.0b013e31817d9c61.
- Stouten, P. F. W. *et al.* (1993) 'How does the switch II region of G-domains work?', *FEBS Letters*, 320(1), pp. 1–6. doi: 10.1016/0014-5793(93)81644-F.
- Su, H. *et al.* (2022) 'A comprehensive investigation on pan-cancer impacts of constitutive centromere associated network gene family by integrating multi-omics data: A CONSORT-compliant article', *Medicine (United States)*, 101(7), p. E28821. doi: 10.1097/MD.00000000000028821.
- Sun, M. and Zhang, X. (2022) 'Current methodologies in protein ubiquitination characterization: from ubiquitinated protein to ubiquitin chain architecture', *Cell and Bioscience*. BioMed Central, 12(1), pp. 1–17. doi: 10.1186/s13578-022-00870-y.
- Tagawa, H. *et al.* (2004) 'MASL1, a candidate oncogene found in amplification at 8p23.1, is translocated in immunoblastic B-cell lymphoma cell line OCI-LY8', *Oncogene*, 23(14), pp. 2576–2581. doi: 10.1038/sj.onc.1207352.
- Tamaoka, S. *et al.* (2023) 'Rare sequence variants associated with the risk of non-syndromic biliary atresia', *Hepatology Research*, 44(0), pp. 2–31. doi: 10.1111/hepr.13946.

- Tang, R., Langdon, W. Y. and Zhang, J. (2019) 'Regulation of immune responses by E3 ubiquitin ligase Cbl-b', *Cellular Immunology*. Elsevier, 340(November 2018), p. 103878. doi: 10.1016/j.cellimm.2018.11.002.
- Tereshko, V. *et al.* (2001) 'Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression', *Nature Structural Biology*, 8(10), pp. 899–907. doi: 10.1038/nsb1001-899.
- Tomkins, J. E. *et al.* (2018) 'Comparative Protein Interaction Network Analysis Identifies Shared and Distinct Functions for the Human ROCO Proteins', *Proteomics*, 18(10), pp. 1–12. doi: 10.1002/pmic.201700444.
- Tomkins, J. E. *et al.* (2020) 'PINOT: An intuitive resource for integrating protein-protein interactions', *Cell Communication and Signaling*, 18(1), pp. 1–11. doi: 10.1186/s12964-020-00554-5.
- Tomkins, J. E. and Manzoni, C. (2021) 'Advances in protein-protein interaction network analysis for Parkinson's disease', *Neurobiology of Disease*. Elsevier Inc., 155(March), p. 105395. doi: 10.1016/j.nbd.2021.105395.
- Toyofuku, T. *et al.* (2015) 'Leucine-Rich Repeat Kinase 1 Regulates Autophagy through Turning On TBC1D2-Dependent Rab7 Inactivation', *Molecular and Cellular Biology*, 35(17), pp. 3044–3058. doi: 10.1128/mcb.00085-15.
- Tu, W. *et al.* (2010) 'DAPK1 Interaction with NMDA Receptor NR2B Subunits Mediates Brain Damage in Stroke', *Cell*. Elsevier Ltd, 140(2), pp. 222–234. doi: 10.1016/j.cell.2009.12.055.
- Udeshi, N. D. *et al.* (2013) 'Refined preparation and use of anti-diglycine remnant (k-ε-gg) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments', *Molecular and Cellular Proteomics*, 12(3), pp. 825–831. doi: 10.1074/mcp.O112.027094.
- Uhlén, M. *et al.* (2015) 'Tissue-based map of the human proteome', *Science*, 347(6220). doi: 10.1126/science.1260419.
- Vancaenenbroeck, R. *et al.* (2012) 'Expression, purification and preliminary biochemical and structural characterization of the leucine rich repeat namesake domain of leucine rich repeat kinase 2', *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., 1824(3), pp. 450–460. doi: 10.1016/j.bbapap.2011.12.009.
- Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1),

pp. D439–D444. doi: 10.1093/nar/gkab1061.

Vaughn, J. L. *et al.* (1977) 'The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae)', *In Vitro*, 13(4), pp. 213–217. doi: 10.1007/BF02615077.

Vetter, I. R. and Wittinghofer, A. (2001) 'The guanine nucleotide-binding switch in three dimensions', *Science*, 294(5545), pp. 1299–1304. doi: 10.1126/science.1062023.

Walker, J. M. (2009) *Homology Modeling: Methods and Protocols, Methods in Molecular Biology*. Available at: <http://books.google.com/books?id=Ku2wPAAACAAJ>.

Walkup, W. G. *et al.* (2015) 'Phosphorylation of synaptic GTPase-activating protein (synGAP) by Ca<sup>2+</sup>/Calmodulin-dependent protein kinase II (CaMKII) and cyclin-dependent kinase 5 (CDK5) alters the ratio of its GAP activity toward Ras and Rap GTPases', *Journal of Biological Chemistry*. © 2015 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 290(8), pp. 4908–4927. doi: 10.1074/jbc.M114.614420.

Wang, H. *et al.* (2018) 'High Glucose Stimulates Expression of MFHAS1 to Mitigate Inflammation via Akt/HO-1 Pathway in Human Umbilical Vein Endothelial Cells', *Inflammation*, 41(2), pp. 400–408. doi: 10.1007/s10753-017-0696-0.

Wang, J. *et al.* (2020) 'Identification of a six-gene prognostic signature for oral squamous cell carcinoma', *Journal of Cellular Physiology*, 235(3), pp. 3056–3068. doi: 10.1002/jcp.29210.

Wang, T. *et al.* (2020) 'The E3 ubiquitin ligase CHIP in normal cell function and in disease conditions', *Annals of the New York Academy of Sciences*, 1460(1), pp. 3–10. doi: 10.1111/nyas.14206.

Wang, Z. *et al.* (2018) 'Meta-analysis of human gene expression in response to *Mycobacterium tuberculosis* infection reveals potential therapeutic targets', *BMC Systems Biology*, 12(1), pp. 1–18. doi: 10.1186/s12918-017-0524-z.

Warø, B. J. and Aasly, J. O. (2018) 'Exploring cancer in LRRK2 mutation carriers and idiopathic Parkinson's disease', *Brain and Behavior*, 8(1), pp. 2–7. doi: 10.1002/brb3.858.

Watanabe, R. *et al.* (2020) 'The In Situ Structure of Parkinson's Disease-Linked LRRK2', *Cell*, 182(6), pp. 1508–1518.e16. doi: 10.1016/j.cell.2020.08.004.

Weng, W. H. *et al.* (2004) 'Characterization of large chromosome markers in a malignant fibrous histiocytoma by spectral karyotyping, comparative genomic hybridization (CGH), and array CGH', *Cancer Genetics and Cytogenetics*, 150(1), pp. 27–32. doi: 10.1016/j.cancergencyto.2003.08.009.



- West, A. B. *et al.* (2005) 'Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity', *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), pp. 16842–16847. doi: 10.1073/pnas.0507360102.
- Westerlund, M. *et al.* (2008) 'Developmental regulation of leucine-rich repeat kinase 1 and 2 expression in the brain and other rodent and human organs: Implications for Parkinson's disease', *Neuroscience*, 152(2), pp. 429–436. doi: 10.1016/j.neuroscience.2007.10.062.
- Whitmarsh, A. J. and Davis, R. J. (2007) 'Role of mitogen-activated protein kinase kinase 4 in cancer', *Oncogene*, 26(22), pp. 3172–3184. doi: 10.1038/sj.onc.1210410.
- van Wijk, S. J. *et al.* (2019) 'Visualizing ubiquitination in mammalian cells', *EMBO reports*, 20(2), pp. 1–18. doi: 10.15252/embr.201846520.
- Winczura, K. *et al.* (2018) 'Characterizing ZC3H18, a Multi-domain Protein at the Interface of RNA Production and Destruction Decisions', *Cell Reports*. Elsevier Company., 22(1), pp. 44–58. doi: 10.1016/j.celrep.2017.12.037.
- Xiang, S. *et al.* (2017) *Activation of Ras by Post-Translational Modifications, Conquering RAS: From Biology to Cancer Therapy*. Elsevier Inc. doi: 10.1016/B978-0-12-803505-4.00006-0.
- Xing, W. *et al.* (2013) 'Targeted disruption of leucine-rich repeat kinase 1 but not leucine-rich repeat kinase 2 in mice causes severe osteopetrosis', *Journal of Bone and Mineral Research*, 28(9), pp. 1962–1974. doi: 10.1002/jbmr.1935.
- Xing, W. R. *et al.* (2017) 'Role and mechanism of action of leucine-rich repeat kinase 1 in bone', *Bone Research*. The Author(s), 5(December 2016), pp. 1–13. doi: 10.1038/boneres.2017.3.
- Xu, L. *et al.* (2017) 'E3 Ubiquitin Ligase Cbl-b Prevents Tumor Metastasis by Maintaining the Epithelial Phenotype in Multiple Drug-Resistant Gastric and Breast Cancer Cells', *Neoplasia*, 19(4), pp. 374–382. doi: 10.1016/j.neo.2017.01.011.
- Xu, L. zhi, Li, B. qiu and Jia, J. ping (2019) 'DAPK1: a Novel Pathology and Treatment Target for Alzheimer's Disease', *Molecular Neurobiology*, 56(4), pp. 2838–2844. doi: 10.1007/s12035-018-1242-2.
- Yang, S. *et al.* (2007) 'Identification of genes with correlated patterns of variations in DNA copy number and gene expression level in gastric cancer', *Genomics*, 89(4), pp. 451–459. doi: 10.1016/j.ygeno.2006.12.001.

Yao, C. *et al.* (2010) 'LRRK2-mediated neurodegeneration and dysfunction of dopaminergic neurons in a *Caenorhabditis elegans* model of Parkinson's disease', *Neurobiology of Disease*. Elsevier Inc., 40(1), pp. 73–81. doi: 10.1016/j.nbd.2010.04.002.

Yu, L. R. *et al.* (2007) 'Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra', *Journal of Proteome Research*, 6(11), pp. 4150–4162. doi: 10.1021/pr070152u.

Yue, Z. (2009) 'LRRK2 in Parkinson's disease: In vivo models and approaches for understanding pathogenic roles', *FEBS Journal*, 276(22), pp. 6445–6454. doi: 10.1111/j.1742-4658.2009.07343.x.

Zalckvar, E., Berissi, H., Mizrachy, L., *et al.* (2009) 'DAP-kinase-mediated phosphorylation on the BH3 domain of beclin 1 promotes dissociation of beclin 1 from Bcl-XL and induction of autophagy', *EMBO Reports*, 10(3), pp. 285–292. doi: 10.1038/embor.2008.246.

Zalckvar, E., Berissi, H., Eisenstein, M., *et al.* (2009) 'Phosphorylation of Beclin 1 by DAP-kinase promotes autophagy by weakening its interactions with Bcl-2 and Bcl-XL', *Autophagy*, 5(5), pp. 720–722. doi: 10.4161/auto.5.5.8625.

Zeng, C. *et al.* (2016) 'Leucine-rich repeat kinase-1 regulates osteoclast function by modulating RAC1/Cdc42 Small GTPase phosphorylation and activation', *American Journal of Physiology - Endocrinology and Metabolism*, 311(4), pp. E772–E780. doi: 10.1152/ajpendo.00189.2016.

Zhan, S., Wang, T. and Ge, W. (2017) 'Multiple functions of the E3 ubiquitin ligase CHIP in immunity', *International Reviews of Immunology*. Taylor & Francis, 36(5), pp. 300–312. doi: 10.1080/08830185.2017.1309528.

Zhang, L., Nephew, K. P. and Gallagher, P. J. (2007) 'Regulation of death-associated protein kinase: Stabilization by HSP90 heterocomplexes', *Journal of Biological Chemistry*, 282(16), pp. 11795–11804. doi: 10.1074/jbc.M610430200.

Zhang, X. *et al.* (2018) 'Activation of JNK signaling in osteoblasts is inversely correlated with collagen synthesis in age-related osteoporosis', *Biochemical and Biophysical Research Communications*. Elsevier Ltd, 504(4), pp. 771–776. doi: 10.1016/j.bbrc.2018.08.094.

Zhao, Y. *et al.* (2023) 'Tissue specific LRRK2 interactomes reveal a distinct striatal functional unit', *PLoS Computational Biology*, 19(1), pp. 1–23. doi: 10.1371/journal.pcbi.1010847.

Zheng, L. *et al.* (2021) 'ELK4 promotes the development of gastric cancer by inducing M2 polarization of macrophages through regulation of the KDM5A-PJA2-KSR1 axis', *Journal of*

*Translational Medicine*, 19(1), pp. 1–13. doi: 10.1186/s12967-021-02915-1.

Zheng, N. and Shabek, N. (2017) 'Ubiquitin ligases: Structure, function, and regulation', *Annual Review of Biochemistry*, 86, pp. 129–157. doi: 10.1146/annurev-biochem-060815-014922.

Zhong, J. *et al.* (2015) 'MFHAS1 Is Associated with Sepsis and Stimulates TLR2/NF- $\kappa$ B Signaling Pathway Following Negative Regulation', *PLoS ONE*, 10(11), pp. 1–17. doi: 10.1371/journal.pone.0143662.

Zhong, J. *et al.* (2017) 'Ubiquitylation of MFHAS1 by the ubiquitin ligase praja2 promotes M1 macrophage polarization by activating JNK and p38 pathways', *Cell Death and Disease*, 9(8), pp. 1–10. doi: 10.1038/cddis.2017.102.

Zhong, J. *et al.* (2018) 'Correction: Ubiquitylation of MFHAS1 by the ubiquitin ligase praja2 promotes M1 macrophage polarization by activating JNK and p38 pathways (Cell Death and Disease (2017) DOI: 10.1038/cddis.2017.102)', *Cell Death and Disease*, 9(8), pp. 1–10. doi: 10.1038/s41419-018-0828-y.

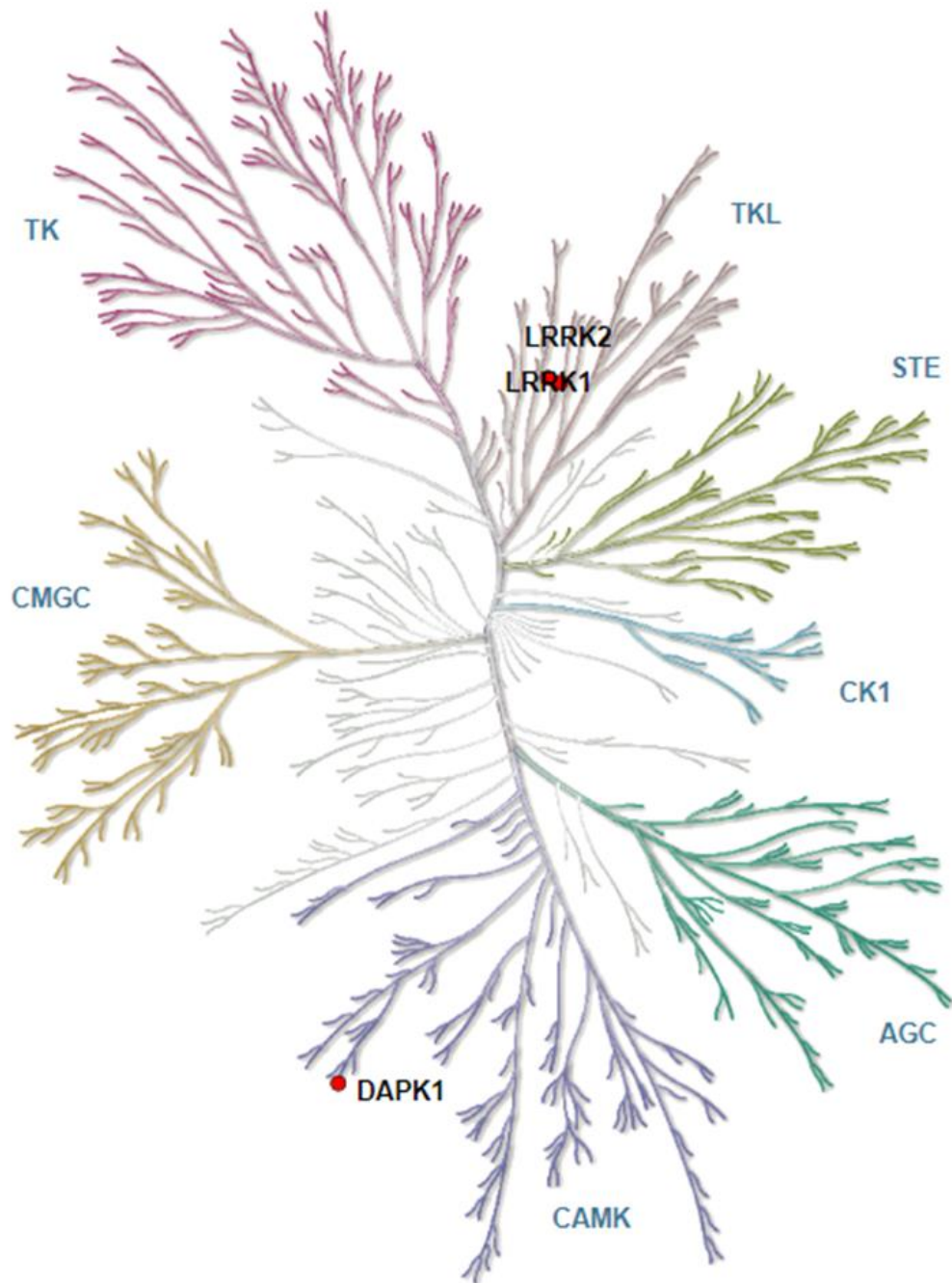
Zhou, H. *et al.* (2021) 'LncRNA-MAP3K4 regulates vascular inflammation through the p38 MAPK signaling pathway and cis-modulation of MAP3K4', *FASEB Journal*, 35(1), pp. 1–20. doi: 10.1096/fj.202001654RR.

Zhu, Y. *et al.* (2017) 'Quantitative and correlation analysis of the DNA methylation and expression of DAPK in breast cancer', *PeerJ*, 2017(3), pp. 1–10. doi: 10.7717/peerj.3084.

Zimprich, A. *et al.* (2004) 'Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology', *Neuron*, 44, pp. 601–607. doi: 10.1016/J.NEURON.2004.11.005.

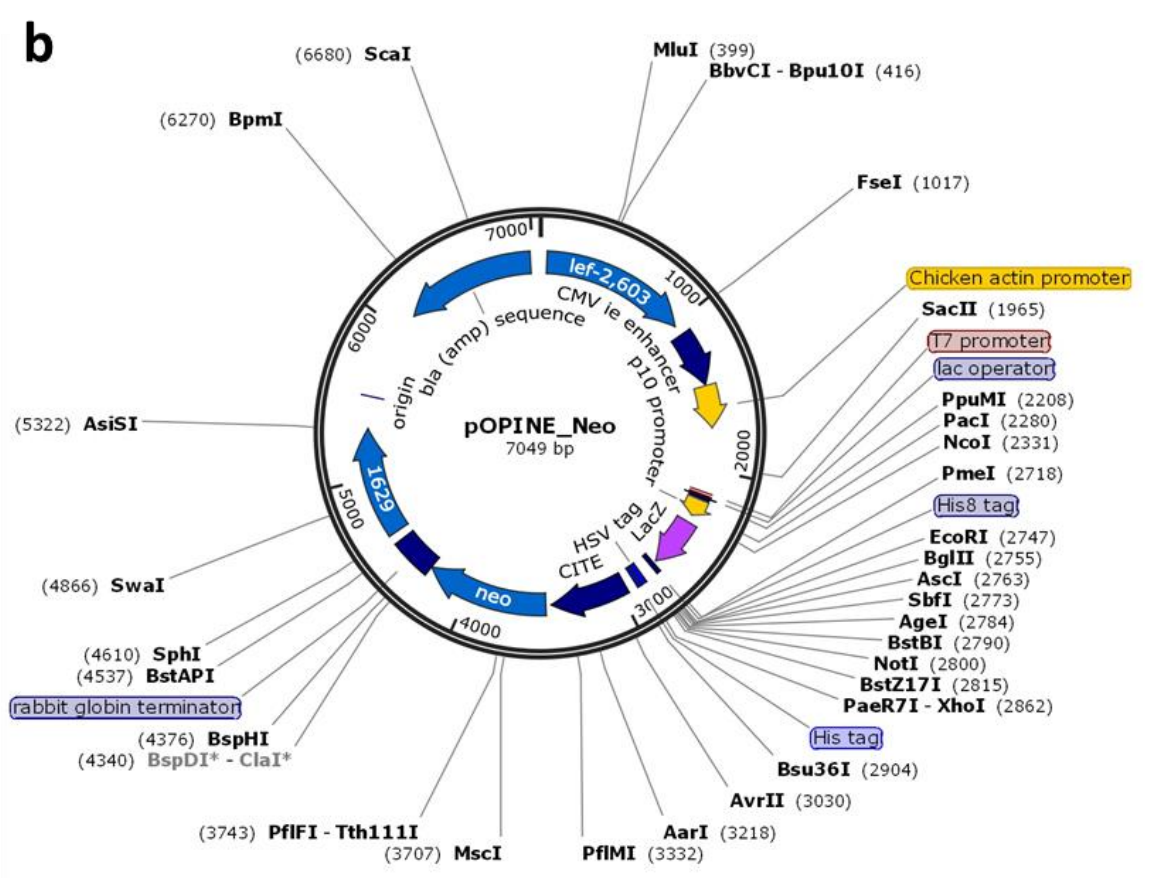
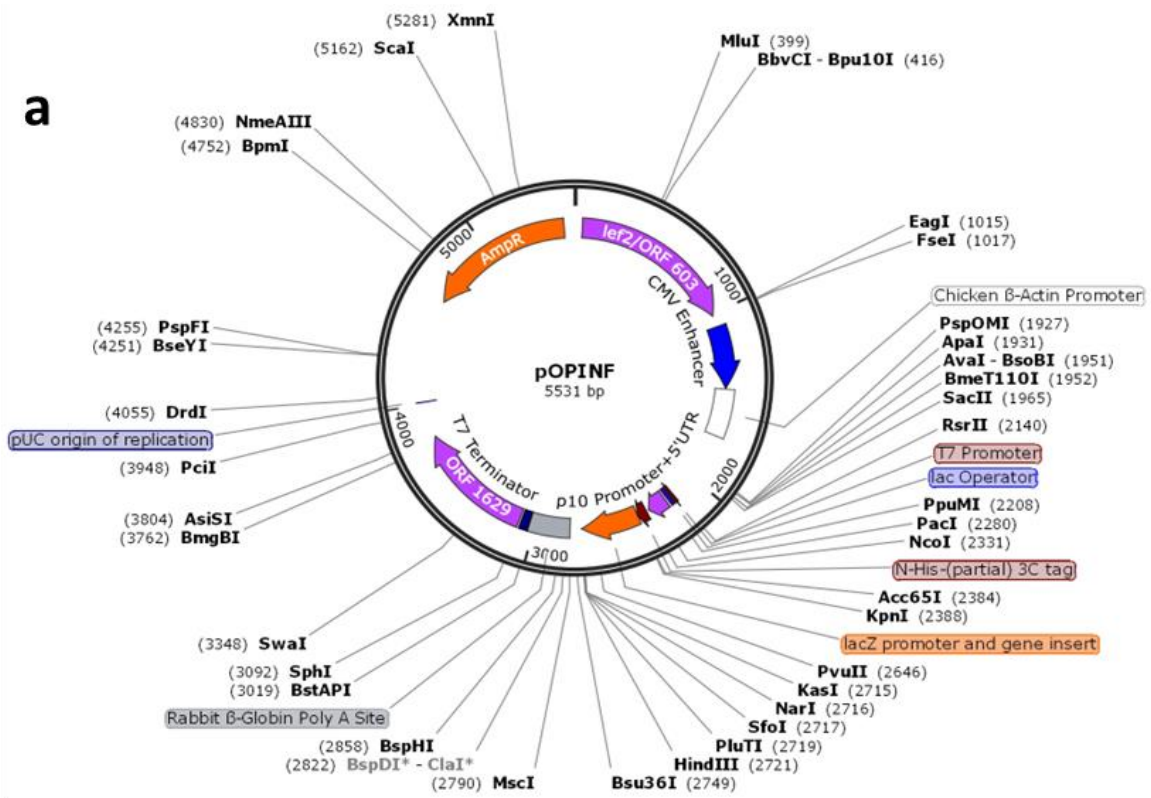
## Supplementary figures

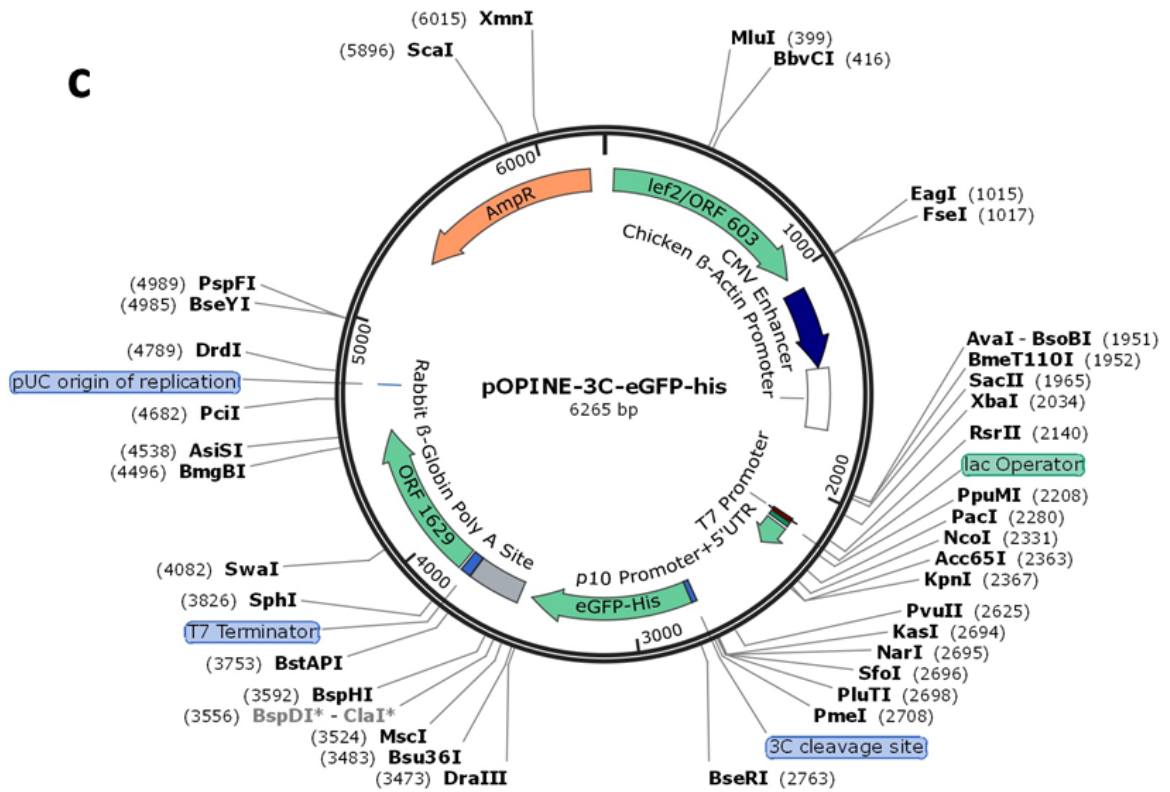
---



"Illustration reproduced courtesy of Cell Signaling Technology, Inc. ([www.cellsignal.com](http://www.cellsignal.com))"

**Supplementary Figure 1:** Representation of the human Kinome. LRRK2, LRRK1 and DAPK1 are shown in red. (Manning *et al.*, 2002). Interactive KinMap from Kinhub (Eid *et al.*, 2017).





Supplementary figure 2: Map of the vectors used in Chapter 2. (a): pOPinF (b):pOPinE\_Neo (c):pOPINE-3C-eGFP-his.

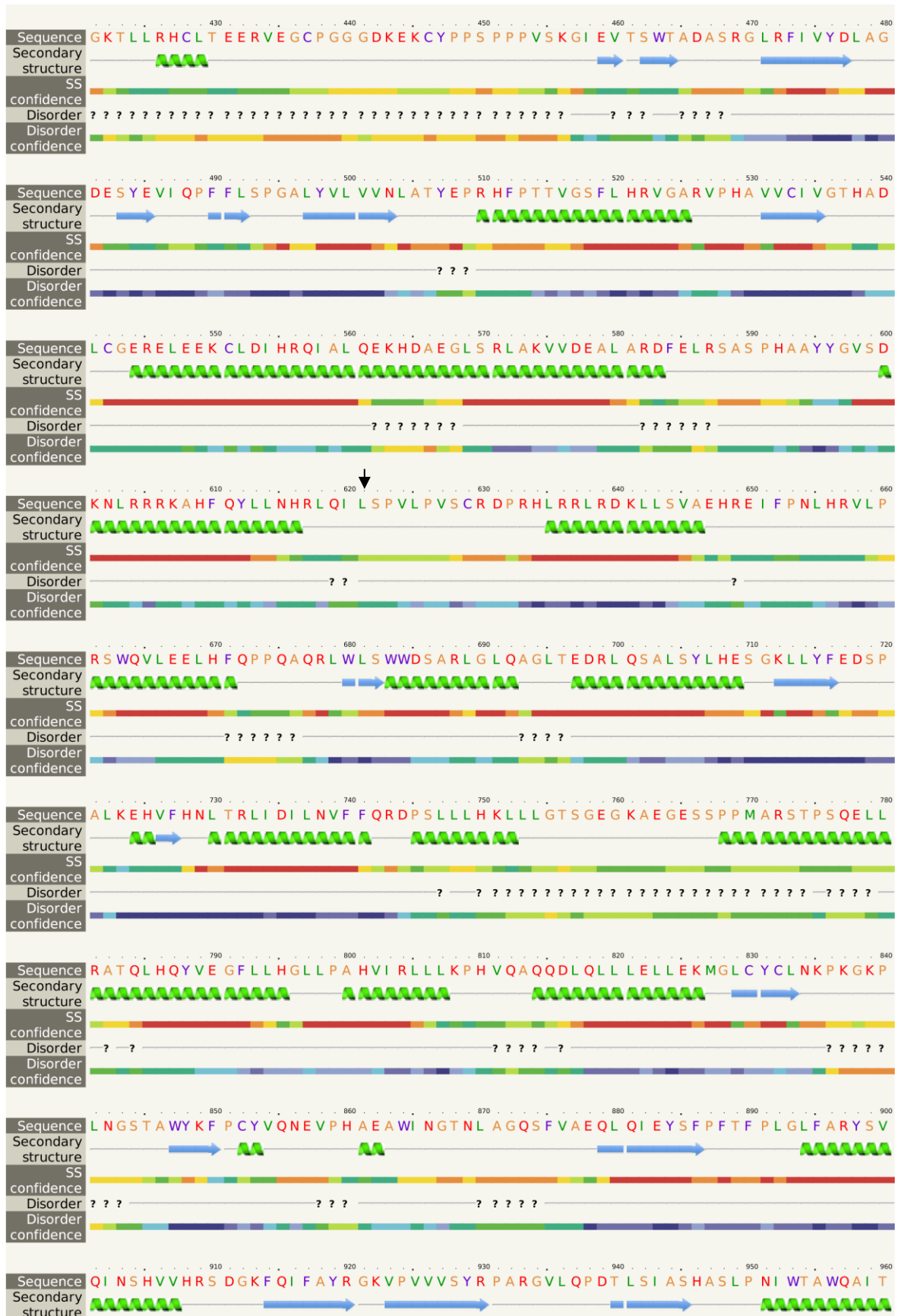
# Phyre2

Email: [redacted]  
Description: MASL1\_full\_length  
Date: Tue Oct 16 11:17:37 BST 2018  
Unique Job ID: 452a047991672e86

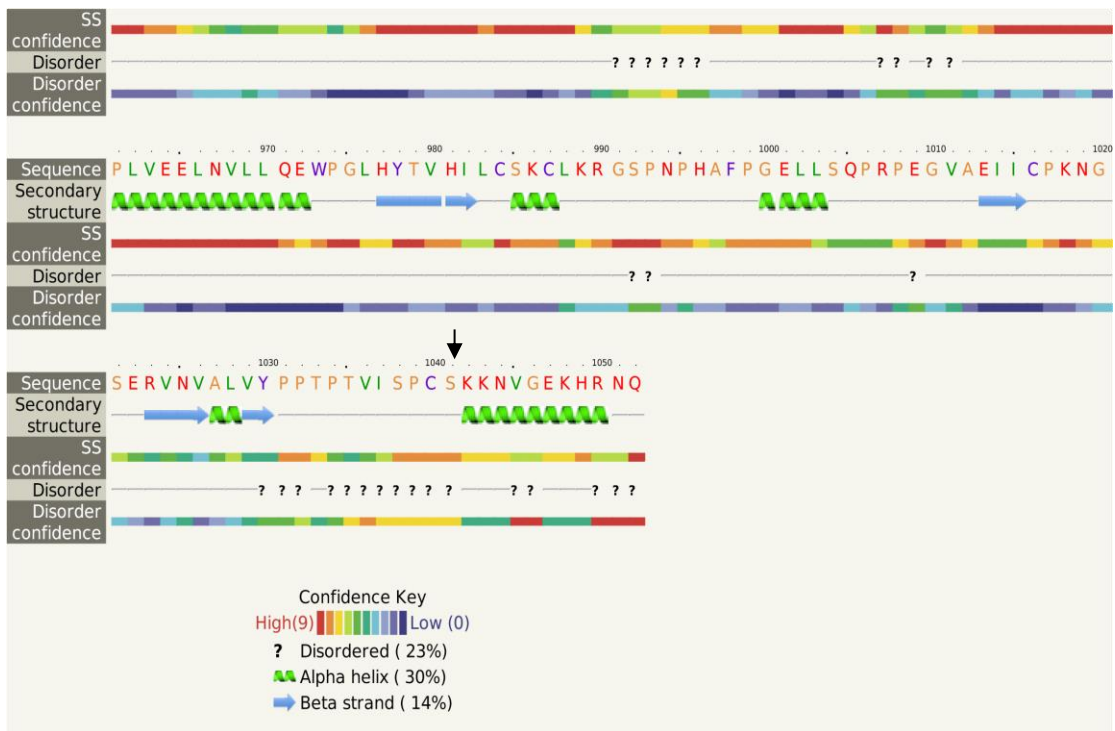
## Secondary structure and disorder prediction











**Supplementary figure 3:** Phyre2 secondary prediction of MFHAS1 protein, red boxes highlight the Leucine Rich Repeat (LRR) motifs. Black arrows show the domain boundaries.

DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	MASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNI	60
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	HVPLLIVLDSYMRVASVQQVGSLLCKLIEVCPGTMQSLMGPDVGNWVGLGVHQILIK	120
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	MLTVHNASVNLVIGLKTLDLLLTSGKITLLILDEESDIFMLIFDAMHSFPANDEVQKLG	180
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	CKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEIVLHVLHCLHSLAIPCNVVE	240
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	VLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNIIVLNEVHEFVVKAVQ	300
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	QYPENAAALQISALSCLALLTETIFLNQDLEKKNENQENDDEGEEDKLFWLEACYKALTWH	360
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	RKNKHVQEAACWALNLLMYQNSLHEKIGDEDGHFPAHREVMSMLMHSSSKEVFQASAN	420
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	ALSTLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEQSNTSLDIM	480
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	AAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDTEFHKLNMVKKQCFKN	540
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	DIHKLVLAAALNRFIGNPGIQKCGLKVISSIVHFPPDALEMLSLEGAMDSVLHTLQMYPDDQ	600
DAPK1	-----	0
CtROCO	-----	0
LRRK1	-----	0
MFHAS1	-----	0
LRRK2	EIQCLGLSLIGYLI TKKNVFIGTGHL LAKILVSSLYRFKDVAEIQTKGFQTI LAI LKLSA	660
DAPK1	-----MTVFRQENVDDYYDT-----	15
CtROCO	-----	0
LRRK1	-----MAGMSQR	7
MFHAS1	-----	0
LRRK2	SFSKLLVHHSFDLVI FHQMSNIMEQKDQQFLNLCKCFARVAMDDYLKNVMLERACDQN	720
DAPK1	-----GEE---LGSGQFAVVKKCREKSTGLQYAAKF- IKKRRTKSSRRG-VSRE	59
CtROCO	-----	0
LRRK1	PPSMYWCV---GPE-----ES---AVCPERA---METLNGAGDTGGKPSTRGGDPAA	50
MFHAS1	-----	0
LRRK2	NSIMVECLLLL GADANQAKEGSSLICQVCEKESPKLVELLNLSGSR-----EQ	769

DAPK1 DIEREVSILKEIQHPNVITLHEVYEN---KTDVILILELVAGGELFDFLAEKESLTEEE 115  
CtROCO ----- 0  
LRRK1 RSRRTREGIRAAAYRRGDRGGARDLLE-----EACDQCASQLEKQQLLSIPAAAYGDLE--- 101  
MFHAS1 -----MAGMDSGNLKTARLWRDAALRA- 22  
LRRK2 DVRK--ALTISIGKGSQIISLLRRLALDVANNISICLGGFCIGKVE--PSWLGPLFPDK 825

DAPK1 ATEFLK-----Q-----I LNVVYVYHLSLQIAHF----- 138  
CtROCO -----MSDLDVIRQIEQELG 15  
LRRK1 -----MVRVYLLSKRLVLELPTPTDDNPAVVAAYFGHTAVVQELLESPL 144  
MFHAS1 ----RK---LRSNLRQL-----TLTAAGACPGAG-----ADALE-- 50  
LRRK2 TSNLRKQTNIASTLARMVIRYQMK-SAVEEGTASGSD-----GNFSEDEVL 869

DAPK1 -----DLKPENIMLLDRNVKPKRIKIDF-GLAHKIDF-----GNEFKNIFGTPEFVA 185  
CtROCO MQLPEVDKWKWYSK-----GYKLDKQ- 37  
LRRK1 GPCSPQRLLNW--MLAL-ACQRGHGVLVLLVTHGADPESYAVRKNFPVIVRPLLYAA 201  
MFHAS1 -PASQLVLPAN-----L---GDIE----- 66  
LRRK2 SKFDEWTFPIPDSSMDSV-FAQSDDL-----SEGSEGSFLVKKKSN----- 909

DAPK1 PEIVNYEPLGLEADMWSIGVI--TYILLSGASPFLGDTKQETLA-NVSAVNYEFEDEYFS 242  
CtROCO -----RVTAIGLYDCGSD-----TLDRIIQPLESL-KSLSLESLSSNQIT 76  
LRRK1 IKSGN-ED--IAIFLLRHGAYFCSYI-----LLDSPDPS-----KH--LLRKYFI 241  
MFHAS1 -----ALNLG-----NNGLEEVPEGLGSALGSLRVLVLRNRFA 100  
LRRK2 -----SISVGEFYRDAV-LQRCSPNLQRHSNSLGPIDHEDLLKRRKRKILS 954  
\* \* : : :

DAPK1 NTSALAKDFIRLLVKDPKKRMTIQDSLQHPWIKPKDTQQAALSRKASAVNMEKFKKFAAR 302  
CtROCO DISPLAS-----LNSLSMLWLDNRNQTIDIAPL-ASLNSL-----S 110  
LRRK1 EASPLPSSYPGKTAL-----RVKWSHLRLPWVLDLWLD 275  
MFHAS1 -----R---LP PAVAE----- 108  
LRRK2 SDDSLRSS-----KL-----QSHMRHSD---SISLAS----- 979

DAPK1 KKWKQSVRLISLCQRLSRFLSRNSMVARSDDTLDEEDSFVMKAI I---HAINDD--- 355  
CtROCO MLWLFNGKISDIAPLESLSLTELQL---SSNQITDIAPLA--SLKSLTELSLSGNNIS 164  
LRRK1 -----ISCQITELDL---SANCLATLPSVIPWGLINLRKLNLSDNHLG 315  
MFHAS1 -----LGHHLTEL DV---SHNRLTALGAEVVSAALRELRLKLNLSHNQLP 148  
LRRK2 -----EREYITSLDL---SANELRDI DALS----- 1001  
: : \* : :

DAPK1 NVPGLOH-----LLGSLSNYDVNQPNKHGTPPLLIAAGCGN 391  
CtROCO DIA-----P--LESLSLTELSSNQITDIAPL-ASLKSLT 198  
LRRK1 ELPGVQSS-----DEI---ICSRLEIDISSNKLSHLPPGFHLHLSKLO 355  
MFHAS1 ALPAQIGALAHLEELDVSNRLAHLPSDL-SCLSRRLTLDVDHNQLTAFPRQLLQVLALE 207  
LRRK2 QKCCISVHLEHLEKLELHQNALTSFPQQLCETLKS LTHLDLHSNKFTSFPSYLLKMSCIA 1061  
\* : :

DAPK1 IQILQLLIKRGSRIDVQDKGGSNAVYAARHGHDVTLKFLSENK----- 435  
CtROCO --ELSSSNQISDIAPLE-----SLKSLTELQLSRNQ 228  
LRRK1 --KLTASKNCLEKLFEE-----EN----- 372  
MFHAS1 --ELDVSSNRLRG-----LPEDISA----- 225  
LRRK2 --NLDVSRNDIGPSVVDPTV-KCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNK 1118  
\* : :

DAPK1 ----CPLDVKDKSGEMALHVAARYGHADVAQLLC-SFGSNPNIQDKEEETPLHCAAWHGY 490  
CtROCO ISDIAPLESLSLTELQLS---SNQITDIAPL--ASLKSLELQLSRNQISDIAP----- 278  
LRRK1 ---ATNWIGLRKLQELD IS---DNKLTLPALFLHSFKLSLNSLNVSRNNLKVFPDPWA-- 424  
MFHAS1 -----LRALKILWLS---GAELGTLPA GFCELAS-LESMLDNNGLQALPAQFS-- 270  
LRRK2 ISGICSPRLKELKILNLS---KNHISLSENFL EACPKVESFSARMNFLAAMPF----- 1170  
: : : : :

DAPK1 YSVAKALCEAGCNVNIK--N-----REGETPLLTASARGYHDI VECLAEHGADLNACDKD 543  
CtROCO ----LE-----S-LNSLSKLWLNQITDI-----APLASLNS---- 306  
LRRK1 --CPLKCKASRNALECLPDKMAVFWKNHLKDVDFSENAL---KEVPLGLFQLDA---- 474  
MFHAS1 --CLQR-----LKMNLSSNLF---EEFPAALLPLAG--- 297  
LRRK2 --LPPS-----MTILKLSQNKF---SCIPEALNLPH--- 1197  
\* : :

DAPK1 GHIALHLAVRRCQMEVIKTLLSQGC-----FVDYQ--DR 575  
CtROCO -LT--ELELSSNQITDIAPLASLKS--LSTLWSSNQISDIA----- 343  
LRRK1 -LM--FLRLQGNQLAALPPQEKWTCRQLKTLDLSRNQLGKNEDGLKTKRIAFFTTGRQR 531  
MFHAS1 -LE--ELYLSRNQLTSVPSLI-----SGL----- 318  
LRRK2 -LR--SLDMSSNDIQYLP GPAHWKSLNRELLF SHNQISILDL----- 1237  
\* : : : :

DAPK1 HGNT-----PLHVACKDG---NMPIVVALCEANCNLDISNKYGRTPHLHAA 618  
CtROCO -----PLASLESLESLSSNQISDISPL-ASLNSLTGFDV-----RR 380  
LRRK1 SGTEAASVLEFPAPLSESLVLCNLDNHLDTVPPSVCLLKSSELYL-----GN 580  
MFHAS1 -----GRLTLWL-----DNNRIRYLPDSIVELTGLEELVL-----Q 350  
LRRK2 -----SEKAYLWSRVEKHLHSHNKLEIPPEIGCLENLTSLDV-----SY 1277  
. : : : :

DAPK1 NNGILDVVRYLCLMGA--SVEALTTDGK---TAEDLARSEQHEHVAGLLARLRKDTHRGL 673  
CtROCO NP- IKRLPETITGFDM EILW NDFSSSGFITFFDNPLESPP--PEI---VKQKGEAVR-Q 432  
LRRK1 NPGLRELPPELGQLGN--LWQ-----LDTEDLTI SNVPAEI---QKEGPKAML-S 624  
MFHAS1 GNQIAVLPDHFGQLSRVGLWK-----IKDNPLIQPPYEV-----CMKGIPIYI-A 393  
LRRK2 NLELRSFPNEMGKLSK--IWD-----LPLDELH-LNFDFKH----IGCKAKDII-R 1320  
. : . : . : \* : \*

DAPK1	FIQQ-----LR-PTQNLQPRIKKLFGHSGSGKTTLVESLKCGLLRSFFRRRRPRLSSTN	727
CtROCO	YFQSIIEEARSKEALVHLQELKVHLIGDGMAGKTSLLKQLIGETFDPKES-----	482
LRRK1	YLRA----QLRK--AEKCKLMKMIIVGPPRQKSTLLEILQTRAPQVVH-----	668
MFHAS1	AYQK----ELAHSQPAVQPRLKLLLMGHKAAGKTLRHLCLTEERVEGCPGG-----GDK	443
LRRK2	FLQQ----RLKK--AVPYNRMKLMIVGNTSGSKTTLQQLMKTKS-----	1360
	: * : : * * * : * *	
DAPK1	SSRFPPSPLASKPTVSVSINNLYPGCENVSVRSRSMFEPGLTKGMLEVFVAPTHHPHCS	787
CtROCO	-----QTHGLNVVTKQAPNIK	498
LRRK1	-----GEATIRTT--KWELQR	682
MFHAS1	EKCYPPSP-----PPVSKGIEVT--SWTAD-	466
LRRK2	----DLGM-----QSATVGDVK--DWPIQI	1380
	:	
DAPK1	ADDQSTKAIDIQNAYLNGVGDVSVWEFSGNPVYFCCYDYFAANDPTS IHVVVFSLEEPYE	847
CtROCO	-----GLENDELKECLFHFWDFGGQIMHASHQFFMTRSSVYMLLLDSRTDS---	546
LRRK1	-----PAGSRAKVESVEFNVDIGGPASMATVNCFFTDKALYVVVWNLALGEEA-	732
MFHAS1	-----ASRGLRFIVYDLAGDESYEVIQPFFLSPGALYVVLVNLATYEPRH	511
LRRK2	-----RD---KRKRDLVLNVWDFAGREEFYSTHPHMTQRALYLAVYDLSKGQAE-	1427
	: : : : * * : :	
DAPK1	IQLNQVIFWLSFLKSLVPVEEPIAFGGKLNPLQVVLVA---THADIMNVRPAGG--E	901
CtROCO	----NKHYWLRHIEKYGGKSPVIVVMNKIDENPSYNIQ-----KKIN	585
LRRK1	--VANLQFVLLNIEAKAPNAVVLVVGTHLDLIEAKFRVERIATLRAYVLALCRSPSGSRA	790
MFHAS1	-FPTTVGSLHHRVGARVPHAVVCIVGTHADLCGERELEEKCLDIHRQIALQ----EKHDA	566
LRRK2	--VDAMKPWLFNIKARASSSPVILVGTHTLVDSEKQRKACMSKITKELLN-----K	1476
	: * : : :	
DAPK1	FGYDKDTSLLKEIRNRFGNLDLHISNKLFLVDAGASGSKDMKVLNRNHL-----	948
CtROCO	ERFPA---I-----ENRFHRI SCKNGDVE-SIAKSLKSAVLHP-----D	621
LRRK1	TGFPD---ITFK-----HLHEISCKSL-----EGQEGLRQLIFHVTCMKDVGSTIGC	835
MFHAS1	EGLSR---LAKVVDEALARDFELRSASPHAAYYGVSDKNLRRRKAHFQYLLN----HRL	618
LRRK2	RGFPA-----IRDYHFVNATEESDALAKLRKTI INESLN--FKIR-----D	1515
	: : :	
DAPK1	QEIRSQIVSVCPPMTH---LCEKII-----STLPSWRKLNGP-----NQLM	986
CtROCO	SIYGTPLAP-----SWIKVKEKLV EATTAQR	647
LRRK1	QRLAGRLIPRSY----LSLQEAVALAEQQR-----SRDDDVQYLT-----R	873
MFHAS1	QILSPVLPVSCRDPHRLRRLRDKLSSVAEHREIFPNLHRVLPSPVQVLEELHFQPPQAQR	678
LRRK2	QLVVGQLIPDCY----VELEKII-----SERKNVPIEFVDIRK-----R	1552
	: :	
DAPK1	SLQQFVYDVQDQLNPLASEEDLRRIAQQQLHSTGEINIMQS--ETVQDVLLLDPRWLCTNV	1044
CtROCO	YLNRTVEKICNDSGITDPGERKTLGLYLNNGIVLYFEA--LDLSEIYVLDPHWVTIGV	705
LRRK1	QLEQLVEQ--TPDNDIKDYEDLQSAISFLIETGTLHFPTSHGRLNLYFLDPIWLSECL	931
MFHAS1	LWLSWWD SARLGLQAGLTEDRLQSALSYLHESGKLLYFEDSPALKEHVFNHNT-RLIDIL	737
LRRK2	L-LQLV---RENQLQDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIM	1607
	: * * : : :	
DAPK1	LGKLLSVETPRALH-----HYRGRYT-	1065
CtROCO	YRI-INSSKTKNH---LNTSALGYI-----LNEEQ-----	732
LRRK1	QRI-FNIKGSRSVA-----K	945
MFHAS1	NVF-FQRDPSLLHKL LLLGTS GEGKAEGESSPPMARS--TPSQELLRATQLHQYVEGFL	794
LRRK2	AQI-----LTV---KVEGCPKHPKGIISRRDVEKFLSK--KRFKPNYM-	1646
	:	
DAPK1	----VEDIQR-----LVPDSVVEELLQILDAMDICARDLS-SG-----TMVDVPA	1105
CtROCO	---IRCDEYD-PAKNNKFTYTLLEQRYLLDIMKQFELCYD---EGK-----GLFIIPS	778
LRRK1	NGVIRAEDLRMLLVGT--GFTQQTEEQYFQFLAKFEIALP---VAN-----DSYLLPH	993
MFHAS1	HGLLPAHVIRLLKPH--VQAQQDLQLLELLEKMGCLCYCLNKPKGKPLNGSTAWYKFC	852
LRRK2	-----SQYFKLLEKFQIAL---PIGE-----EYLLVPS	1671
	: : : : * :	
DAPK1	LIKTDN-----LHRSWADEEDEV--VYGGVRI VPEHLT-PFPCGIFHKVQVNLCRWI	1156
CtROCO	NLPTQIDN--EPEI-----TEGEPLRFIMKYD-YLPSTIIPRLMIAMQHQI	821
LRRK1	LLPSK-----PGLDTHGMRHPTANTIQRVFKMS-FVPVGFWRFIARMLISL	1039
MFHAS1	YVQN----EVPHAEAWINGTNLAGQ-SFVAEQLEIYSFPFTFPLGLFARYSVQINSHV	906
LRRK2	SLSDHRPVIELPHCEN-----SEIIIRLYEMP-YFPMGFWSRLINRLEIS	1716
	: : * : : :	
DAPK1	HQQST-----EGD-----ADIRLWVN--GCKLANRGA	1181
CtROCO	-----LDRMQWRY--GMVLSQDH	838
LRRK1	AEMDLQLFENKNTKSRNRKVTIYSFTGNQRNRCSTFRVKRNQTIYWQE--GLLVTFDGG	1097
MFHAS1	VHRS-----DG-----KFQI---FAYRGKVPVVVSYPRA	932
LRRK2	PYML-----SG-----RERALRPNRMYWRQ--GIYLNWSPE	1745
	: :	
DAPK1	ELLVLL-----VNHGQGIEVQVRGLETEKIKCCLLLD-----SVC	1216
CtROCO	EGALAKVVA---ETKDDSTITAIQGEPRCKREYLSI IWYEIKKINANFTNLDVKEFIPLP	895
LRRK1	YLSVESSDVNWKKKSGGMKIV---CQSEVRDFSAMA-----FIT	1134
MFHAS1	RGVLQP-DTLS----IASHAS---LPNIWTAQAIT-----PLV	963
LRRK2	AYCLVGSEVLD-NHPESFLKIT--VPSCKRGCILG-----QVV	1781
	: :	
DAPK1	STIENVMATTLPGLLTVKHYLSPQQLREHHEPVMIIYQPRDFFRAQTLKETS LTNTMGGYK	1276
CtROCO	GHPDELVE--YKELLGLEKMGDR-----E-----	917
LRRK1	DHVNSLIDQWFPALTATESDGT-----PLM-----E---	1160
MFHAS1	EELNVLLQE-----	972
LRRK2	DHIDSLMEEWFPGLLEIDICGEG-----ETLL-----K---	1809
	: :	

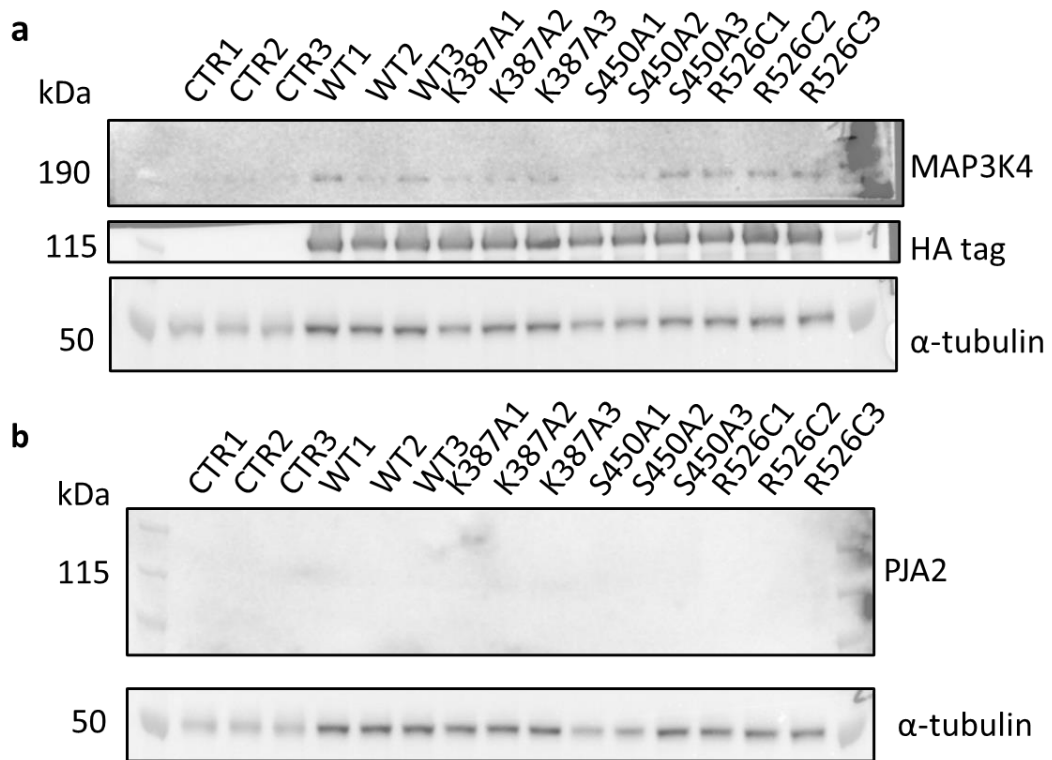
DAPK1	ESFSSIMCFGCHDVYSQASLGMIDHASDLNLLTRRKLRSRLDPPDPLGKDWCLLAMNLGL	1336
CtROCO	----YVSGKLE-KVFSVSKMLDSVI-----SKEERNKERLMG-	949
LRRK1	---QYVPCPVCEATAWAQ---HTDPS-----EKSEDEVQYFDME-	1191
MFHAS1	-----WPG---LHYTV-----HILCSKCLKRG-	991
LRRK2	-----KWAL---YSFND-----GEEHQKILLD--	1828
	:	
DAPK1	PDLVAKYNTSNGAPKDFLPSPLHALLRE---WTT-----YPESTVGTLMMSKLRRELG	1384
CtROCO	-DI----NIKLENIGNPTIPIHQVQEVNVSQETVQHVENLQGGFFENLKADIL-REAELE	1002
LRRK1	-DC----VLTAIERDFISCPRHDPDLVPLQELVP-----ELFMTDFPARLFLENSKLE	1239
MFHAS1	-SP----NPHAFPGEELLSQPRPEG-----	1010
LRRK2	-DL----MKKAEEGDLLVNPDPRLTIPISQIAP-----DLILADLPRNIMLNDELE	1876
	*	
DAPK1	RRDAADFLKASSVFKINLDGNGQEAYASSCNS-----	1417
CtROCO	IDDPKER-KRLANELELAENAITK--MDAAVKSNGKNLKPDPVKDRLEGEFID-----	1050
LRRK1	HSEDEG-----SVLGQGGSGTVIYRARYQQQPVAVKRFHIKKFKNFANVPADTMLRH	1291
MFHAS1	-----	1010
LRRK2	FEQAPE-----FLLGDGSFGS-VYRAAYEGEEVAVKIFNKHT-----	1912
DAPK1	-----	1417
CtROCO	-----NLANENS-	1057
LRRK1	LRATDAMKNFSEFRQEASMLHALQHPCIVALIGISIHPLCFALELAPLSSLNTVLSENAR	1351
MFHAS1	-----	1010
LRRK2	-----SLRLLRQELVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDK-	1963
DAPK1	-----GTSYNSISSVVS-----	1430
CtROCO	-----RLRKGIALVM-NGAEKVQ-----	1074
LRRK1	DSSFIPLGHMLTQKIAYQIASGLAYLHKKNIIFCDLKSDNIIVWSLDVKEHINIKLSDYG	1411
MFHAS1	-----	1010
LRRK2	---ASLTRLQHRIALHVADGLRYLHSAMI IYRDLKPHNVLLFTLYPNAAI IAKIADYG	2019
DAPK1	-----	1430
CtROCO	-----KLARYNNVAPFFDLPSVPPVLL-GKEKT-----	1102
LRRK1	ISRQSFHEGALGVEGTPGYQAPFIR-PRIVYDEKVDMSYGMVLYELLSGQRPALGH---	1467
MFHAS1	VAEIIICPKNGS-----ERVN-----	1025
LRRK2	IAQYCCRMGIKTSEGTGPFRAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKF	2079
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	----HQLQIAKLSKGIKIRPVLGQPEEVQFRRLQALMMECWDTKPEKRPLALS SVVSQMKDP	1523
MFHAS1	-----VALVYPTPTVISPCSKKNVGEKHRNQ-----	1052
LRRK2	PNEFDELEIQGKLPDPVKE----YGCAPWPMVEKLIKQCLKENPQERPTSQVFDILNSA	2135
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	TFATFMVELCCGKQT---AFFSSQGEYTVVFWDGKEESRN----YTVVNT-----KGLM	1572
MFHAS1	-----	1052
LRRK2	ELVCLTRRILLPKNVIVECMVATHHNSRNASIWLGCGHTRDQGLSFLDLNTEGYTSEEVA	2195
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	EVQRMCCPGMKVSCQLQVQSLWTATEDQKIYIYTLKGMCP LNTPPQALDTPAVVTCFLA	1632
MFHAS1	-----	1052
LRRK2	D-SRILCLA---LVHLPVEKESWIVSGTQSGTLLVINTEDGK-KRHTLEKMTDSVTCLYC	2250
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	VPV--IKKNSYLVLAGLADGLVAVFPVVRGTPKDSCSYLCSHTANRSKFSIADEDARQNP	1690
MFHAS1	-----	1052
LRRK2	NSFSKQSKQKNFLLVGTADGKLAI FEDKTVKLKGA-----	2285
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	YPVKAMEVVNSGSEVWVYNGPGLLVIDCASLEICRRLEPYMAPSMVTSVVCSSSEG---RG	1747
MFHAS1	-----	1052
LRRK2	APLKI LNI-----GNVSTPLMCLSESTNSTE	2311
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	EEVVWCLDDKANSVVMYHSTTYQLCARYFCGVPSPLRDMFPVRPLDTEPPAASHTANPKV	1807
MFHAS1	-----	1052
LRRK2	RNVMMGG-----CGTKIFSF---SNDFTIQKLIETRTRTSQL-FSYAAF	2349
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	PEGDS---IADVSIMYSEELGTQILIHQESLTDYCSMSSYSSPPRQAARS PSSLPSSPA	1864
MFHAS1	-----	1052
LRRK2	SDSNIITVVVDTALYIAKQNSPVVEVWDDKTEKLCGL-----	2386

DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	SSSSVPFSTDCEDSDMLHTPGA---A-SDRSEHDLTPMDGETFSQHLQAVKILAVRDLIW	1920
MFHAS1	-----	1052
LRRK2	-----IDCVHFLREVMVKENKESKHK-----MSYSGRVK-TLCLQKNTALW	2426
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	VPRRGDVIVIGLEKDSGAQRGRVIAVLK----ARELTPHGVLVDAAVV---AKDTVVCT	1973
MFHAS1	-----	1052
LRRK2	IGTGGGHILLLDLSTRRLIR---VIYNFCNSVRVMMTAQLGSLKNVMLVVLGYNRNKTEGT	2483
DAPK1	-----	1430
CtROCO	-----	1102
LRRK1	FENENTEWCLAVWRGWGAREFDIFYQSYEELGRLEACTRKRR----	2015
MFHAS1	-----	1052
LRRK2	QKQKEIQSCLTVWDINLPHEVQNLEKHIEVRKEL--AEKMRRTSVE	2527

**Supplementary figure 4:** Clustal alignment of the human ROCO proteins: LRRK2, LRRK1, DAPK1, MFHAS1 and the CtROCO from *C. tepidum*.

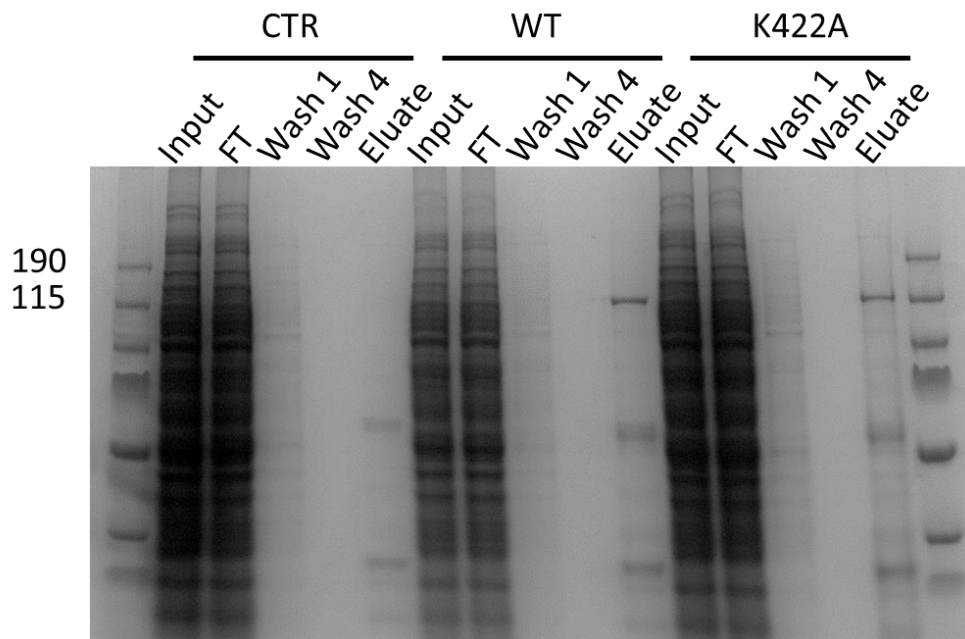
<b>MFHAS1 first layer interactome pathways using g:profiler</b>
Insulin signalling pathway
GnRH signalling pathway
Cellular senescence
C-type lectin receptor signalling pathway
Phototransduction
Glucagon signalling pathway
Long-term potentiation
Glioma
Apelin signalling pathway
Rap1 signalling pathway
Lipid and atherosclerosis
Oxytocin signalling pathway
Ras signalling pathway
Melanogenesis
Pathways in cancer
Neurotrophin signalling pathway
Kaposi sarcoma-associated herpesvirus infection
Renin secretion
Amphetamine addiction
Pertussis
Gastric acid secretion
Human immunodeficiency virus 1 infection
Estrogen signalling pathway

**Supplementary figure 5:** Kyoto Encyclopedia of Genes and Genomes pathways (KEGG) pathway of the functional enrichment of MFHAS1 from g:profiler.



**Supplementary figure 6:** Immunoblot of replicate eluates sample from the immunoprecipitation experiment. Control (CTR) Wild Type (WT) and mutant S450A MFHAS1. MFHAS1: Malignant Fibrous Histiocytoma-Amplified Sequence 1. MAP3K4: Mitogen-activated protein kinase kinase 4. Praja2: E3 ubiquitin-protein ligase Praja-2 also called PJA2. HA: hemagglutinin.





**Supplementary figure 7:** SDS-page gel in Coomassie blue. Immunoprecipitation experiment of the K422A MFHAS1 mutation, control (CTR) and wild type (WT). FT: Flow-through.