Universidade de Aveiro Departamento de Química
Ano 2013

Lúcia Isabel Paisano Sabala

Alterações metabolómicas induzidas pela DJ-1 e seus mutantes

Universidade de Aveiro Departamento de Química
Ano 2013

Lúcia Isabel Paisano Sabala

Alterações metabolómicas induzidas pela DJ-1 e seus mutantes

DJ-1 mutants metabolomics: finding Parkinson's disease biomarkers

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, Ramo de Biotecnologia Molecular, realizada sob a orientação científica do Doutor Bruno José Fernandes Oliveira Manadas, Investigador Auxiliar do Centro de Neurociências e Biologia Celular da Universidade de Coimbra, e do Doutor António Carlos Matias Correia, Professor Catedrático do Departamento de Biologia da Universidade de Aveiro.

Financiado por Fundos FEDER através do Programa Operacional Factores de Competitividade – COMPETE e por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito dos projectos PEst-C/SAU/LA0001/2013-2014 e PTDC/NEU-NMC/0205/2012.

Agradecimentos

Gostaria de agradecer, em primeiro lugar, ao Doutor Bruno Manadas pela oportunidade que me deu de integrar e desenvolver este projecto. Obrigada pela disponibilidade sempre revelada, por todo o conhecimento transmitido, conselhos e críticas. Obrigada por sempre ter acreditado (até mesmo mais do que eu) e nunca ter desistido de mim. Foi crucial sentir sempre esse apoio e confiança para dar um passo de cada vez. Obrigada por tudo.

Agradeço também ao Doutor António Correia por ter aceitado de imediato co-orientar este projecto, possibilitando a sua realização.

Gostaria também de agradecer ao Doutor Pedro Castanheira por me ter recebido na Unidade de Biotecnologia Molecular do Biocant, por todos os conhecimentos que me transmitiu e pelo acompanhamento contínuo do meu trabalho.

Agradeço também ao Doutor Mário Grãos por me ter recebido na Unidade de Biologia Celular do Biocant e por toda a atenção prestada ao longo do desenvolvimento do meu trabalho.

Gostaria de agradecer ainda ao Doutor Rui Brito e ao Doutor Tiago Faria pela sua disponibilidade em me receber no Departamento de Química da Universidade de Coimbra para a realização do estudo da estrutura secundária das proteínas por Dicroísmo Circular.

Um agradecimento a todos os meus colegas da Unidade de Proteómica e Metabolómica do Biocant que sempre me acompanharam nas minhas alegrias e tristezas.

À Liliana Loureiro que passou comigo por todos os obstáculos e vitórias que tivemos ao longo do desenvolvimento do trabalho. À Sandra Anjo por tudo o que me ensinou, por ter estado sempre presente, sempre preocupada, sempre atenta ao meu trabalho, e que sempre teve uma palavra animadora quando era preciso. À Matilde Melo por me ter recebido de braços abertos e me ter acompanhado todo o tempo sem nunca ter hesitado. À Cátia Santa que sempre que eu precisava estava lá, mesmo não sabendo para o que ia, capaz de me deixar sempre bem-disposta a qualquer momento. À Vera Mendes pela paciência que sempre teve comigo e por tudo o que me ensinou sobre espectrometria de massa e metabolómica. À Margarida Coelho por me tentar ensinar Estatística e pelas saídas para almoçar quando não trazíamos comida. À Joana Pinto por todos os dias de canções que tínhamos, era a única que me acompanhava nessas “maluqueiras”. Á Mariana Marcos pelos momentos de diversão e de riso. À Susana Saraiva por me animar sempre que diz alguma coisa. Com ela é impossível não soltar uma gargalhada.

Gostaria também de agradecer a todos os colegas da Unidade de Biotecnologia Molecular e da Unidade de Biologia Celular que sempre mostraram disponibilidade para ajudar.

Um especial agradecimento à Manuela Lago por ter sido a amiga que esteve sempre comigo em todas as horas e nunca me deixou “cair”. Sem ela não teria sido possível chegar ao fim.

Um obrigada a todos os meus amigos que mesmo estando longe estão presentes no meu coração.

Obrigada prima Lília, “sobrinho” Filipe e tios algarvios pelo precioso retiro espiritual que me proporcionaram nas férias de Verão permitindo recarregar as energias necessárias para continuar o trabalho.

Por último, obrigada às pessoas mais importantes da minha vida, pais e mana, que sempre estiveram e sempre vão estar ao meu lado nos bons e maus momentos para me ouvir e amparar. Obrigada por me amarem e aceitarem como sou.

Muito obrigada!

palavras-chave

Doença de Parkinson, stress oxidativo, DJ-1, mutações da DJ-1, Metabolómica, biomarcadores.

resumo

A doença de Parkinson (DP) é a segunda doença neurodegenerativa mais comum caracterizada pela perda massiva de neurónios dopaminérgicos. Apesar de décadas de pesquisa a causa da DP esporádica ainda é desconhecida. A DP é uma doença multifactorial complexa, que provavelmente resulta de uma interacção elaborada na sua maioria de factores desconhecidos: vários genes, efeitos modificadores de alelos de susceptibilidade, exposições ambientais e interacções gene-ambiente e o seu impacto directo sobre o desenvolvimento e envelhecimento do cérebro. A descoberta de genes relacionados com a doença tem contribuído substancialmente para a compreensão dos mecanismos moleculares envolvidos na patogénese da DP. Sabe-se que um conjunto de acontecimentos que conduzem à morte da célula, incluindo o stress oxidativo, contribui para a patogénese da DP. Entre os vários genes mutados na DP familiar, apenas o gene *DJ-1*, um gene autossómico recessivo causador de DP familiar de início precoce, desempenha um papel directo nos mecanismos de defesa oxidativa da *substantia nigra pars compacta*. O estudo da biologia da DJ-1 pode fornecer informações importantes para vias celulares alteradas na DP. Assim, a compreensão de como as mutações da DJ-1 interferem com a estrutura e função da proteína é de crucial importância. Mutações no gene *DJ-1* podem levar à perda da função neuroprotectora da proteína. Deste modo, pode ocorrer um desequilíbrio homeostático no sistema e nos metabolitos celulares, que podem ser utilizados como marcadores celulares de condições de stress. Portanto, o objectivo deste estudo foi comparar várias condições biológicas para identificar metabolitos que são significativamente alterados em condições de repouso e de stress oxidativo, e avaliar também o efeito da adição da DJ-1 WT e mutantes recombinantes à linha celular SH-SY5Y sob condições normais e condições de stress oxidativo. A fim de atingir esse objectivo, diferentes proteínas mutantes recombinantes foram produzidas e caracterizadas estruturalmente para avaliar a sua acção na modulação dos metabolitos. Uma vez adicionadas às células, uma análise não direccionada de espectrometria de massa dos metabolitos foi realizada a fim de encontrar potenciais metabolitos de interesse. Este foi o primeiro estudo para o perfil metabolómico do stress oxidativo com a adição exógena de DJ-1 WT e mutantes recombinantes, e permitiu a descoberta de oito possíveis biomarcadores de stress oxidativo. No futuro, estes resultados devem ser validados numa análise direccionada, para identificação, quantificação, interpretação funcional e análise das vias dos metabolitos, para tentar compreender a sua modulação pela DJ-1 e o seu potencial uso como marcadores de stress oxidativo, e em último caso como biomarcadores da doença de Parkinson. Assim, estes resultados podem contribuir para estratégias futuras para o tratamento e prevenção da doença e oferecer novos rumos para o reconhecimento de indicadores bioquímicos específicos da doença.

keywords

Parkinson's disease, oxidative stress, DJ-1, DJ-1 mutations, Metabolomics, biomarkers.

abstract

Parkinson's disease (PD) is the second most common neurodegenerative disorder characterized by massive loss of dopaminergic neurons. Despite of decades of research the cause of sporadic PD is still unknown. PD is a complex multifactorial disorder, which probably results from an elaborate interplay of mostly unknown factors: several genes, modifying effects by susceptibility alleles, environmental exposures and gene-environment interactions, and their direct impact on the developing and aging brain. The discovery of disease-related genes has contributed substantially to the understanding of the molecular mechanisms involved in PD pathogenesis. It is known that a cascade of events leading to cell death, including the oxidative stress, contributes for the pathogenesis of PD. Among several genes mutated in familial PD, only *DJ-1, an* autosomal recessive gene causative of familial early onset PD, plays a direct role in oxidative defense mechanisms of *substantia nigra pars compacta*. The study of DJ-1 biology can provide important clues to altered cellular pathways in PD. Thus, understanding how the causative DJ-1 mutations interfere with the structure and function of DJ-1 protein is of critical importance. Mutations in *DJ-1* gene may lead to loss of neuroprotective function of the protein. In this way it may occur a homeostatic imbalance in cell system and metabolites, which can be used as cellular markers of stress conditions. Therefore, the aim of this study was to compare multiple biological conditions to identify the metabolites that are significantly altered in resting and oxidative stress conditions, and access also the effect of the addition of the recombinant DJ-1 WT and mutants to SH-SY5Y cell line under normal and oxidative stress conditions. In order to achieve this goal, different recombinant protein mutants were produced and structurally characterized to access their rule in metabolite modulation. Once added to cells, an untargeted mass spectrometry analysis of metabolites was conducted in order to find potential and putative metabolites of interest. This was the first study for oxidative stress metabolomics profiling with the exogenous addition of recombinant DJ-1 WT and mutants and allowed the finding of eight possible oxidative stress biomarkers. In the future, these results must be validated in a targeted analysis, for metabolite ID verification, quantitation, functional interpretation, and pathway analysis, to try to understand their modulation by DJ-1 and their potential use as oxidative stress markers and latter as Parkinson´s disease biomarkers. Hence, these findings may contribute to future strategies for the treatment and prevention of the disease and offer new directions for recognizing disease-specific biochemical indicators.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

°C	degree Celsius
ACN	acetonitrile
AD	autosomal dominant
ADP	adenosine diphosphate
ANK	ankyrin repeat
ANOVA	analysis of variance
APCI	atmospheric pressure chemical ionization
APPI	atmospheric pressure photoionization
AR	autosomal recessive
ARM	Armadillo
ATP	adenosine triphosphate
BBB	blood brain barrier
BSA	bovine serum albumin
C18	octadecylsilane bonded silica
CCD	charged-coupled device
CD	circular dichroism
cDNA	complementary DNA
CNS	central nervous system
CNV	copy number variation
CO_2	carbon dioxide
COMT	catechol-*O*-methyltransferase
COR	COOH terminal of Roc
DA	dopamine
Da	Dalton
DAT	dopamine transporter
DMEM	Dulbecco's modified Eagle medium
dNTP's	deoxyribonucleotide triphosphates
DPBS	Dulbecco's phosphate buffered saline
dsDNA	double stranded DNA
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EO	early-onset
ESI	electrospray ionization
FA	formic acid
FBS	fetal bovine serum
FDR	false discovery rate

g	gram
g	gravity force
GD	Gaucher's disease
GTP	guanosine-5'-triphosphate
H_2O_2	hydrogen peroxide
HCl	hydrochloric acid
HPLC	high-performance liquid chromatography
IBR	in-between-ring
IDA	information dependent acquisition
Indels	insertions and deletions
IPTG	isopropyl β-D-1-thiogalactopyranoside
JO	juvenile-onset
K_2HPO_4	dipotassium phosphate
KCl	potassium cloride
KRS	Kufor Rakeb syndrome
L-DOPA	L-3,4-dihydroxyphenylalanine or levodopa
LB medium	Luria-Bertani medium
LBs	Lewy bodies
LC-MS	liquid chromatography coupled to mass spectrometry
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
LO	late-onset
LRR	leucine-rich repeat
LRRK2	leucine-rich repeat kinase 2
m/z	mass-to-charge ratio
MAO-B	monoamine oxidase B
mETC	mitochondrial electron transport chain
$MgSO_4$	magnesium sulfate
MMTS	S-methyl methanethiosulfonate
MPP^+	1-methyl-4-phenylpyridinium ion
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
mtDNA	mitochondrial DNA
MTS	mitochondrial targeting signal
MW	molecular weight
$(NH_4)_2SO_4$	ammonium sulfate
NAC	nonamyloid component
NaCl	sodium chloride
$NaH_2PO_4 \cdot H_2O$	sodium phosphate monobasic monohydrate

NH_4HCO_3	ammonium bicarbonate
NMR	nuclear magnetic resonance
$OH^{\cdot}$	hydroxyl radical
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PD	Parkinson's disease
Pen-Strep	penicillin-streptomycin
PINK1	PTEN-induced kinase 1
PLS-DA	partial least square-discriminant analysis
PTEN	phosphatase and tensin homolog
RING	really interesting new gene
Roc	Ras of complex proteins: GTPase
ROI	regions of interest
ROS	reactive oxygen species
RT	retention time
S.E.M	standard error of the mean
SCA2	spinocerebellar ataxia type 2
SCX	strong cation exchange
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SNCA	α-synuclein
SNpc	*substantia nigra pars compacta*
SWATH	Sequential Windowed data independent Acquisition of the Total High-resolution Mass Spectra
TCEP	*tris*(2-carboxyethyl)phosphine
TEAB	triethylammonium bicarbonate
T_m	melting temperature
TM	transmembrane
TOF	Time-of-Flight
Tris	tris(hydroxymethyl)aminomethane phosphate
UBL	ubiquitin-like
UPS	ubiquitin-proteasomal sydtem
UV	ultraviolet
WT	wild-type

❶ INTRODUCTION

1. INTRODUCTION

1.1 Parkinson's Disease

Parkinson's disease (PD) is a neurodegenerative disorder first described by James Parkinson in 1817 [1]. PD is the most common movement disorder and the second most common neurodegenerative disorder after Alzheimer's disease, affecting 7 to 10 million people worldwide, according to the Parkinson's Disease Foundation [2-4]. Disease prevalence is age-associated, with approximately 1% of the population being affected over the age of 65 years, increasing to 4-5% in 85-year-olds [3, 5, 6].

PD results mainly from progressive degeneration of nigrostriatal dopaminergic neurons in the *substantia nigra pars compacta* (SNpc) and other monoaminergic cell groups in the brainstem and cortical areas (Figure 1.1). It also shows increased microglial activation and accumulation of eosinophilic, intracellular proteinaceous inclusions in surviving dopaminergic neurons, known as Lewy bodies (LBs) and Lewy neurites (collectively known as Lewy-related pathology) (Figure 1.1). The basic components of these inclusions are α-synuclein, neurofilament proteins, and ubiquitin [2, 3, 6-8], however some monogenic forms of PD lack this typical Lewy body pathology [6].

The main clinical phenotype of PD is parkinsonism, a movement disorder that is characterized by motor dysfunctions, such as bradykinesia, resting tremor, rigidity and postural instability [1-3, 5-7, 9]. In addition to the motor disturbances PD is characterized by numerous non-motor symptoms, such as autonomic insufficiency, cognitive impairment, depression, olfactory deficits, psychosis and sleep disturbance [2, 3, 6].

Motor dysfunction is thought to arise from progressive loss of dopaminergic neurons within the *substantia nigra pars compacta* and becomes evident when approximately 80% of striatal dopamine (DA), neurotransmitter that normally sends signals in the brain to control body movement, and 50% of nigral neurons are lost. Thus, the population of undiagnosed asymptomatic patients is probably large [2, 3, 7].

It is not possible to diagnose parkinsonism with neuropathologic methods. It is only possible to describe histologic, neurochemical, and molecular findings that are frequently associated with parkinsonism [9]. The diagnosis of PD can only be confirmed at

autopsy because more than forty different neurological diseases can show signs of parkinsonism. Therefore, the clinical diagnosis of PD is typically based on the presence of cardinal motor features, absence of atypical findings suggestive of an alternate diagnosis, and response to levodopa (L-DOPA) [3, 7].

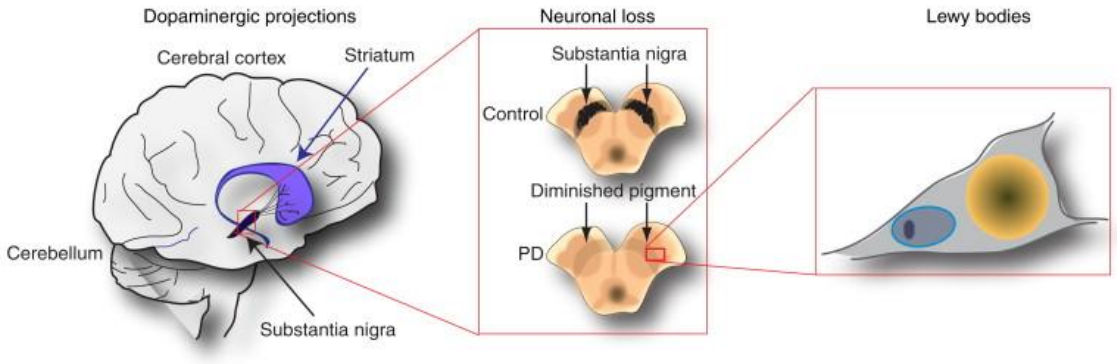


Figure 1.1 - *Substantia nigra* degeneration and Lewy Bodies in Parkinson's disease. View of the main neuropathological events in PD at three levels from left to right. At the level of the brain, a major pathway is degeneration of the dopaminergic projections from the *substantia nigra* (in black) to the striatum (in purple), both of which are in the midbrain under the cerebral cortex. At the level of *substantia nigra*, the neurons that form the presynaptic portion of this pathway are normally melanized and are easily identified by this pigment in control brains (upper panel). In contrast, the loss of neurons in this region is so substantial that the whole area becomes depigmented in PD cases (lower panel). Of the few remaining cells, many show pathological changes, including the accumulation of proteins and lipids in Lewy bodies [10].

Patients have a response to dopamine replacement therapy using the metabolic precursor of dopamine, L-DOPA, which is the main treatment for PD. However, its use is complicated by the emergence of motor fluctuations and dyskinesias [3, 5]. Other treatment options include dopamine agonists, anticholinergics, amantadine, monoamine oxidase B (MAO-B) and catechol-*O*-methyltransferase (COMT) inhibitors, and deep brain stimulation [1, 3]. However, neuroprotective treatment that delays or prevents neurodegeneration in PD remains an unrealized goal. Levodopa and dopamine agonists only relieve symptoms [2, 3].

Due to demographic changes and increased life expectancy the prevalence of PD will further increase dramatically worldwide in the coming decades in the absence of a neuroprotective treatment, preventive interventions or cure [6, 11]. A better understanding of disease pathogenesis may lead to the development of a neuroprotective treatment.

1.1.1 PD Etiology

Some parkinsonism disorders are chronic and progressive and caused by an unknown degenerative disease process, whereas others may have clear genetic cause or can be transient and caused by effects of toxins, metabolic disturbances, or drugs (Figure 1.2) [9].

The cause of sporadic PD, which occurs in the absence of genetic linkage and accounts for more than 90% of all diagnosed cases, is still unknown [7]. Decades of research have not found a single cause for PD and therefore a single factor is unlikely to appear [1, 2, 5]. However, aging is the principle risk factor for Parkinson's disease. In addition to age, gender also influences the incidence of PD as several of the studies found evidence of a higher incidence in men than in women [2, 8, 12, 13].

PD is a complex multifactorial disorder, which probably results from an elaborate interplay of mostly unknown factors: several genes, modifying effects by susceptibility alleles, environmental exposures and gene-environment interactions, and their direct impact on the developing and aging brain [1, 3, 6, 14, 15].

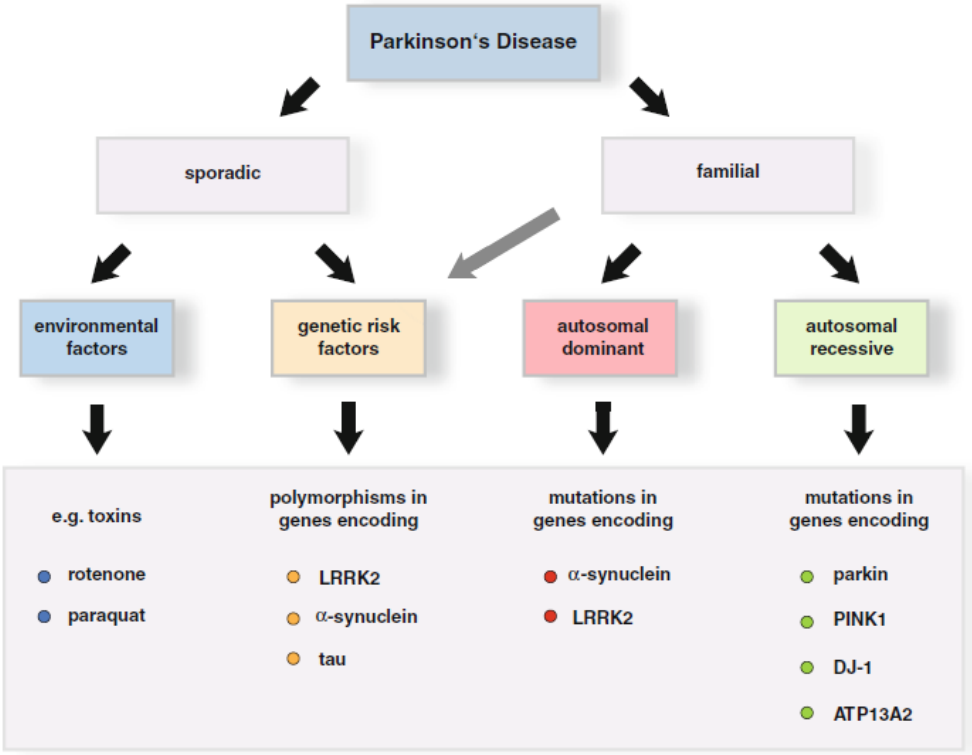


Figure 1.2 - Etiology of Parkinson's disease [16].

1.1.1.1 Environmental Factors

Many environmental factors may increase the risk of developing PD, such as exposure to pesticides and herbicides, rural living, well-water consumption, heavy metal (copper, manganese, lead and iron) and solvent exposure, certain occupations, including welding and mining, and electromagnetic fields [5, 13, 17, 18]. By contrast, some protective effects have been described for cigarette smoking, caffeine intake, alcohol and use of anti-inflammatory agents. Nevertheless, it is not clear how these agents influence disease risk [5, 6, 13, 17].

A number of exogenous toxins have been associated with the development of parkinsonism, including trace metals, cyanide, lacquer thinner, organic solvents, carbon monoxide and carbon disulfide, as well as there has also been interest in the possible role of endogenous toxins such as tetrahydroisoquinolines and beta-carbolines [9, 19]. These toxins can be administered either systemically or locally, depending on the type of agent used and the species involved [18].

The most convincing evidence for an environmental factor in PD relates to the toxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a by-product of the illicit manufacture of a synthetic meperidine derivative, with selective toxicity for dopaminergic neurons (Figure 1.3) [18, 19].

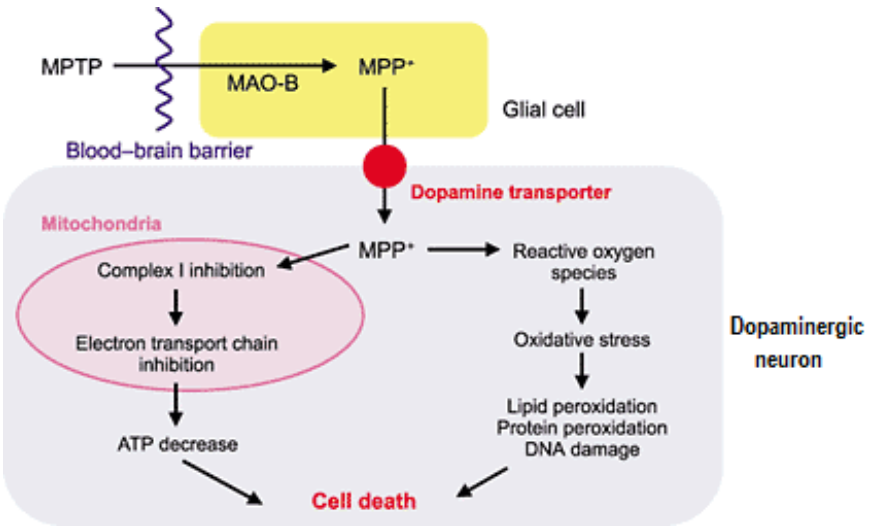


Figure 1.3 - MPTP neurotoxic process [20]. It was demonstrated that MPTP crosses the blood-brain barrier (BBB) and is taken up by glial cells where it is oxidized by monoamine oxidase B into a pyridinium species, before being converted into the neurotoxin 1-methyl-4-phenylpyridinium ion (MPP^+) by further oxidation [8, 21]. MPP^+ is then taken up by dopamine transporter (DAT) into dopaminergic neurons of the *substantia nigra*, where it blocks mitochondrial complex I of the respiratory chain, leading to impaired energy metabolism, oxidative stress, proteasomal dysfunction and, eventually, death of dopaminergic neurons, similar to that found in PD [8, 19, 22].

When administered, MPTP causes a L-3,4-dihydroxyphenylalanine (L-DOPA) responsive parkinsonian syndrome, characterized by all of the basic symptoms of PD, thus representing the best PD-like phenotype obtainable in experimental animals [18]. A major limitation of the MPTP for studying PD, which is shared by the majority of toxins, is that the SNpc lesion is not accompanied by the formation of LB-like cytoplasmic inclusions, a crucial neuropathological hallmark of PD [18].

MPTP, MPP^+ and certain pesticides, in particular paraquat (*N*,*N*'-dimethyl-4,4'-bipyridinium dichloride) have structural similarities (Figure 1.4), although the mechanisms of action are quite different [18].

Paraquat is a non-selective bipyridyl contact herbicide. Because it is a charged molecule, paraquat does not cross the BBB [18]. In the cytosol, paraquat generates a lot of oxidative stress by acting as a redox cycling compound through of superoxide anion formation, as well as by impairing recycling of oxidized glutathione (GSH) to its reduced form, which hampers the efficiency of intracellular antioxidant systems [18]. Since this toxin has low affinity to mitochondrial complex I, dysfunction of this complex does not appear to play a significant role in neurotoxicity induced by paraquat [18].

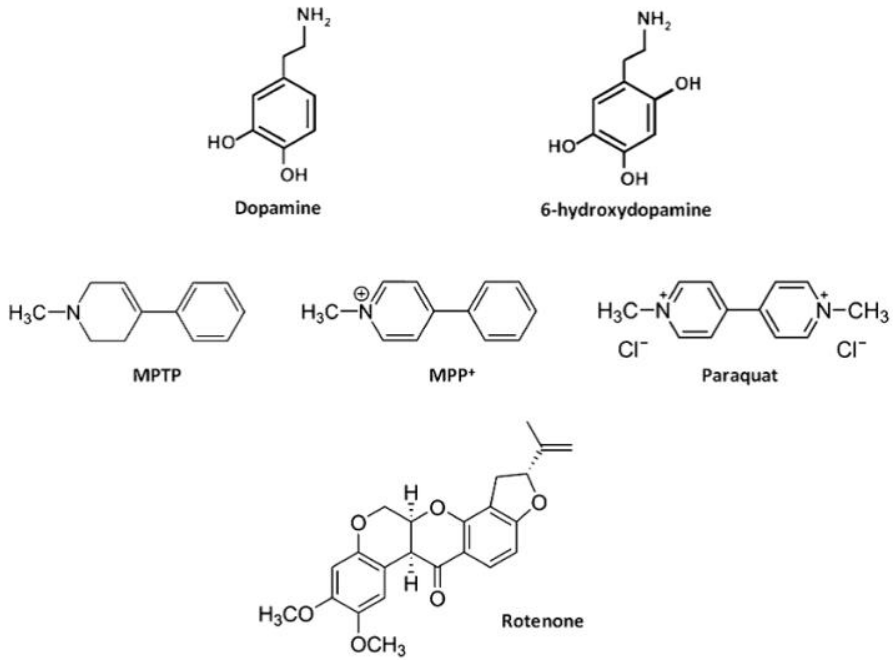


Figure 1.4 - Structures of dopamine and major dopaminergic toxins used to replicate features of Parkinson's disease in animal models [18].

Rotenone (Figure 1.4) is a flavonoid found in the roots and stems of several plants and used as a broad-spectrum pesticide [18]. Due to its high lipophilicity, rotenone crosses the BBB, and once in the cell, it blocks complex I activity, causing formation of reactive oxygen species (ROS), and inhibits proteasome activity, generating proteolytic stress [18, 21]. In rats, rotenone caused the selective degeneration of nigrostriatal dopaminergic neurons and formation of LB-like cytoplasmic inclusions. A finding seen for the first time in the field of PD toxic models [21].

Current research in PD is mainly performed on animal models of PD induced by intoxication with MPTP, paraquat, rotenone and models of post-encephalitic parkinsonism. Although neither of which has fully reproduced the clinical and pathological features of true PD [1, 5].

1.1.1.2 Genetic Factors

The mapping and identification of the first mutations responsible for Parkinson's disease in 1996 showed that PD may be hereditary [14]. In the last 2 decades, molecular genetic analyses in PD families provided important insights in disease mechanisms related with PD pathology [15, 23, 24]**.** However, only about 5-10% of patients report a positive familial history of the total PD patients [14].

To date 28 different chromosomal regions more or less convincingly related to PD are known. In the current PD genetics nomenclature, 18 specific chromosomal locus, are termed PARK, and numbered in chronological order of their identification [14]. PARK loci are mostly associated with rare forms of younger onset disease that have clinical features distinguishing familial from sporadic cases, and many putative genetic risk factors (Table 1.1) [6, 13]. For all of the loci, the causative gene has not yet been identified, nor do all of the identified genes contain causative or disease-determining mutations [14].

It is of note that some of the loci have been identified by genetic linkage analysis in large families, some based on the known function of the protein product of the gene they contain, yet others have been established by genome wide association studies performed on a population level [14].

Table 1.1 - Parkinson's disease-associated loci and genes [8, 14].

PARK loci	Gene	Map Position	Inheritance	Disease Onset	Status and Remarks	Mutations
PARK1/ PARK4	*SNCA*	4q21	AD; rarely sporadic	EO	Confirmed	A30P, E46K, A53T and genomic duplications/ triplications
PARK2	*Parkin*	6q25-q27	AR; sporadic	JO; EO	Confirmed	~170 mutations (point mutations, exonic rearrangements)
PARK3	Unknown	2p13	AD	LO	Unconfirmed; may represent a risk factor	Not identified
PARK5	*UCHL1*	4p14	AD	LO	Unconfirmed	One mutation in a single PD sibling pair
PARK6	*PINK1*	1p35-p36	AR	EO	Confirmed	~50 point mutations, rare large deletions
PARK7	*DJ-1*	1p36.23	AR	EO	Confirmed	~15 point mutations and large deletions
PARK8	*LRRK2*	12q12	AD; sporadic	LO	Confirmed	> 80 missense variants, >7 of them pathogenic, including the common G2019S
PARK9	*ATP13A2*	1p36	AR	Juvenile KRS, EOPD	Confirmed	~10 point mutations
PARK10	Unknown	1p32	Unclear	LO	Confirmed susceptibility locus	Not identified
PARK11	*GIGYF2*	2q36-q37	AD	LO	Not independently confirmed; possibly represents a risk factor	7 missense variants
PARK12	Unknown	Xq21-q25	Unclear	LO	Confirmed susceptibility locus; possibly represents a risk factor	Not identified
PARK13	*Omi/HTRA2*	2p13	Unclear	LO	Unconfirmed	2 missense variants
PARK14	*PLA2G6*	22q12-q13	AR	Juvenile L-dopa-responsive dystonia-parkinsonism	Confirmed	2 missense mutations
PARK15	*FBXO7*	22q12-q13	AR	EO pakinsonian-pyramidal syndrome	Confirmed	3 point mutations
PARK16	Unknown	1q32	Unclear	LO	Confirmed susceptibility locus	Not identified
PARK17	*VPS35*	16q11.2	AD	LO	Confirmed	
PARK18	*EIF4G1*	3q27.1	AD	LO	Unconfirmed	
Not assigned	*SCA2*	12q24.1	AD for SCA2	Unclear		Low-range interrupted CAG expansions in *SCA2*
Not assigned	*GBA*	1q21	AR for GD	Unclear	Confirmed susceptibility locus	

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; EO, early-onset; LO, late-onset; JO, juvenile-onset; GD, Gaucher's disease; SCA2, spinocerebellar ataxia type 2; KRS, Kufor Rakeb syndrome.

It is believed that identifying the consequences of PD genes mutations will lead to new therapeutic strategies, perhaps ones specifically designed for a particular mutation [25].

The continued study of the cellular functions of each of the PD-related genes has indicated that protein misfolding, as well as dysfunction in the protein degradation systems, may play a crucial role in the cascade of deleterious events implicated in the neurodegenerative process of PD [7].

1.1.2 PD related Genes

Only six of the 28 specific regions contain genes with mutations that conclusively cause monogenic PD. Monogenic PD is a form of the disease for which a mutation in a single gene is sufficient to cause the phenotype. These forms of PD are relatively rare, even collectively, mutations in these six genes explain only a limited number (3-5%) of disease occurrences [3, 14].

The discovery of disease-related genes, *α-synuclein (SNCA), Leucine-rich repeat kinase 2 (LRRK2), Parkin (PARK2), PTEN-induced kinase 1 (PINK1),* and *DJ-1 (PARK7),* has contributed substantially to the understanding of the molecular mechanisms involved in PD pathogenesis [6, 7].

1.1.2.1 Autosomal Dominant PD

In autosomal-dominant disorders, one mutated allele of the gene is enough to cause the disease [14]. Similar to sporadic PD, autosomal dominant PD is principally related with mutations in *SNCA* and *LRRK2*. In these cases, the age of onset and pathology may quite-closely resemble idiopathic PD [14, 26, 27]. The pathogenic role of other dominant genes in PD, *UCHL1*, *GIGYF2* and *HTRA2*, is still controversial, because they have not been found in other patients or appear to act as genetic risk factors (Table 1.1) [8].

The *SNCA* and *LRRK2* genes and their corresponding proteins will be described in the following subsections.

1.1.2.1.1 α-Synuclein (SNCA, PARK1/4*)*

SNCA gene (114 kb), located on chromosome 4q21, has 6 exons coding for a small neuronal and abundant 140-amino acid (14 kDa) protein, α-synuclein (Figure 1.5) [3, 14, 15, 28]. α-Synuclein consists of three domains: the amino-terminal region; a central hydrophobic domain; and an acidic, negatively charged carboxy-terminal domain, which contains an aggregation inhibition region [3, 14].

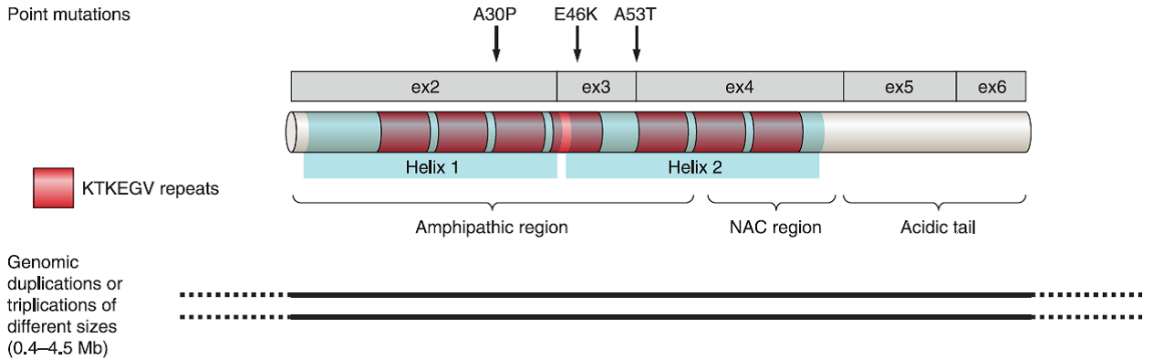


Figure 1.5 - **Schematic representation of the α-synuclein protein**. The three pathogenic missense mutations are indicated above the protein organization and duplications or triplications of different genomic sizes below the protein. *SNCA* mutations might reduce the affinity of the protein for lipids, thus increasing the intracellular pool of proteins that, along with duplication and triplication of the gene, accentuates the tendency of the protein to form oligomers and later fibrillar aggregates. NAC, nonamyloid component [8].

α-Synuclein is expressed throughout the mammalian brain and enriched in presynaptic nerve terminals [3]. It is typically found as a natively unfolded, soluble protein in the cytoplasm or associated with lipid membranes [3, 15].

Without genetic changes, α-synuclein is an abundant protein and a main component of LBs in idiopathic PD [6, 15]. The exact biological function of α-synuclein in brain is still not fully understood, although there is evidence that implicates α-synuclein in neurotransmitter release and synaptic vesicle turnover at the presynaptic terminals [15].

SNCA was the first gene with mutations reported to cause autosomal-dominant PD [14]. These mutations have been linked to rare cases of familial PD, only explaining disease in about 2.5% of known unrelated affected carriers [3, 14, 15]. Patients with *SNCA* mutations usually have early-onset PD (EOPD), being age of onset below 50 years, with an initially good response to levodopa treatment [3, 6, 14].

However, the disease has a rapid progression and often presents with dementia and cognitive decline and sometimes with atypical clinical features such as prominent cognitive deterioration, central hypoventilation, myoclonus and severe postural hypotension [3, 6, 14]. Lewy bodies are present and spread through the *substantia nigra* and locus ceruleus, hypothalamus and cerebral cortex [14].

Only three different missense mutations as well as duplications and triplications of the entire gene have been reported [3, 14, 28]. The first missense mutation identified, A53T, seems to be the most frequent one and was found in one Italian, eight Greek, two

Korean, and one Swedish family as well as in one apparently sporadic PD patient of Polish origin [3, 6, 14, 15, 28]. A second missense mutation A30P was described in a German family with a typical PD phenotype [3, 6]. Finally, the E46K mutation was identified in a Basque family with parkinsonism that progressed to dementia with LBs, visual hallucinations and fluctuations in consciousness [6, 14, 28].

Seventeen duplications of the entire coding region of *SNCA* have been reported to date, 13 in PD families and 4 in sporadic cases; and triplications of the *SNCA* gene were also found in three independent families [14, 15].

While the three missense mutations in *SNCA* are very rare, multiplication of the gene appears to be a more common cause of PD [6, 15]. Gene triplication leads to earlier onset, more severe phenotype and faster progression of disease than duplication, indicating that disease severity is dependent on α-synuclein expression levels [6, 14, 28].

The mechanism by which common *SNCA* variants modify susceptibility for PD is not yet known [3]. However, there is evidence to suggest that *SNCA* alleles associated with increased PD risk are correlated with higher α-synuclein expression *in vitro* and *in vivo* [3].

All three missense mutations impair the amino-terminal domain and *in vitro* show an increased propensity for self-aggregation and toxic oligomerization into protofibrils and fibrils, compared with protein wild-type [3, 14, 15]. Therefore, the missense *SNCA* mutations possibly cause PD through a toxic gain of function, and LBs may represent the attempt to eliminate the cell of toxic damaged α-synuclein [14, 28].

1.1.2.1.2 Leucine-rich repeat kinase 2 (LRRK2, **PARK8***)*

LRRK2 is a large gene (144 kb) located on chromosome 12q12, consisting of 51 exons coding for the LRRK2 protein (Figure 1.6) [3, 14, 15]. LRRK2 is a large 2527-amino acid (286 kDa) protein leucine-rich repeat kinase 2 that consists of a leucine-rich repeat toward the amino terminus of the protein and a kinase domain toward the carboxyl terminus with various conserved functional domains in between [14]. LRRK2 has guanosine-5'-triphosphate (GTP)–regulated serine/threonine kinase activity and a scaffold protein function implied by the multiple protein–protein interaction regions [15, 28].

LRRK2 protein is found in the cytosol and mitochondrial outer membrane, plasma membrane, lysosomes, endosomes, transport vesicles, Golgi apparatus and synaptic vesicles, and its expression has been described in the central nervous system, heart, kidney, lung, liver, and peripheral leukocytes [3].

The exact biological function of LRRK2 remains unknown, because no physiological substrates have been identified so far, although it has been identified as a tyrosine kinase-like protein [3, 15].

LRRK2 was the second causal gene linked to autosomal dominant inherited PD and *LRRK2* mutations are the most frequent known cause of late-onset autosomal dominant and sporadic PD, with a mutation frequency ranging from 2 to 40% in different populations [13, 14, 24]. The high mutation frequency in both familial and sporadic patients makes *LRRK2* the most frequently mutated gene of the six major PD genes [15].

Clinically, *LRRK2*-linked PD usually shows mid-to-late onset, with an average age of onset in the 50-60 years and slow progression. The clinical signs and symptoms are heterogeneous but can resemble sporadic PD with a more benign prognosis, favorably response to levodopa therapy, and less dementia [13, 14]. The disease can show LB pathology, amyloid and tau pathology, neuronal loss without intracellular inclusions and motor neuron disease [3, 6, 14].

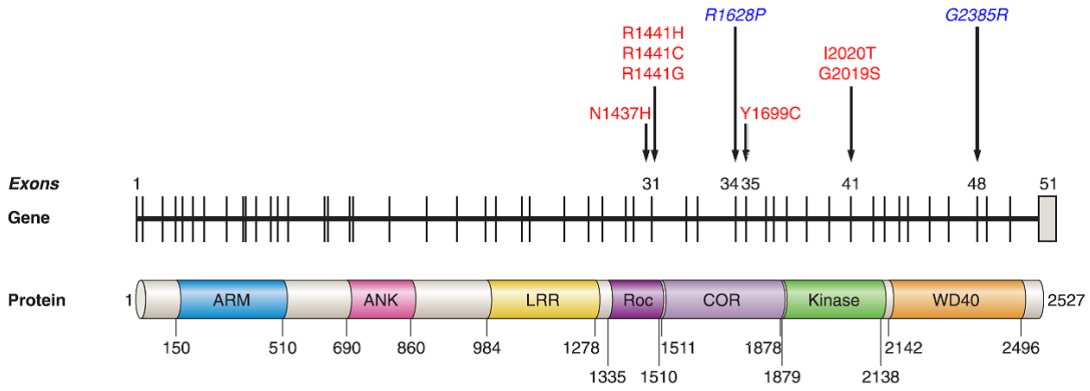


Figure 1.6 - Schematic representation of the *LRRK2* gene and protein. The seven confirmed pathogenic mutations are indicated in red and the two Asian-specific risk factors in blue, above the protein organization. ARM, Armadillo; ANK, ankyrin repeat; LRR, leucine-rich repeat; Roc, Ras of complex proteins: GTPase; COR, COOH terminal of Roc. [8].

PD-causing *LRRK2* missense mutations were initially discovered in a German-Canadian and American family as well as in four Basque families [6]. To date, more than 80 variations in *LRRK2* have been described, but only 7 mutations have been proven to be pathogenic based on co-segregation with the disease: R1441G, R1441C, R1441H, Y1699C, G2019S, I2020T and N1437H [6, 8, 15, 29].

The most frequent and best studied *LRRK2* missense mutation is the G2019S, which occurs in 1-7% of PD patients of European origin, about 20% of Ashkenazi Jewish patients, and approximately 40% of North African Arabs with PD [6, 14, 28]. The mutation is present in sporadic and familial cases and can be clinically indistinguishable from each other [6, 28]. The most prominent phenotype of *LRRK2* G2019S carriers is an asymmetric levodopa-responsive parkinsonism with common presence of tremor and a lower risk for cognitive impairment and olfactory dysfunction in comparison with non-carrier PD patients. However, atypical clinical symptoms can also be seen, such as marked autonomic dysfunction and dementia [6].

The remaining six pathogenic *LRRK2* mutations are less frequent worldwide [6]. The R1441 codon constitutes a mutation hotspot with three different codon substitutions: R1441C, R1441G, R1441H [15].

Pathogenic changes in *LRRK2* are clustered in 10 exons, mostly encoding the carboxy terminal region of the protein. For example, G2019S and I2020T mutants affect its kinase activity [14]. However, pathogenic mechanism leading to PD caused by *LRRK2* mutations is still unclear [14].

1.1.2.2 Autosomal Recessive PD

In autosomal-recessive disorders two mutations, one on each gene copy (allele), are necessary to cause the phenotype [14].

Homozygous or compound heterozygous mutations in the recessive genes *Parkin, PINK1* and *DJ-1* are associated with heritable parkinsonism with early age at onset and, generally, no atypical signs [8]. These genes share similar clinical and pathological features and may serve similar cellular functions [30]. A fourth gene, *ATP13A2*, might also play a role in rare cases with early-onset PD. However, to date, *Parkin* and *PINK1* are the genes most frequently associated with autosomal recessive early onset parkinsonism [8].

Though the pathogenesis of the PD is unknown, it is likely that autosomal recessive PD is more critical in origin and perhaps principally focused on the mitochondria rather than a progressive protein folding disorder, and must thus be addressed carefully with regards to therapeutic intervention [30].

The *Parkin*, *PINK1,* and *DJ-1* genes and their corresponding proteins will be described in the following subsections.

*1.1.2.2.1 Parkin (*PARK2*)*

Parkin is the second largest gene (1.38 Mb) in human genome. It is located on chromosome 6q25-q27 and has 12 exons coding for the Parkin protein, a 465-amino acid (52 kDa) protein E3-ubiquitin ligase (Figure 1.7) [3, 14].

Parkin is predominantly a cytosolic protein but also co-localizes to synaptic vesicles, the Golgi complex, endoplasmic reticulum, and the mitochondrial outer membrane [3].

Parkin was described to act as an E3-ubiquitin ligase that targets cytoplasmic dysfunctional or excessive proteins for proteasomal degradation and plays a role in receptor trafficking [1, 15]. Further, it was shown that under physiological conditions Parkin is involved in mitochondrial maintenance and might induce subsequent autophagy of dysfunctional mitochondria [15].

Parkin was the second identified PD gene and the first gene causing an autosomal recessive form of the disorder [14]. *Parkin* gene was identified in autosomal recessive forms of familial juvenile onset parkinsonism (AR-JP) in Japanese families [1, 31].

Mutations in the *Parkin* gene are the most frequent known cause of early-onset PD (<40–50 years). They account for 10-20% of sporadic early-onset PD patients, over 50% of recessive familial forms, about 80% in those with onset before age 20 and they are very rare in those with onset after 50 years of age [1, 6, 8]. Together, homozygous and compound heterozygous mutations of *Parkin* are responsible for about 1.3-8.2% of early onset PD [6].

The disease is usually slowly progressive with an excellent response to dopaminergic treatment [14]. Although patients with *Parkin*-related parkinsonism exhibit loss of pigmented nigral dopamine neurons and gliosis, LBs are usually not observed [3, 6,

14, 28]. Dystonia is frequently present as well as hyperreflexia, sleep benefit, diurnal fluctuations and early development of dyskinesias. Less frequent atypical pathological findings in *Parkin*-associated PD include neurofibrillary tangles in the neocortex, brainstem and hippocampus [3, 6].

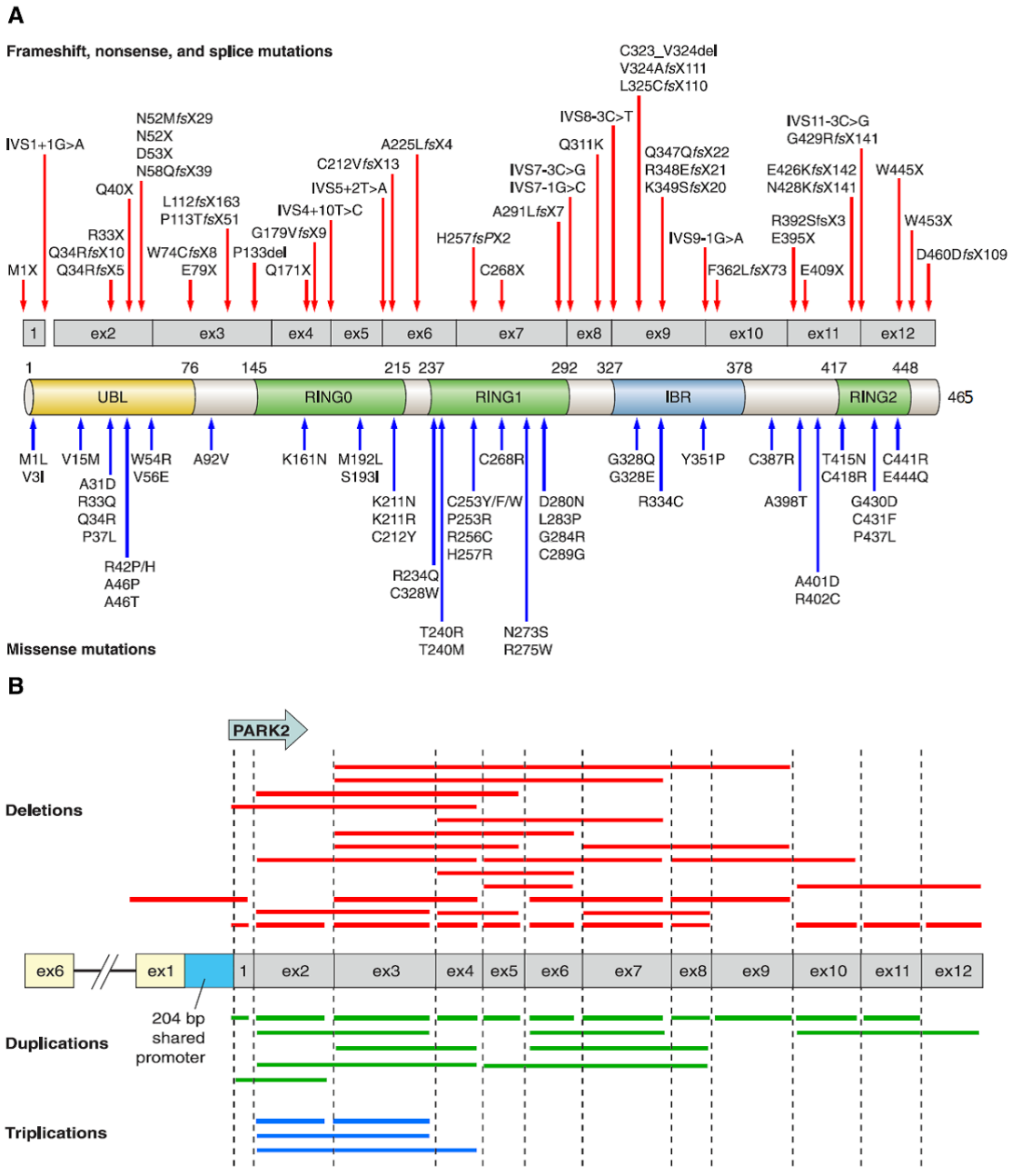


Figure 1.7 - Schematic representation of *Parkin* and the functional domains of the Parkin protein. **(A)** Pathogenic frameshift mutations are indicated above the transcript and protein organizations and missense mutations below; **(B)** Exonic deletions are represented above the transcript (red lines) and exonic duplications (green lines) or triplications (blue lines) below the transcript. UBL, ubiquitin-like; RING, really interesting new gene; IBR, in-between-ring. [8].

More than 170 different mutations have been identified throughout the sequence of this large gene, including alterations of all 12 exons, such as simple mutations like nonsense, missense and splice site mutations, small insertions/deletions (indels), as well as copy number variations (CNVs) of the promoter region and single or multiple exons [8, 14, 15]. The most common mutations are deletions of exons 3 and 4, a point mutation in exon 7 (924C>T) and a single base pair deletion in exon 2 (255/256delA) [3].

Most *Parkin* mutations, including insertions and deletions, impact its E3-ubiquitin ligase activity leading to a loss of Parkin function [15, 28]. Further, many PD-linked point mutations alter Parkin wild-type cellular localization, solubility, or propensity to aggregate [3].

1.1.2.2.2 PTEN-induced kinase 1 (PINK1, **PARK6***)*

PINK1 gene (18 kb) is located on chromosome 1p35-p36, has 8 exons and encodes a 581-amino acid (63 kDa) phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (Figure 1.8) [3, 14]. It consists of an amino-terminal mitochondrial targeting motif, a conserved serine–threonine kinase domain, and a carboxy-terminal auto-regulatory domain [14]. It is a putative serine/threonine kinase and a mitochondrial protein located in the matrix and in the intermembrane space that is ubiquitously expressed in the brain and systemic organs [1, 3].

The PINK1 protein is involved in mitochondrial response to cellular and oxidative stress. This response is likely mediated by regulation of the calcium efflux, influencing processes such as mitochondrial trafficking, reactive oxygen species formation, mitochondrial respiration efficacy, and opening of the mitochondrial permeability transition pore as well as by interaction with cell death inhibitors and chaperones [15]. In addition, PINK1 is an important player in the alleged PINK1/Parkin pathway, regulating mitochondrial morphology and functionality in response to stressors [15]. PINK1 is stabilized on mitochondria with lower membrane potential, and as such, it recruits Parkin from the cytosol to mitochondria, where it becomes enzymatically active and initiates the autophagic clearance of mitochondria by lysosomes [14].

PINK1 is the second most frequent causative gene of autosomal recessive early-onset parkinsonism after *Parkin* and may play an important role in sporadic PD [1]. The

frequency of *PINK1* mutations is in the range of 1–9%, with considerable variation across different ethnic groups [14].

Although age of onset for *PINK1*-related PD is usually between 40-50 years, clinical features are similar to late onset PD, with slow progression, excellent response to levodopa, development of dyskinesias and in some instances dementia [3, 6]. Clinical features, such as prominent dystonia, sleep benefit and hyperreflexia, which were thought to be indicative of the phenotype associated with mutations in the *Parkin* gene, are also frequently reported in *PINK1*-associated parkinsonism [6]. Psychiatric comorbidity, especially anxiety and depression, is overrepresented in *PINK1*-associated PD [6]. Neuronal loss in the *substantia nigra* and LB pathology in *PINK1*-associated PD were comparable to idiopathic PD [6].

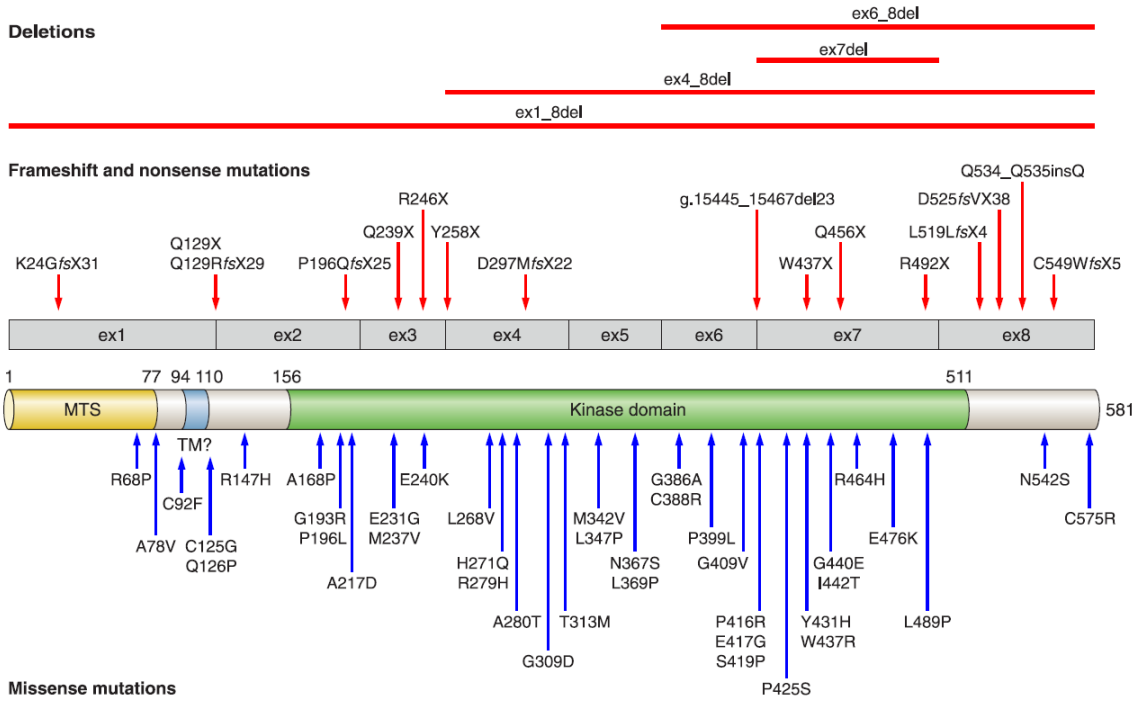


Figure 1.8 - Schematic representation of *PINK1* and the functional domains of the PINK1 protein. Pathogenic frameshift mutations are represented above the transcript and protein organizations and missense mutations below the protein; rare deletions (above, red lines) including a deletion of the whole gene are also represented. MTS, mitochondrial targeting signal; TM, transmembrane. [8].

The first mutations discovered were the G309D homozygous missense and a W437X truncating mutation found in the families of Spanish and Italian descent, respectively [3]. Since then, several mutations have been identified including point

mutations, indels, whole-gene or single/multiple exon CNVs located across the entire gene and truncating mutations [1, 3, 8, 15]. The largest total number of mutations was found in exon 7 and the most frequent mutation is Q456X [14]. Most mutations are isolated except for nine homozygous *PINK1* mutations (Q129fs, A168P, R246X, Y258X, T313M, W437X, Q456X, D525fs and deletion of exons 6–8) for which recurrence in different families has now been observed [6, 15].

PINK1 mutations may cause loss of protein function in patients with recessively inherited forms of PD because most mutations are located in a highly conserved amino acid position in the protein kinase domain, demonstrating the importance of PINK1's enzymatic activity in the pathogenesis of PD [1]. PINK1 wild-type appears to be important in neuroprotection against mitochondrial dysfunction and proteasome induced apoptosis, whereas the G309D mutation impairs this protective effect, possibly by interfering with adenosine diphosphate (ADP) binding and thus inhibiting kinase activity [3]. The E240K and L489P mutants disrupt PINK1's protectivity by either enhancing the instability of the protein or disrupting the kinase activity of the protein [3]. Functional studies indicate that cells transfected with *PINK1* mutants have disrupted mitochondrial membrane potential under stressful conditions [3].

*1.1.2.2.3 DJ-1 (***PARK7***)*

DJ-1 gene (24 kb) is located on chromosome 1p36.23 and has 7 exons coding for a highly conserved 189-amino acid (20 kDa) protein (Figure 1.9) [3, 32, 33]. DJ-1 is a homodimer that belongs to the peptidase C56 family of proteins [34]. It is a cytosolic protein but can also localizes in the nucleus and translocate into the mitochondria [35, 36]. DJ-1 is ubiquitously and abundantly expressed in most mammalian tissues including in the brain where it is found in both neuronal and glial cells [37]. However, DJ-1 does not appear to be an essential component of LBs [37].

DJ-1 was initially identified as an oncogene and described in association with male infertility in rats [8]. It is a member of the ThiJ/Pfp1 family of molecular chaperones, which are induced under oxidative stress conditions, playing a neuroprotective role [8, 28]. DJ-1 might act as a redox-sensor protein, which can prevent the aggregation of α-synuclein or as a reactive oxygen species scavenger through auto-oxidation, behaving as

an antioxidant [33, 38-43]. These proposed functions for DJ-1 could be particularly important in nigral dopamine neurons that are exposed to particularly high levels of oxidative stress [3].

Though the link between DJ-1 and oxidative stress may have been as early as 2001, it was not until 2003 that a genetic linkage to PD was established [32]. *DJ-1* is the third gene associated with early-onset autosomal recessive PD, and it is mutated in about 1–2% of all EOPD cases, therefore it is the least common of the known causes of autosomal recessive parkinsonism [8, 29]. Given that *DJ-1*-linked PD seems to be rare and very few patients have been reported in the literature, there is limited knowledge on the clinical features, neuropathology, and genotype-phenotype correlation for *DJ-1*-related PD [3, 6, 14, 29, 44, 45]. Although *DJ-1* mutations are rare even in early onset PD, recent studies suggest that DJ-1 protein might play an important role in sporadic late-onset PD [3]. Because of the rarity of mutations in *DJ-1*, most studies have not analyzed their PD patient groups, making it highly likely that putative pathogenic mutations have been missed and that the current mutation frequency of *DJ-1* is an underestimate [15].

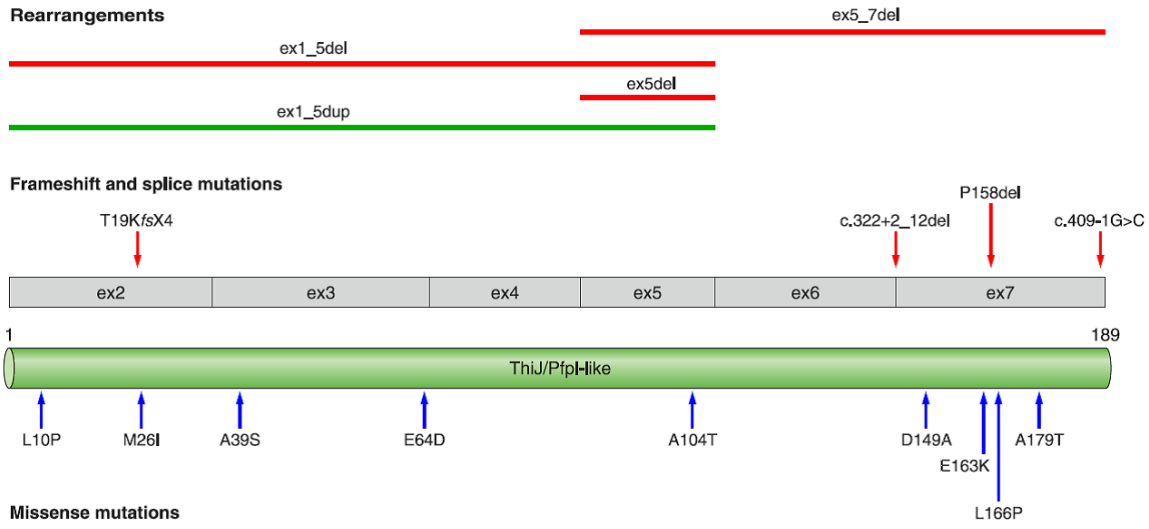


Figure 1.9 - Schematic representation of *DJ-1* and the functional domains of the DJ-1 protein. Pathogenic frameshift mutations are indicated above the transcript and protein organization and missense mutations below the protein; rare deletions (red lines) and duplications (green lines) (above) are also represented [8].

The onset age for *DJ-1*-related PD ranges from 20 to 40 years [41, 44, 45]. The *DJ-1*-related phenotype seems to be comparable though to other forms of autosomal

recessive parkinsonism (*Parkin*- and *PINK1*-related PD) [6]. Patients with *DJ-1*-linked parkinsonism have shown a slow disease progression, dystonic features, including blepharospasm and laterocollis, brisk tendon reflexes, dementia with behavioral abnormalities, amyotrophic lateral sclerosis, dystonic features, psychiatric abnormalities, short statue, and brachydactyly [3, 6, 32, 46, 47].

Mutation analyses in patients identified homozygous, compound heterozygous as well as heterozygous missense mutations in coding and promoter regions, splice site mutations and exonic deletions, that collectively appear to result in loss of protein stability and subsequent function of the DJ-1 protein [8, 32, 41, 45, 48]. To date, about 15 different point mutations and exonic deletions have been described, mostly in the homozygous or compound-heterozygous state [8]. A large homozygous deletion of *DJ-1* exons 1-5 and a homozygous missense mutation, L166P, were first identified in a Dutch and an Italian family, respectively [32]. Heterozygous *DJ-1* missense mutations have also been reported in PD patients, but their role in PD pathogenesis is not clear [41, 48-50].

The mutated proteins are frequently not properly folded, unstable, and quickly degraded by the proteasome [14]. Thus, their neuroprotective function and antioxidant activity are reduced [51, 52].

1.1.3 Parkinson's Disease Pathogenesis

The pathogenic mechanisms that underlie Parkinson's disease remain unknown. However, it is known that a cascade of events leading to cell death contributes for the pathogenesis of PD [13]. This cascade includes oxidative stress, impaired mitochondrial function, protein misfolding and aggregation due to ubiquitin–proteasomal system (UPS) dysfunction, impaired lysosome and chaperone-mediated autophagy and other pathogenic dysfunctions [2, 7, 13].

The pathogenic factors cited above are not mutually exclusive, and one of the key aims of current PD research is to elucidate the sequence in which they act and whether points of interaction between these pathways are key to the death of SNpc dopaminergic neurons [53].

The understanding of PD pathogenesis has been greatly advanced by studies of toxic PD models and by genetic studies of disease-causing mutations in cell and animal

models [28]. Studies of genes linked to autosomal recessive PD, namely *Parkin, PINK1* and *DJ-1*, demonstrate that mitochondrial dysfunction, oxidative stress and an unbalance in protein homeostasis (characterized by an increase in protein misfolding and aggregation accompanied by an impaired removal of misfolded proteins due to impairment of the UPS) may represent the principal molecular pathways that commonly lead to neuronal degeneration and clinical PD [4, 21, 22, 28, 54]. These processes are being intensively examined, partly in the hope that they will shed light on PD [22].

1.1.3.1 Mitochondrial Impairment

Mitochondrial homeostasis is essential for maintaining neuronal function, such as neuronal signaling, plasticity and transmitter release, and there are different pathways to maintain its homeostasis [55].

The dysfunction of mitochondria has been implicated in the pathogenesis of a wide range of neurodegenerative diseases, including PD [22, 56, 57]. However, the link between mitochondrial dysfunction and PD is not yet clearly elucidated, and therefore further studies are necessary to elucidate the origin of the mitochondrial dysfunction found in PD [58].

The idea that mitochondrial dysfunction may be central to PD emerged with the discovery of the mechanism of action of the environmental toxin MPTP [4, 8, 21]. Environmental toxins, such as MPTP and rotenone, inhibit complex I of the mitochondrial electron transport chain (mETC) and its administration *in vivo* mimics the pathological and behavioral hallmarks of PD and *in vitro* alters mitochondrial dynamics [21, 22, 59]. A mitochondrial complex I defect could contribute to neuronal degeneration and lead to cell damage caused by free radicals generated directly at this site or by way of a compensatory increase in respiration at complex II as well as to apoptosis through decreased ATP synthesis and a bioenergetic defect [19, 53, 54].

Complex I activity were reported to be reduced, in the range of 20-30%, in post-mortem *substantia nigra*, frontal cortex and platelets of PD patients [4, 8, 21]. Mitochondrial dysfunction may also affect other peripheral tissues, including skeletal muscle, and other non-neuronal tissues. Further, it may not be restricted to complex I [4,

8]. Apart from dopaminergic endotoxins there are several other explanations proposed in the literature for a decrease in complex I activity in PD [21].

A recent meta-analysis that combined gene expression data from numerous PD studies found that the expression of ten sets of genes differs between PD and control DAergic SNpc cells [21]. Another possibility that could underlie the mitochondrial defects seen in PD is the accumulation of point mutations and deletions in complex I genes in the mitochondrial DNA (mtDNA) [21, 56].

A host of studies have addressed the functions of proteins linked to PD in maintaining mitochondria homeostasis in cultured cells, flies and mice, which further supports a causal relationship between abnormal mitochondrial homeostasis and PD [55, 57].

Increased PD risk has been linked to mutations in *SNCA*, *LRRK2, Parkin, PINK1,* and *DJ-1* genes, and the proteins encoded by these genes have been, directly or indirectly, implicated in maintaining mitochondrial homeostasis, including membrane potential, calcium (Ca^{2+}) homeostasis, cristae structure, respiratory activity, mtDNA integrity and autophagy-dependent scavenging of dysfunctional mitochondria [55-58].

Current evidence suggests a connection between α-synuclein and mitochondria. Complex I inhibition both *in vitro* and *in vivo* leads to the accumulation of LB-like α-synuclein-positive inclusions, which suggests that α-synuclein aggregation may be a consequence of mitochondrial dysfunction and might be an effector of neuronal cell death [54]. Nevertheless, whether mitochondrial dysfunction is the cause or effect of protein aggregation in PD is still controversial [58].

PINK1 is the first gene to directly link mitochondria to PD, and it can partially protect against mitochondrial dysfunction induced by proteasome inhibition [54]. In *Drosophila* models of PINK1, several studies strongly suggested that PINK1 and Parkin act in a common pathway that influences mitochondrial integrity in a subset of tissues [22]. A coherent hypothesis is that these two proteins might act directly at the mitochondria, through their respective phosphorylation or ubiquitination activities [22].

There is less evidence for a direct role of DJ-1 in mitochondrial function. Nevertheless, the fact that neurons with reduced levels of endogenous DJ-1 and

Drosophila DJ-1 mutants exhibit increased sensitivity to environmental mitochondrial toxins associated with PD may indicate a role for DJ-1 in mitochondrial function [22]. Further, in human dopaminergic neuroblastoma cells, DJ-1 knock-down leads to mitochondrial depolarization and fragmentation, also suggesting that DJ-1 is important for maintaining the integrity and function of the mitochondria [60, 61]. DJ-1 is necessary for mitochondrial complex I to exert its enzyme activity but DJ-1 alone is not sufficient to maintain activity of complex I [62]. Since DJ-1 possesses no mitochondrial import signal, it is thought that DJ-1 is translocated into complex I in association with another protein(s) that possesses mitochondrial import signals [62]. The DJ-1 protein has been recently proposed to act in parallel with PINK1 and Parkin to control mitochondrial function, morphology, and mitophagy [58].

Additional studies are needed to probe the molecular bases of DJ-1 protection against mitochondrial dysfunction and cell death and to examine the relationship between DJ-1, PINK1, and Parkin in the context of PD pathogenesis [28].

Although it remains to be shown whether mitochondrial alterations in PD constitute a primary or a secondary event, or are just part of a larger multifactorial pathogenic process, both toxin models and genetic links of PD are providing clues to the interplay of mitochondrial function, mitochondrial dynamics, and PD pathogenesis. The targeting of mitochondrial dysfunction holds promise for the development of novel therapeutic strategies aimed at stopping or slowing down the progression of dopaminergic neurodegeneration in this currently incurable neurodegenerative disorder [59, 63].

1.1.3.2 Oxidative Stress

Oxidative stress is considered to compromise the integrity of vulnerable neurons and thus to contribute to several neurodegenerative disorders [54, 64]. It is well accepted that the redox imbalance in PD is not a secondary end-stage epiphenomenon but a driving force of the onset and progression of the disease [64].

The source of oxidative stress, which could develop in the SNpc, is unclear, but may include mitochondrial dysfunction; increased dopamine metabolism that can yield excess hydrogen peroxide (H_2O_2) and other reactive oxygen species; a deficiency in the

antioxidant glutathione, thereby diminishing the capacity of the brain to remove H_2O_2; and an increase in reactive iron, which can generate highly toxic hydroxyl radicals ($OH^{\cdot}$) via interaction with H_2O_2 [19, 21, 54].

Indeed, post-mortem studies in PD brains demonstrate increased iron, decreased antioxidant GSH, and high levels of oxidative damage to lipids, proteins, and DNA, suggesting that the SNpc is in a state of oxidant stress [19, 21, 22, 54, 64]. Therefore, if ROS are not efficiently eliminated by cellular antioxidants, the cellular components become damaged by oxidation, which in turn leads to further oxidative stress, cellular dysfunction, and even cell death [58, 65].

Dopaminergic neurons in the *substantia nigra* have high levels of pro-oxidant iron, that can promote ROS production, and low levels of glutathione, hence they are particularly vulnerable to oxidative stress and oxidative stress-induced somatic mtDNA mutations [56].

Mitochondria consumes about 85–95% of the oxygen inspired during respiration, and powerful oxidants are normally produced as by-products, including H_2O_2 [53, 58]. Thus, the electron transport chain in mitochondria is a major source of ROS in eukaryotic cells [4, 21, 64]. Increased formation of mitochondrial ROS and/or defective ROS removal by mitochondrial defense systems results in oxidative damage to mtDNA, proteins and lipids and perturbs redox signaling pathways [4]. Oxidative damage to mtDNA may compromise respiratory chain subunits encoded by mtDNA, thereby establishing a vicious circle of oxidative stress and bioenergetics failure [4, 53]. Therefore, mitochondrial dysfunction, more specifically inhibition of complex I, is probably the leading source of increased oxidative and nitrosative stress observed in the brain of PD patients [53, 54, 58].

Mitochondria-related energy failure may disrupt vesicular storage of DA, causing the free cytosolic concentration of DA to rise and allowing harmful DA-mediated reactions to damage cellular macromolecules [53, 64]. Thus, DA may be essential to make SNpc dopaminergic neurons particularly susceptible to oxidative stress [53].

Glutathione depletion is one of the earliest oxidative signs detected in the course of PD and indicates that oxidative damage can occur even before complex I deficiency [58].

Recent evidence has revealed that ROS are not simply subproducts of mitochondrial metabolism. Under normal conditions they have important physiological functions, as signaling molecules in diverse cellular pathways [4, 21, 58]. Moreover, a mild increase in ROS formation may activate protective stress response pathways [4]. Hence, the common paradigm that mitochondrial dysfunction is intimately linked to increased oxidative stress which promotes neurodegenerative diseases is not unequivocally accepted [4].

Oxidative stress appears to provide a critical link between exposure to environmental factors, such as drugs and pesticides or heavy metals, and genetic factors predisposing to PD [64]. Environmental toxins, such as paraquat, rotenone, and MPTP, are capable of generating reactive intermediates with the ability to directly react with biological macromolecules in processes such as thiol alkylation, carbonylation, nitration, and lipid peroxidation [64, 65].

Among several genes mutated in familial PD, only *DJ-1* plays a direct role in oxidative defense mechanisms of SNpc [64]. Oxidative stress causes an acidic shift in the isoelectric point of DJ-1 suggesting self-oxidation, therefore it functions as an antioxidant [22, 33, 43]. Embryonic stem cells deficient in DJ-1 display increased sensitivity to oxidative stress [66]. Perhaps, as a secondary consequence of increased ROS production, DJ-1 regulates mitochondrial morphology and function and mitigates cell death through direct scavenging of H_2O_2, because DJ-1 has peroxiredoxin-like peroxidase activity [57, 67].

Other PD-related gene products, like Parkin and PINK1, may indirectly control oxidative status of the cell by removing dysfunctional mitochondria, that if not removed may in turn lead to an increase in cellular ROS levels and to the bioenergetics defects frequently observed in PD [58, 64].

Future analysis of post-mortem brain tissue from monogenic forms of PD and genetic animal models, including DJ-1 or PINK1 knockout mice, may help to clarify further

the molecular pathway linking PD with mitochondrial dysfunction and oxidative stress [54].

1.2 Role of DJ-1

DJ-1 is an evolutionarily conserved, 189-residue protein with approximately 20 kDa containing eight α-helices (α1-α8) and eleven β-strands (β1-β11), adopting a helix-strand-helix sandwich structure [52]. It is a member of the DJ-1/ThiJ/PfpI superfamily of proteins, which are conserved in a wide variety of organisms [52].

DJ-1 protein forms a dimeric structure under physiologic conditions. Although most of the residues involved in the dimerization are highly conserved, DJ-1 and its closest homologues share a peculiar dimerization pattern in the superfamily, which is partially determined by the presence of an additional C-terminal helix [68, 69]. A putative active site has been identified close to the dimer interface, with some similarities to the active site of cysteine proteases and the residues C106, H126, and E18 being likely involved [68].

DJ-1 is ubiquitously and highly expressed, particularly in liver, skeletal muscle, kidney, pancreas, testis and heart. It is detected at slightly lower levels in placenta and brain. Moreover, it can be detected in astrocytes, Sertoli cells, spermatogonia, spermatids and spermatozoa. In the brain, its expression is also ubiquitous, with higher levels of the transcript in subcortical regions, such as the caudate nucleus, the thalamus, the *substantia nigra*, and the hippocampus, that are more affected in PD [32, 35, 37].

Under normal conditions, DJ-1 is located predominantly in the cytoplasm and, to a lesser extent, in the nucleus and mitochondria. In response to oxidative stress, DJ-1 translocates to the mitochondria and also to the nucleus and exerts an increased cytoprotective effect against oxidative damage [33, 70].

The exact function of DJ-1 remains unknown. However, it has been involved in processes as different as cell cycle regulation and oncogenesis, male fertility, control of gene transcription, regulation of mRNA stability, and response to cell stress [68]. Functional studies and mass spectrometry indicated that DJ-1 is an atypical peroxiredoxin-like peroxidase that scavenges reactive oxygen species through oxidation

of C106 to cysteine-sulfinic acid (C106-SO_2^-), behaving as an antioxidant [67, 71]. This modification plays a key role in regulating the protective function of DJ-1 [72, 73]. Cysteine can form three different species: cysteine-sulfenic (-SOH), -sulfinic (-SO_2H), and –sulfonic (-SO_3H) acid [72]. Of the three oxidation states of cysteine, only cysteine-sulfenic acid is readily reduced to the thiol under physiological conditions [72]. Because cysteine can be oxidized to three distinct species, each with different structural and chemical properties, cysteine oxidation is a way for reactive oxygen species to alter the DJ-1 activity, and therefore DJ-1 has special importance in understanding the role of regulatory cysteine oxidation in neuronal survival [72]. In this way, C106 controls the neuroprotective function of DJ-1 [33, 74, 75]. The oxidation of C106 to sulfinic acid is essential for the ability of DJ-1 to respond to oxidative stress by binding to mitochondria or by suppressing α-synuclein fibrillization via a chaperone mechanism [76, 77].

It was also shown that DJ-1 undergoes cleavage of a C-terminal peptide that leads to activation of protease activity in response to oxidative stress and enhanced cytoprotective action against oxidative stress-induced apoptosis [78]. In addition, it was shown that DJ-1 may also function as a redox-sensitive chaperone [79]. Therefore, *DJ-1* mutations in autosomal recessive early-onset PD lead to loss of the DJ-1 function contributing to neurodegeneration [68].

DJ-1 might be involved in the cellular stress response at multiple levels. It might directly react to stress signals, as redox changes and misfolded proteins, by a chemical shift and/or change in multimerization state. It might also modulate the gene-expression of the stress response at post-transcriptional level, by the known interaction with RNA-binding protein complexes [80]. Moreover, DJ-1 might translocate to the nucleus in response to stress signals, where it might interact with PIASx-alfa, a modulator of the nuclear androgen receptor, or other co-factors, and modulate the gene expression at the transcriptional level [68, 81].

DJ-1 protects against neuronal oxidative stress, and although loss of DJ-1 alone may not be sufficient to produce parkinsonism, it may confer hypersensitivity to dopaminergic insults when challenged [82].

A more detailed understanding of the mechanism by which DJ-1 can protect against stresses is still needed to better understand the activities of this protein and its role in neurodegenerative Parkinson's disease [83].

1.3 *DJ-1* Mutations

Mutations in the DJ-1 protein are associated with a recessive form of early onset familial PD [84]. Despite the rare incidence of *DJ-1* mutations in PD, the study of DJ-1 biology can provide important clues to altered cellular pathways in PD. Thus, understanding how the causative *DJ-1* mutations interfere with the structure, function, and localization of DJ-1 protein is of critical importance [84].

For PD-linked *DJ-1* mutations that lead to changes in the amino acid sequence of the intact protein, functional effects are likely to be fundamentally mediated by associated changes in the physicochemical properties of the mutant proteins [51]. In general, mutations can interfere with protein function through a variety of mechanisms [51]. At one extreme, mutations can lead to global destabilization and unfolding of a protein, which is usually associated with complete loss of function, at least for proteins that require a well folded conformation to carry out their biological activities [51]. At the other extreme, mutations can lead to subtle changes on the protein's surface that can modulate its interactions with binding partners in ways that compromise its function without perturbing its structure [51]. Intermediate effects are also possible, where the conformation of a protein can be locally (or globally) altered in ways that do not lead to complete global unfolding but nevertheless perturb structural (or dynamic) properties that are crucial for function [51]. The known PD-linked *DJ-1* mutations cover the entire range of potential effects on the structure of the protein [51].

To date, a variety of pathogenic homozygous, compound heterozygous and heterozygous *DJ-1* mutations have been reported (Table 1.2), including a 14 kb deletion englobing the first five exons, as well as the L166P, M26I, E163K, A104T, and E64D missense mutations [85, 86]. Among these, homozygous and compound heterozygous mutations may trigger a loss-of-function, the presumed cause of their pathogenicity, and

they are clearly associated with early onset PD, while it is unclear if heterozygous mutations are causative of PD [84, 86]. *DJ-1* point mutations are distributed throughout the DJ-1 structure [91]. Therefore, even minor perturbations to the structure of DJ-1 can result in impairment of function sufficient to cause disease [91].

Table 1.2 – **Mutations occurring in the *DJ-1* gene** [45, 47-49, 84, 87-90].

Mutation	Inheritance	Average Age of disease onset	Ethnic group affected	Effect	Year reported
14 kb deletion	Homozygous	32.6	Dutch	Loss of protein	2003
L166P	Homozygous	30	Italian	Protein instability	2003
M26I	Homozygous	39	Ashkenazi Jewish	Decreased stability	2003
A104T	Heterozygous	35	Latino	Unknown	2003
D149A **G78G**	Compound Heterozygous	36	Afro-Caribbean	Unknown	2003
R98Q	Heterozygous	<50	Global	Polymorphism	2003
c.56delC **c.57G→A** **IVS6-1 G-C**	Compound Heterozygous	24	Hispanic	Frame shift Altered transcript	2003
E64D	Homozygous	34	Turkish	Unknown	2004
E163K	Homozygous	31.6	Italian	Altered activity	2005
L10P	Homozygous	18.5	Chinese	Decreased stability and impaired homodimer formation	2008
P158DEL	Homozygous	33.8	Dutch	Decreased stability and impaired homodimer formation	2009
A179T	Heterozygous	47	Dutch	Unknown	2009
Exon1-5 duplication	Heterozygous	46	Dutch	Unknown	2009
g.168_185 deletion	Heterozygous		Global	Polymorphism	2003
g.168_185 duplication	Homozygous	31.6	Italian	Unknown	2005
c.253-322 deletion	Heterozygous	45	Serbian	Unknown	2004
Exon 5-7 deletion	Heterozygous		Italian	Altered transcript	2004
IVS5+2-12 deletion	Heterozygous		Russian	Altered transcript	2004
A107P **IVS2-2 A-G**		<31	Iranian	Unknown	2010
A171S	Heterozygous	68	African American	Unknown	2004

The L166P, E64D, M26I, A104T, and D149A mutations have been shown to create structural perturbations of DJ-1 protein that lead to global destabilization, unfolding of the protein structure, heterodimer formation, or reduced antioxidant activity [51, 52, 92]. Some *DJ-1* mutations associated with disease or artificial mutants deficient in oxidative signaling have the common effect of being unable to protect cells against oxidative damage [93].

Despite notable progress made in the characterization of some pathological mutations in *DJ-1*, detailed studies of the folding and conformational dynamics of DJ-1, as well as, a detailed structural explanation for the observed deleterious effects of disease-associated missense mutations on *DJ-1* are still lacking [91]. So, it will be necessary to further characterize in greater detail the structure and dynamics of the mutants, and to characterize more fully their functional effects *in situ* and *in vivo* [51]. A detailed biophysical characterization of DJ-1 should be done to realize the impact of disease-associated mutations on the protective function of the protein, as well as, an improved understanding of how DJ-1 inactivation takes place in, and contributes to PD should emerge [51, 91].

1.3.1 Natural Mutations

This study will focus on three naturally occurring homozygous PD-linked *DJ-1* mutations: L166P, M26I and E163K, further described in the following subsections.

1.3.1.1 Leu166Pro (L166P)

In a consanguineous Italian family with autosomal recessive early-onset Parkinson disease, it was identified a homozygous 497T-C transition in exon 7 of the *DJ-1* gene, resulting in the substitution of a highly conserved leucine residue for a proline at position 166 (L166P) of the DJ-1 protein [32]. The clinical phenotype was characterized by a slow disease progression with additional behavioral and psychic disturbances, and dystonic features, including blepharospasm [32].

L166P is one of the most extensively investigated and deleterious DJ-1 polymorphs [52]. L166 is located in the center of α-helix 7 of the protein, near the dimer interface [34]. This helix is part of a hydrophobic core formed by three helices (two contributed by

the C-terminal and one by the N-terminal part of the monomer), which is involved in the dimerization of the protein [68]. The L166P substitution significantly destabilizes the dimer interface, interrupting over 100 intermolecular contacts that are important for dimer formation [52]. Therefore, this mutation is predicted to lead to the unfolding of the C-terminal domain of DJ-1 as well as to disrupt dimer formation due to the strong helix breaking properties of the substituted proline [34, 52, 68, 93].

In vitro, this mutation leads to a drastic reduction in the secondary structure content critical for protein stability and dimerization, namely α-helical content, suggesting that the structure of the mutant protein is globally and severely perturbed [51]. L166P also disrupts α1, α5, α6 and α8 helices with α8 undergoing particularly severe disruption [52].

The L166P substitution also led to major perturbations in the region of a highly conserved cysteine residue (C106) that participates in dimerization and that is critical for a proposed chaperone function of DJ-1 [52]. Indeed, L166P has been shown to abrogate DJ-1 chaperone activity, as a consequence of defective dimerization and reduced stability [51, 52]. C106 is located nearly to the residue 166 and the N-terminal region of α1, a key mediator of dimer formation, demonstrating that structural disruptions propagate throughout the whole protein [52]. These findings indicate that the structural effects of the L166P substitution in DJ-1 are not confined to the vicinity of the substitution, but propagate rapidly throughout the protein [52].

The DJ-1 L166P is incapable of forming a stable homodimer with itself and a heterodimer with DJ-1 WT, and since helices α7 and α8 are known to help mediate dimerization, the failure to form a dimer in solution is consistent with the finding that L166P perturbs the structure of one or both of these helices [52].

Because L166P mutant does not appear to exist as homodimers in solution, contrary to the native DJ-1, but only as rapidly degraded monomers, it is suggested that protein destabilization accounts for the dysfunction of this mutant *in vitro* and *in vivo* [69, 76, 93, 94].

In cell culture, the L166P mutant is drastically unstable. It is expressed at much lower steady-state levels than the protein wild-type and exhibits a greatly increased

turnover rate [69, 84, 86, 93, 95, 96]. The structural perturbation of DJ-1 induced by the L166P substitution causes the protein to be ubiquitinated, implying a role for the ubiquitin-proteasome system in the turnover of this mutant, although degradation by non-proteasomal pathways are likely involved as well, significantly reducing the half-life of the protein *in vivo.* However, these effects are partially, but not fully abrogated by proteasome inhibitors [51, 52].

The L166P mutation severely perturbs DJ-1 protein structure, resulting in the formation of a spontaneously unfolded protein [84]. DJ-1 L166P also forms high molecular-weight complexes containing DJ-1 oligomers or aggregates with other proteins in cell culture, probably because of its inability to adopt a stable, dimeric structure [52, 76]. Structural disruptions of DJ-1 L166P are associated with a large increase in the total hydrophobic surface area relative to protein wild-type, possibly explaining the tendency of the mutant protein to aggregate [52].

The L166P mutant appears to adopt an altered cytoplasmic distribution. It has shown reduced nuclear localization, since it partially mislocalizes or is sequestered to mitochondria, suggesting a possible link to oxidative stress [32, 69, 80, 94].

A number of studies also demonstrated that the pathogenic DJ-1 L166P mutant insufficiently protects against oxidative stress, since it has a reduced antioxidative activity [83, 92]. Indeed, DJ-1 L166P is ineffective in eliminating H_2O_2 within cells or protecting cells from death induced by H_2O_2 [66, 82, 91, 92, 95]. Therefore, probably consequent to instability, L166P reduces the neuroprotective function of DJ-1 [43]. It was also observed that the L166P mutation in *DJ-1* originates fragmented mitochondria and elevated markers of autophagy [71].

Although this severe destabilization of the L166P leads to an increased rate of degradation *in vivo*, the primary cause of the loss of function associated with this mutant is the loss of the protein's native structure, therefore the rapid degradation of the protein *in vivo* is a secondary effect [51].

1.3.1.2 Met26Ile (M26I)

In an Ashkenazi Jewish patient with early-onset Parkinson's disease, it was identified a homozygous 78G-A transition in exon 2 of the *DJ-1* gene, resulting in a

methionine to isoleucine substitution at position 26 (M26I) of the DJ-1 protein [41]. The clinical phenotype was characterized by right-sided rigidity, bradykinesia, classical pill rolling rest tremor, dystonia, dyskinesia and psychological disturbance, particularly anxiety [41].

M26 is a conserved residue located in α-helix 1, which is a central feature of the DJ-1 dimer interface [76]. Thus, a perturbation of structure in this helix disrupts the contacts at this interface and leads to the decrease in the stability of the dimer [76]. Since M26I resides in the core of the protein it must exert its pathogenic effect by directly altering the properties of DJ-1 [91].

The M26I mutation does not result in a significant change in side chain conformation at the site of the substitution [91]. The modest disruption of hydrophobic packing contacts around I26 is a consequence of introducing a β-branched amino acid in the tightly packed hydrophobic core of the protein [91]. In addition, the M26I substitution creates a small cavity with consequent loss of optimal packing contacts in the interior of the protein due to the loss of the Cε and Sδ atoms of M26 [91].This cavity permits greater conformational flexibility in the core of DJ-1 and likely contributes to the modest thermal destabilization of DJ-1 M26I [91].

In addition, the M26I mutant has a decrease in secondary structure that might be expected to result from an alteration of the C terminus of α-helix 1 and the N-terminal regions of the three affected β-strands, as well as an alteration in the C-terminal region of α-helix 7 and a part of helix 8 [51].

The M26I mutant is able to adopt the dimeric wild-type fold, although less stable than the wild-type [84, 86]. The instability of these dimers leads to aggregation/precipitation as well as significantly increased turnover rate of this mutant in cells relative to all known DJ-1 mutants, except the highly unfolded L166P variant, which probably may be the fundamental cause of the loss of function of this DJ-1 mutant [51, 91, 93]. In addition, the expression levels of the M26I mutant are dramatically decreased in cell lines, though to a lesser degree than the L166P mutant [84].

However, there are conflicting data regarding the stability of the M26I mutant. While some studies showed that DJ-1 M26I is well-structured in solution and in the

crystal, another studies showed that DJ- 1 M26I has reduced secondary structural content and displays a pronounced tendency to aggregate [91].

As there is no direct structural perturbation near C106, this mutant may be functional. However, chemical shift changes in charged residues near M26 may reflect alterations that influence C106 oxidation and thereby underlie the reported effects of this mutation on the stability of the oxidized form of the protein [76]. It is known that the M26I mutation may preferentially destabilize more extensively oxidized forms of DJ-1, namely DJ-1 containing cysteine-sulfonic acid and methionine sulfoxide, thereby enhancing the loss of properly folded DJ-1 under conditions of cellular oxidative stress [91].

The M26I mutant also displays reduced protection against oxidation stress despite the ability to form stable dimers [91]. However, it is unclear if the reduced protection against oxidative stress by the M26I mutation is simply associated with reduced half-life or perhaps with an increased propensity for the M26I mutant protein to aggregate [83].

1.3.1.3 Glu163Lys (E163K)

In 3 affected sibs from a consanguineous southern Italian family with early-onset parkinsonism, it was identified a homozygous 3385G-A transition in exon 7 of the *DJ1* gene, resulting in the substitution of a highly conserved glutamic acid by a lysine at position 163 of the DJ-1 protein [47]. Patients develop parkinsonism, dementia, severe amyotrophic lateral sclerosis, weakness and muscle atrophy, speech deficits, and cognitive impairment [47].

Very little is known about the effects of the E163K point mutation on *DJ-1* [91]. E163 is in close proximity of L166, and both residues are located in α-helix 7, which is one of the helices in the dimer interface of DJ-1 and it is critical in forming stably folded protein [83]. However, unlike L166P, this mutation remains the ability of DJ-1 to dimerize [83]. The E163K mutation clearly does not cause the dramatic structural changes of the L166P mutation, but the fact that E163 is highly conserved in DJ-1 across species accentuates the importance of this residue [83].

It has been previously shown that some pathogenic mutants of DJ-1 like M26I are structurally similar to DJ-1 WT in the ability to form a homodimer, but can demonstrate

reduced stability compared to the WT protein [83]. However, the DJ-1 E163K mutant shows comparable stability to WT and also does not lead to changes in solubility [83].

DJ-1 WT as well as pathogenic mutants may localize to the mitochondria and its localization may increase in response to mitochondrial stress [83]. However, the DJ-1 E163K does not alter its subcellular localization [83].

Many studies demonstrate that DJ-1 influences several neuroprotective pathways but that the E163K mutation specifically impairs the oxidative stress protective mechanism and even increases sensitivity to oxidative insult both in cell culture and in animal models [83]. However the E163K mutant retains the ability to protect N2A cells against proteasome inhibition as well as mitochondrial stress through mitochondrial complex I and III inhibition [83].

Under oxidative stress, human DJ-1 WT in N2A cells can relocate in close proximity to the mitochondria, but does not enter these organelles, while DJ-1 E163K is impaired in this property, which may play a role in its inability to protect against oxidative stress [83]. It is possible that simply the change from a negatively to positively charged residue and/or the subtle structural effects of this mutation may prevent the interaction of DJ-1 with other proteins that may be involved in the redistribution of DJ-1 under conditions of oxidative stress [83].

In DJ-1 WT, the carboxylate side chain of E163 in α-helix 7 makes a salt bridge with the guanidinium side chain of R145, which in turn links the C-terminus of the other DJ-1 monomer at the dimer interface by donating two hydrogen bonds to the peptide carbonyl oxygen of V186 [91].

The very high degree of conservation of both E163 and R145 and the electrostatically conservative character of sequence variations at these two positions in homologues of DJ-1 suggests that this salt bridge interaction is likely important [91].

The crystal structure of DJ-1 E163K shows that substitution of glutamic acid for lysine disrupts this salt bridge and results in increased mobility of R145 due to repulsive electrostatic interactions and loss of hydrogen bonding potential between K163 and R145 as well as consequently interferes with a network of hydrogen bonds that involves L186 and spans the dimer interface [91].

The salt bridge between R145 and E163 in DJ-1 WT is stabilized by both charge complementarity and hydrogen bonding and the E163K mutation eliminates both types of interaction, leading to destabilization of the protein [91].

The loss of DJ-1 protective function can occur without evident biochemical changes on the protein and demonstrate that alterations to specific residues in this protein can specifically affect individual functions indicating that DJ-1 is likely involved in multiple cellular pathways [83]. Hence, in contrast to other mutations in *DJ-1*, the effect of the E163K mutation on impairing DJ-1 function is specific to oxidative stress [83].

The findings that this mutation appears to exacerbate the response to oxidation and diminish the redistribution of DJ-1 towards the mitochondria suggest that both processes may be related, but further studies will be needed to substantiate a direct association [83].

Because of the early onset and more extensive and diverse phenotype associated with E163K mutation, its analysis and an understanding of how it can lead to disease may provide new insights into the function of DJ-1 [83].

1.4 Metabolomics

Metabolomics is the comprehensive analysis of all metabolites in a biological system [97]. It is a rapidly evolving tool providing a quantitative assessment of low molecular weight analytes that define the metabolic status of a biological system, therefore complementing transcriptomics and proteomics [97, 98].

Metabolomic investigations have been applied extensively in various research areas including environmental and biological stress studies, functional genomics, biomarker discovery for disease diagnosis and better understanding of the pathophysiology of diseases, and integrative systems biology [97-99]. Those studies facilitate understandings of biochemical fluxes and discoveries of metabolites which are indicative of unusual biological or environmental perturbations [97].

Through the process of homeostasis, the body automatically attempts to maintain a constant internal environment, even when disease, drugs, or toxins affect concentrations and fluxes of endogenous metabolites. To maintain this constant internal

environment, increased levels of endogenous metabolites and any exogenous metabolites are eliminated [100].

Metabolic analysis is characterized by two complementary methods: targeted and untargeted. The targeted approach focuses on identifying and quantifying selected metabolites, or metabolite classes, such as substrates of an enzyme, direct products of a protein, a particular class of compound or members of a particular pathway [97]. In the targeted approach, the chemical properties of the investigated compounds are known, and sample preparation can be designed to reduce matrix effects and interference from associated compounds [97]. On the other hand, in the untargeted approach all the metabolites of a biological system are measured, if possible [97].

A typical metabolic study aims at comparing multiple biological groups to identify metabolites that are significantly altered. It starts with an untargeted analysis to screen potential and putative metabolites of interest and then subjects these metabolites to a targeted analysis for metabolite ID verification, quantitation, functional interpretation, and pathway analysis (Figure 1.10) [97].

To achieve appropriate coverage of the metabolome, several analytic platforms with complementary features may be needed. Currently, the majority of metabolomics measurements are performed using nuclear magnetic resonance (NMR) and liquid chromatography-mass spectrometry (LC-MS). However, the choice of metabolomics analytical instrumentation and software depends on the specific questions and focus of the experiment, since each type of instrument have certain strengths [101, 102].

Excellent sensitivity and resolution for a wide range of molecule types, the ability to handle a large range of concentrations (from pM to mM) for different molecular types, the ability to identify and quantify different molecules, short analysis time, to enable the measurement of many samples without sample degradation during the measurement, and reproducible measurement across different centers and in time are the general requirements for metabolomics instruments [102].

Among analytical tools for metabolomics studies, LC−MS is one of the most commonly used, because it offers good separation and accurate detection of metabolites in complex specimens with high sensitivity and resolution [98]. LC−MS has been widely

used for both comprehensive metabolomics profiling and targeted quantitation of metabolites [98]. This technique is often used to obtain the largest possible biochemical profile information subset. It is a tool that can be used to characterize, identify, and quantify a large number of compounds in a biological sample where metabolite concentrations might cover a broad range of information with regard to disease pathophysiology [101].

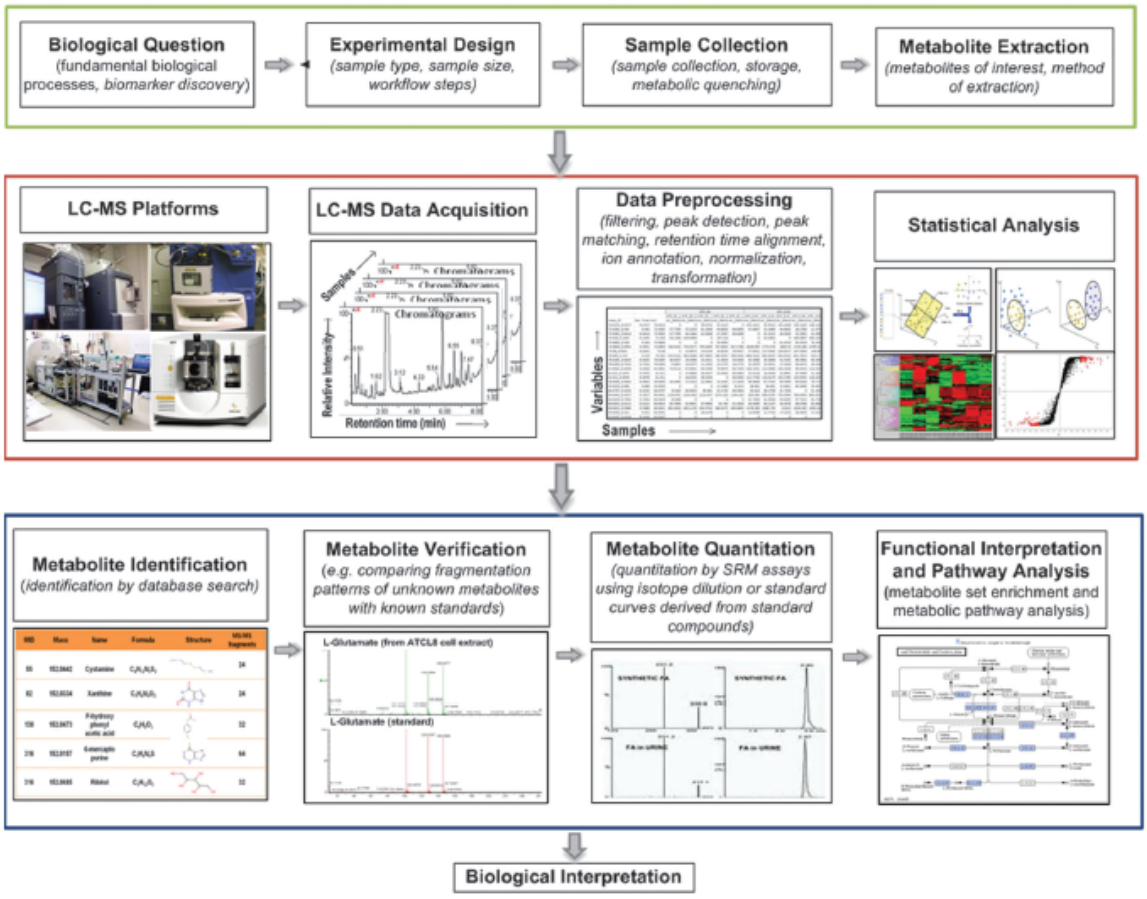


Figure 1.10 - Typical workflow of a metabolomics study using a liquid chromatography-mass spectrometry (LC-MS) based platform [97].

A mass spectrometer is typically composed of three major parts: ion source, mass analyzer, and detector. The ion source converts sample molecules into ions that will be resolved by the mass analyzer before they are measured by the detector [97]. There are several types of ion sources including electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), among others [97]. Due to the diverse chemical properties of metabolites, it is often necessary to

analyze the biological sample in both +ve (positive) and -ve (negative) ionization modes under a scan range of m/z 50–1000 to maximize metabolome coverage [97].

ESI is the method of choice in LC-MS-based metabolomics studies, because its ‘‘soft ionization’’ capability produces a large number of ions through charge exchange in solution and often forms intact molecular ions which support initial identification [97]. However, nowadays instruments with dual ionization capabilities (e.g. ESI and APCI or ESI and APPI) result in an increased coverage of the metabolome [97].

Most metabolomics studies use a separation method before mass spectrometric analysis. For instance, high performance liquid chromatography (HPLC) allows separation of compounds of a wide range of polarity [97]. Therefore, LC-ESI-MS is becoming a method of choice for profiling metabolites in complex biological samples. Chromatographic separation can reduce sample complexity and alleviate matrix effects during ionization as well as can improve the sensitivity of MS detection and also results in better MS data quality due to reduced background noise [97].

To convert the raw LC-MS data into a peak list which can be easily interpreted and compared across runs, multiple pre-processing steps must be performed, such as peak detection, peak matching and retention time alignment, normalization of peak intensities, among others, using software tools [97].

A statistical analysis must be made to detect the peaks whose intensity levels are significantly altered between distinct biological groups [97]. The statistical analysis methods can be classified as univariate and multivariate analysis. The univariate approach assesses the statistical significance of each peak separately. Commonly used univariate techniques include *t*-test, fold-change analysis, Wilcoxon rank-sum test, analysis of variance (ANOVA), among others [97]. Multivariate analysis considers the combinatorial effect of multiple variables. It can be further classified as unsupervised and supervised techniques. Unsupervised approach refers to methods that identify hidden structures in the data without knowing the class labels. One of the most popular unsupervised techniques in the LC-MS-based metabolomics study is principal component analysis (PCA). PCA has been extensively used in multiple studies to find indicative metabolites for diseases [97]. Unlike unsupervised, supervised approach uses the class label information

to construct a model to interpret the LC-MS data. Partial least square-discriminant analysis (PLS-DA) is a supervised technique widely used in LC-MS-based metabolomics data analysis [97].

One of the major challenges in metabolomics studies is the identification of metabolites. In human body, it is estimated that there are 2000 major metabolites. However, the total number of possible metabolites in nature can reach up to 1 000 000. Quantitation of metabolites can help to evaluate metabolic changes and further check the results from semi-quantitative analysis [97].

Applications of metabolomics for the study of central nervous system (CNS) diseases, including Parkinson's disease, can include additional information about mechanisms of disease, identification of prognostic, diagnostic, and substitute markers for a disease state, the ability to identify disease based on metabolic profiles, identification of biomarkers for disease, for disease progression or for response to therapy, and lastly can also include new tools in the process of drug discovery and drug development [101].

Although some promising biomarker candidates, such as antibodies against neuromelanin, pathological forms of α-synuclein, DJ-1, and patterns of gene expression, metabolomics and protein profiling have been reported there is still a lack of unique biomarkers (Table 1.3) [103].

Table 1.3 - **Biochemical biomarker candidates and their potential utilities in PD [104].**

Biochemical marker(s)	Prodromal diagnosis	Confirming PD at its onset	Differentiation from atypical parkinsonism	Monitoring progression	Sensitivity and specificity for PD diagnosis
α-Synuclein	Not likely in LRRK2 cases	Possibly useful	Possibly useful-potentially more useful if PS-129 is measured or it is used with the ratio of p-tau/tau	Possibly useful	Measuring oligomers 75% and 87.5%; improves to 89.3% and 90.6% when measuring the ratio of oligomers:total
DJ-1	Possibly useful in G2019S LRRK2 cases	Possibly useful	Unknown	Possibly useful	
Aβ	Possibly useful	Possibly useful	Unknown	Possibly useful if used in conjunction with other protein(s), for example tau, fractalkine	
Tau	Possibly useful	Possibly useful	Possibly useful if used in a p-tau/tau ratio with α-synuclein	Possibly useful if used in conjunction with Aβ	
Uric acid	Unknown	Possibly useful	Unknown	Possibly useful	
Glutathione	Unknown	Possibly useful	Unknown	Unknown	

A major challenge is to identify early biochemical changes and signatures that are unique to patients at the earliest stages of Parkinson's disease, or even before symptoms appear [101].

❷ OBJECTIVES

2. OBJECTIVES

As mentioned above, mutations in *DJ-1* gene may lead to loss of neuroprotective function of the protein. In this way it may occur a homeostatic imbalance in cell system and metabolites, which can be used as cellular markers of stress conditions.

Therefore, the aim of this study is to compare multiple biological conditions to identify the metabolites that are significantly altered in resting and oxidative stress conditions, and access also the effect of the addition of recombinant DJ-1 WT and mutants to SH-SY5Y cell line under normal and oxidative stress conditions.

In order to achieve this goal, different recombinant protein mutants production and characterization is required to access their rule in metabolite modulation. Once added to cells, an untargeted mass spectrometry analysis of metabolites can highlight potential and putative metabolites of interest. These require further validation in a targeted analysis, for metabolite ID verification, quantitation, functional interpretation, and pathway analysis.

Overall it is expected to find possible oxidative stress biomarkers, their modulation by DJ-1 and their potential use as oxidative stress markers and latter as Parkinson´s disease biomarkers. PD biomarkers are required to enhance therapeutics research and to understand PD pathogenesis, and this project aims to contribute to this field of research.

❸ METHODS

3. METHODS

3.1 Recombinant DJ-1 Mutants Production

3.1.1 Human *DJ-1* Cloning

The DJ-1_pSKB-3 construct was produced and kindly provided by Matilde Melo (Figure 3.1).

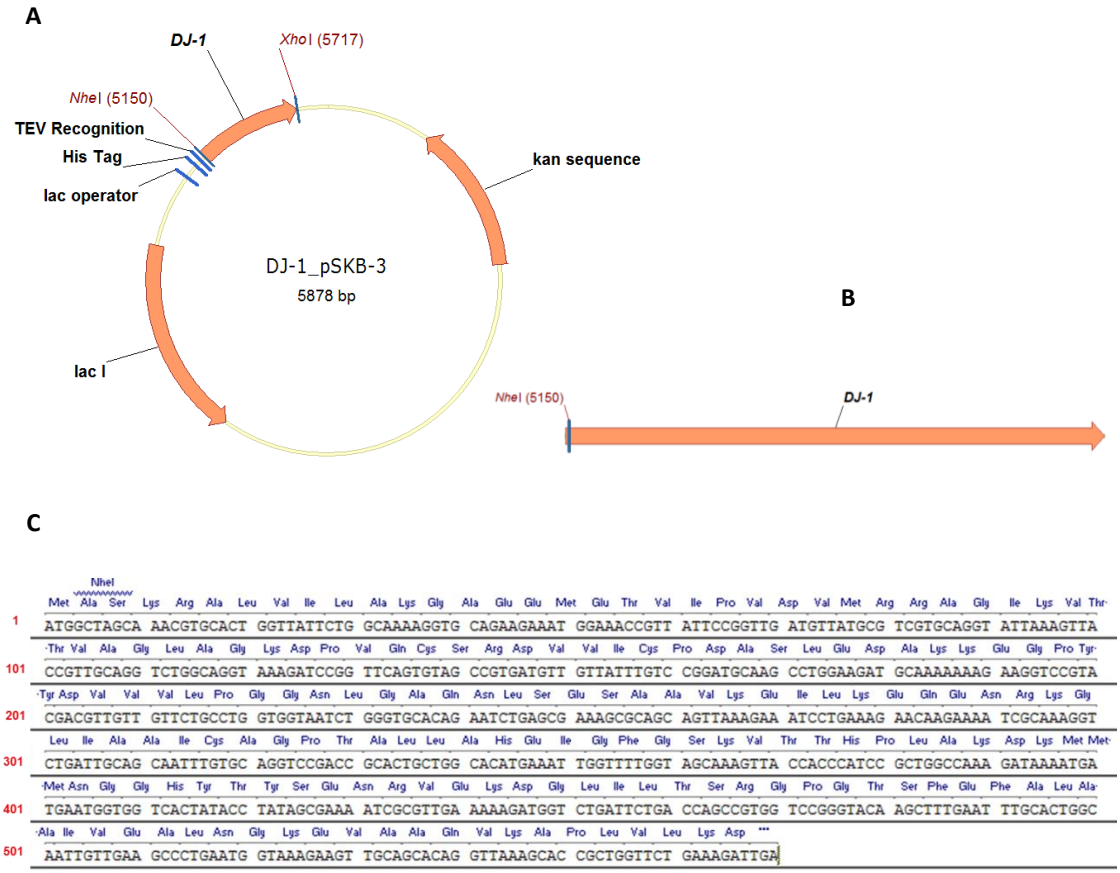


Figure 3.1 - Map of DJ-1_pSKB-3 Construct. A - Representation of the DJ-1_pSKB-3 construct which contains: lac I – lac repressor gene; His Tag – hexahistidine tag; TEV – tobaco etch virus protease recognition site; kan – kanamycin resistance gene; the *DJ-1* cDNA (optimized for *E. coli* expression) cloned between Nhe I and Xho I restriction endonuclease sites. The vector pSKB-3 corresponds to pET-28a where thrombin recognition site was modified to TEV recognition site. **B -** Representation of *DJ-1* gene. **C -** *DJ-1* sequence (570 base pair) with amino acid sequence (189 amino acids) above.

3.1.2 *In Vitro* Site-Directed Mutagenesis of DJ-1_pSKB-3 Construct

In vitro site-directed mutagenesis of DJ-1_pSKB-3 construct was performed according to the QuikChange® Site-Directed Mutagenesis protocol (Stratagene) with some modifications which are presented below.

3.1.2.1 Mutagenic Primer Design

The mutagenic oligonucleotide primers were designed individually according to the desired mutation using Vector NTI Advance® Software (Invitrogen). The following considerations were made for designing mutagenic primers: (i) both of the mutagenic primers contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid; (ii) primers are between 25 and 45 bases in length, with a melting temperature (T_m) of ≥78°C. The following equation (3.1) was used for estimating the T_m of primers:

$$T_m = 81.5 + 0.41(\%GC) - \frac{675}{N} - \%mismatch \qquad \textbf{(3.1)}$$

Where, N is the primer length in bases and values for %GC and %mismatch are whole numbers.

(iii) The desired mutation is in the middle of the primer with ~10–15 bases of correct sequence on both sides; and (iv) Primers have a minimum GC content of 40% and terminate in one or more C or G bases.

3.1.2.2 Mutant Strand Synthesis Reaction (Thermal Cycling)

Two complementary oligonucleotides containing the desired mutation were synthesized (Integrated DNA Technologies) (Table 3.1).

The sample reaction was prepared in a volume of 50 µL, in the presence of 50 ng of dsDNA template (DJ-1_pSKB-3 construct), 10 pmoles of forward and reverse primers, 2.5 Units of *Pfu* DNA Polymerase, 10 nmoles of dNTP's and 5 µL of 10x Cloned *Pfu* Reaction Buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-HCl (pH 8.8), 20 mM $MgSO_4$, 1% Triton® X-100, 1 mg/mL nuclease-free bovine serum albumin (BSA)). The PCR

was performed using the following cycling parameters: a first step of 30 seconds at 95 °C, followed by 18 cycles of 30 seconds at 95 °C, 1 minute at 55 °C, and 12 minutes at 68 °C.

Table 3.1 – Complementary oligonucleotides containing the desired mutation (yellow highlight).

	Primers	
Mutation	**Forward**	**Reverse**
L166P	5- CAA GCT TTG AAT TTG CAC CGG CAA TTG TTG AAG CCC -3	5- GGG CTT CAA CAA TTG CCG GTG CAA ATT CAA AGC TTG -3
M26I	5- GTT ATT CCG GTT GAT GTT ATT CGT CGT GCA GGT ATT AAA G -3	5- CTT TAA TAC CTG CAC GAC GAA TAA CAT CAA CCG GAA TAA C -3
E163K	5- GTC CGG GTA CAA GCT TTA AAT TTG CAC TGG CAA TTG -3	5- CAA TTG CCA GTG CAA ATT TAA AGC TTG TAC CCG GAC -3

3.1.2.3 *Dpn* I Digestion of the Amplification Products

The amplified parental (nonmutated) supercoiled dsDNA was digested with 20 Units of *Dpn* I restriction enzyme at 37 °C for 4 hours.

3.1.2.4 Transformation of XL1-Blue or DH5-α Supercompetent Cells

To perform the cell transformation 10 µL of the *Dpn* I-treated DNA (DJ-1_mutants_pSKB-3 constructs) from each sample reaction were transferred to 100 µL aliquots of XL1-Blue supercompetent *E. coli* cells or DH5-α competent *E. coli* cells. The transformation reactions were gently swirled to mix and incubated on ice for 30 minutes, then a heat pulse was performed at 42 °C for 90 seconds and the reactions were placed on ice for 2 minutes. Liquid Luria-Bertani (LB) medium (800 µL) without antibiotic was added to the transformation reactions, which were incubated at 37 °C for 2 hours with shaking at 180 rpm. The reactions were then centrifuged at 20,000*g* for 3 minutes. The supernatant was discarded (~800 µL) and the pellet was resuspended in the remaining volume (~100 µL). This volume was plated on LB/agar plates supplemented with kanamycin (50 µg/mL) and the transformation plates were incubated at 37 °C for about 16 hours. One or two colonies were picked and grown in liquid LB supplemented with kanamycin (50 µg/mL) at 37 °C with shaking at 180 rpm for about 16 hours. The DNA was

isolated using the PureLinK® Quick Plasmid Miniprep Kit (Invitrogen) and the DNA sequence was confirmed by Sanger DNA Sequencing (STAB Vida or Macrogen).

3.1.3 DJ-1 Mutants Expression

The DJ-1_mutants_pSKB-3 constructs were transformed into competent *E. coli* BL21star (DE3) strain and plated into LB/agar supplemented with 50 µg/mL kanamycin. One colony was used to inoculate 50 mL, 70 mL or 125 mL of LB supplemented with 50 µg/mL kanamycin that was grown overnight at 37 °C with shaking, from where 25 mL, 50 mL (2 x 25 mL) or 100 mL (4 x 25 mL) were used to inoculate 1 L, 2 L or 4 L of LB, respectively, supplemented with 50 µg/mL kanamycin (depending on protein construct and batch). The cells were allowed to grow at 37 °C with shaking, and the temperature was decreased to 18 °C when the optical density at 600 nm reached 0.5. One hour later the protein expression was induced by the addition of IPTG to a final concentration of 1 mM. The protein expression was allowed to occur for about 16 hours.

3.1.4 DJ-1 Mutants Purification

The cell suspension was centrifuged (20 minutes; 4,000*g*; 4 °C), the cellular pellet was suspended in 60 mL of 20 mM sodium phosphate, 500 mM NaCl, 20 mM Imidazole, pH 7.4, and disrupted through a high pressure homogenizer EmulsiFlex-C3 (AVESTIN), with 3 passages at 1000 bar. The cellular extract was clarified by centrifugation (20 minutes; 19,000*g*; 4 °C) and the supernatant (protein) was applied to a 5 mL HisTrap HP column (GE Healthcare) using a Bio-Rad BioLogic LP (Bio-Rad) low-pressure chromatography system. After column loading, the column was extensively washed with binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM Imidazole, pH 7.4) and protein elution was obtained by stepwise increasing of imidazole concentration (50, 100, 300 and 500 mM).

With exception for the first L166P mutant batch production, the fraction which contained the highest amount of DJ-1 (the most intense chromatographic peak) was applied to an HiLoad 26/600 Superdex 200 prep grade column (GE Healthcare) and the

protein was eluted by passing PBS Buffer (8 mM K_2HPO_4, 2 mM $NaH_2PO_4 \cdot H_2O$, 150 mM NaCl). For the first L166P mutant batch, the HisTrap fraction which contained the highest amount of DJ-1 was first applied to a 5 mL HiTrap Desalting column (GE Healthcare) to exchange the buffer from imidazole to PBS, and then the resulting protein sample was applied to an HiLoad 26/600 Superdex 200 prep grade column and eluted as described above for the other proteins.

Protein was concentrated (with exception of E163K mutant) using a 10 kDa MWCO centrifugal concentrator (Millipore) and quantified by measuring absorption at 280 nm using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific) [extinction coefficient of the monomeric DJ-1 mutant forms, $\varepsilon = 3.4\ (\mu g/\mu L)^{-1}\ cm^{-1}$]. Proteins were stored at about 2 μg/μL in PBS with 10% glycerol at -80 °C until further use.

The eluted fractions of affinity and gel filtration chromatographies containing DJ-1 were analyzed by SDS-PAGE (12.5% polyacrylamide gel homemade) stained with Coomassie Brilliant Blue G-250 (Bio-Rad). In each lane 10 μL of protein sample were applied to observe the protein content profile. For the first L166P mutant produced and E163K mutant the fractions containing DJ-1 were quantified by using the 2-D Quant Kit (GE Healthcare) and in each lane 15 μg of protein sample were applied.

3.2 Structural Characterization of DJ-1 and DJ-1 Mutants

DJ-1 mutant proteins produced in this project as well as DJ-1 wild-type (kindly provided by Matilde Melo) were structurally characterized through LC-MS/MS, LC-MS, HPLC-Size Exclusion Chromatography, Circular Dichroism Spectroscopy and Thermal Shift assay.

3.2.1 Liquid Digestion of Proteins

The stored DJ-1 proteins (10 μg) were resuspended in a 0.5 M triethylammonium bicarbonate (TEAB) solution to a final volume of 45 μL. Then 4 μL of 50 mM *tris*(2-carboxyethyl)phosphine (TCEP), a reducing agent, were added to the samples which were sonicated [750 W Sonicator with cuphorn (VibraCell – Sonics®)] at 20% of amplitude for 1

minute (1 second on 1 second off cycle). Cysteine residues were alkylated with 2 µL of 200 mM S-methyl methanethiosulfonate (MMTS) and samples remained at room temperature for 10 minutes. After this step, 47 µL of 0.5 M TEAB and 2 µL of 0.1 µg/µL trypsin (Roche) were added to samples which were allowed to react overnight at 37 °C. The digestion reactions were stopped with 2 µL of formic acid (FA) and peptides mixtures were concentrated by rotary evaporation, using the Concentrator Plus (Eppendorf) at 60 °C. The samples were resuspended to 100 µL with 20% acetonitrile (ACN) and 0.4% FA and sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle).

Samples were cleaned using Strong Cation Exchange (SCX) OMIX tip columns (Agilent technology). SCX tips were hydrated with four solutions: (i) 50 µL of 1 M NaCl, 20% ACN and 0.4% FA; (ii) 150 µL of 20% ACN and 0.4% FA; (iii) 150 µL of 100% ACN; and (iv) 50 µL of 1 M NaCl, 20% ACN and 0.4% FA. Then SCX tips were equilibrated with 200 µL of 20% ACN and 0.4% FA. Each sample was passed through the columns five times and washed with 200 µL of 20% ACN and 0.4% FA solution. Peptides were eluted to new tubes with 300 µL of 500 mM NH_4HCO_3 and 20% ACN. Eluates were concentrated using the Concentrator Plus at 60 °C, resuspended to 100 µL in a solution of 2% ACN and 1% FA and sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle).

Peptides were desalted using C18 Bond Elut OMIX tips (Agilent technology). Briefly, tip columns were hydrated with 200 µL of 50% ACN and equilibrated with 300 µL of 2% ACN and 1% FA. Peptides were loaded to the columns and this step was repeated five times, followed by a washing step with 100 µL of 2% ACN and 1% FA solution. Peptides were eluted to new tubes with 400 µL of 70% ACN and 0.1% FA and eluates were concentrated using the Concentrator Plus at 60 °C. Sample volume was adjusted to 30 µL in a solution of 2% ACN and 0.1% FA. Samples were sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle), centrifuged at 14,000*g* for 5 minutes and the supernatants were analyzed by LC-MS/MS.

3.2.2 In-Gel Digestion of Proteins

The stored DJ-1 samples (10 µg) were analyzed by SDS-PAGE [4-20% polyacrylamide gel (Bio-Rad)] stained either with Coomassie Brilliant Blue G-250 or silver.

In order to alkylate cysteine residues acrylamide [40% acrylamide/bis solution (37.5:1) (Bio-Rad)] was added to samples (5 µL of acrylamide per 30 µL of sample) and allowed to react for 10 minutes before running the gels. Coomassie stained gel bands were excised, sliced in small pieces, and transferred to microcentrifuge tubes with 1 mL of de-ionized water (to prevent gel bands dehydration). Gel slices were then destained by removing water and adding 1 mL of destaining solution (50 mM ammonium bicarbonate and 30% acetonitrile). The destaining step was performed in a thermomixer (comfort, Eppendorf) at 850 rpm for 15 minutes at 25 °C. The destaining solution was removed, and the process was repeated if the gel pieces remained blue. Otherwise 1 mL of water was added and the tubes were shaken in the thermomixer at 850 rpm for 10 minutes at 25 °C. After this washing step, the water was removed and the gel bands were dehydrated on Concentrador Plus for 1 hour at 60 °C. Then, enough volume of trypsin (10 ng/µL trypsin in 10 mM ammonium bicarbonate solution) was added to cover the dried gel bands and incubated for 10 minutes, on ice, in order to rehydrate the gel. After this period, 10 mM ammonium bicarbonate was added to cover gel bands again, and incubated overnight at room temperature in the dark to perform the in-gel digestion. The tryptic solution (containing trypsin and peptides) was collected to low binding microcentrifuge tubes (Eppendorf) and the remaining peptides were sequentially extracted from gel pieces by adding 200 µL of 30%, 50%, and 98% ACN in 1% FA, with agitation in the thermomixer at 1050 rpm for 15 minutes at 25 °C for each extraction solution. Each solution was collected to the tube containing the initial tryptic solution. Peptide mixtures were concentrated by rotary evaporation, using the Concentrator Plus at 60 °C, resuspended to 100 µL in a solution of 2% ACN and 1% FA and sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle).

Peptides were desalted using C18 Bond Elut OMIX tips (Agilent technology), as described in subsection 3.2.1, and concentrated using the Concentrator Plus at 60 °C. Sample volume was adjusted to 30 µL in a solution of 2% ACN and 0.1% FA. Samples were sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle), centrifuged at 14,000*g* for 5 minutes and the supernatants were analyzed by LC-MS/MS.

3.2.3 LC-MS/MS of Digested Proteins

Peptides were resolved by liquid chromatography (NanoLC Ultra 2D, Eksigent) on a ChromXP™ C18 reverse phase column (300 µm ID × 15 cm length, 3 µm particles, 120 Å pore size, Eksigent) at 5 µL/min. Peptides were eluted into the mass spectrometer with an acetonitrile gradient in 0.1% FA (5% to 35% ACN for peptides originated by liquid digestion procedure and 2% to 35% ACN for peptides generated by in-gel digestion procedure, for 24 minutes), using an electrospray ionization (ESI) source (DuoSpray™ Source, ABSciex) in positive mode. The mass spectrometer (Triple TOF™ 5600 System; ABSciex) was programmed for information dependent acquisition (IDA) scanning full spectra (350-1250 m/z), followed by 20 MS/MS on multiple charged ions (+2 to +5) and performed one MS/MS before adding those ions to the exclusion list for 15 seconds (mass spectrometer operated by Analyst® TF 1.6, ABSciex).

Peptide identification was performed using Protein Pilot software v4.5 (ABSciex). Search parameters used were the following: UniProt_SwissProt database with and without recombinant DJ-1 WT and mutants sequence database against all species and UniProt_SwissProt database against *E. coli*; trypsin digestion; MMTS as cysteine alkylating reagent (for peptides originated by liquid digestion procedure) and acrylamide as cysteine alkylating reagent (for peptides originated by in-gel digestion procedure); ID focus: biological modifications; thorough ID search effort; and 0.05 Unused ProtScore (10% confidence score) as detected protein threshold.

Data analysis was based on an independent False Discovery Rate analysis (FDR) using the target-decoy approach. Positive identifications were considered when proteins present 95% confidence (5% local FDR) with more than one peptide hit with individual confidence above 95% or with a single peptide hit with an individual confidence above 95% and a minimum sequence tag of 3 amino acids (4 consecutive peaks in the MS/MS spectrum). When a FDR analysis was not possible (few proteins identified) an Unused above 1.3 was considered as positive identification with more than one peptide with individual confidence above 95% or with one peptide and a minimum sequence tag of 3 amino acids (4 consecutive peaks in the MS/MS spectrum).

3.2.4 LC-MS of Intact Proteins

The purified DJ-1 WT and DJ-1 mutants (400 pmoles) were precipitated with cold acetone (-20 °C) in a volume ratio of 1:6 (sample:acetone). The sample tubes were inverted several times to mix and incubated 20 minutes at -20 °C. Then samples were centrifuged for 20 minutes at 20,000*g* at 4 °C, the acetone was discarded and the pellet was resuspended in 30 µL of 2% ACN and 1% FA. Finally, samples were sonicated at 20% of amplitude for 1 minute (1 second on 1 second off cycle) and the supernatants were analyzed by LC-MS.

Proteins were resolved by liquid chromatography (NanoLC Ultra 2D, Eksigent) on a ChromXP™ C18 reverse phase trap-column (350 µm ID × 0.5mm length, 3 µm particles, 120 Å pore size, Eksigent) at 7 µL/min. Proteins were eluted into the mass spectrometer with an acetonitrile gradient (2% to 50% ACN, for 5 minutes) in 0.1% FA, using an ESI source (DuoSpray™ Source, ABSciex) in positive mode. The mass spectrometer (Triple TOF™ 5600 System; ABSciex) was programmed for scanning full spectra (600-1250 m/z) using the Intact Protein Mode, which enhances the analysis of intact proteins of high molecular weight and charge states through the optimization of key parameters of TOF MS scan types (mass spectrometer operated by Analyst® TF 1.6, ABSciex).

Intact mass was determined using BioAnalyst™ Software (ABSciex). Briefly, the deconvolution of the spectra was obtained by Baesian Protein Reconstruct algorithm of the average of the maximum spectra possible, with the following parameters: mass range from 15 to 30 kDa (determined according with the expected protein molecular weight), spectrum limit range from 650 to 1000 m/z (determined according with protein charge envelope profile), signal to noise threshold of approximately 10 (according with the ratio from the less intense peak used in calculation to the noise determined by a script present in Analyst). From the deconvoluted spectrum it was chosen a peak from the isotopic distribution for modeling the data that was then used to calculate the molecular weight of the protein.

3.2.5 HPLC-Size Exclusion Chromatography

High-performance liquid chromatography (HPLC)-Size exclusion chromatography (SEC) was performed using a Prominence Shimadzu system (Shimadzu Scientific Instruments) and a Superdex 200 5/150 GL column (GE Healthcare); and data were collected and analyzed by LC Solution Software (Shimadzu Scientific Instruments). Purified DJ-1 WT and DJ-1 mutants (2 µL at about 2 µg/µL) were loaded on column and eluted with PBS containing 10% glycerol (buffer in which the purified proteins were stored) at a flow rate of 0.4 mL/min, monitoring absorbance at 214 nm. The column and retention times were calibrated under identical running conditions with the following molecular mass standards: Aprotinin (6.5 kDa) (AppliChem), Ribonuclease A (13.7 kDa), Carbonic anhydrase (29 kDa), Ovalbumin (43 kDa), Conalbumin (75 kDa), Aldolase (158 kDa), Ferritin (440 kDa), and Blue Dextran (≈2000 kDa) (all from GE Healthcare). Approximated molecular weights of eluted proteins were determined using a calibration curve established with the standards.

3.2.6 Circular Dichroism Spectroscopy

The secondary structure of DJ-1 wild-type and mutant proteins was evaluated by Circular Dichroism (CD) spectroscopy. The proteins were dissolved in PBS with 10% glycerol at concentrations around 2 µg/µL (except for L166P-1 sample which concentration was 1.3 µg/µL) as determined by UV absorbance measurements at 280 nm. Far UV CD spectra were acquired on an Olis DSM 20 circular dichroism spectropolarimeter continuously purged with nitrogen, equipped with a Quantum Northwest CD 150 temperature-controlled cuvette and controlled by the GlobalWorks software.

Spectra were recorded at 25 °C (for all proteins) and 37 °C (except for L166P mutants) between 190 and 260 nm at 1 nm intervals using a 0.05 mm pathlength cuvette. Two scans obtained with an integration time of 6 seconds were averaged and corrected by subtracting a baseline spectrum acquired under the same experimental conditions. The results are expressed as the mean residue molar ellipticity, $[\Theta]_{MRW}$ (deg.cm^2/dmol) defined in the following equation (3.2) as:

$$[\Theta]_{MRW} = \frac{\Theta_{obs}\, x\, 100\, x\, MW}{l\, x\, C\, x\, n} \qquad \textbf{(3.2)}$$

Where Θ_{obs} is the observed ellipticity (mdeg), MW is the protein molecular weight (g/mol), l is the cuvette pathlength (cm), C is the protein concentration (mg/ml) and n is the number of amino acid residues of the protein.

The secondary structure contents were calculated with the software CONTIN and GlobalWorks using the CONTILL, CDSSTR and Selcon3 algorithms against the CLSTR reference basis set which contains soluble and denatured proteins with known secondary structure.

3.2.7 Protein Thermal Shift Assay

To monitor protein unfolding, the fluorescent dye Sypro Orange (Sigma-Aldrich) was used. Sypro Orange fluoresces strongly when located in hydrophobic regions exposed in unfolded proteins. Increase in fluorescence is used to monitor the protein-unfolding transition.

The thermal shift assay was conducted in a 7500/7500 Fast Real-Time PCR System (Applied Biosystems). Solutions of 5 µL of 50x Sypro Orange and 0.5 mg/mL protein (in a final volume of 50 µL) were added to the wells of a MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems). PBS with 10% glycerol buffer was added in the control samples. The plates were sealed with MicroAmp™ Optical Adhesive Film (Applied Biosystems) and stepwise heated from 25 °C to 95 °C. The reaction volume was 30 µL and the overall heating rate was 1 °C/min, with a hold step for fluorescence reading every 0.37 °C. As the reaction mixture is heated, the protein unfolds and there is a consequent increase in fluorescence. Fluorescence changes in the wells of the plate were measured 190 times (readings) per well and fluorescence intensity was measured with excitation and emission of 490 and 580 nm, respectively. For each protein and control samples, data from four independent wells were averaged and plotted.

All samples wells were illuminated with a tungsten halogen lamp. Light from this lamp passed through five excitation filters before reaching sample wells. Fluorescence emission was then detected through five emission filters to a charge-coupled device (CCD) camera. Emission filters were optimized for use with FAM™/SYBR® Green I, VIC®/JOE™, NED™/TAMRA™/Cy3™, ROX™/Texas Red®, and Cy5™ fluorescent dyes.

After the experiment, the 7500 software used regions of interest (ROI), optical, dye, and background calibration data to determine the location and intensity of the fluorescence signals in each read, the dye associated with each fluorescence signal, and the significance of the signal.

A real-time melt experiment generates the negative first derivative plot of the melting curve raw fluorescence data, from which the melting temperature, T_m, (the midpoint of the protein unfolding transition or the inflection point of the melting curve) can be determined as the minimum peak.

All data were processed using 7500 System Software (Applied Biosystems).

3.3 Role of DJ-1 in Neuroprotection

3.3.1 SH-SY5Y Cell Culture

Human neuroblastoma SH-SY5Y cells were cultured in Dulbecco's modified Eagle medium (DMEM) with Glutamax™ and low glucose (1 g/L) (Gibco), supplemented with 10% fetal bovine serum (FBS) (Gibco), 1.25 µg/ml amphotericin B solution (Invitrogen) and 1% penicillin-streptomycin solution (Pen-Strep) (Cambrex).

For cell passage, cells were washed with Dulbecco's phosphate buffered saline (DPBS) (Cambrex, Charles City, IA) and detached with trypsin-EDTA (0.05% solution in phosphate buffered saline (PBS)) (Invitrogen).

Cells were maintained at 37 °C, 5% CO_2/95% air in a humidified incubator (Shel Lab 3517-2) (Sheldon Manufacturing, Inc.).

3.3.2 Cell Viability Assessment under Oxidative Stress and DJ-1 Stimuli

SH-SY5Y cells were seeded at 93.75×10^3 cells/cm^2 in DMEM with 10% FBS in 96-well plates (Corning) at 37 °C, 5% CO_2/95% air in a humidified incubator (Sheldon Manufacturing, Inc.).

Four hours after plating, the culture medium was totally removed and cells were stimulated with 100 µM and 200 µM H_2O_2 (Sigma-Aldrich) in DMEM with 0.1% FBS (freshly made) in the presence or absence of recombinant DJ-1 wild-type (1 µM) or the corresponding vehicle (PBS with 10% glycerol) for 24 hours at 37 °C under a gas phase of 95% air/5% CO_2 in a humidified incubator. Control condition consisted in exchanging the culture medium to DMEM with 0.1% FBS.

Cell viability was assessed by using the Cell Titer-Glo® Luminescent assay (Promega) in white opaque 96-well plates (Corning). The luminescent signal was detected by a LUMIstar Galaxy automated microplate luminescence reader (BMG Labtech), according to the manufacturer's instructions.

3.3.3 Statistical Analysis

Statistical analysis of results from cell viability was performed using SPSS (Statistical Package for the Social Sciences) version 21.0 (IBM®) and GraphPad PRISM®5. Data normality was tested using Shapiro-Wilk Test and statistical evaluation was performed with a one-way ANOVA analysis. To compare the different H_2O_2 concentrations with Control a Student's *t*-test was performed. The results were reported as statistically significant when $p < 0.05$. Data were expressed as mean ± standard error of the mean (S.E.M.). Every experimental condition was tested in three sets of independent experiments.

3.4 LC-MS/MS-based DJ-1 WT and Mutants Metabolomics Study

3.4.1 Cell Culture, and Oxidative Stress and DJ-1 Stimuli

SH-SY5Y cells were seeded at 93.75×10^3 cells/cm^2 in DMEM with 10% FBS in 100 mm Petri dishes (Corning) at 37 °C under a gas phase of 95% air/5% CO_2 in a humidified incubator (Sheldon Manufacturing, Inc.).

Four hours after plating, the culture medium was totally removed and cells were stimulated with 200 µM H_2O_2 (Sigma-Aldrich) in DMEM with 0.1% FBS (freshly made) in the presence of recombinant DJ-1 WT (1 µM), DJ-1 M26I (1 µM), DJ-1 E163K (1 µM) or the corresponding vehicle (PBS with 10% glycerol) for 24 hours at 37 °C under a gas phase of 95% air/5% CO_2 in a humidified incubator. Control condition consisted in exchanging the culture medium to DMEM with 0.1% FBS and adding the protein vehicle. Every experimental condition was tested in three sets of independent experiments.

3.4.2 Intracellular Metabolite Quenching and Extraction

To stop stimulation the culture medium was completely removed from the culture dishes, which were put on ice, and cells were quickly washed with PBS at 37 °C. Cells were then readily quenched and intracellular metabolites extracted using methanol:water (80:20) solution (-20 °C) while cells were detaching from the culture dishes by scraping using rubber cell scrapers (TPP, Switzerland). Cellular extracts were transferred into 15 mL centrifuge tubes (VWR), sonicated [750 W Sonicator with Tapered Microtip 3 mm (VibraCell – Sonics®)] at 30% of amplitude for 2 minutes on ice (1 second on 3 seconds off cycle), transferred again into microcentrifuge tubes and centrifuged at 20,000*g* for 30 minutes at 4 °C. The supernatants were collected to new microcentrifuge tubes and metabolites were concentrated by rotary evaporation, using the Concentrator Plus (Eppendorf) at 60 °C. The evaporated samples were resuspended to 100 µL in a solution of 2% ACN and 1% FA and sonicated [750 W Sonicator with cuphorn (VibraCell – Sonics®)] at 20% of amplitude for 2 minutes (1 second on 1 second off cycle). Metabolites were desalted using C18 Bond Elut OMIX tips (Agilent technology), as described in subsection 3.2.1, and concentrated using the Concentrator Plus at 60 °C. Samples volume was

adjusted to 15 µL in a solution of 2% ACN and 0.1% FA. Samples were sonicated at 20% of amplitude for 2 minutes (1 second on 1 second off cycle), centrifuged at 14,000*g* for 5 minutes and the supernatants were analyzed by LC-MS/MS.

3.4.3 LC-MS/MS Analysis of Intracellular Metabolites

Samples were analyzed on a NanoLC Ultra 2D separation system (Eksigent) coupled to an electrospray ionization source (DuoSpray™ Source, ABSciex) operated in positive mode, and a Triple TOF™ 5600 System mass spectrometer (ABSciex). Metabolites were separated onto a Halo C18 column (0.3 x 150 mm, 2.7 µm, 90 Å, Eksigent) at 5 µL/min, with an acetonitrile gradient (2% to 61% ACN, for 42 minutes) in 0.1% FA. Using the same chromatographic conditions, mass spectrometer was programmed for two different forms of data acquisition: information dependent acquisition (IDA) and information independent acquisition SWATH analysis.

For IDA, a full mass spectra (30-1250 m/z) was acquired, followed by 11 MS/MS of ions with +1 to +4 charges and one MS/MS was performed before adding those ions to the exclusion list for 15 seconds.

For SWATH experiments, the mass spectrometer was operated in a looped product ion mode. The instrument was specifically tuned to allow a quadrupole resolution of 50 m/z mass selection. Using an isolation width of 51 m/z (containing 1 m/z for the window overlap), a set of 17 overlapping windows was constructed covering the precursor mass range of 50–900 m/z. A 250 milliseconds survey scan (50-1500 m/z) was acquired at the beginning of each cycle for instrument calibration and SWATH MS/MS spectra were collected from 50–1500 m/z for 120 milliseconds resulting in a cycle time of 2.34 seconds from the precursors ranging from 50 to 900 m/z. The collision energy for each window was determined according to the calculation for a charge 1+ ion centered upon the window with a collision energy spread of 15. The mass spectrometer was operated by Analyst® TF 1.6, ABSciex.

3.4.4 LC-MS/MS Data Processing

The data files acquired by the mass spectrometer were directly imported by the MarkerView™ software (ABSciex, version 1.2.1.1). Peak detection used an algorithm which processes each of the mass spectra of a sample by order of increasing scan number. Masses belonging to the same peak 'cluster' are merged together with a resulting area equal to the sum of all intensities for each mass value in the cluster.

For peak detection a set of parameters was applied in order to detect peaks which have (i) retention time above 10 minutes and below 46 minutes; (ii) minimum spectral peak width of 50 ppm, where peaks narrower than this value are presumed to be noise; (iii) noise threshold superior to 100 counts; and (iv) minimum retention time peak width of 6 scans.

The alignment step was useful to decide if two peaks [m/z, RT] found in two samples represent the same chemical component or not. If their retention times and m/z values were both within the specified tolerances they were assumed to be the same "feature". The two parameters applied in peak alignment were the retention time tolerance (RT tolerance) and mass tolerance. RT tolerance was set to 0.5 minutes meaning that two or more peaks from the same sample or different samples having the same m/z value were considered the same if the retention time did not exceeded 0.5 minutes. The mass tolerance considered was of 50 ppm, where two or more peaks from the same sample or different samples which differ only in 50 ppm are considered to be the same peaks. Subsequently, the MultiQuant™ software (version 2.1.1) was used to confirm the peak area of the most promising features, and these area values were used for further analysis.

The Pareto scaling method, the *t*-test and the analysis of the principal components (PCA) were performed in the MarkerView™ software (version 1.2.1.1).

The results were reported as statistically significant when $p < 0.05$. Data were expressed as mean ± standard error of the mean (S.E.M.). Every experimental condition was tested in three sets of independent experiments.

❹ RESULTS

4. RESULTS

Since the discovery of the linkage of gene mutations of DJ-1 to autosomal recessive familial PD, much attention has been paid to the role of this molecule in the pathogenesis of PD [105-108]. DJ-1 has also been linked to sporadic Parkinson's disease and appears to be involved in cellular oxidative stress rescue mechanisms during neurodegeneration [105-109]. Recent, large-scale omics studies have generated terabytes of data but not yet met the goal of developing biomarkers suitable for clinical use in PD [110]. Once oxidative stress is increased in PD, some specific markers of it may be useful for tracking disease progression in PD [111, 112].

The first task of this project consisted in the production of three recombinant DJ-1 mutant proteins. These mutants, as well as DJ-1 WT, were then structurally characterized and added to neuronal cells, under normal or oxidative stress conditions, to have a more comprehensive understanding of the role that DJ-1 plays in oxidative stress, and therefore in Parkinson's disease.

4.1 Recombinant DJ-1 Mutants Production

4.1.1 *In Vitro* Site-Directed Mutagenesis of DJ-1_pSKB-3 Construct

The DJ-1 construct (Figure 3.1) used in this study (kindly provided by Matilde Melo) contains the cloned human *DJ-1* cDNA between the NdeI and XhoI sites of pSKB-3 and produces a tobaco etch virus (TEV)-cleavable N-terminal hexahistidine tagged protein.

All *DJ-1* point mutations (L166P, M26I and E163K) were generated by site-directed mutagenesis and the mutations were verified using Sanger DNA sequencing (Figure 4.1, Figure 4.2 and Figure 4.3, respectively).

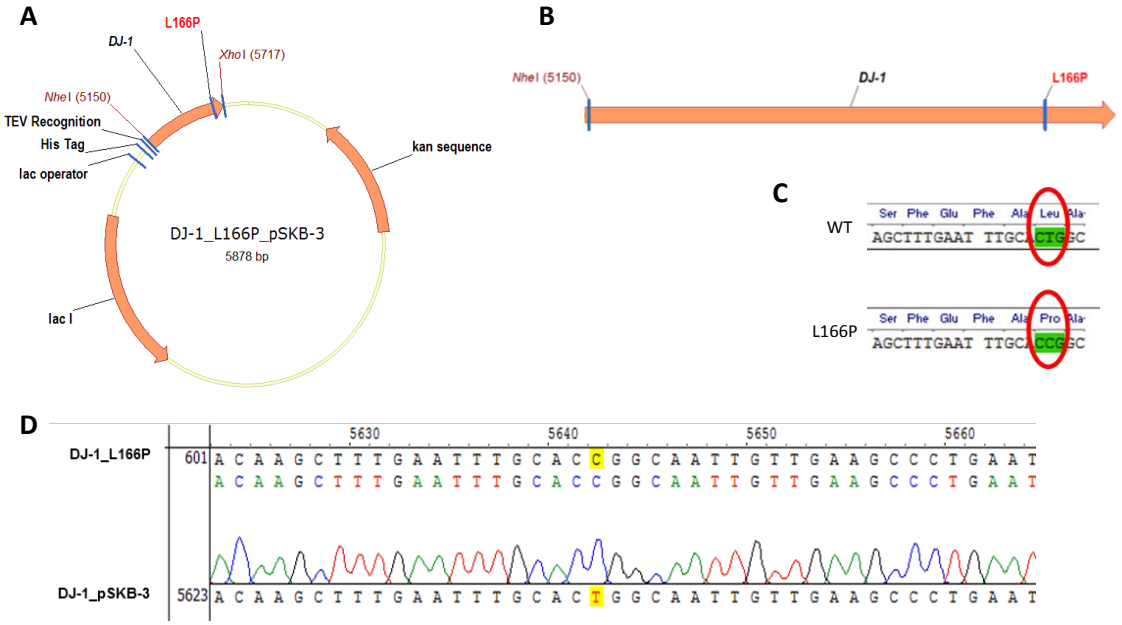


Figure 4.1 - Map of DJ-1_L166P_pSKB-3 Construct. A - Representation of the DJ-1_L166P_pSKB-3 construct which contains: lac I – lac repressor gene; His Tag – hexahistidine tag; TEV – tobaco etch virus protease recognition site; kan – kanamycin resistance gene; the *DJ-1* cDNA (optimized for *E. coli* expression) cloned between Nhe I and Xho I restriction endonuclease sites; the *DJ-1* L166P point mutation local highlighted. The vector pSKB-3 corresponds to pET-28a where thrombin recognition site was modified to TEV recognition site. **B -** Representation of *DJ-1* gene with the L166P point mutation local highlighted. **C -** Part of *DJ-1* WT and *DJ-1* L166P sequences where it is represented the change of a thymine by a cytosine nucleotide base responsible for the mutation (translation of a proline instead a leucine amino acid residue). **D -** Sanger sequencing shows (yellow highlight) the mutation confirmation (substitution of a thymine for a cytosine nucleotide base).

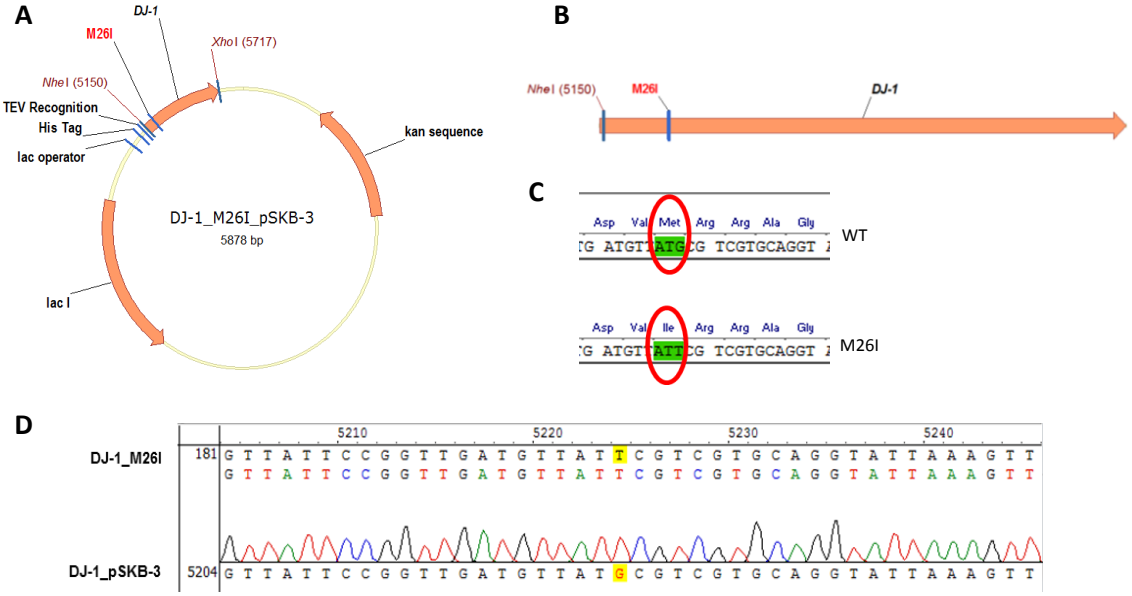


Figure 4.2 - Map of DJ-1_M26I_pSKB-3 Construct. A - Representation of the DJ-1_M26I_pSKB-3 construct which contains: lac I – lac repressor gene; His Tag – hexahistidine tag; TEV – tobaco etch virus protease recognition site; kan – kanamycin resistance gene; the *DJ-1* cDNA (optimized for *E. coli* expression) cloned between Nhe I and Xho I restriction endonuclease sites; the *DJ-1* M26I point mutation local highlighted. The vector pSKB-3 corresponds to pET-28a where thrombin recognition site was modified to TEV recognition site. **B -** Representation of *DJ-1* gene with

the M26I point mutation local highlighted. **C -** Part of *DJ-1* WT and *DJ-1* M26I sequences where it is represented the change of a guanine by a thymine nucleotide base responsible for the mutation (translation of an isoleucine instead a methionine amino acid residue). **D -** Sanger sequencing shows (yellow highlight) the mutation confirmation (substitution of a guanine for a thymine nucleotide base).

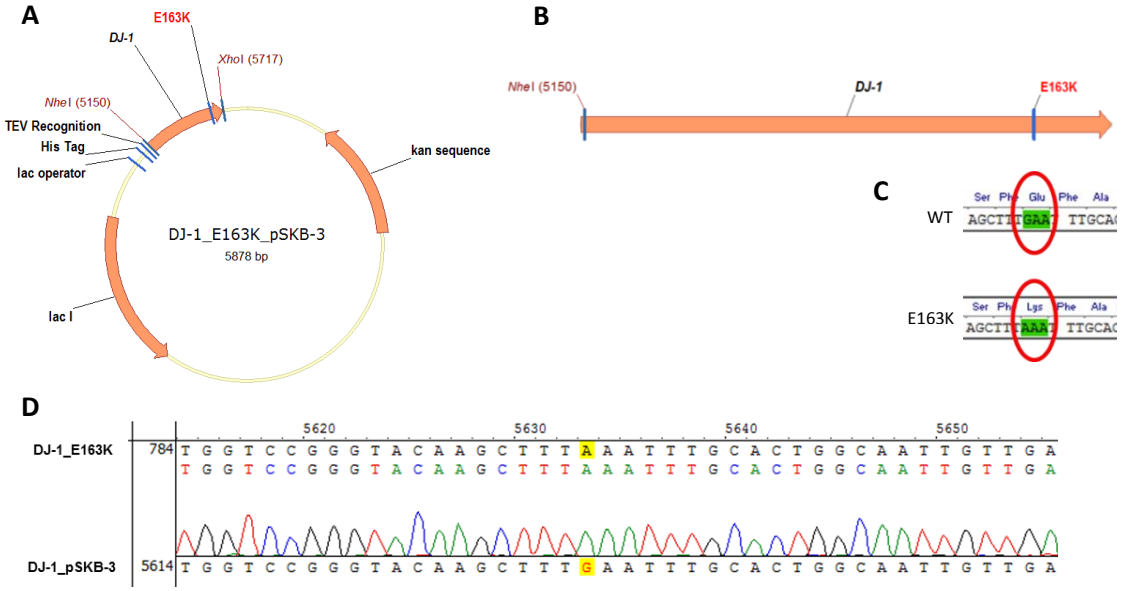


Figure 4.3 - Map of DJ-1_E163K_pSKB-3 Construct. A - Representation of the DJ-1_E163K_pSKB-3 construct which contains: lac I – lac repressor gene; His Tag – hexahistidine tag; TEV – tobaco etch virus protease recognition site; kan – kanamycin resistance gene; the *DJ-1* cDNA (optimized for *E. coli* expression) cloned between Nhe I and Xho I restriction endonuclease sites; the *DJ-1* E163K point mutation local highlighted. The vector pSKB-3 corresponds to pET-28a where thrombin recognition site was modified to TEV recognition site. **B -** Representation of *DJ-1* gene with the E163K point mutation local highlighted. **C -** Part of *DJ-1* WT and *DJ-1* E163K sequences where it is represented the change of a guanine by an adenine nucleotide base responsible for the mutation (translation of a lysine instead a glutamate amino acid residue). **D -** Sanger sequencing shows (yellow highlight) the mutation confirmation (substitution of a guanine for an adenine nucleotide base).

4.1.2 DJ-1 Mutants Expression and Purification

Recombinant DJ-1 mutants were expressed in *E. coli* BL21star (DE3) strain and purified by affinity chromatography, using an HisTrap column since the produced proteins contain an hexahistidine tag (6His-tag) (Figure 4.4 A, Figure 4.6 A, Figure 4.8 A, Figure 4.10 A, and Figure 4.12 A). The most concentrated protein fraction of each DJ-1 mutant (corresponding to the highest chromatographic peak which eluted at 300 mM imidazole) was purified by gel filtration chromatography, using an HiLoad 26/600 Superdex 200 prep grade column (Figure 4.4 B, Figure 4.6 B, Figure 4.8 B, Figure 4.10 B, and Figure 4.12 B), with an intermediate step for the L166P-1 mutant before its gel filtration chromatography (see methods subsection 3.1.4 for details). Proteins were stored at -80 °C in PBS with 10%

glycerol at a final concentration of about 2 µg/µL. The protein fractions from DJ-1 mutants expression and purification steps were electrophoretically separated and stained with Colloidal Coomassie Brilliant Blue G-250 (Bio-Rad) to observe the protein content profile (Figure 4.5, Figure 4.7, Figure 4.9, Figure 4.11, and Figure 4.13).

The L166P mutant was produced in three different batches due to the low protein content obtained. So, the number after the name of this mutant corresponds to the batch in which it was produced: DJ-1 L166P-1 (Figure 4.4 and Figure 4.5) was the first, DJ-1 L166P-2 (Figure 4.6 and Figure 4.7) the second, and DJ-1 L166P-3 (Figure 4.8 and Figure 4.9) the third L166P mutant produced. The M26I (Figure 4.10 and Figure 4.11) and E163K (Figure 4.12 and Figure 4.13) mutants were produced only once.

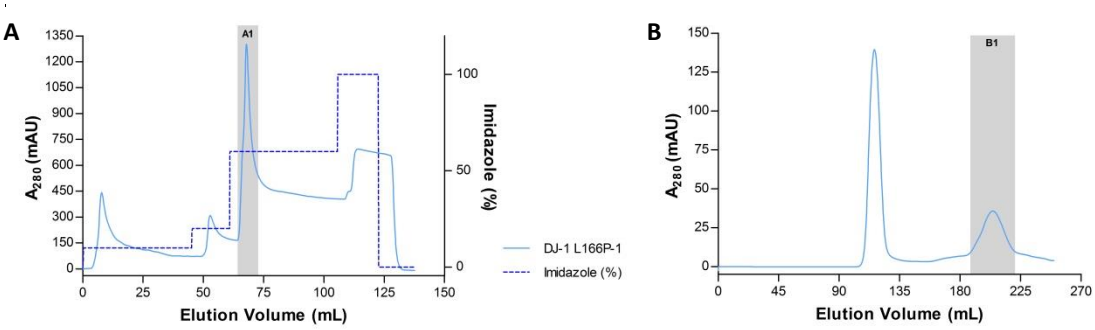


Figure 4.4 - Chromatograms of DJ-1 L166P-1 Purification. **A -** Affinity chromatography performed in an HisTrap column. Fraction A1 - eluted fraction with 300 mM imidazole. **B -** Gel filtration chromatography. Fraction A1 from the affinity chromatography was applied on an HiLoad 26/600 Superdex 200 prep grade column after buffer exchange to PBS on an HiTrap Desalting column. Fraction B1 - eluted fraction with PBS.

The L166P-1 mutant purification (Figure 4.4) results indicate that before the protein expression induced with IPTG, there is not a highly intense band at ≈23 kDa (expected mass of a DJ-1 monomer) (Figure 4.5, lane BI). On the other hand, after the induction of DJ-1 L166P-1 expression, a higher intensity band appears at this molecular weight in the insoluble fraction (lane AI I), but not in the soluble fraction (lane AI S), indicating that the L166P-1 mutant protein is only slightly soluble. All the analyzed eluted fractions (lanes A1, B1, B1 (3d) and B1 AC) as well as the stored protein (lane DJ-1 L166P-1) present this band. However, new bands appear after protein concentration (lane B1 AC) and protein storage (lane DJ-1 L166P-1), indicating either the degradation of the protein or/and the increased concentration of other contaminants. In the gel filtration chromatography of this protein (Figure 4.4 B) only the fraction eluted between 180 mL

and 225 mL (fraction B1) was collected because at this elution volume the dimeric form of DJ-1 is eluted (dimeric form seems to be the functionally relevant form of the protein [113-115]). Thus, the previous peak (eluted between 90 mL and 135 mL) corresponds to a higher molecular weight form of protein and for that reason it was not collected.

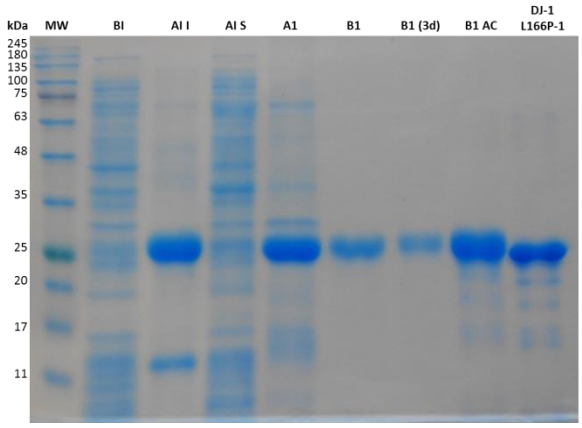


Figure 4.5 – SDS-PAGE followed by Coomassie staining of different fractions of DJ-1 L166P-1 Purification. MW – molecular weight marker: NZY Colour Protein Marker II (NZYTech); BI – before induction of protein expression; AI I – insoluble fraction after induction of protein expression; AI S – soluble fraction after induction of protein expression; A1 – eluted fraction A1 (Figure 4.4 A) followed by buffer exchange to PBS on an HiTrap Desalting column; B1 – eluted fraction B1 (Figure 4.4 B); B1 (3d) – same sample than B1 after 3 days of storage at 4 °C; B1 AC – fraction B1 (3d) after concentration on a 10 kDa MWCO centrifugal concentrator; DJ-1 L166P-1 – purified DJ-1 L166P-1 stored in PBS with 10% glycerol at -80°C. In each lane 15 µg of protein were applied.

A second batch of L166P (L166P-2) was purified (Figure 4.6) and a highly intense band in the insoluble fraction after the induction of this mutant expression at ≈23 kDa (Figure 4.7 A, lane AI I) can also be seen, however in the soluble fraction (lane AI S) this band is much less intense, indicating that the L166P-2 mutant is poorly soluble, as it has already been observed for the first production of this mutant (Figure 4.5, lanes AI I and AI S). The fraction loaded on the HisTrap column (lane Ld) and the resulting flow through (lane FT A), present similar protein profiles which means that most of the protein was not captured by the column, probably because this mutant is found mostly in the insoluble fraction) (Figure 4.7 A, lane AI I). Nevertheless, all the analyzed eluted fractions (lanes A1, A2, A3, A4, B1, B2 and B2 AC) as well as the stored protein (lane DJ-1 L166P-2) present a band at ≈23 kDa, indicating the presence of this protein mutant (Figure 4.7 A and B). However, new bands appear after protein concentration (Figure 4.7 B, lane B2 AC) and protein storage (Figure 4.7 B, lane DJ-1 L166P-2), indicating either the degradation of the protein and/or the increased concentration of contaminants. In the gel filtration

chromatography of this protein (Figure 4.6 B) the fractions eluted approximately at 200 mL (dimeric form of DJ-1) (fraction B2) and between 100 mL and 150 mL (higher molecular weight form of protein) (fraction B1) were collected but only the fraction corresponding to the dimeric form of DJ-1 (fraction B2) was stored.

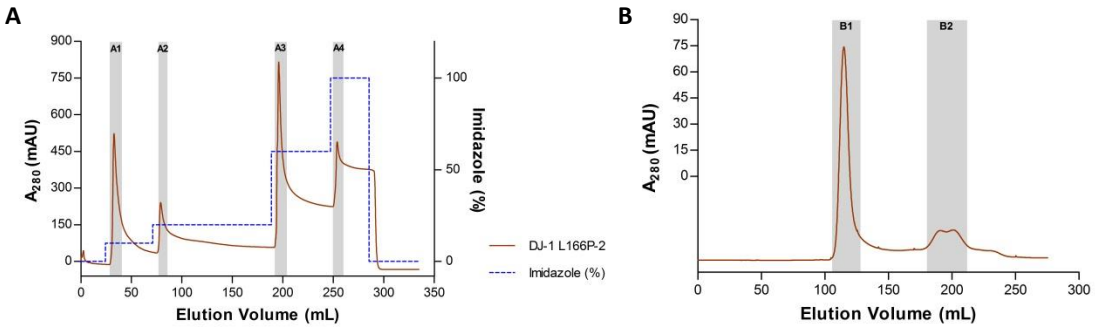


Figure 4.6 - Chromatograms of DJ-1 L166P-2 Purification. **A -** Affinity chromatography performed in an HisTrap column. Fraction A1 - eluted fraction with 50 mM imidazole; fraction A2 – eluted fraction with 100 mM imidazole; fraction A3 – eluted fraction with 300 mM imidazole; fraction A4 – eluted fraction with 500 mM imidazole. **B -** Gel filtration chromatography. Fraction A3 from the affinity chromatography was applied on an HiLoad 26/600 Superdex 200 prep grade column. Fraction B1 and B2- eluted fractions B1 and B2 with PBS.

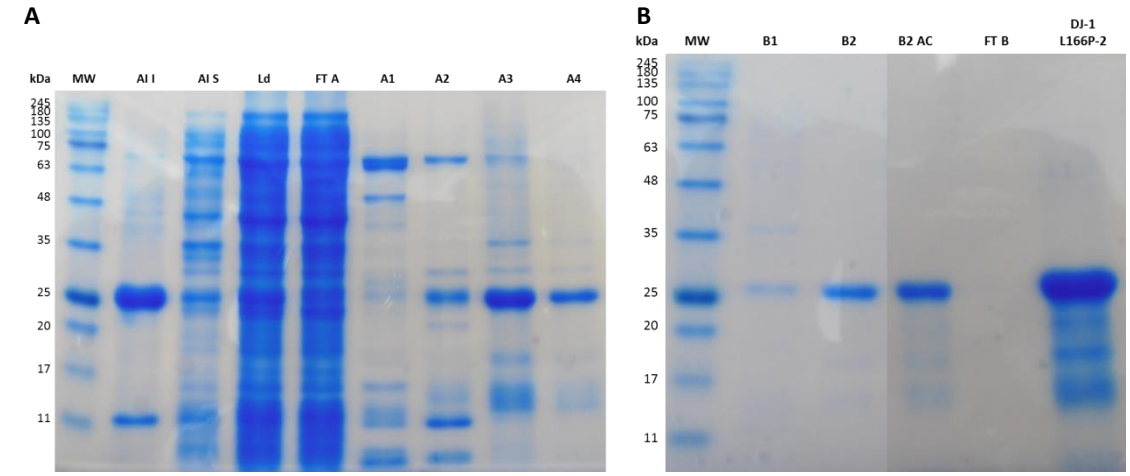


Figure 4.7 – SDS-PAGE followed by Coomassie staining of different fractions of DJ-1 L166P-2 Purification. MW – molecular weight marker: NZY Colour Protein Marker II (NZYTech). **A –** SDS-PAGE of fractions from protein expression and affinity purification (Figure 4.6 A). BI – before induction of protein expression; AI I – insoluble fraction after induction of protein expression; AI S – soluble fraction after induction of protein expression; Ld - loaded on the HisTrap column; FT A- flow-through of the HisTrap column; A1 – eluted fraction A1; A2 – eluted fraction A2; A3 – eluted fraction A3; A4 – eluted fraction A4. **B -** SDS-PAGE of fractions from gel filtration purification (Figure 4.6 B). B1 – eluted fraction B1; B2 - eluted fraction B2; B2 AC – eluted fraction B2 after concentration on a 10 kDa MWCO centrifugal concentrator; FT B - flow-through of the 10 kDa MWCO centrifugal concentrator; DJ-1 L166P-2 – purified DJ-1 L166P-2 stored in PBS with 10% glycerol at -80°C. In each lane 10 µL of protein sample were applied.

For the last production of the L166P mutant (L166P-3) (Figure 4.8), the results indicate that before the protein expression there is not a highly intense band at ≈23 kDa (Figure 4.9 A, lane BI) which only appears after the induction of L166P-3 mutant expression in the insoluble fraction (lane AI I), but not as much as in the soluble fraction (lane AI S). This fact confirms once again that this mutant is only slightly soluble. Most of the protein was not captured by the affinity purification column as it can be seen by the protein profile of the fraction loaded on the HisTrap column (lane Ld) and the resulting flow through (lane FT A). All the analyzed eluted fractions (lanes A1, A2, A3, A4, B1 and B1 AC) as well as the stored protein (lane DJ-1 L166P-3) present a band at ≈23 kDa, indicating the presence of this protein mutant, although severe protein loss is observed during the purification and concentration steps (Figure 4.9 A and B). In the gel filtration chromatography of this protein (Figure 4.8 B) the fraction eluted between 250 mL and 300 mL (dimeric form of DJ-1) presented low intensity and it was not collected. Thus, only the fraction eluted between 100 mL and 150 mL (higher molecular weight form of protein) (fraction B1) was collected.

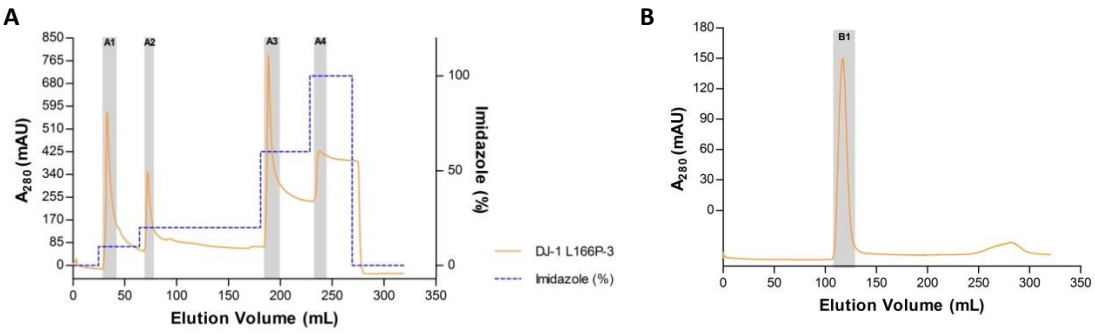


Figure 4.8 - Chromatograms of DJ-1 L166P-3 Purification. **A -** Affinity chromatography performed in an HisTrap column. Fraction A1 - eluted fraction with 50 mM imidazole; fraction A2 – eluted fraction with 100 mM imidazole; fraction A3 – eluted fraction with 300 mM imidazole; fraction A4 – eluted fraction with 500 mM imidazole. **B -** Gel filtration chromatography. Fraction A3 from the affinity chromatography was applied on an HiLoad 26/600 Superdex 200 prep grade column. Fraction B1 - eluted fraction B1 with PBS.

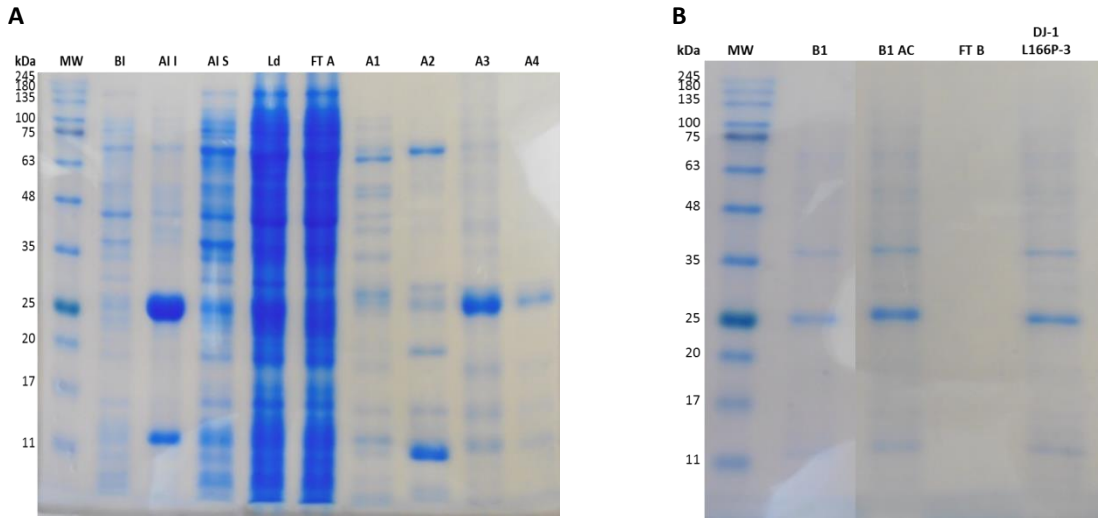


Figure 4.9 - SDS-PAGE followed by Coomassie staining of different fractions of DJ-1 L166P-3 purification. MW – molecular weight marker: NZY Colour Protein Marker II (NZYTech). **A –** SDS-PAGE of fractions from protein expression and affinity purification (Figure 4.8 A). BI – before induction of protein expression; AI I – insoluble fraction after induction of protein expression; AI S – soluble fraction after induction of protein expression; Ld - loaded on the HisTrap column; FT A- flow-through of the HisTrap column; A1 – eluted fraction A1; A2 – eluted fraction A2; A3 – eluted fraction A3; A4 – eluted fraction A4. **B -** SDS-PAGE of fractions from gel filtration purification (Figure 4.8 B). B1 – eluted fraction B1; B1 AC – eluted fraction B1 after concentration on a 10 kDa MWCO centrifugal concentrator; FT B - flow-through of the 10 kDa MWCO centrifugal concentrator; DJ-1 L166P-3 – purified DJ-1 L166P-3 stored in PBS with 10% glycerol at -80°C. In each lane 10 µL of protein sample were applied.

The M26I mutant purification (Figure 4.10) results indicate that before the protein expression there is not a highly intense band at ≈23 kDa (Figure 4.11 A, lane BI). On the other hand, after the induction of this mutant expression, a higher intensity band appears at this molecular weight in the soluble fraction (lane AI S), also appearing in the insoluble fraction (lane AI I) but with lower intensity. The fraction loaded on the HisTrap column (lane Ld) also present an intense band, disappearing then in the resulting flow through (lane FT A), which indicates that almost all the recombinant protein was captured by the column. All the analyzed eluted fractions (lanes A1, A2, A3, A4, B1 and B1 AC) as well as the stored protein (lane DJ-1 M26I) present this band (Figure 4.11 A and B). In the gel filtration chromatography of this protein (Figure 4.10 B), only the fraction corresponding to the dimeric form of DJ-1 (fraction B1) was collected.

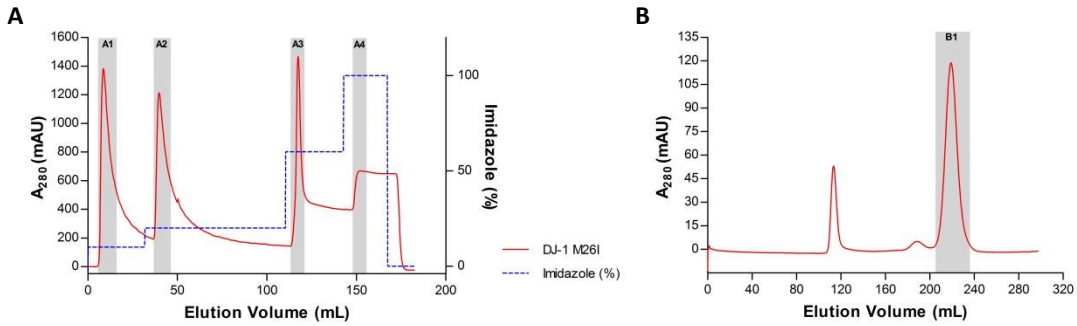


Figure 4.10 - Chromatograms of DJ-1 M26I Purification. **A -** Affinity chromatography performed in an HisTrap column. Fraction A1 - eluted fraction with 50 mM imidazole; fraction A2 – eluted fraction with 100 mM imidazole; fraction A3 – eluted fraction with 300 mM imidazole; fraction A4 – eluted fraction with 500 mM imidazole. **B -** Gel filtration chromatography. Fraction A3 from the affinity chromatography was applied on an HiLoad 26/600 Superdex 200 prep grade column. Fraction B1 - eluted fraction B1 with PBS.

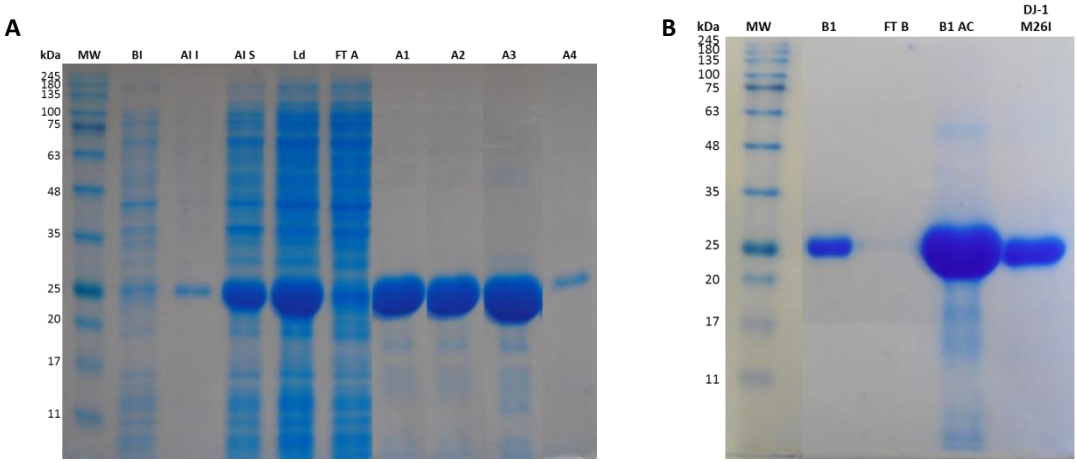


Figure 4.11 - SDS-PAGE followed by Coomassie staining of different fractions of DJ-1 M26I Purification. MW – molecular weight marker: NZY Colour Protein Marker II (NZYTech). **A –** SDS-PAGE of fractions from protein expression and affinity purification (Figure 4.10 A). BI – before induction of protein expression; AI I – insoluble fraction after induction of protein expression; AI S – soluble fraction after induction of protein expression; Ld - loaded on the HisTrap column; FT A- flow-through of the HisTrap column; A1 – eluted fraction A1; A2 – eluted fraction A2; A3 – eluted fraction A3; A4 – eluted fraction A4. **B -** SDS-PAGE of fractions from gel filtration purification (Figure 4.10 B). B1 – eluted fraction B1; B1 AC – eluted fraction B1 after concentration on a 10 kDa MWCO centrifugal concentrator; FT B - flow-through of the 10 kDa MWCO centrifugal concentrator; DJ-1 M26I – purified DJ-1 M26I stored in PBS with 10% glycerol at -80°C. In each lane 10 µL of protein sample were applied.

Finally, the E163K mutant was also purified (Figure 4.12) and the results indicate that before the protein expression there is not a highly intense band at ≈23 kDa (Figure 4.13, lane BI). On the other hand, after the induction of this mutant expression, a higher intensity band appears at this molecular weight in the insoluble fraction (lane AI I), also appearing in the soluble fraction (lane AI S) but with lower intensity, indicating that this mutant is only slightly soluble. All the analyzed eluted fractions (lanes A1 and B1) as well as the stored protein (lane DJ-1 E163K) present this band (Figure 4.13). In the gel filtration

chromatography of this protein (Figure 4.10 B), only the fraction corresponding to the dimeric form of DJ-1 (fraction B1) was collected. This mutant was the only one that did not need a protein concentration step before being stored.

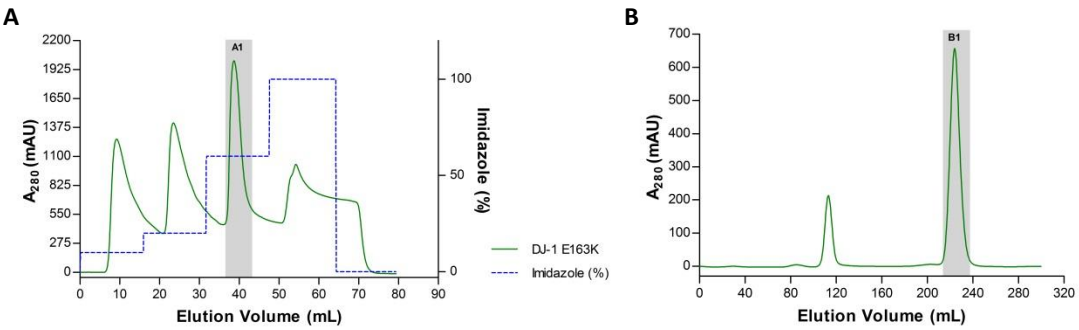


Figure 4.12 - Chromatograms of DJ-1 E163K Purification. **A -** Affinity chromatography performed in an HisTrap column. Fraction A1 - eluted fraction with 300 mM imidazole. **B -** Gel filtration chromatogram. Fraction A1 from the affinity chromatography was applied on an HiLoad 26/600 Superdex 200 prep grade column. Fraction B1 - eluted fraction with PBS.

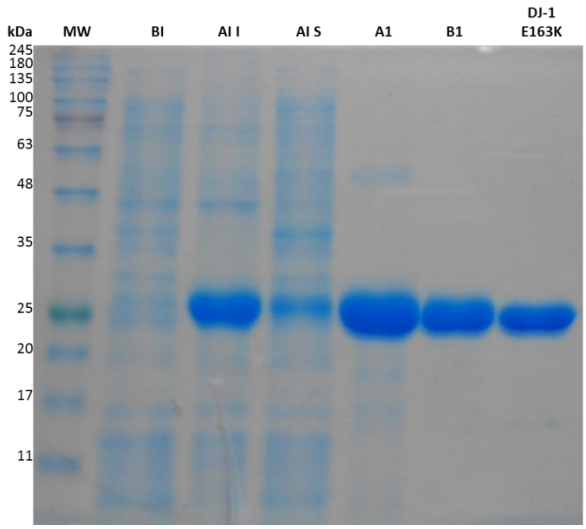


Figure 4.13 – SDS-PAGE followed by Coomassie staining of different fractions of DJ-1 E163K Purification. MW – molecular weight marker: NZY Colour Protein Marker II (NZYTech); BI – before induction of protein expression; AI I – insoluble fraction after induction of protein expression; AI S – soluble fraction after induction of protein expression; A1 – eluted fraction A1 (Figure 4.12 A); B1 – eluted fraction B1 (Figure 4.12 B); DJ-1 E163K – purified DJ-1 E163K stored in PBS with 10% glycerol at -80°C. In each lane were applied 15 µg of protein.

4.2 Structural Characterization of DJ-1 and DJ-1 Mutants

In order to assess whether the produced recombinant proteins are with its characteristic structure various techniques of structural characterization, presented below, were performed.

4.2.1 Protein Identification by LC-MS/MS

The purified and stored proteins (including DJ-1 WT) were resolved by SDS-PAGE and a more sensitive staining (silver staining) was used to visualize potential contaminants (Figure 4.14). In order to understand what was the content of the purified DJ-1 proteins besides themselves, gel lanes with DJ-1 proteins [from Coomassie staining gel (data not shown)] and stored liquid DJ-1 proteins were digested and analyzed by LC-MS/MS.

The human DJ-1 is a 189-residue protein[32], however, the recombinant DJ-1 WT and DJ-1 mutants have 213 amino acids (excluding the initial methionine) because the N-terminal of these recombinant proteins was engineered to contain an hexahistidine tag (6-His-tag) and a TEV cleavage sequence (Figure 3.1, Figure 4.1, Figure 4.2 and Figure 4.3).

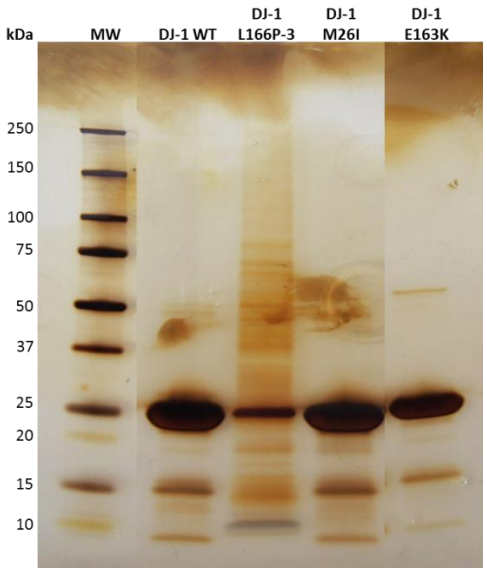


Figure 4.14 - SDS-PAGE followed by Silver staining of purified DJ-1 samples. MW – molecular weight marker: Precision Plus Protein All Blue Standards (Bio-Rad); DJ-1 WT – purified DJ-1 WT; DJ-1 L166P-3 - purified DJ-1 L166P-3; DJ-1 M26I - purified DJ-1 M26I; DJ-1 E163K – purified DJ-1 E163K. In each lane 10 µg of protein were applied.

The recombinant DJ-1 WT, DJ-1 M26I and DJ-1 E163K sequences were identified with 99.5% coverage (Figure 4.15 A, C and D). The sequence coverage of recombinant DJ-1 WT, DJ-1 M26I and DJ-1 E163K is not 100% due to the lack of the initial methionine in the final sequences, probably removed by *E. coli* methionyl-aminopeptidase, since *E. coli* was the overexpressing system used.

The recombinant DJ-1 L166P-3 sequence was identified with 83.2% coverage (Figure 4.15 B). Despite the sequence of this mutant having a lower % coverage than the others DJ-1 proteins, the mutant was identified by the amino acid at the corresponding position 166 of human DJ-1, a proline (P) (Figure 4.15 B, highlight), instead the characteristic leucine (L) of DJ-1 WT (Figure 4.15 A, the last highlighted amino acid).

A

MGSSHHHHHHDYDIPTTENLYFQGH|MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTVAGLAGKDPVQCS
RDVVICPDASLEDAKKEGPYDVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG
SKVTTHPLAKDKMMNGGHYTYSENRVEKDGLILTSRGPGTSFEFALAIVEALNGKEVAAQVKAPLVLKD

B

MGSSHHHHHHDYDIPTTENLYFQGH|MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTVAGLAGKDPVQCS
RDVVICPDASLEDAKKEGPYDVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG
SKVTTHPLAKDKMMNGGHYTYSENRVEKDGLILTSRGPGTSFEFAPAIVEALNGKEVAAQVKAPLVLKD

C

MGSSHHHHHHDYDIPTTENLYFQGH|MASKRALVILAKGAEEMETVIPVDVIRRAGIKVTVAGLAGKDPVQCS
RDVVICPDASLEDAKKEGPYDVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG
SKVTTHPLAKDKMMNGGHYTYSENRVEKDGLILTSRGPGTSFEFALAIVEALNGKEVAAQVKAPLVLKD

D

MGSSHHHHHHDYDIPTTENLYFQGH|MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTVAGLAGKDPVQCS
RDVVICPDASLEDAKKEGPYDVVVLPGGNLGAQNLSESAAVKEILKEQENRKGLIAAICAGPTALLAHEIGFG
SKVTTHPLAKDKMMNGGHYTYSENRVEKDGLILTSRGPGTSFKFALAIVEALNGKEVAAQVKAPLVLKD

Figure 4.15 – Sequence Coverage of recombinant DJ-1 WT and DJ-1 Mutants. Green - residues identified in peptides with at least 95% confidence; gray - unidentified residues; yellow – residues identified in peptides with at least 50% and less than 95% confidence; red – residues identified in peptides with less than 50% confidence; vertical black line – separation of the hexahistidine tag and TEV cleavage sequence N-terminal (left side of the line) from the 189-residue sequence (right side of the line). The amino acid counting is performed from the vertical black line. **A – Sequence coverage of DJ-1 WT**. Yellow highlight – amino acids (methionine (M), glutamate (E) and leucine (L)) that were changed in DJ-1 mutants: M26I, E163K and L166P, respectively. **B – Sequence coverage of DJ-1 L166P-3**. Yellow highlight – amino acid (proline (P)) that was changed in DJ-1 L166P. **C – Sequence coverage of DJ-1 M26I**. Yellow highlight – amino acid (isoleucine (I)) that was changed in DJ-1 M26I. **D – Sequence coverage of DJ-1 E163K**. Yellow highlight – amino acid (lysine (K)) that was changed in DJ-1 E163K.

The produced recombinant proteins have some protein contaminants (Figure 4.14). The analysis of these digested gel lanes and the liquid digestion of stored proteins

by LC-MS/MS (Table 4.1) shows that most of the protein contaminants are bovine proteins, probably contained in the LB medium used in bacterial culture, and *Escherichia coli* proteins, which was the host for the cloning and expression of the desired proteins.

Table 4.1 – DJ-1 and Contaminants identification by LC-MS/MS.

Protein ID				DJ-1 WT		DJ-1 M26I		DJ-1 E163K	
Name	**Accession**	**Species**	**Mass (kDa)**	**UP**	**US**	**UP**	**US**	**UP**	**US**
Protein DJ-1	**Q99497|PARK7_HUMAN**	***Homo sapiens***	**19.891**	**134**	**203.16**	**171**	**160.64**	**112**	**165.29**
Putative molybdate metabolism regulator	P33345|MOLR_ECOLI	*Escherichia coli*	140.939					1	2.00
Probable hypoxanthine oxidase XdhD	Q46814|XDHD_ECOLI	*Escherichia coli*	103.519					1	2.00
Putative two-component response regulator-like APRR6	Q9C9F6|APRR6_ARATH	*Arabidopsis thaliana*	86.182	1	0.24				
Cyclomaltodextrin glucanotransferase	P31797|CDGT_GEOSE	*Geobacillus stearothermophilus*	78.923	1	0.44				
Serum albumin	P02769|ALBU_BOVIN	*Bos taurus*	69.293					3	3.26
Aldehyde dehydrogenase 5, mitochondrial	P40047|ALDH5_YEAST	*Saccharomyces cerevisiae*	56.693					1	2.00
Glutathione reductase	P06715|GSHR_ECOLI	*Escherichia coli*	48.773			1	2.00	2	2.65
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	Q5GRK4|ISPG_WOLTR	*Wolbachia sp. subsp. Brugia malayi*	47.381					1	2.00
Protochlorophyllide reductase B, chloroplastic	Q42850|PORB_HORVU	*Hordeum vulgare*	42.148	1	2.00				
Farnesyl diphosphate synthase	P22939|ISPA_ECOLI	*Escherichia coli*	32.160					1	2.00
Uncharacterized protein yffS	P76550|YFFS_ECOLI	*Escherichia coli*	29.751					1	2.00
Transcriptional regulatory protein CusR	P0ACZ8|CUSR_ECOLI	*Escherichia coli*	25.395	1	2.00				
Beta-casein	P02666|CASB_BOVIN	*Bos taurus*	25.107					1	2.00
Alpha-S1-casein	P02662|CASA1_BOVIN	*Bos taurus*	24.529			1	2.00	2	3.77
Transcriptional regulator YqjI	P64588|YQJI_ECOLI	*Escherichia coli*	23.401			2	2.02		
High-molecular weight cobalt-containing nitrile hydratase subunit alpha	P21219|NHA1_RHORH	*Rhodococcus rhodochrous*	22.835			1	2.00	1	0.88
Glycerol-3-phosphate acyltransferase	A6Q218|PLSY_NITSB	*Nitratiruptor sp.*	22.430	1	2.00				
Uncharacterized lipoprotein yceB	P0AB26|YCEB_ECOLI	*Escherichia coli*	20.500			1	2.03		

Uncharacterized protein yaiL	P51024|YAIL_ECOLI	*Escherichia coli*	19.923	1	2.00				
Beta-lactoglobulin	P02754|LACB_BOVIN	*Bos taurus*	19.883					1	1.38
Putative lipocalin 1-like protein 1	Q5VSP4|LC1L1_HUMAN	*Homo sapiens*	17.918					1	0.80
Ferric uptake regulation protein	P0A9A9|FUR_ECOLI	*Escherichia coli*	16.795			2	4.00	5	10.00
Lysozyme C, milk isozyme	Q6B411|LYSM_BOVIN	*Bos taurus*	16.783					1	2.00
50S ribosomal protein L11P	A3CSJ5|RL11_METMJ	*Methanoculleus marisnigri*	16.317	1	2.00				
Superoxide dismutase [Cu-Zn]	Q96VL0|SODC_CLAPU	*Claviceps purpurea*	15.839			2	3.89	2	1.20
50S ribosomal protein L9	P0A7R1|RL9_ECOLI	*Escherichia coli*	15.769			1	2.00		
Fatty acid-binding protein, epidermal	Q01469|FABP5_HUMAN	*Homo sapiens*	15.164					2	4.00
Thioredoxin-1	P0AA25|THIO_ECOLI	*Escherichia coli*	11.807	5	8.88	4	8.02	4	8.08
Dermcidin	P81605|DCD_HUMAN	*Homo sapiens*	11.284					2	3.70
Glutaredoxin-1	P68688|GLRX1_ECOLI	*Escherichia coli*	9.685	2	3.00	2	3.70	2	3.12
50S ribosomal protein L28	P0A7M2|RL28_ECOLI	*Escherichia coli*	9.006			1	2.00		

Abbreviations: UP – Unique Peptides; US – Unused Score

Although the DJ-1 proteins have some contaminants, the number of unique peptides found to identify the DJ-1 proteins is much larger than those found to identify the contaminants (Table 4.1).

The contaminants of the L166P-3 mutant are not presented here because this mutant originated a large number of protein contaminants which it is not comparable with that of the other DJ-1 proteins. This large number of contaminants (data not shown) may explain the lower sequence coverage of this mutant when compared with the sequence coverage of the other mutants (Figure 4.15).

4.2.2 LC-MS of Intact Proteins

To determine the molecular weights of the monomeric form of DJ-1 WT, DJ-1 M26I and DJ-1 E163K the purified proteins were analyzed by LC-MS (Figure 4.16). For the recombinant DJ-1 L166P-3 was not possible to determine the molecular weight of its monomeric form through direct analysis of the protein most likely due to its aggregated state.

The molecular weights calculated using the mass spectra of protein charge envelops of intact DJ-1 WT, DJ-1 M26I and DJ-1 E163K and the BioAnalyst™ Software (Table 4.2) are very close to their theoretical (average) molecular weights, without the initial methionine, calculated using the Mass Calculator tool (Bioinformatics Solutions, Inc.), with molecular weight shifts between 0.4 and 1.1 Da.

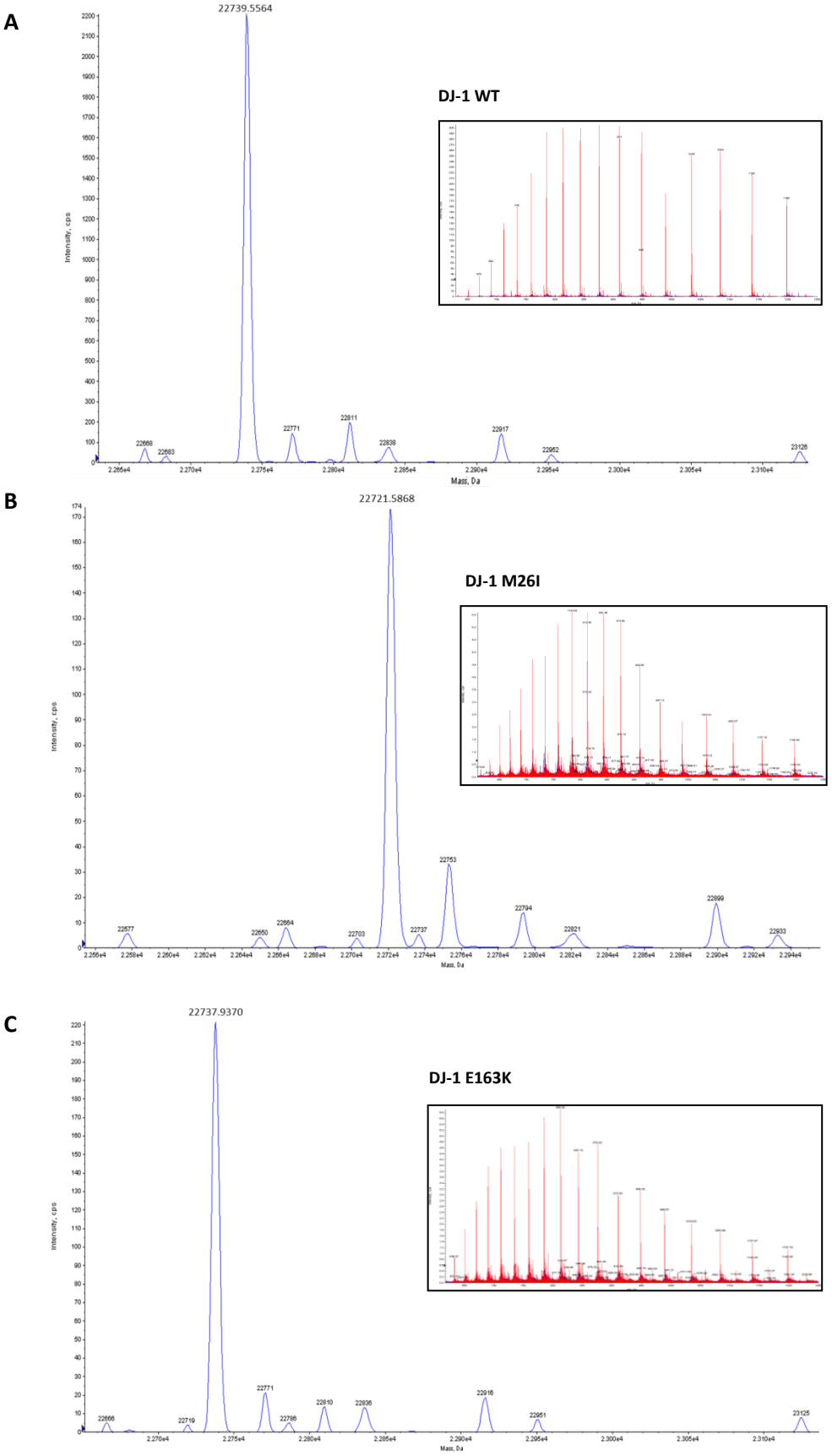


Figure 4.16 – LC-MS analysis of intact DJ-1 WT, DJ-1 M26I and DJ-1 E163K. Deconvoluted mass spectra of intact DJ-1 WT **(A)**, DJ-M26I **(B)** and DJ-1 E163K **(C)** using BioAnalyst™. Inserts show obtained charged envelope.

Table 4.2 – Average and Calculated Molecular Weights for the purified DJ-1 proteins.

Protein	Average Molecular Weight (Da)	Calculated Molecular Weight (Da)	Molecular Weight Shift (Da)
WT	22739.998	22739.5564	0.4416
M26I	22721.965	22721.5868	0.3782
E163K	22739.057	22737.9370	1.1200

4.2.3 HPLC-Size Exclusion Chromatography

DJ-1 is found as homodimer and this seems to be the functionally relevant form of the protein [113-115]. In order to assess the form of the purified proteins, an HPLC-Size Exclusion chromatography was performed. The retention times of standards were used to perform a calibration curve and the molecular weights of purified DJ-1 proteins were determined using their retention times (Figure 4.17 A and B and Supplementary Figure 8.1).

The molecular weights of the purified DJ-1 WT, DJ-1 M26I and DJ-1 E163K are ≈41 kDa, ≈42 kDa and ≈37 kDa, respectively, which correspond to the dimer form [116].

The three batches of the L166P mutant produced were also evaluated by this analysis along with the above mentioned proteins (Figure 4.17 A and B). The molecular weights of the purified DJ-1 L166P-1 and DJ-1 L166P-2 are ≈91 kDa, and ≈86 kDa, respectively, which do not correspond to the proteins dimer form. The molecular weight of the purified DJ-1 L166P-3 could not be calculated because this protein was eluted in the column void volume determined by the elution volume for Blue Dextran (≈2000 kDa) and for this reason it is out of the HPLC-Size Exclusion Chromatography calibration curve (Supplementary Figure 8.1). The results indicate that this mutant is in an aggregated state or it exists in the form of protein oligomer or even protein polymer. While in the L166P-1 and L166P-2 gel filtration purification the fraction corresponding to the protein dimeric form was collected and stored (Figure 4.4 B and Figure 4.6 B, respectively), in the L166P-3 gel filtration purification the fraction corresponding to a higher protein molecular weight

form was collected (Figure 4.8 B). Hence the differences found in the molecular weight of this mutant for the three batches produced.

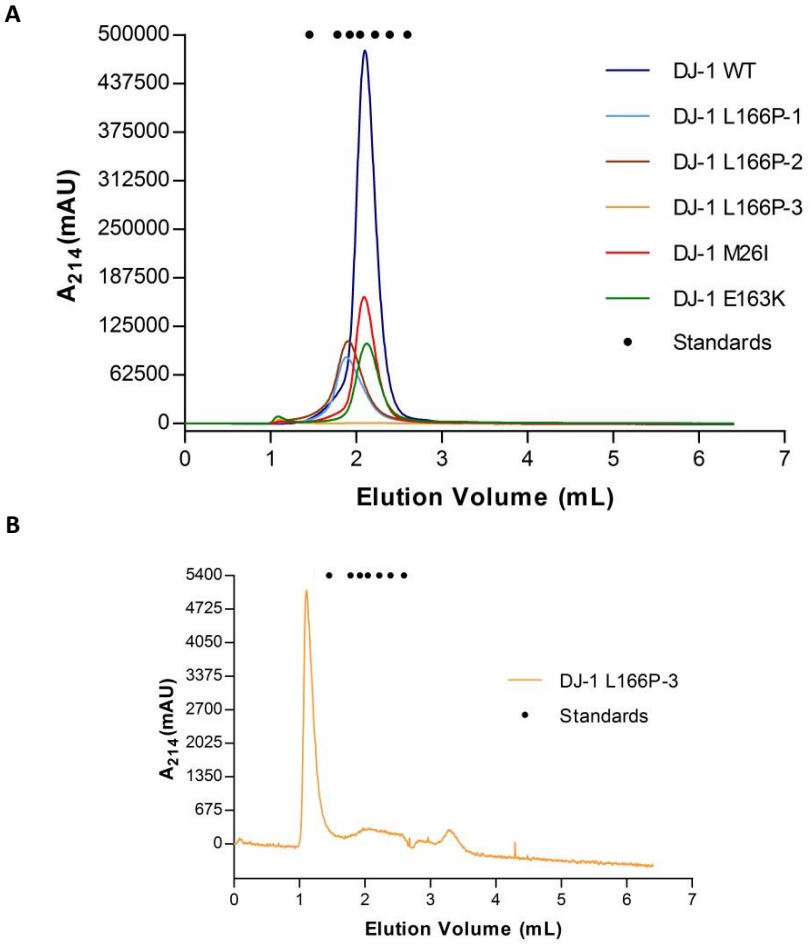


Figure 4.17 - HPLC-Size Exclusion Chromatography of purified DJ-1 proteins. Molecular Weight Standards from left to right: Ferritin (440 kDa); Aldolase (158 kDa); Conalbumin (75 kDa); Ovalbumin (43 kDa); Carbonic Anhydrase (29 kDa); Ribonuclease A (13.7 kDa); and Aprotinin (6.5 kDa). **A –** HPLC-Size Exclusion chromatogram of DJ-1 WT and DJ-1 mutants. **B -** HPLC-Size Exclusion chromatogram of DJ-1 L166P-3 showing its elution in the column void volume.

4.2.4 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a widely used technique in protein structure analysis. This technique is commonly used to investigate the secondary structure because each secondary structural element presents a particular profile. A protein CD spectrum is the result of the secondary structure composition and as a consequence, CD is useful to study the variation of the protein folding [117].

DJ-1 WT, M26I and E163K CD spectra were acquired at 37 °C. The CD spectrum of the DJ-1 WT protein is typical of a well-structured protein containing both α-helical and β-sheet secondary structure (Figure 4.18, blue line). The far-UV CD spectra of the M26I and E163K mutants are similar in shape to that of the protein wild-type but exhibit decreased signal amplitude, with the effect being most pronounced for the E163K mutant.

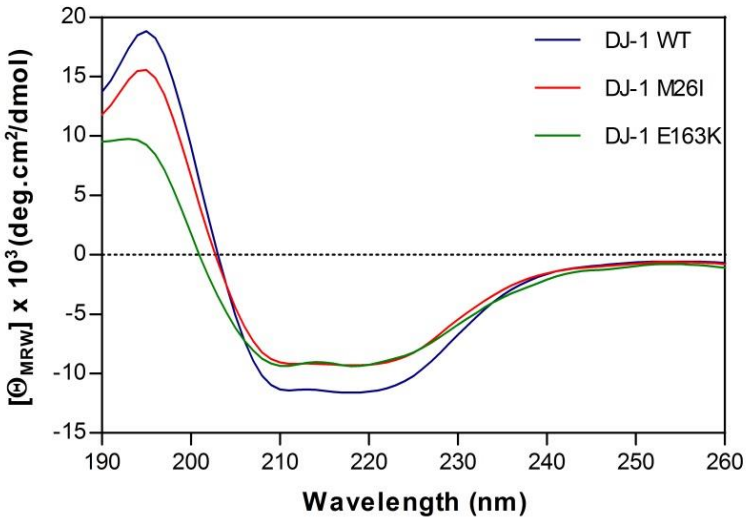


Figure 4.18 - Circular Dichroism (CD) spectra of DJ-1 WT, DJ-1 M26I and DJ-1 E163K. The secondary structure of DJ-1 WT and mutant proteins was evaluated by CD spectroscopy at 37 °C.

For DJ-1 WT, CD data gave 37% (CONTIN) or 38% (GlobalWorks) of α-helix, 25% (CONTIN) or 16% (GlobalWorks) of β-sheet structure, and 38% (CONTIN) or 46% (GlobalWorks) of no regular secondary structure (Table 4.3). For the M26I and E163K mutants, CD data retrieved a lower content of α-helix and higher β-sheet content. For DJ-1 M26I, CD data gave 30% (CONTIN and GlobalWorks) of α-helix, 38% (CONTIN) or 21% (GlobalWorks) of β-sheet structure, and 32% (CONTIN) or 49% (GlobalWorks) of no regular secondary structure. For DJ-1 E163K, CD data gave 23% (CONTIN) or 27% (GlobalWorks) of α-helix, 28% (CONTIN) or 22% (GlobalWorks) of β-sheet structure, and 49% (CONTIN) or 51% (GlobalWorks) of no regular secondary structure.

The X-ray structures analysis (Table 4.3) reveals that DJ-1 WT, M26I and E163K have about 40% of α-helix structure and 20% of β-sheet structure. This is in agreement with the results obtained by CD for DJ-1 WT. However, the CD results for M26I and E163K do not reproduce these amounts of secondary structure.

Table 4.3 – CD data (at 37 °C) obtained for purified DJ-1 proteins and X-Ray Crystallography data for DJ-1 proteins available in PDB (see Supplementary Table 8.2-8.4 for more detailed information).

	CD data								
	CONTIN program			GlobalWorks software (Average of 3 algorithms)			X-Ray Crystallography data (Average of all structures from PDB)		
Protein	α-helix %	β-sheet %	Remainder %	α-helix %	β-sheet %	Remainder %	α-helix %	β-sheet %	Remainder %
WT	37	25	38	38	16	46	40	20	40
M26I	30	38	32	30	21	49	36	19	45
E163K	23	28	49	27	22	51	36	20	44

DJ-1 L166P-1 and L166P-3 CD spectra were only acquired at 25 °C in a first preliminary analysis, as well as, for the DJ-1 WT, M26I and E163K (Supplementary Figure 8.2). Once the biological assays were performed at 37 °C, the CD analysis was later performed at this temperature (Figure 4.18). As the analysis at 25 °C of L166P-1 and L166P-3 CD spectra (Supplementary Figure 8.2) showed that these proteins did not have a defined secondary structure, when the analysis of the DJ-1 proteins was performed at 37 °C the L166P mutant was not included.

4.2.5 Protein Thermal Shift Assay

Other form to evaluate potential significant changes in protein structure and stability is by Thermal Shift Assay, where through negative first derivative plots of the thermal melting curves, it is possible to determine the melting temperatures of proteins (corresponding to the minimum peak) by evaluating the amount of dye that binds to the proteins.

The results indicate that the DJ-1 WT is the most stable protein, with a melting temperature of approximately 63.5 °C, and the E163K mutant the least stable protein, with a melting temperature of approximately 54.1 °C (Figure 4.19 A and B). The M26I mutant is less stable than DJ-1 WT but more stable than E163K mutant, by presenting a melting temperature of approximately 60.0 °C (Figure 4.19 A and B).

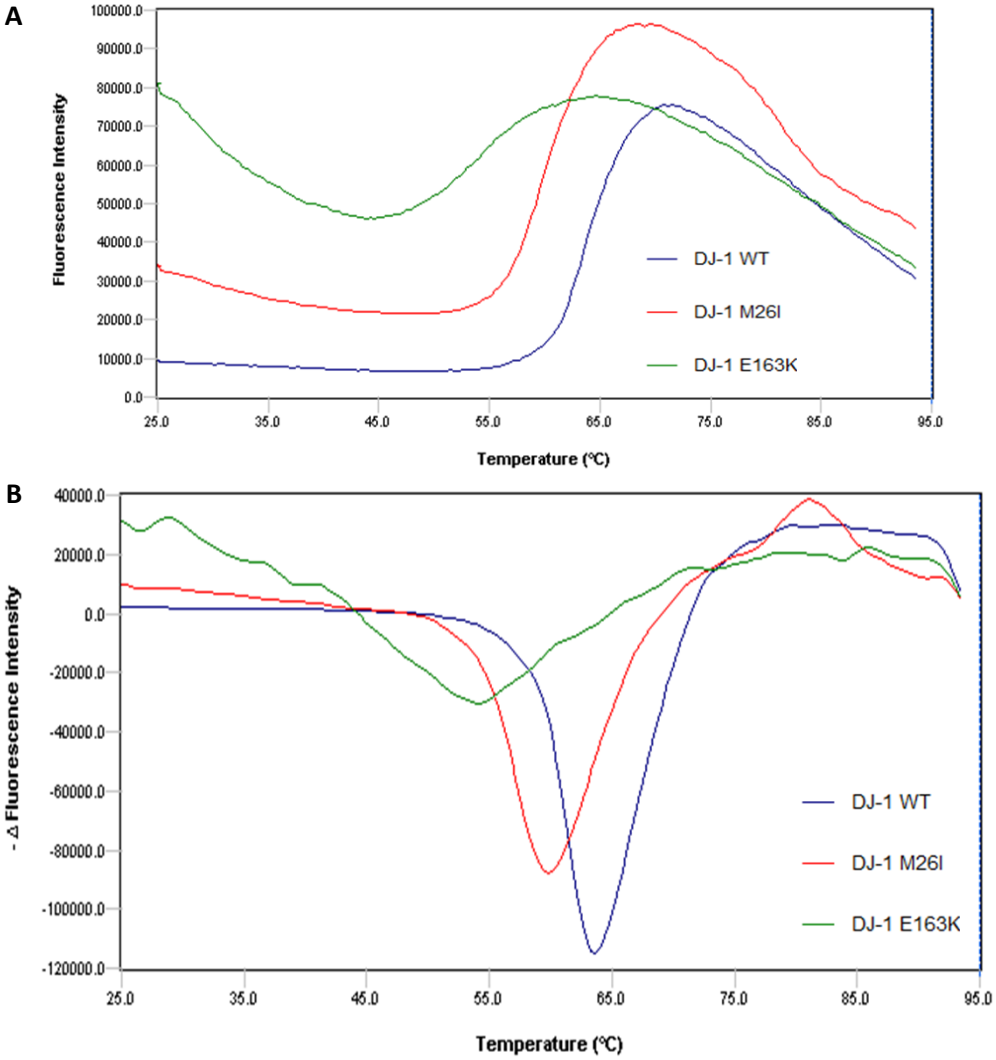


Figure 4.19 - Thermal Shift Assay for DJ-1 WT, DJ-1 M26I and DJ-1 E163K. A - Thermal melting curves (fluorescence). **B -** Negative first derivative plots of the thermal melting curves: the melting temperature, T_m, can be determined as the minimum peak (see text for details).

For L166P-1 and L166P-3 mutants it was not possible to determine their melting temperatures (there is not a minimum peak) (Figure 4.20, lower panel) since they did not present an usual profile of denaturation (Figure 4.20, upper panel). Therefore the L166P mutant is thermally unstable and it has melting profiles of a typical aggregated protein.

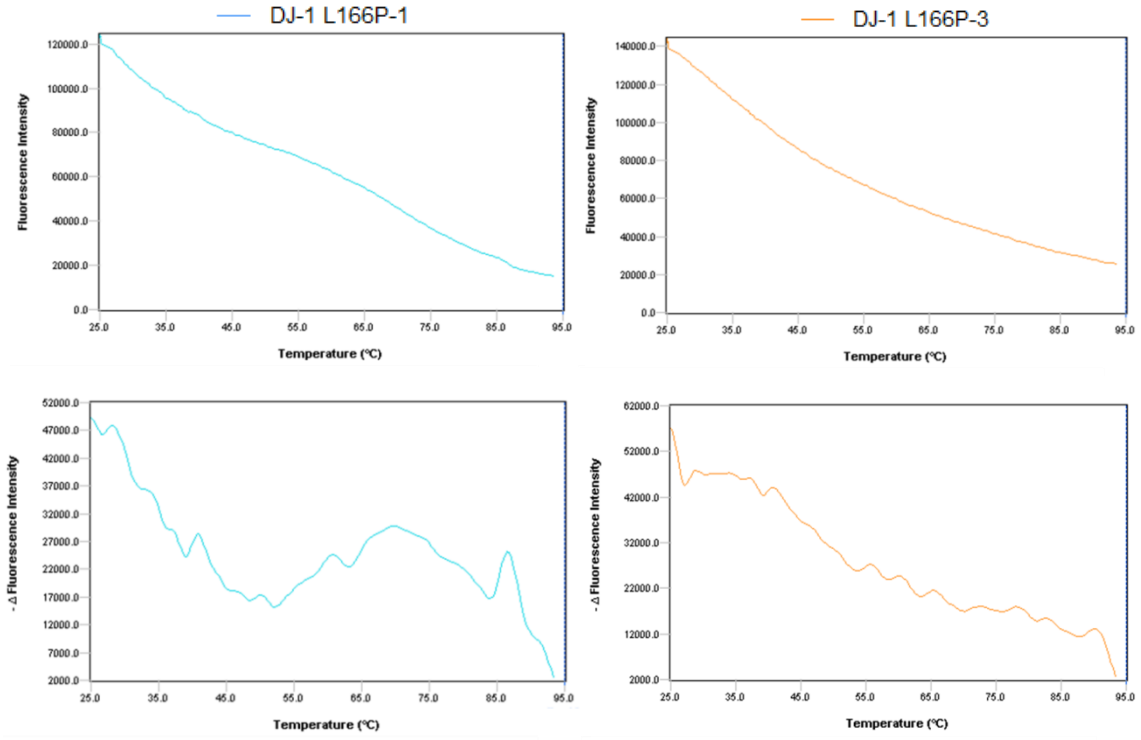


Figure 4.20 – Thermal Shift Assay for DJ-1 L166P-1 (left panel) and DJ-1 L166P-3 (right panel). Thermal melting curves (fluorescence) are represented on the upper panel and Negative first derivative plots of the thermal melting curves are represented on the lower panel.

4.3 Role of DJ-1 in Neuroprotection

The produced recombinant DJ-1 WT and the M26I and E163K mutants have the predicted molecular mass of the protein (subsection 4.2.2), they are present, in solution, as dimers (subsection 4.2.3), they have a secondary structure well defined (subsection 4.2.4), and they are stable, in solution, at 37 °C (subsection 4.2.5). Thus, these three proteins can be used in the biological assays. On the contrary, none of the recombinant L166P mutants produced can be used in the biological assays, once none of them achieved the criteria mentioned above for the other recombinant DJ-1 proteins (subsection 4.2.2-4.2.5).

In order to address if the recombinant DJ-1 WT is functional, its neuroprotective capacity against oxidative stress was tested.

The human neuroblastoma SH-SY5Y cell line has been widely used as an *in vitro* model of dopaminergic neurons for Parkinson's disease [118] and, is known to be responsive to oxidative stress caused by hydrogen peroxide [119]. The inhibition of H_2O_2-

induced cell death by the exogenous addition of recombinant DJ-1 WT protein was already shown [119]. So, to address if this recombinant DJ-1 WT is effective to play this function, cells were treated with different concentrations of H_2O_2 in the presence of 1 µM DJ-1 WT (or the corresponding amount of vehicle).

The results indicate that SH-SY5Y cells were sensitive to the oxidative stress-inducer hydrogen peroxide (Figure 4.21, blank bars), as expected, once the treatment of cells with 100 and 200 µM H_2O_2 induced cell death. However, an inhibition of cell death, as result of the presence of 1 µM His-tagged recombinant human DJ-1 WT protein, was not observed (gray bars). The DJ-1 vehicle (PBS with 10% glycerol) seems to have similar effect on cells than DJ-1 (black bars). Thus, the inhibition of H_2O_2-induced cell death by the exogenous addition of this recombinant DJ-1 WT protein was not confirmed in this study.

As the treatment of cells with 200 µM H_2O_2 showed to be statistically effective to induce cell death, this H_2O_2 concentration was used in the next task of this project to generate oxidative stress.

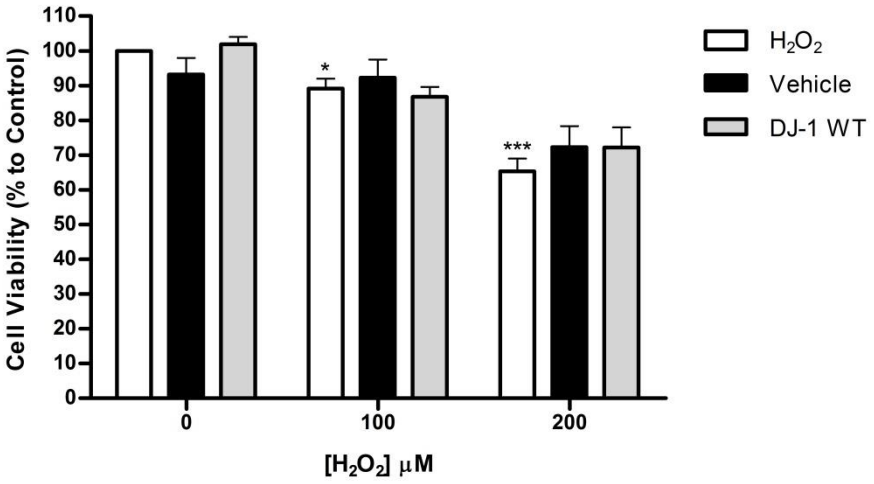


Figure 4.21 – Cell viability assessment under H_2O_2-induced oxidative stress. SH-SY5Y cells were treated with H_2O_2 (100 and 200 µM) in the presence or absence of recombinant His-tagged human DJ-1 WT protein (1 µM) or vehicle, 4h after plating. After 24 h, cell viability was assessed by the Cell Titer-Glo assay. Data present mean±SEM of three determinations, based on untreated cultures as 100% of viability (Control). Significance (Student's two-tailed *t*-test): **p* < 0.05, ****p* < 0.001 treatment with H_2O_2 vs. Control (0 µM H_2O_2).

4.4 LC-MS/MS-based DJ-1 WT and Mutants Metabolomics Study

In order to identify the metabolites that are significantly altered in resting and oxidative stress conditions, and also access the effect of the addition of DJ-1 WT and mutants to the SH-SY5Y cell line under such conditions, a metabolomics study was carried out. To achieve this objective, 8 different biological conditions were tested: (1) control (Vehicle) [cells with the addition of protein vehicle (PBS with 10% glycerol)]; (2) cells stimulated with 200 µM H_2O_2 and the addition of vehicle (Vehicle_ H_2O_2); (3) cells with the addition of DJ-1 WT (WT); (4) cells stimulated with 200 µM H_2O_2 and the addition of DJ-1 WT (WT_ H_2O_2); (5) cells with the addition of DJ-1 M26I (M26I); (6) cells stimulated with 200 µM H_2O_2 and the addition of DJ-1 M26I (M26I_ H_2O_2); (7) cells with the addition of DJ-1 E163K (E163K); and (8) cells stimulated with 200 µM H_2O_2 and the addition of DJ-1 E163K (E163K_ H_2O_2).

4.4.1 LC-MS/MS Analysis of Intracellular Metabolites and Data Processing

Firstly, SH-SY5Y cells were plated, and after 4 hours of incubation they were stimulated with 200 µM H_2O_2 in the presence of recombinant DJ-1 WT (1 µM), DJ-1 M26I (1 µM), DJ-1 E163K (1 µM) or the corresponding vehicle (PBS with 10% glycerol). After 24 hours of stimulation, cells were quenched to stop their metabolism and intracellular metabolites were extracted using methanol:water (80:20) solution.

Three replicates of each biological condition were analyzed by LC-MS/MS SWATH untargeted analysis to screen potential and putative metabolites of interest. The full-scan spectra of 24 samples were acquired. However, the LC-MS/MS analysis for some of the samples did not result: for the first and the third replicates of WT; for the first and second replicates of WT_H_2O_2; and for the first replicate of E163K. Therefore, a total of 19 acquired files were analyzed.

The MarkerView™ Software is designed to compare data from several samples, including direct analysis of LC-MS data, so that differences can be identified. Univariate analysis (*t*-test) and Multivariate analysis (PCA) can be performed in order to find possible

biomarkers. Data files (wiff extension) are imported by MarkerView™, and peak detection, alignment and scaling steps are performed in the software prior to data analysis. The manual inspection of the raw mass spectra is an essential step to select the best values of each parameter used in peak detection and alignment. Peak detection works in the individually raw data files to generate a data table containing the peaks detected with the respective m/z value, RT and peak area.

Based on only the analysis between the Vehicle and Vehicle_H_2O_2 conditions, once they represent the basis of normal condition and oxidative stress condition, respectively, the search of potential and putative metabolites of interest was initiated, through the finding of ions that vary between these two conditions.

The 6 files acquired for these samples were imported to MarkerView™ for LC-MS/MS data processing. Different combinations of the parameter values for peak detection and alignment and their influence on the peaks detected and on the results obtained by PCA analysis were tested. The parameters values chosen for peak detection were (i) retention time above 10 minutes and below 46 minutes; (ii) minimum spectral peak width of 50 ppm; (iii) noise threshold superior to 100 counts; and (iv) minimum retention time peak width of 6 scans. And the parameter values chosen for peak alignment were (i) retention time tolerance of 0.5 minutes; and (ii) mass tolerance of 50 ppm.

For these 6 files acquired, with peak detection and alignment performed using the parameter values mentioned above, 6446 peaks were detected. The number of peaks detected was reduced to 109 peaks, by showing only peaks statistically different ($p < 0.05$, *t*-student) between Vehicle and Vehicle_H_2O_2. Manual validation was conducted in a few representative peaks, and overall they revealed to be correctly detected and aligned. The presence or absence of an ion in a sample can be examined by looking to the extracted ion chromatogram (XIC), which indicates the ion current detected for a specific m/z along the time (data not shown).

The large number of variables involved (i.e. the number of peaks detected by the LC-MS/MS) requires the use of multivariate techniques, as PCA. However, prior to the

analysis of the principal components it is important to define how peak responses are scaled in order to adjust the relative importance of the variables. The scaling method used in this study was the Pareto scaling, where the mean-centered values are divided by the square root of the standard deviation.

The PCA analysis, for the 109 peaks detected (Figure 4.22), and the manual validation of these ions through the observation of their XICs (data not shown), allowed the finding of 8 different interesting variables which can be potential ions responsible for the distinction of Vehicle from Vehicle_H_2O_2 conditions (Table 4.4). The score values (representation of samples, Figure 4.22, upper pannel) for Vehicle and Vehicle_H_2O_2 conditions represented in the two first principal components (PC1 and PC2) using the Pareto scaling, explain 94.6% of the group variance. Variables (features with a specific m/z value and retention time) responsible for the formation of the observed groups are represented in the loadings plot (Figure 4.22, lower panel). The scores plot (upper panel) shows the evident separation of the two groups along the first principal component (PC1), and the loadings plot shows ions which have the largest contribution along the PC1 and which explain the separation of the two groups. The features 398.2/35.5 and 376.3/35.6 (lower panel, in orange) seem to be the main contributors for the separation of the Vehicle_H_2O_2 condition located in the negative side of the PC1 (upper panel). While features 790.2/21.4, 590.1/21.1, 410.1/20.6, 381.1/21.1, 363.1/21.1 and 335.1/21.2 (lower panel, in black) seem to be the main contributors for the separation of the Vehicle condition located in the positive side of the PC1 (upper panel).

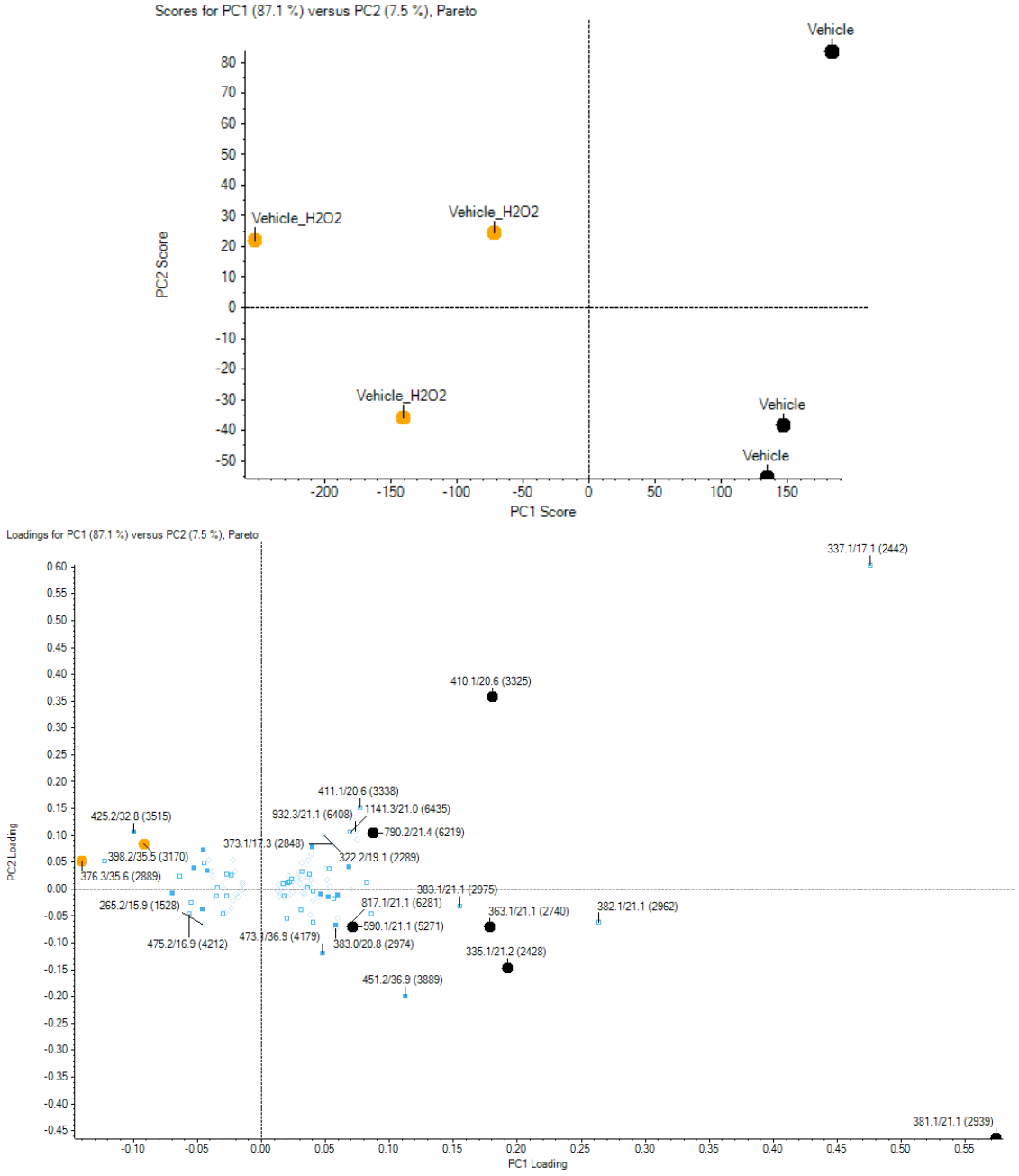


Figure 4.22 – Principal component analysis (PCA) of Vehicle and Vehicle_H_2O_2 conditions data acquired by LC-MS/MS and processed using MarkerView™ software and Pareto scaling. PCA was performed using the 109 variables statistically different ($p < 0.05$) obtained from *t*-test. The scores plot (upper panel) shows the formation of two main groups along the PC1, whose separation is best explained by the ions most represented along the PC1 on the loadings plot (lower panel). Orange: Vehicle condition. Black: Vehicle_H_2O_2 condition.

Table 4.4 - Interesting features which can differentiate the Vehicle condition from the Vehicle_H_2O_2 condition.

	Vehicle		Vehicle_H_2O_2		
Feature (m/z/RT)	**Area Mean**	**Area SD**	**Area Mean**	**Area SD**	***p*-value**
335.1/21.2	516.8	182.7	3218.3	834.9	0.005
398.2/35.5	656.1	209.3	9.7	16.8	0.006
376.3/35.6	1097.5	369.8	0.0	0.0	0.007
363.1/21.1	1107.3	646.9	2876.5	55.0	0.009
790.2/21.4	5.9	10.2	519.1	212.8	0.014
381.1/21.1	13474.6	7884.7	32064.5	2192.3	0.017
590.1/21.1	190.9	198.4	524.4	97.6	0.044
410.1/20.6	803.8	235.5	3079.3	1379.6	0.048

For each of the 8 interesting features found, the fragmentation spectra were acquired (Figure 4.23 and Supplementary Figure 8.3-8.9) and some fragments (indicated by red arrows) of the 8 precursors were chosen to confirm the quantification and the profile of the respective precursor ion in the biological samples.

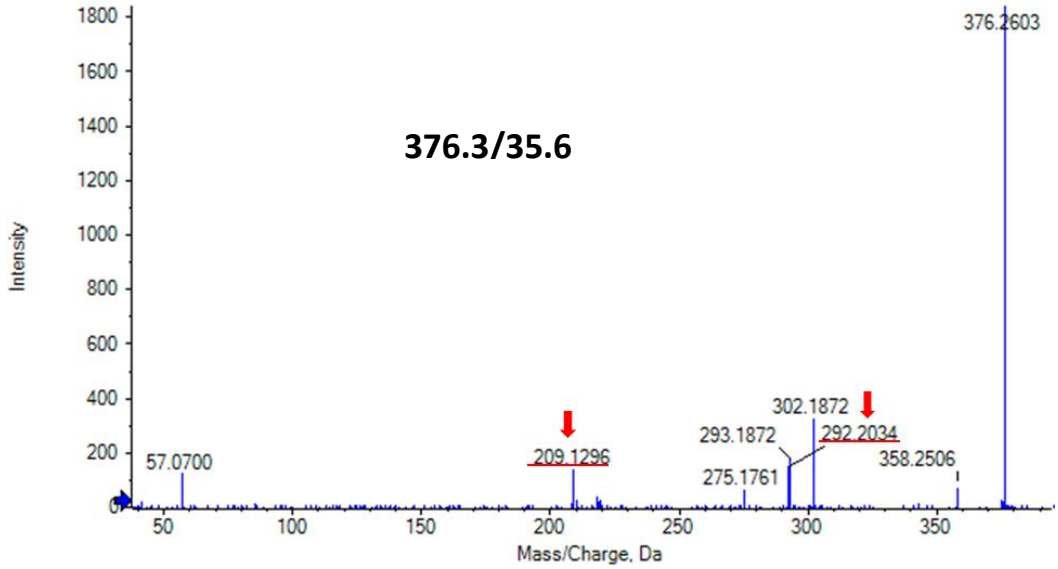


Figure 4.23 – Fragmentation spectrum of the precursor ion with m/z value of 376.3 and RT of 35.5 minutes. Red arrows represent the fragments selected for quantification.

The peak integration of the precursor ions and their fragments was performed using the MultiQuant™ software, but now all the 19 files acquired for the intracellular

metabolite samples were used. A total of 24 ions (8 precursors and 16 fragments) were imported to the software and their peak integration was performed in each one of the 19 samples. For each interesting ion and its fragments, the mean peak area with the respective standard error of the mean (SEM), were calculated in 6 biological conditions (1) Vehicle; (2) Vehicle_ H_2O_2; (5) M26I; (6) M26I_ H_2O_2; (7) E163K and (8) E163K_ H_2O_2. For the remaining 2 conditions (3) WT and (4) WT_ H_2O_2 this was not possible because there is only one replicate of each. However, the peak area of the ions in these conditions was calculated and graphically represented along with the other conditions (Figure 4.24-4.31).

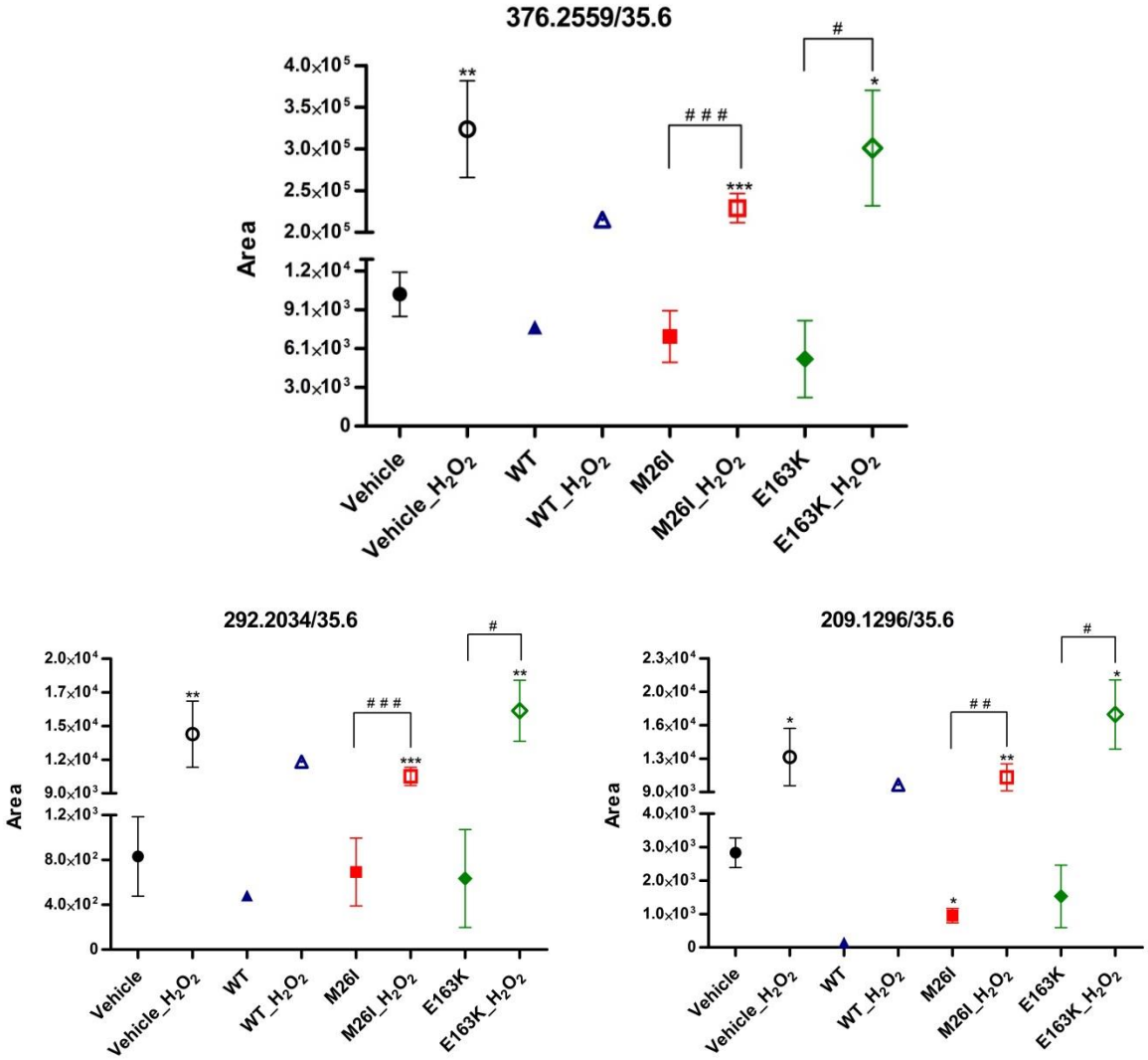


Figure 4.24 - Peak areas of the precursor ion with m/z value of 376.3 and RT of 35.6 minutes (upper panel) and its fragments ions with m/z value of 292.2 (lower panel, left side) and 209.1 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ when compared to Vehicle. #$p < 0.05$, ##$p < 0.05$, ###$p < 0.05$ treatment with 200 µM H_2O_2 and M26I or E163K vs. corresponding treatment with M26I or E163K.

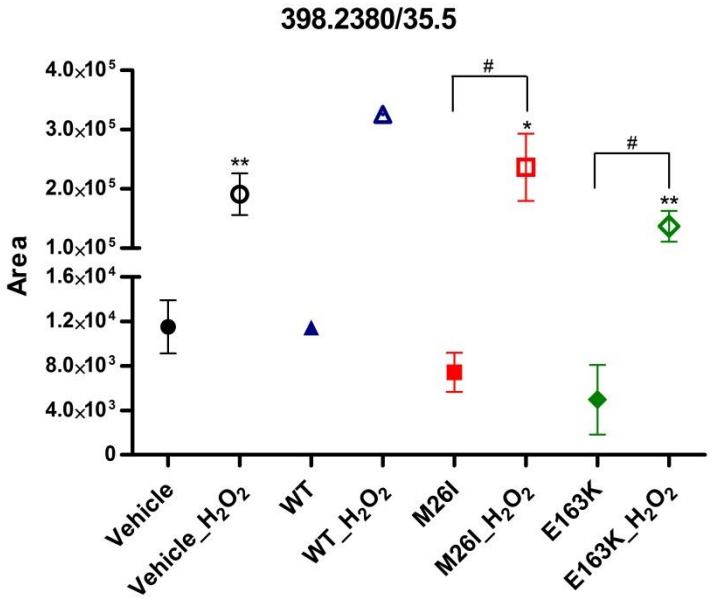


Figure 4.25 - Peak areas of the ion with m/z value of 398.2 and RT of 35.5 minutes in the 8 different biological conditions tested. For this ion there was no fragment to be monitored. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$ when compared to Vehicle. #$p < 0.05$ treatment with 200 µM H_2O_2 and M26I or E163K vs. corresponding treatment with M26I or E163K.

The results show the division of the ions in two main groups: a group of ions more intense in the conditions where the cells were stimulated with 200 µM H_2O_2 ((2) Vehicle_ H_2O_2; (4) WT_ H_2O_2; (6) M26I_ H_2O_2; and (8) E163K_ H_2O_2) (Figure 4.24 and 4.25), and a group of ions more intense in the conditions where the cells were not under oxidative stress ((1) Vehicle; (3) WT; (5) M26I; and (7) E163K) (Figure 4.26-4.31). This trend is confirmed by the quantification of the fragments of the precursor ions. The precursor ion with m/z value of 398.2 was not fragmented as can be seen by its fragmentation spectrum (Supplementary Figure 8.3), probably because the collision energy applied was not enough.

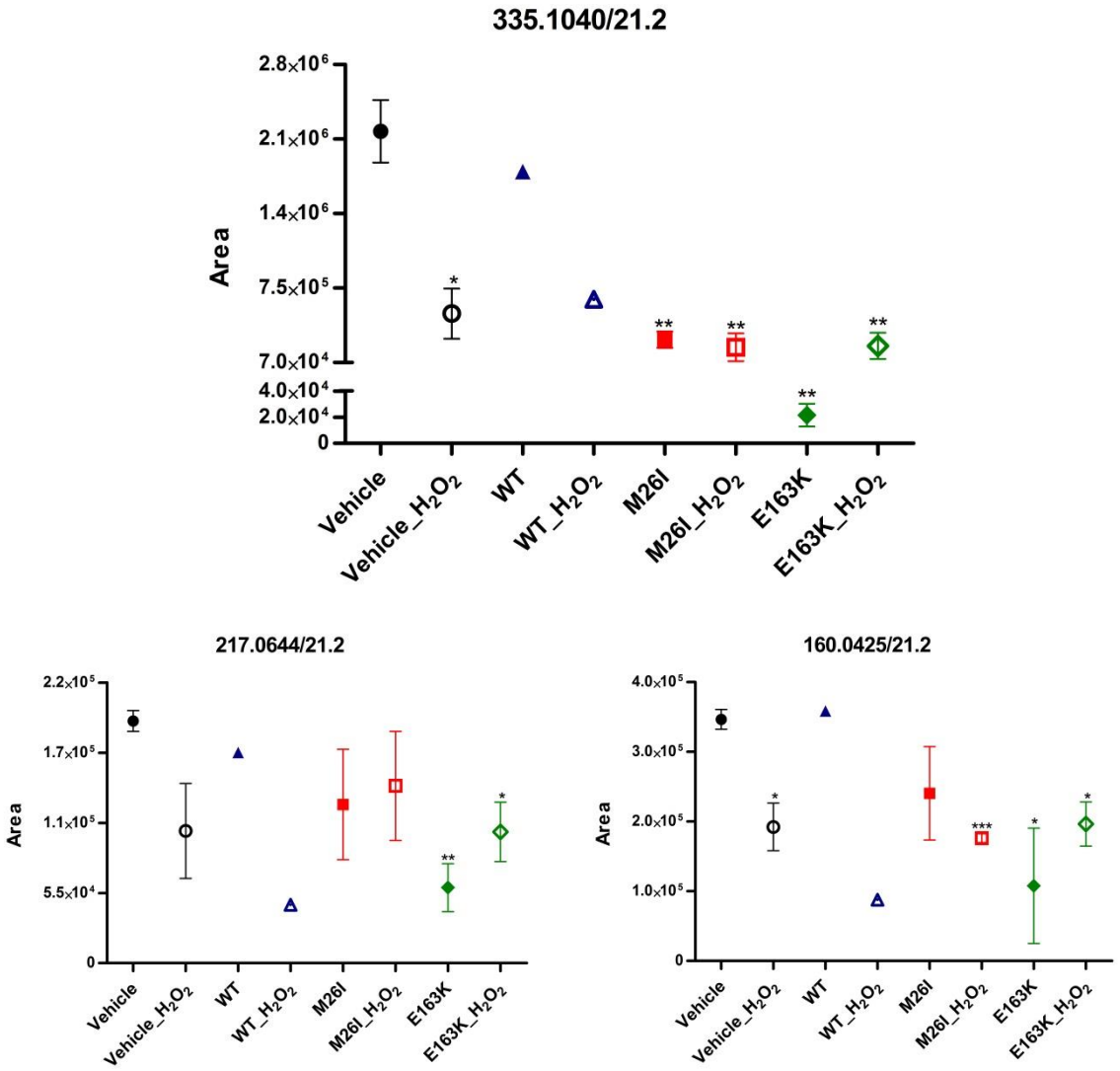


Figure 4.26 - Peak areas of the precursor ion with m/z value of 335.1 and RT of 21.2 minutes (upper panel) and its fragments ions with m/z value of 217.1 (lower panel, left side) and 160.0 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ when compared to Vehicle.

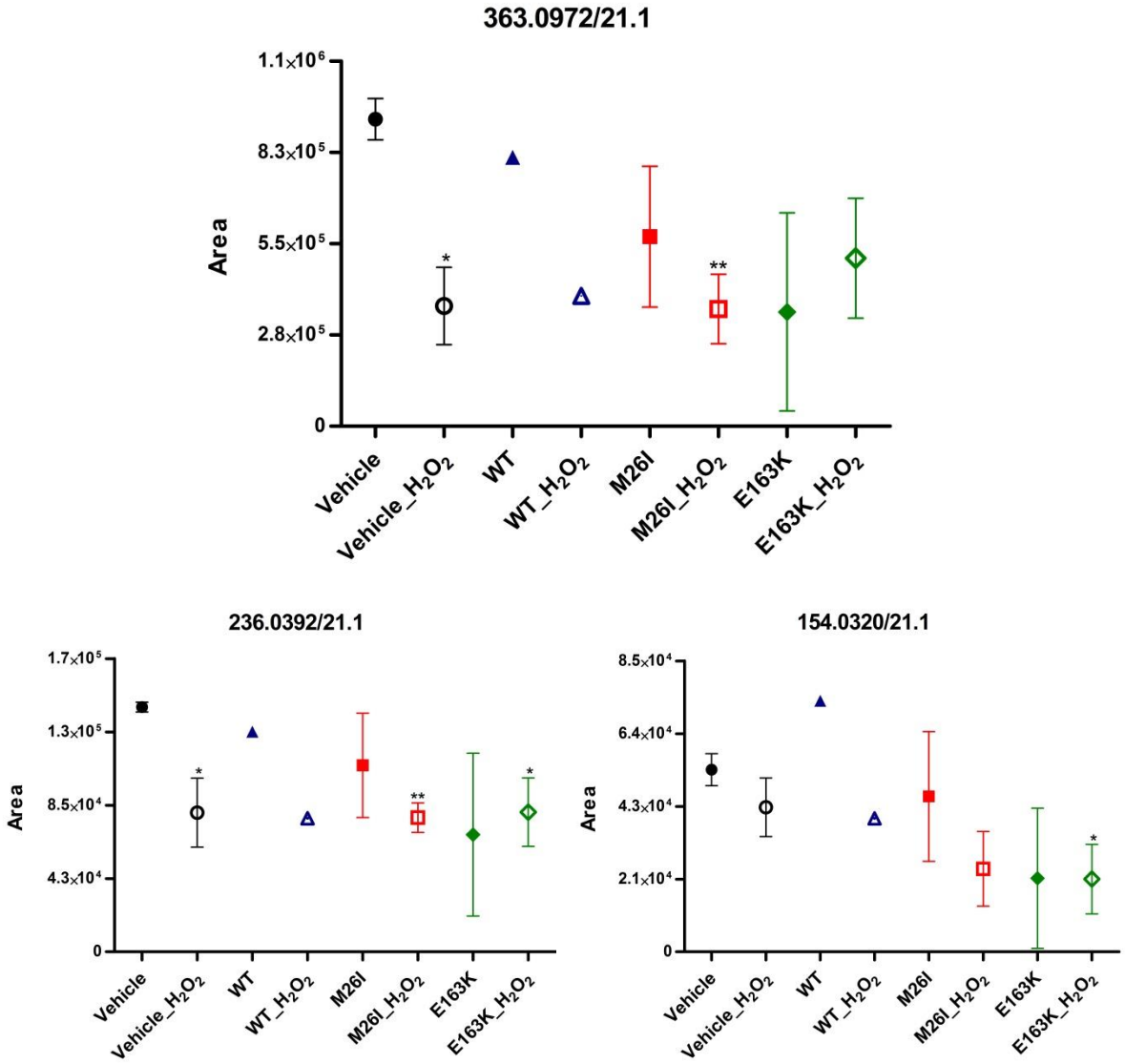


Figure 4.27 - Peak areas of the precursor ion with m/z value of 363.1 and RT of 21.1 minutes (upper panel) and its fragments ions with m/z value of 236.0 (lower panel, left side) and 154.0 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$ when compared to Vehicle.

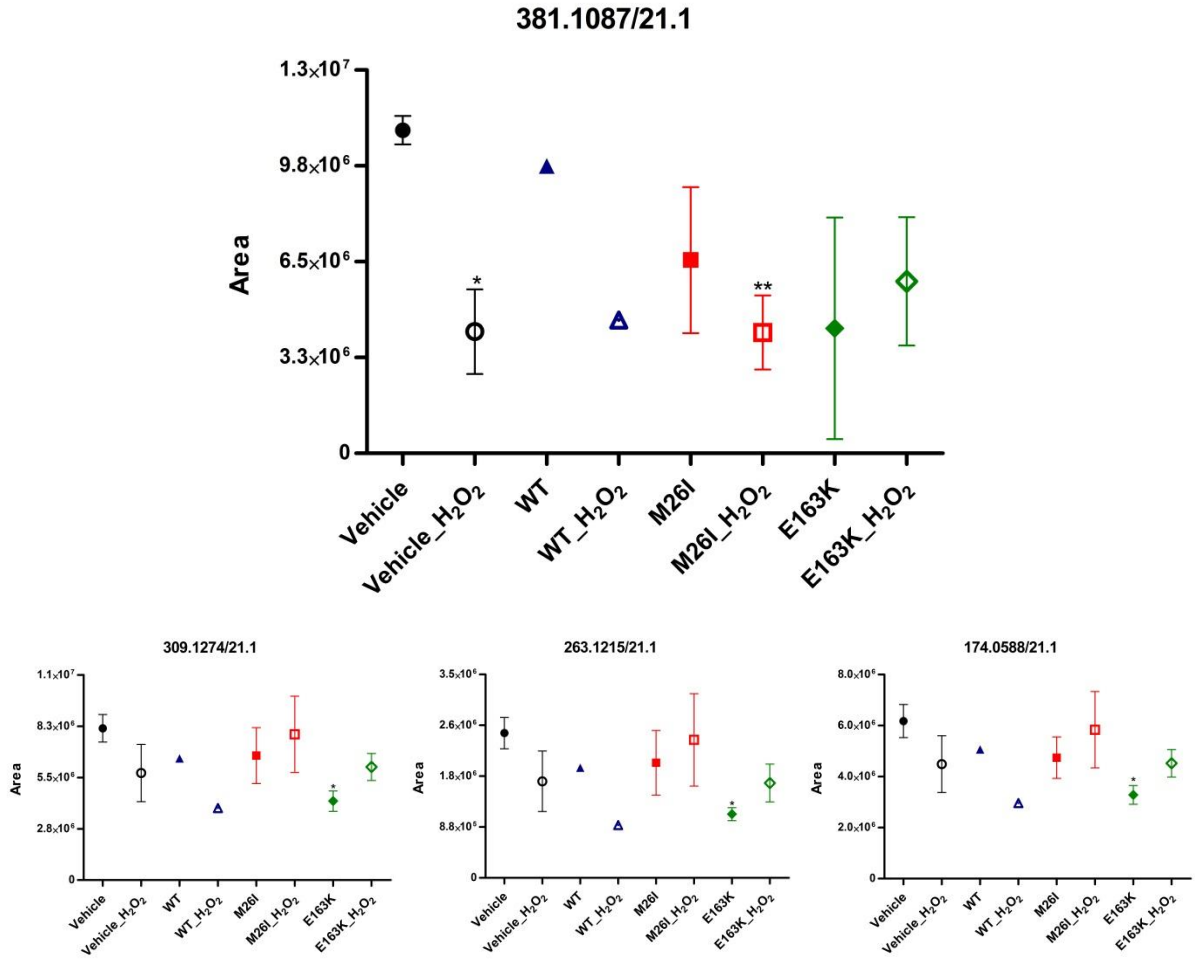


Figure 4.28 - Peak areas of the precursor ion with m/z value of 381.1 and RT of 21.1 minutes (upper panel) and its fragments ions with m/z value of 309.1 (lower panel, left side), 263.1 (lower panel, middle) and 174.1 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$ when compared to Vehicle.

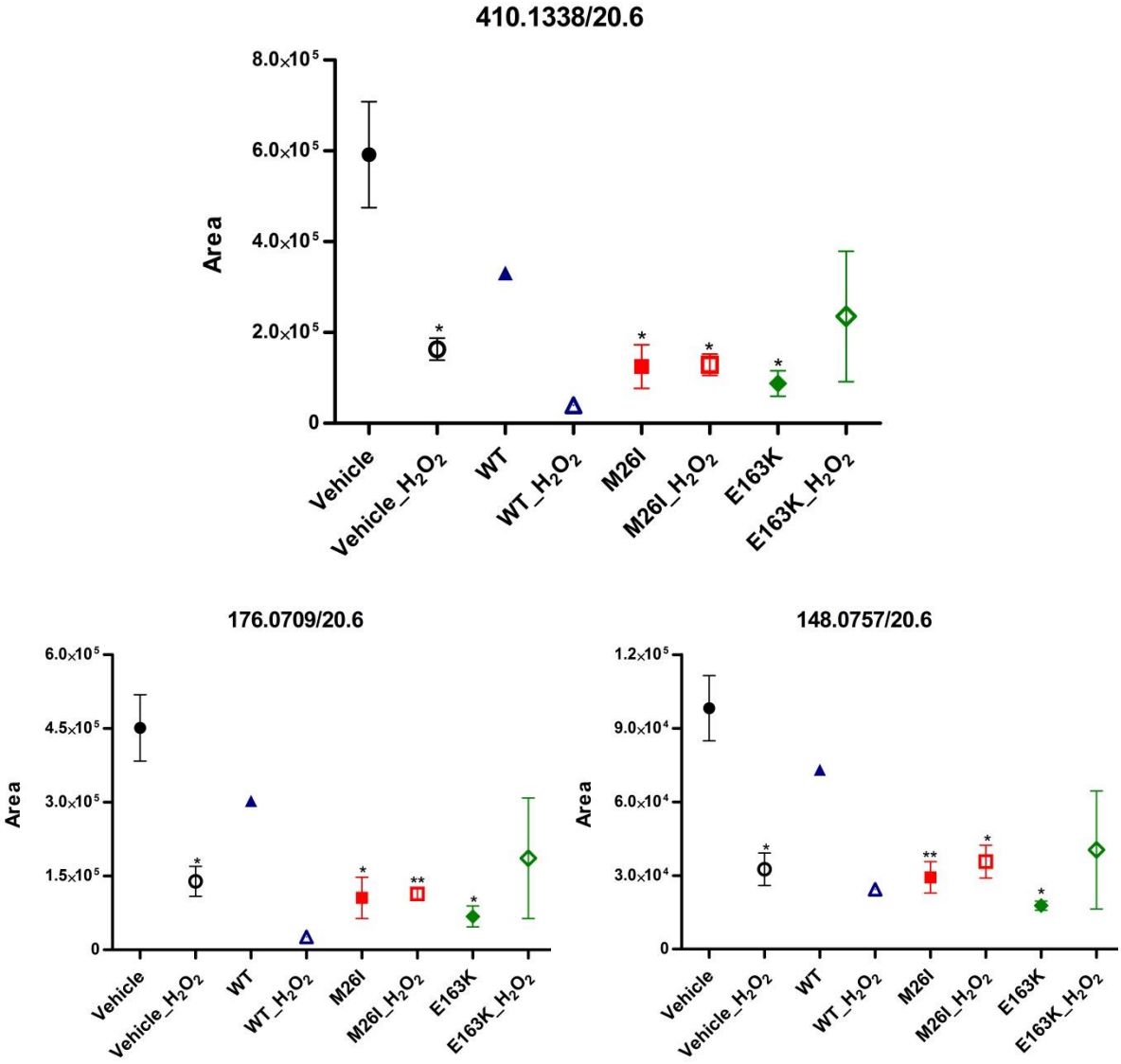


Figure 4.29 - Peak areas of the precursor ion with m/z value of 410.1 and RT of 20.6 minutes (upper panel) and its fragments ions with m/z value of 176.1 (lower panel, left side) and 148.1 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): *$p < 0.05$, **$p < 0.01$ when compared to Vehicle.

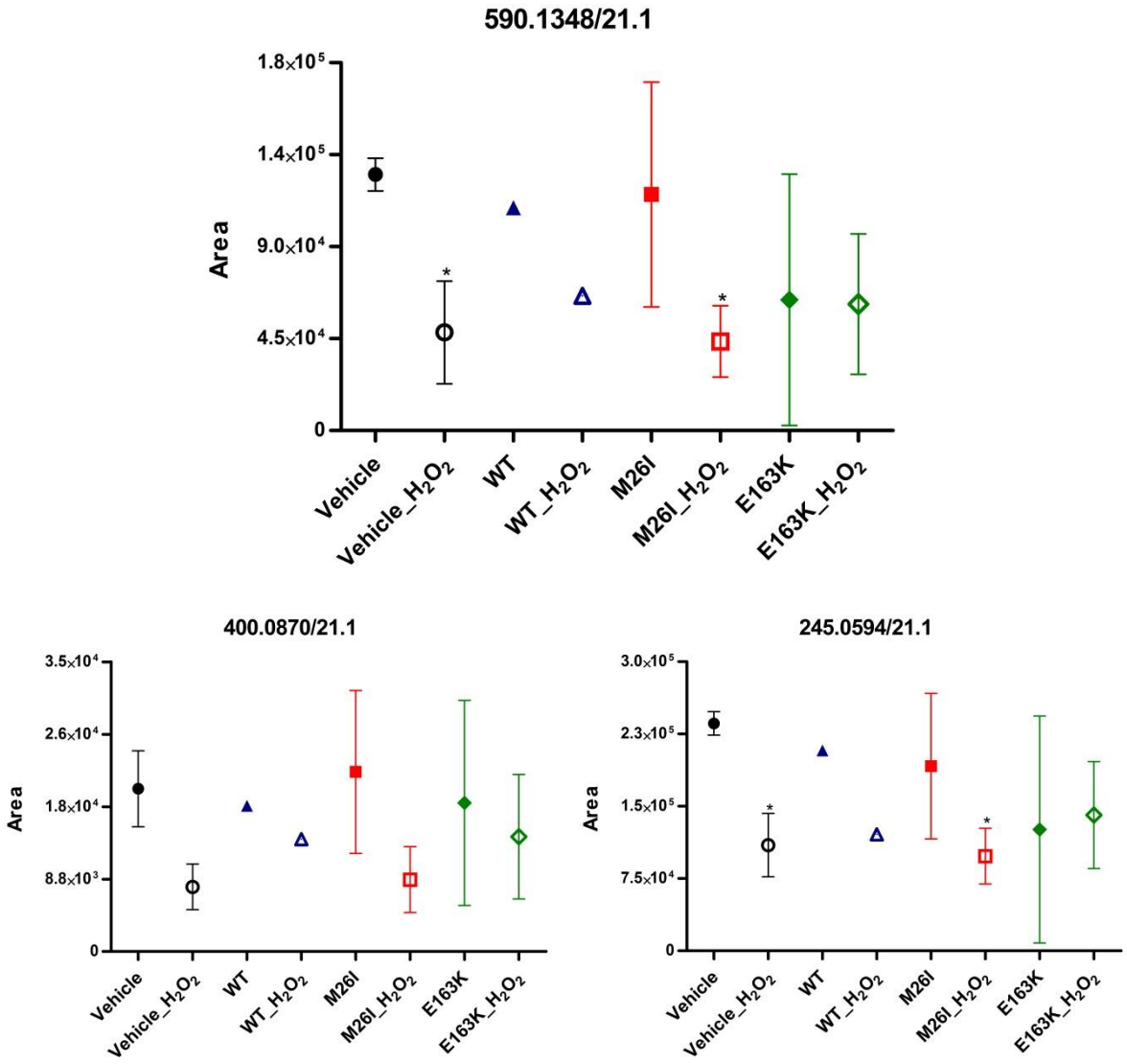


Figure 4.30 - Peak areas of the precursor ion with m/z value of 590.1 and RT of 21.1 minutes (upper panel) and its fragments ions with m/z value of 400.1 (lower panel, left side) and 245.1 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): **p* < 0.05 when compared to Vehicle.

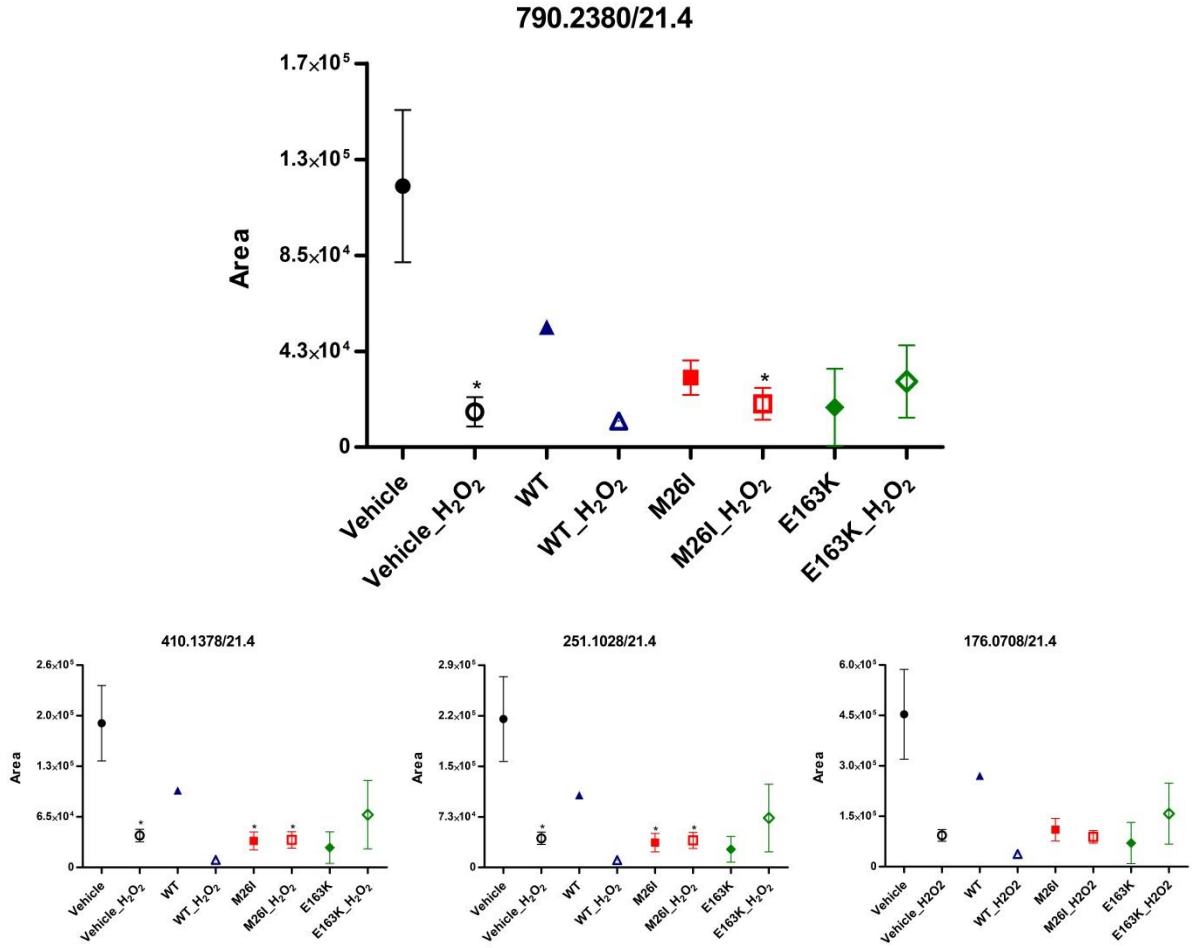


Figure 4.31 - Peak areas of the precursor ion with m/z value of 790.2 and RT of 21.4 minutes (upper panel) and its fragments ions with m/z value of 410.1 (lower panel, left side), 251.1 (lower panel, middle) and 176.1 (lower panel, right side) in the 8 different biological conditions tested. Data represent the mean±SEM of three independent experiments. Significance (Student's two-tailed *t*-test): $*p < 0.05$ when compared to Vehicle.

When a PCA analysis was performed with only these 24 features for all the 19 samples (data not shown), the representation of scores showed that no evident separation was achieved between the 8 different biological conditions tested in this study: (1) Vehicle; (2) Vehicle_ H_2O_2; (3) WT; (4) WT_ H_2O_2; (5) M26I; (6) M26I_ H_2O_2; (7) E163K; and (8) E163K_ H_2O_2. There is a huge variability between the replicates of each biological condition and also between different biological conditions as can be observed by the SEM of the the majority of the samples. Therefore, the replicate number for each sample must be increased also to try to find more specific ion groups within the different biological samples.

❺ DISCUSSION

5. DISCUSSION

Parkinson's disease, a neurodegenerative movement disorder, is associated with selective degeneration of nigrostriatal dopamine neurons [120]. Although the underlying mechanisms contributing to neurodegeneration in PD seem to be multifactorial, oxidative stress is widely considered to be central to many forms of the disease [120]. The *DJ-1 gene,* a causative gene for familial PD, expresses the multifunctional DJ-1 protein. DJ-1 is a stress sensor and some DJ-1 mutations lead to loss of this function [121]. If this happens it may occur a homeostatic imbalance in cell system and metabolites, which can be used as cellular markers of stress conditions.

Taking these facts together, this study intended to give some insights about the metabolites that are significantly altered in resting and oxidative stress conditions, and access also the effect of the addition of recombinant DJ-1 WT and mutants to SH-SY5Y cell line, widely used as an *in vitro* model of dopaminergic neurons for Parkinson's disease studies, under normal and oxidative stress conditions [118].

Three recombinant DJ-1 mutants were produced in order to be used in further studies of DJ-1-mediated neuroprotection against H_2O_2-induced cell death and DJ-1 metabolomics study.

All *DJ-1* point mutations (L166P, M26I and E163K) were generated by site-directed mutagenesis of the sequence of DJ-1 coding DNA in the pSKB-3 vector (Figure 3.1) and the mutations were confirmed by Sanger DNA sequencing (Figure 4.1, Figure 4.2 and Figure 4.3, respectively). Until positive mutation confirmation, all the steps involved in this mutagenesis technique had to be optimized.

The three recombinant mutant proteins were expressed and purified by affinity and gel filtration chromatographies (Figure 4.4, Figure 4.6, Figure 4.8, Figure 4.10, and Figure 4.12). The L166P mutant was produced in three different batches due to the low protein content obtained. The difficulty to express and purify this mutant has been reported in many studies [69, 84, 86, 93, 95, 96]**.**

The recombinant DJ-1 WT, DJ-1 M26I and DJ-1 E163K amino acid sequences only lack the initial methionine (Figure 4.15 A, C and D), due to a natural post-translational

process. On the other hand the recombinant DJ-1 L166P-3 sequence was also identified but with lower sequence coverage (Figure 4.15 B).

The produced recombinant proteins have some protein contaminants (Figure 4.14), but these are mainly proteins from the LB medium, used for bacterial cultures and proteins from the host (*E. coli*) used to clone and express the proteins of interest (Table 4.1). Moreover, when the number of unique peptides as well as the unused score of the purified recombinant DJ-1 WT and mutants are compared with those of their contaminants any effect of them on the function of the DJ-1 proteins can be devalued. The contaminants found to be from *Homo sapiens* can be probably from *Bos taurus* due to the bovine genome not being fully sequenced yet and the UniProt_SwissProt protein database performs a protein search based on the existing sequenced genomes translated to observed and/or predicted protein sequences.

DJ-1 WT, DJ-1 M26I and DJ-1 E163K presented the predicted molecular weight of its monomers (≈23 kDa) (Figure 4.16 and Table 4.2). For the recombinant DJ-1 L166P-3 it was not possible to determine the molecular weight of its monomeric form because it was in an aggregate form.

The molecular weights of the purified DJ-1 WT, DJ-1 M26I and DJ-1 E163K were ≈41 kDa, ≈42 kDa and ≈37 kDa, respectively (Figure 4.17), corresponding to their dimeric form [116], which is an important functional feature of the protein [113-115]. There is a small discrepancy in the molecular weight of the three proteins that can be explained by the chromatographic technique used. The HPLC-Size Exclusion Chromatography is a technique of low resolution when compared with LC-MS (used in determining the molecular weight of the protein monomers) and therefore it does not provide as accurate results as LC-MS. Thus, the molecular weights of the three recombinant proteins were considered near of those corresponding to the dimeric form. The molecular weights of the purified DJ-1 L166P-1 and DJ-1 L166P-2 were ≈91 kDa, and ≈86 kDa, respectively (Figure 4.17), which do not correspond to the proteins dimer form. The molecular weight of the purified DJ-1 L166P-3 could not be calculated because this protein was eluted in the column void volume determined by the elution volume for Blue Dextran (≈2000 kDa).and for this reason it is out of the HPLC-Size Exclusion Chromatography calibration

curve (Supplementary Figure 8.1). The results indicated that this mutant was in an aggregated state or it existed in the form of protein oligomer or even protein polymer, as it has been reported [52]**.**

DJ-1 WT protein showed to be a well-structured protein containing both α-helical and β-sheet secondary structure at 37 °C, as well as M26I and E163K mutants (Figure 4.18 and Table 4.3). However, overall the mutants showed a lower content of secondary structure with the negative effect being most pronounced for the E163K mutant. The X-ray structures analysis previously reported (Table 4.3) was in agreement with the results obtained by CD for DJ-1 WT but not with the CD results for M26I and E163K mutants. For L166P-1 and L166P3 mutants, the CD analysis at 25 °C revealed that the proteins had a very low content of defined secondary structure and a high content of no regular structure, as it has been reported [51, 96, 122], with the negative effect being most pronounced for the L166P-3 (Supplementary Figure 8.2 and Supplementary Table 8.6). No X-ray crystallographic structures were found for L166P mutant**.** As the analysis at 25 °C of L166P-1 and L166P-3 CD spectra (Supplementary Figure 8.2) showed that these proteins did not have a defined secondary structure, when the analysis of the DJ-1 proteins was performed at 37 °C the L166P mutant was not included.

DJ-1 WT was the most stable protein, with a melting temperature of approximately 63.5 °C (Figure 4.19 A and B). Various studies have reported different DJ-1 WT melting temperatures ranging from 60 °C to 77 °C, so the melting temperature calculated for the DJ-1 WT in this project was as expected [51, 76, 86, 91, 123]. The E163K mutant is the least stable protein, with a melting temperature of approximately 54.1 °C, that it was in accordance with the only melting temperature study performed for this mutant (55.1 °C) [91]**.** The M26I mutant was less stable than DJ-1 WT but more stable than E163K mutant, by presenting a melting temperature of approximately 60.0 °C. Some studies have reported melting temperatures for this mutant ranging from 52 °C to 69 °C, therefore, the melting temperature calculated in this project was in the expected range. Because the biological assays were performed at 37 °C and the melting temperatures of the three proteins were found to be higher than 37 °C, the three proteins were considered to be stable. For L166P-1 and L166P-3 mutants it was not possible to

determine their melting temperatures (Figure 4.20, lower panel) since they did not present an usual profile of denaturation (Figure 4.20, upper panel). Therefore the L166P mutant was thermally unstable and it had melting profiles of a typical aggregated protein [52, 84, 96]**.**

Based on the structural characterization described for the recombinant DJ1-M26I and E163K mutants, as well as the DJ-1 WT, all proteins were included in further biological assays. However the L166P mutant did not assemble the necessary structural conditions to proceed to the biological assays.

SH-SY5Y cells were sensitive to the oxidative stress-inducer hydrogen peroxide (Figure 4.21, blank bars), as expected, once the treatment of cells with 100 and 200 µM H_2O_2 induced cell death (as previously observed in our research group, but with higher cellular viability – data not shown). However, an inhibition of cell death, as result of the presence of recombinant DJ-1 WT protein, was not observed (gray bars). Thus, the inhibition of H_2O_2-induced cell death by the exogenous addition of this recombinant DJ-1 WT protein was not confirmed in this study (in contrary to the results previously observed – data not shown). As the treatment of cells with 200 µM H_2O_2 showed to be statistically effective to induce cell death, this H_2O_2 concentration was used in the next task of this project to generate oxidative stress.

In order to identify the metabolites that were significantly altered in resting and oxidative stress conditions, and also access the effect of the addition of DJ-1 WT and mutants to the SH-SY5Y cell line under such conditions, a metabolomics study was carried out. The analysis between the Vehicle and Vehicle_H_2O_2 conditions, once they represent the basis of normal condition and oxidative stress condition, respectively, allowed the detection of 109 statistically different peaks from where only eight were considered to be responsible by the separation of these two groups. This was the first study for oxidative stress metabolomics profiling with the exogenous addition of DJ-1 WT and mutants, and showed the division of the features in two main groups: a group of features more intense in the conditions where the cells were stimulated with 200 µM H_2O_2 (Vehicle_ H_2O_2; WT_ H_2O_2; M26I_ H_2O_2; and E163K_ H_2O_2) (Figure 4.24 and 4.25), and a group of features more intense in the conditions where the cells were not under oxidative stress (Vehicle; WT;

M26I; and E163K) (Figure 4.26-4.31). However, the effect of the addition of recombinant DJ-1 WT and mutants to cells under normal and oxidative stress conditions was inconclusive so far. So, further studies are required to better elucidate the role of DJ-1 mediated metabolic changes in neuronal protection mechanisms.

❻ CONCLUSION

6. CONCLUSION

The main goal of this study was to give some clues about the role that DJ-1 plays in oxidative stress, one of the most relevant insults in Parkinson's disease pathology, by studying mutant forms of this protein. DJ-1 mutations may lead to a homeostatic imbalance in cell system and metabolites, which can be used as cellular markers of stress conditions. To access the rule that DJ-1 mutants perform in metabolite modulation, three recombinant DJ-1 mutants were produced and exogenously added to cultured SH-SY5Y cells, a model cell line to study Parkinson's disease.

This research project allowed the production and structural characterization of three recombinant DJ-1 mutant forms (DJ-1 L166P, DJ-1 M26I and DJ-1 E163K). However, the use of L166P mutant in the biological assays was not possible since it was not in its dimeric (functional) form. The other two mutants and WT protein were found in the dimeric and functional form, and their use in biological assays was allowed.

This study did not reveal any evidence for the neuroprotection of SH-SY5Y cells conferred by the exogenous addition of recombinant DJ-1 WT under H_2O_2-induced oxidative stress conditions. However, H_2O_2 showed to be an efficient oxidative stress agent causing significant cell death in this cell line.

The metabolomics study developed allowed the finding of eight interesting features responsible for the existence of two main groups: a group with two more intense features in the conditions where a H_2O_2 oxidative stress stimulus was applied and another group with six more intense features in the conditions where no oxidative stress stimulus was applied. However, the effect of the addition of recombinant DJ-1 WT and mutants to cells under normal and oxidative stress conditions was inconclusive. So, further studies are required to evaluate the role of DJ-1 in metabolomic modulation under oxidative stress conditions.

This study, the first for oxidative stress metabolomics profiling with the exogenous addition of DJ-1 WT and mutants, allowed the finding of eight possible oxidative stress biomarkers. In the future, these results must be validated in a targeted analysis, for metabolite ID verification, quantitation, functional interpretation, and pathway analysis,

to try to understand their modulation by DJ-1 and their potential use as oxidative stress markers and latter as Parkinson´s disease biomarkers. Hence, these findings may contribute to future strategies for the treatment and prevention of the disease and offer new directions for recognizing disease-specific biochemical indicators.

❼ REFERENCES

7. REFERENCES

1. Zhou, J.-y., et al., *Parkinson's Disease: Insights from the Laboratory and Clinical Therapeutics*, in *Senescence*, T. Nagata, Editor 2012.
2. Lesage, S. and A. Brice, *Parkinson's disease: from monogenic forms to genetic susceptibility factors.* Human Molecular Genetics, 2009. **18**(R1): p. R48-R59.
3. Bekris, L.M., I.F. Mata, and C.P. Zabetian, *The Genetics of Parkinson Disease.* Journal of Geriatric Psychiatry and Neurology, 2010. **23**(4): p. 228-242.
4. Winklhofer, K.F. and C. Haass, *Mitochondrial dysfunction in Parkinson's disease.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2010. **1802**(1): p. 29-44.
5. Farrer, M.J., *Genetics of Parkinson disease: paradigm shifts and future prospects.* Nat Rev Genet, 2006. **7**(4): p. 306-18.
6. Crosiers, D., et al., *Parkinson disease: Insights in clinical, genetic and pathological features of monogenic disease subtypes.* Journal of Chemical Neuroanatomy, 2011. **42**(2): p. 131-141.
7. Cook, C., C. Stetler, and L. Petrucelli, *Disruption of Protein Quality Control in Parkinson's Disease.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(5).
8. Corti, O., S. Lesage, and A. Brice, *What Genetics Tells us About the Causes and Mechanisms of Parkinson's Disease.* Physiological Reviews, 2011. **91**(4): p. 1161-1218.
9. Dickson, D.W., *Parkinson's Disease and Parkinsonism: Neuropathology.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(8).
10. Cookson, Mark R., *Evolution of Neurodegeneration.* Current Biology, 2012. **22**(17): p. R753-R761.
11. Davie, C.A., *A review of Parkinson's disease.* British Medical Bulletin, 2008. **86**(1): p. 109-127.
12. Collier, T.J., N.M. Kanaan, and J.H. Kordower, *Ageing as a primary risk factor for Parkinson's disease: evidence from studies of non-human primates.* Nat Rev Neurosci, 2011. **12**(6): p. 359-66.
13. Hindle, J.V., *Ageing, neurodegeneration and Parkinson's disease.* Age and Ageing, 2010. **39**(2): p. 156-161.
14. Klein, C. and A. Westenberger, *Genetics of Parkinson's Disease.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(1).
15. Nuytemans, K., et al., *Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update.* Human Mutation, 2010. **31**(7): p. 763-780.
16. Pilsl, A. and K. Winklhofer, *Parkin, PINK1 and mitochondrial integrity: emerging concepts of mitochondrial dysfunction in Parkinson's disease.* Acta Neuropathologica, 2012. **123**(2): p. 173-188.
17. Migliore, L. and F. Coppedè, *Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2009. **667**(1–2): p. 82-97.
18. Blandini, F. and M.-T. Armentero, *Animal models of Parkinson's disease.* FEBS Journal, 2012. **279**(7): p. 1156-1166.
19. Olanow, C.W. and W.G. Tatton, *ETIOLOGY AND PATHOGENESIS OF PARKINSON'S DISEASE.* Annual Review of Neuroscience, 1999. **22**(1): p. 123-144.
20. Muramatsu, Y. and T. Araki, *Glial Cells as a Target for the Development of New Therapies for Treating Parkinson's Disease.* Drug News Perspect, 2002. **15**(9): p. 586-590.
21. Hauser, D.N. and T.G. Hastings, *Mitochondrial dysfunction and oxidative stress in Parkinson's disease and monogenic parkinsonism.* Neurobiology of Disease, 2012(0).

22. Fitzgerald, J.C. and H. Plun-Favreau, *Emerging pathways in genetic Parkinson's disease: Autosomal-recessive genes in Parkinson's disease – a common pathway?* FEBS Journal, 2008. **275**(23): p. 5758-5766.
23. Cookson, M.R., *Chapter 3 - Unravelling the role of defective genes*, in *Progress in Brain Research*, B. Anders and M.A. Cenci, Editors. 2010, Elsevier. p. 43-57.
24. Wider, C., T. Foroud, and Z.K. Wszolek, *Clinical implications of gene discovery in Parkinson's disease and parkinsonism.* Movement Disorders, 2010. **25**(S1): p. S15-S20.
25. Surmeier, D.J., et al., *Physiological Phenotype and Vulnerability in Parkinson's Disease.* Cold Spring Harbor Perspectives in Medicine, 2012.
26. Haugarvoll, K., et al., *Lrrk2 R1441C parkinsonism is clinically similar to sporadic Parkinson disease.* Neurology, 2008. **70**(16 Pt 2): p. 1456-60.
27. Chartier-Harlin, M.-C., et al., *α-synuclein locus duplication as a cause of familial Parkinson's disease.* The Lancet, 2004. **364**(9440): p. 1167-1169.
28. Martin, I., V.L. Dawson, and T.M. Dawson, *Recent Advances in the Genetics of Parkinson's Disease.* Annual Review of Genomics and Human Genetics, 2011. **12**(1): p. 301-325.
29. Pankratz, N., et al., *Mutations in DJ-1 are rare in familial Parkinson disease.* Neuroscience Letters, 2006. **408**(3): p. 209-213.
30. Kitada, T., et al., *Considerations regarding the etiology and future treatment of autosomal recessive versus idiopathic Parkinson disease.* Curr Treat Options Neurol, 2012. **14**(3): p. 230-40.
31. Kitada, T., et al., *Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.* Nature, 1998. **392**(6676): p. 605-8.
32. Bonifati, V., et al., *Mutations in the DJ-1 Gene Associated with Autosomal Recessive Early-Onset Parkinsonism.* Science, 2003. **299**(5604): p. 256-259.
33. Canet-Avilés, R.M., et al., *The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(24): p. 9103-9108.
34. Moore, D.J., et al., *A missense mutation (L166P) in DJ-1, linked to familial Parkinson's disease, confers reduced protein stability and impairs homo-oligomerization.* Journal of Neurochemistry, 2003. **87**(6): p. 1558-1567.
35. Zhang, L., et al., *Mitochondrial localization of the Parkinson's disease related protein DJ-1: implications for pathogenesis.* Human Molecular Genetics, 2005. **14**(14): p. 2063-2073.
36. Wilhelmus, M.M.M., et al., *Involvement and interplay of Parkin, PINK1, and DJ1 in neurodegenerative and neuroinflammatory disorders.* Free Radical Biology and Medicine, 2012. **53**(4): p. 983-992.
37. Bandopadhyay, R., et al., *The expression of DJ-1 (PARK7) in normal human CNS and idiopathic Parkinson's disease.* Brain, 2004. **127**(2): p. 420-430.
38. Zhou, W. and C.R. Freed, *DJ-1 Up-regulates Glutathione Synthesis during Oxidative Stress and Inhibits A53T α-Synuclein Toxicity.* Journal of Biological Chemistry, 2005. **280**(52): p. 43150-43158.
39. Batelli, S., et al., *DJ-1 modulates alpha-synuclein aggregation state in a cellular model of oxidative stress: relevance for Parkinson's disease and involvement of HSP70.* PLoS ONE, 2008. **3**(4): p. e1884.
40. Nagakubo, D., et al., *DJ-1, a Novel Oncogene Which Transforms Mouse NIH3T3 Cells in Cooperation withras.* Biochemical and Biophysical Research Communications, 1997. **231**(2): p. 509-513.
41. Abou-Sleiman, P.M., et al., *The role of pathogenic DJ-1 mutations in Parkinson's disease.* Annals of Neurology, 2003. **54**(3): p. 283-286.
42. Moore, D.J., et al., *Association of DJ-1 and parkin mediated by pathogenic DJ-1 mutations and oxidative stress.* Human Molecular Genetics, 2005. **14**(1): p. 71-84.

43. Taira, T., et al., *DJ-1 has a role in antioxidative stress to prevent cell death.* EMBO Rep, 2004. **5**(2): p. 213-8.
44. Bonifati, V., et al., *DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism.* Neurological Sciences, 2003. **24**(3): p. 159-160.
45. Hague, S., et al., *Early-onset Parkinson's disease caused by a compound heterozygous DJ-1 mutation.* Annals of Neurology, 2003. **54**(2): p. 271-274.
46. van Duijn, C.M., et al., *PARK7, a Novel Locus for Autosomal Recessive Early-Onset Parkinsonism, on Chromosome 1p36.* The American Journal of Human Genetics, 2001. **69**(3): p. 629-634.
47. Annesi, G., et al., *DJ-1 mutations and parkinsonism-dementia-amyotrophic lateral sclerosis complex.* Annals of Neurology, 2005. **58**(5): p. 803-807.
48. Macedo, M.G., et al., *Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson's disease.* Movement Disorders, 2009. **24**(2): p. 196-203.
49. Clark, L.N., et al., *Analysis of an early-onset Parkinson's disease cohort for DJ-1 mutations.* Movement Disorders, 2004. **19**(7): p. 796-800.
50. Hedrich, K., et al., *The R98Q variation in DJ-1 represents a rare polymorphism.* Annals of Neurology, 2004. **55**(1): p. 145-145.
51. Malgieri, G. and D. Eliezer, *Structural effects of Parkinson's disease linked DJ-1 mutations.* Protein Science, 2008. **17**(5): p. 855-868.
52. Anderson, P.C. and V. Daggett, *Molecular basis for the structural instability of human DJ-1 induced by the L166P mutation associated with Parkinson's disease.* Biochemistry, 2008. **47**(36): p. 9380-93.
53. Dauer, W. and S. Przedborski, *Parkinson's Disease: Mechanisms and Models.* Neuron, 2003. **39**(6): p. 889-909.
54. Moore, D.J., et al., *MOLECULAR PATHOPHYSIOLOGY OF PARKINSON'S DISEASE.* Annual Review of Neuroscience, 2005. **28**(1): p. 57-87.
55. Sai, Y., et al., *The Parkinson's disease-related genes act in mitochondrial homeostasis.* Neuroscience & Biobehavioral Reviews, 2012. **36**(9): p. 2034-2043.
56. Yan, M.H., X. Wang, and X. Zhu, *Mitochondrial defects and oxidative stress in Alzheimer disease and Parkinson disease.* Free Radic Biol Med, 2012.
57. Itoh, K., et al., *Mitochondrial dynamics in neurodegeneration.* Trends in Cell Biology, 2012(0).
58. Morán, M., et al., *Mitochondrial respiratory chain dysfunction: Implications in neurodegeneration.* Free Radical Biology and Medicine, 2012. **53**(3): p. 595-609.
59. Van Laar, V.S. and S.B. Berman, *The interplay of neuronal mitochondrial dynamics and bioenergetics: Implications for Parkinson's disease.* Neurobiology of Disease, 2012(0).
60. McCoy, M.K. and M.R. Cookson, *DJ-1 regulation of mitochondrial function and autophagy through oxidative stress.* Autophagy, 2011. **7**(5): p. 531-532.
61. Thomas, K.J., et al., *DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy.* Human Molecular Genetics, 2010. **20**(1): p. 40-50.
62. Hayashi, T., et al., *DJ-1 binds to mitochondrial complex I and maintains its activity.* Biochem Biophys Res Commun, 2009. **390**(3): p. 667-72.
63. Perier, C. and M. Vila, *Mitochondrial biology and Parkinson's disease.* Cold Spring Harb Perspect Med, 2012. **2**(2): p. a009332.
64. Janda, E., et al., *Defective autophagy in Parkinson's disease: role of oxidative stress.* Mol Neurobiol, 2012. **46**(3): p. 639-61.
65. Bandopadhyay, R. and J. de Belleroche, *Pathogenesis of Parkinson's disease: emerging role of molecular chaperones.* Trends in Molecular Medicine, 2010. **16**(1): p. 27-36.
66. Martinat, C., et al., *Sensitivity to Oxidative Stress in DJ-1-Deficient Dopamine Neurons: An ES- Derived Cell Model of Primary Parkinsonism.* PLoS Biol, 2004. **2**(11): p. e327.

67. Andres-Mateos, E., et al., *DJ-1 gene deletion reveals that DJ-1 is an atypical peroxiredoxin-like peroxidase.* Proceedings of the National Academy of Sciences, 2007. **104**(37): p. 14807-14812.
68. Bonifati, V., *Autosomal recessive, early-onset Parkinson's disease*2003.
69. Macedo, M.G., et al., *The DJ-1L166P mutant protein associated with early onset Parkinson's disease is unstable and forms higher-order protein complexes.* Human Molecular Genetics, 2003. **12**(21): p. 2807-2816.
70. Junn, E., et al., *Mitochondrial localization of DJ-1 leads to enhanced neuroprotection.* Journal of Neuroscience Research, 2009. **87**(1): p. 123-129.
71. Irrcher, I., et al., *Loss of the Parkinson's disease-linked gene DJ-1 perturbs mitochondrial dynamics.* Human Molecular Genetics, 2010. **19**(19): p. 3734-3746.
72. Blackinton, J., et al., *Formation of a Stabilized Cysteine Sulfinic Acid Is Critical for the Mitochondrial Function of the Parkinsonism Protein DJ-1.* Journal of Biological Chemistry, 2009. **284**(10): p. 6476-6485.
73. Wilson, M.A., *The role of cysteine oxidation in DJ-1 function and dysfunction.* Antioxid Redox Signal, 2011. **15**(1): p. 111-22.
74. Ren, H., et al., *Oxidized DJ-1 Interacts with the Mitochondrial Protein BCL-XL.* Journal of Biological Chemistry, 2011. **286**(40): p. 35308-35317.
75. Foti, R., et al., *Parkinson Disease-associated DJ-1 Is Required for the Expression of the Glial Cell Line-derived Neurotrophic Factor Receptor RET in Human Neuroblastoma Cells.* Journal of Biological Chemistry, 2010. **285**(24): p. 18565-18574.
76. Hulleman, J.D., et al., *Destabilization of DJ-1 by Familial Substitution and Oxidative Modifications: Implications for Parkinson's Disease†.* Biochemistry, 2007. **46**(19): p. 5776-5789.
77. Waak, J., et al., *Oxidizable Residues Mediating Protein Stability and Cytoprotective Interaction of DJ-1 with Apoptosis Signal-regulating Kinase 1.* Journal of Biological Chemistry, 2009. **284**(21): p. 14245-14257.
78. Chen, J., L. Li, and L.-S. Chin, *Parkinson disease protein DJ-1 converts from a zymogen to a protease by carboxyl-terminal cleavage.* Human Molecular Genetics, 2010. **19**(12): p. 2395-2408.
79. Shendelman, S., et al., *DJ-1 Is a Redox-Dependent Molecular Chaperone That Inhibits α-Synuclein Aggregate Formation.* PLoS Biol, 2004. **2**(11): p. e362.
80. Xu, J., et al., *The Parkinson's disease-associated DJ-1 protein is a transcriptional co-activator that protects against neuronal apoptosis.* Human Molecular Genetics, 2005. **14**(9): p. 1231-1241.
81. Takahashi, K., et al., *DJ-1 positively regulates the androgen receptor by impairing the binding of PIASx alpha to the receptor.* J Biol Chem, 2001. **276**(40): p. 37556-63.
82. Kim, R.H., et al., *Hypersensitivity of DJ-1-deficient mice to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyrindine (MPTP) and oxidative stress.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(14): p. 5215-5220.
83. Ramsey, C.P. and B.I. Giasson, *The E163K DJ-1 mutant shows specific antioxidant deficiency.* Brain Research, 2008. **1239**(0): p. 1-11.
84. Repici, M., et al., *Parkinson's disease-associated mutations in DJ-1 modulate its dimerization in living cells.* J Mol Med (Berl), 2012.
85. Ramsey, C.P., *Studies of DJ-1, parkin and alpha-synuclein give insights into plausible mechanisms for Parkinson's disease pathogenesis*, 2010.
86. Logan, T., L. Clark, and S.S. Ray, *Engineered Disulfide Bonds Restore Chaperone-like Function of DJ-1 Mutants Linked to Familial Parkinson's Disease.* Biochemistry, 2010. **49**(27): p. 5624-5633.

87. Kahle, P.J., J. Waak, and T. Gasser, *DJ-1 and prevention of oxidative stress in Parkinson's disease and other age-related disorders.* Free Radical Biology and Medicine, 2009. **47**(10): p. 1354-1361.
88. Ramsey, C.P. and B.I. Giasson, *L10p and P158DEL DJ-1 mutations cause protein instability, aggregation, and dimerization impairments.* Journal of Neuroscience Research, 2010. **88**(14): p. 3111-3124.
89. Ghazavi, F., et al., *PRKN, DJ-1, and PINK1 screening identifies novel splice site mutation in PRKN and two novel DJ-1 mutations.* Movement Disorders, 2011. **26**(1): p. 80-89.
90. Djarmati, A., et al., *Detection of Parkin (PARK2) and DJ1 (PARK7) mutations in early-onset Parkinson disease: Parkin mutation frequency depends on ethnic origin of patients.* Hum Mutat, 2004. **23**(5): p. 525.
91. Lakshminarasimhan, M., et al., *Structural Impact of Three Parkinsonism-Associated Missense Mutations on Human DJ-1†,‡.* Biochemistry, 2008. **47**(5): p. 1381-1392.
92. Takahashi-Niki, K., et al., *Reduced anti-oxidative stress activities of DJ-1 mutants found in Parkinson's disease patients.* Biochemical and Biophysical Research Communications, 2004. **320**(2): p. 389-397.
93. Blackinton, J., et al., *Effects of DJ-1 mutations and polymorphisms on protein stability and subcellular localization.* Molecular Brain Research, 2005. **134**(1): p. 76-83.
94. Miller, D.W., et al., *L166P Mutant DJ-1, Causative for Recessive Parkinson's Disease, Is Degraded through the Ubiquitin-Proteasome System.* Journal of Biological Chemistry, 2003. **278**(38): p. 36588-36595.
95. Görner, K., et al., *Differential Effects of Parkinson's Disease-associated Mutations on Stability and Folding of DJ-1.* Journal of Biological Chemistry, 2004. **279**(8): p. 6943-6951.
96. Olzmann, J.A., et al., *Familial Parkinson's Disease-associated L166P Mutation Disrupts DJ-1 Protein Folding and Function.* Journal of Biological Chemistry, 2004. **279**(9): p. 8506-8515.
97. Zhou, B., et al., *LC-MS-based metabolomics.* Mol Biosyst, 2012. **8**(2): p. 470-81.
98. Xiao, J.F., et al., *LC-MS based serum metabolomics for identification of hepatocellular carcinoma biomarkers in Egyptian cohort.* J Proteome Res, 2012. **11**(12): p. 5914-23.
99. Melamud, E., L. Vastag, and J.D. Rabinowitz, *Metabolomic Analysis and Visualization Engine for LC–MS Data.* Analytical Chemistry, 2010. **82**(23): p. 9818-9826.
100. Cubbon, S., et al., *Metabolomic applications of HILIC–LC–MS.* Mass Spectrometry Reviews, 2010. **29**(5): p. 671-684.
101. Kaddurah-Daouk, R. and K.R. Krishnan, *Metabolomics: a global biochemical approach to the study of central nervous system diseases.* Neuropsychopharmacology, 2009. **34**(1): p. 173-86.
102. Čuperlović-Culf, M., et al., *Cell culture metabolomics: applications and future directions.* Drug Discovery Today, 2010. **15**(15–16): p. 610-621.
103. Gerlach, M., et al., *Biomarker candidates of neurodegeneration in Parkinson's disease for the evaluation of disease-modifying therapeutics.* Journal of Neural Transmission, 2012. **119**(1): p. 39-52.
104. Wang, J., et al., *Biomarkers of Parkinson's disease: current status and future perspectives.* Drug Discovery Today, 2012(0).
105. Waragai, M., et al., *α-Synuclein and DJ-1 as Potential Biological Fluid Biomarkers for Parkinson's Disease.* International Journal of Molecular Sciences, 2010. **11**(11): p. 4257-4266.
106. Salvesen, L., et al., *The DJ-1 concentration in cerebrospinal fluid does not differentiate among parkinsonian syndromes.* Parkinsonism & Related Disorders, 2012. **18**(7): p. 899-901.
107. Wu, Y., W. Le, and J. Jankovic, *PReclinical biomarkers of parkinson disease.* Archives of Neurology, 2011. **68**(1): p. 22-30.

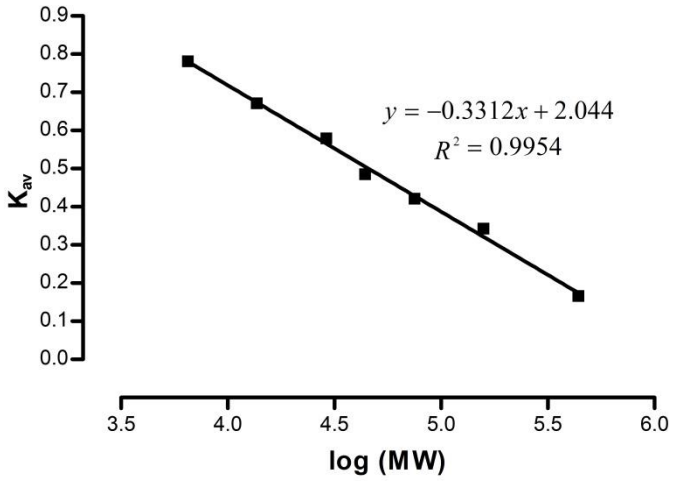
108. Lin, X., et al., *DJ-1 isoforms in whole blood as potential biomarkers of Parkinson disease.* Sci Rep, 2012. **2**: p. 954.
109. Choi, J., et al., *Oxidative Damage of DJ-1 Is Linked to Sporadic Parkinson and Alzheimer Diseases.* Journal of Biological Chemistry, 2006. **281**(16): p. 10816-10824.
110. Mellick, G.D., et al., *Exploiting the potential of molecular profiling in Parkinson's disease: current practice and future probabilities.* Expert Review of Molecular Diagnostics, 2010. **10**(8): p. 1035-1050.
111. Caudle, W.M., et al., *Using 'omics' to define pathogenesis and biomarkers of Parkinson's disease.* Expert Review of Neurotherapeutics, 2010. **10**(6): p. 925-942.
112. Morgan, J., S. Mehta, and K. Sethi, *Biomarkers in Parkinson's Disease.* Current Neurology and Neuroscience Reports, 2010. **10**(6): p. 423-430.
113. Lee, S.J., et al., *Crystal structures of human DJ-1 and Escherichia coli Hsp31, which share an evolutionarily conserved domain.* J Biol Chem, 2003. **278**(45): p. 44552-9.
114. Tao, X. and L. Tong, *Crystal Structure of Human DJ-1, a Protein Associated with Early Onset Parkinson's Disease.* Journal of Biological Chemistry, 2003. **278**(33): p. 31372-31379.
115. Wilson, M.A., et al., *The 1.1-Å resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease.* Proceedings of the National Academy of Sciences, 2003. **100**(16): p. 9256-9261.
116. Honbou, K., et al., *The Crystal Structure of DJ-1, a Protein Related to Male Fertility and Parkinson's Disease.* Journal of Biological Chemistry, 2003. **278**(33): p. 31380-31384.
117. Sturlese, M., *The anti-apoptotic proteins DJ-1 and Mcl-1: molecular basis of different protein-ligand interactions leading to apoptosis.*, 2011.
118. Xie, H.R., L.S. Hu, and G.Y. Li, *SH-SY5Y human neuroblastoma cell line: in vitro cell model of dopaminergic neurons in Parkinson's disease.* Chin Med J (Engl), 2010. **123**(8): p. 1086-92.
119. Yanagida, T., et al., *Oxidative stress induction of DJ-1 protein in reactive astrocytes scavenges free radicals and reduces cell injury.* Oxid Med Cell Longev, 2009. **2**(1): p. 36-42.
120. Sanders, L.H. and J. Timothy Greenamyre, *Oxidative damage to macromolecules in human Parkinson disease and the rotenone model.* Free Radical Biology and Medicine, 2013. **62**(0): p. 111-120.
121. Ariga, H., et al., *Neuroprotective Function of DJ-1 in Parkinson’s Disease.* Oxid Med Cell Longev, 2013. **2013**: p. 9.
122. Alvarez-Castelao, B., et al., *Reduced protein stability of human DJ-1/PARK7 L166P, linked to autosomal recessive Parkinson disease, is due to direct endoproteolytic cleavage by the proteasome.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2012. **1823**(2): p. 524-533.
123. Girotto, S., et al., *Dopamine-derived Quinones Affect the Structure of the Redox Sensor DJ-1 through Modifications at Cys-106 and Cys-53.* Journal of Biological Chemistry, 2012. **287**(22): p. 18738-18749.

❽ SUPPLEMENTARY DATA

8. SUPPLEMENTARY DATA

8.1 HPLC-Size Exclusion Chromatography

In order to assess if the purified DJ-1 has a homodimer form, which is a functional important feature of the protein [113-115], an HPLC-Size Exclusion chromatography was performed. The retention times of standards were used to perform a calibration curve (Supplementary Figure 8.1) and the molecular weights of purified DJ-1 WT and mutants were determined using their retention times.



Supplementary Figure 8.1 – HPLC-Size Exclusion Chromatography Calibration Curve. Molecular Weight Standards from right to left: Ferritin (440 kDa); Aldolase (158 kDa); Conalbumin (75 kDa); Ovalbumin (43 kDa); Carbonic Anhydrase (29 kDa); Ribonuclease A (13.7 kDa); and Aprotinin (6.5 kDa). The mobile phase was PBS with 10% of glycerol (purified recombinant DJ-1 proteins buffer).

K_{av} is the partition coefficient, which is calculated from the measured elution volume (V_e) of each standard protein or proteins of interest, using the equation (8.1):

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \qquad \textbf{(8.1)}$$

Where the V_0 is the column void volume – elution volume for Blue Dextran 2000 –, and V_c is the geometric column volume.

After having the calibration curve of K_{av} versus log molecular weight (Supplementary Figure 8.1), the molecular weights of the proteins of interest can be inferred, using their calculated K_{av}.

8.2 Circular Dichroism Spectroscopy

Supplementary Table 8.1 - Data of DJ-1 WT and Mutants for calculating the Mean Residue Molar Ellipticity, $[\Theta]_{MRW}$, for the CD analysis at 37 °C.

Protein	Molar Mass (g/mol)	Number of residues	Extinction Coefficient ($(mg/ml)^{-1}cm^{-1}$)	A_{280}	Concentration (mg/ml)
WT	22980.78	213	3.4	0.550	1.87
M26I	22962.75	213	3.4	0.609	2.07
E163K	22979.84	213	3.4	0.471	1.60

Supplementary Table 8.2 - CD data (at 37 °C) obtained for purified DJ-1 proteins using the CONTIN program.

	CONTIN program		
Protein	% α-helix	% β-sheet	% Remainder
WT	36.8 ± 4.8	24.9 ± 4.9	38.5 ± 7.2
M26I	30.3 ± 1.3	37.4 ± 6.8	32.1 ± 6.9
E163K	21.5 ± 1.2	25.8 ± 2.8	45.9 ± 3.8

Supplementary Table 8.3 - CD data (at 37 °C) obtained for purified DJ-1 proteins using the three algorithms of the GlobalWorks software and the CLSTR library.

	GlobalWorks software											
Algorithm	CONTILL			CDSSTR			Selcon3			**Average**		
Protein	% α-helix	% β-sheet	% Remainder	% α-helix	% β-sheet	% Remainder	% α-helix	% β-sheet	% Remainder	% α-helix	% β-sheet	% Remainder
WT	37	15	47	40	17	43	38	15	48	**38**	**16**	**46**
WT Rep	36	18	45	38	18	44	38	15	50			
M26I	30	21	49	31	22	44	29	20	50	**30**	**21**	**49**
E163K	24	26	50	31	17	52	26	23	52	**27**	**22**	**51**

Abbreviation: Rep, replicate

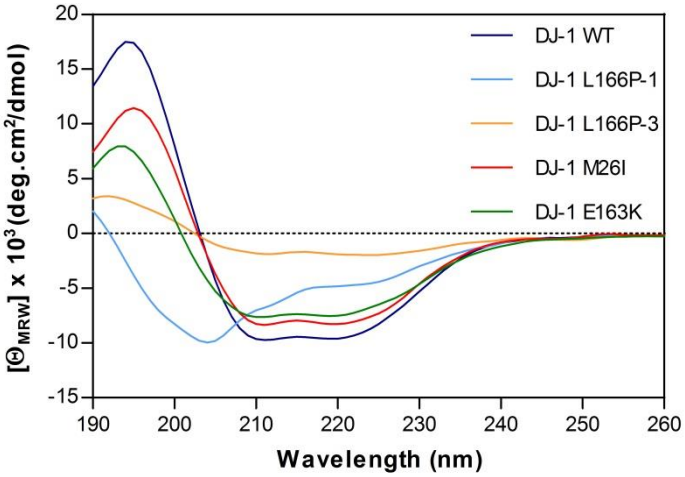
Supplementary Table 8.4 - X-Ray Crystallography Data from PDB of DJ-1 WT and mutants secondary structure.

		X-Ray Crystallography data	
Protein	**PDB entry**	% α-helix	% β-sheet
WT	1J42	44	20
	1P5F	37	20
	1PDW (8 chains)	40	20
		46	20
		35	20
		35	19
		42	20
		35	20
		39	19
		47	20
	Average	**40.0 ± 4.6**	**19.8 ± 0.4**
M26I	2RK4	36	19
E163K	2RK6	36	20

For L166P mutant no entries were found

Supplementary Table 8.5 – Data of DJ-1 WT and Mutants for calculating the Mean Residue Molar Ellipticity, $[\Theta]_{MRW,}$ for the CD analysis at 25 °C.

Protein	Molar Mass (g/mol)	Number of residues	Extinction Coefficient ($(mg/ml)^{-1}cm^{-1}$)	A_{280}	Concentration (mg/ml)
WT	22980.78	213	3.4	0.643	2.19
L166P-1	22964.73	213	3.4	0.380	1.29
L166P-3	22964.73	213	3.4	0.555	1.89
M26I	22962.75	213	3.4	0.586	1.99
E163K	22979.84	213	3.4	0.555	1.89



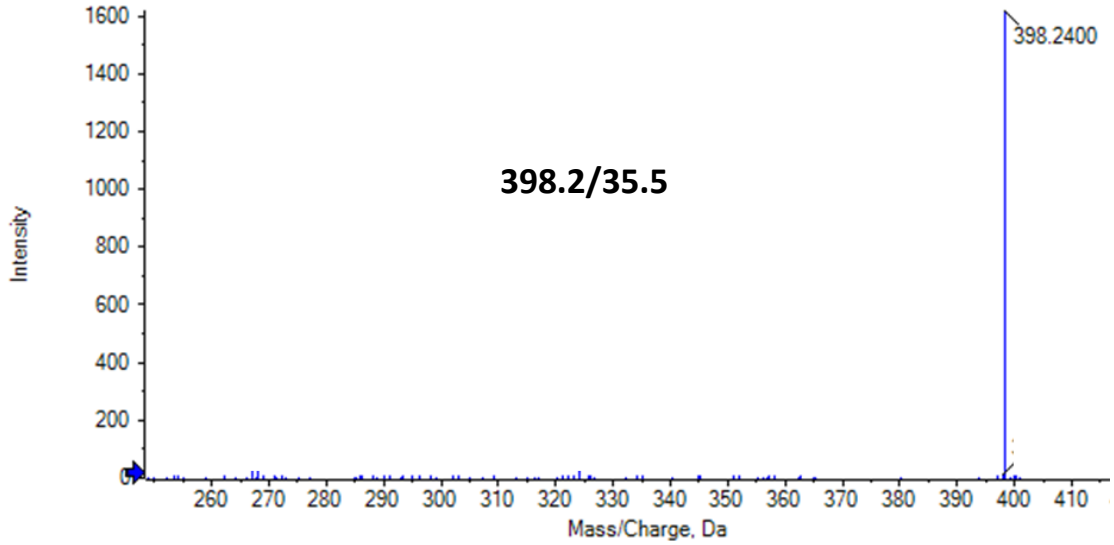
Supplementary Figure 8.2 - Circular Dichroism (CD) spectra (at 25 °C) of DJ-1 WT and Mutants. The secondary structure of DJ-1 WT and mutant proteins was evaluated by CD spectroscopy at 25 °C.

DJ-1 L166P-1 and DJ-1 L166P-3 spectra are clearly different from spectra obtained for the other proteins (Supplementary Figure 8.2). For L166P-1 mutant, the analysis of secondary structure revealed that the protein has 9% to 13% of α-helix structure, 25% to 37% of β-sheet structure and 50% to 66% of no regular structure (Supplementary Table 8.6). For L166P-3, the results are 4% to 8% of α-helix structure, 37% to 51% of β-sheet structure and 41% to 58% of no regular secondary structure (Supplementary Table 8.6) (no PDB entries were found for L166P mutant).

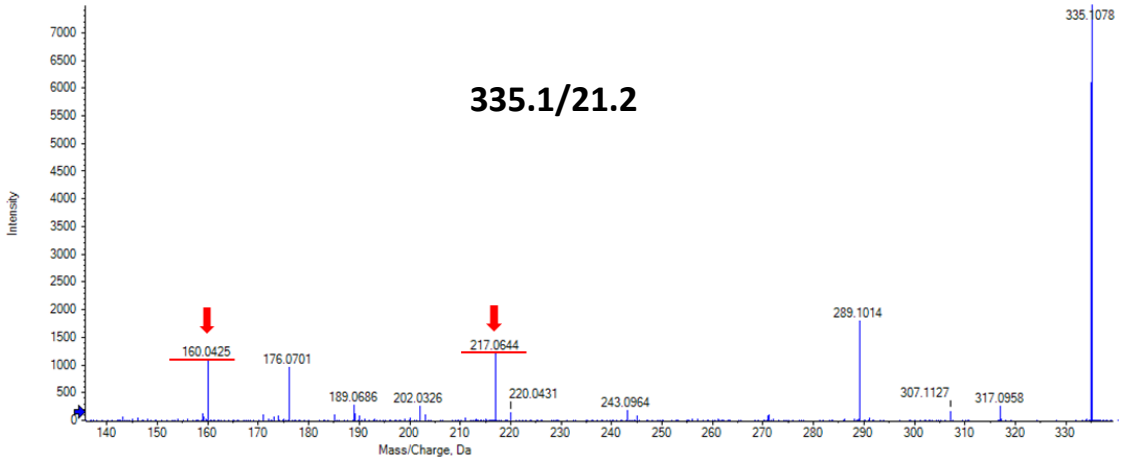
Supplementary Table 8.6 - CD data (at 25 °C) obtained for purified DJ-1 proteins and X-Ray Crystallography data for DJ-1 proteins available in PDB.

	CD data								
	CONTIN Library			GlobalWorks Library (Average of 3 algorithms)			X-Ray Crystallography data (Average of all structures from PDB)		
Protein	α-helix %	β-sheet %	Remainder %	α-helix %	β-sheet %	Remainder %	α-helix %	β-sheet %	Remainder %
WT	32	32	36	32	21	47	40	20	40
L166P-1	13	37	50	9	25	66	-	-	-
L166P-3	8	51	41	4	37	58	-	-	-
M26I	27	32	41	25	26	49	36	19	45
E163K	23	33	44	22	27	53	36	20	44

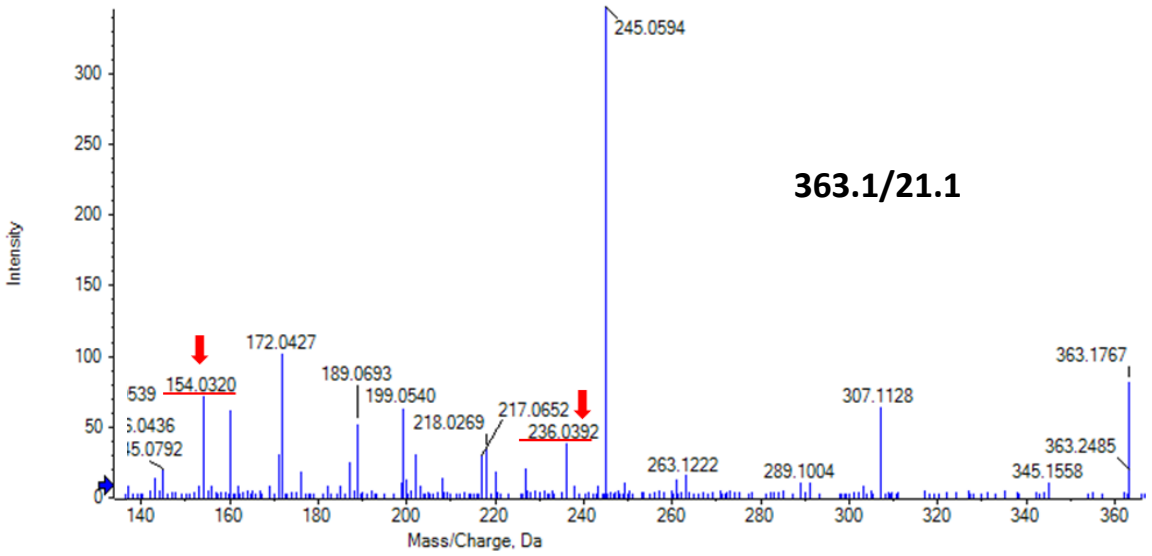
8.3 LC-MS/MS Analysis of Intracellular Metabolites and Data Processing



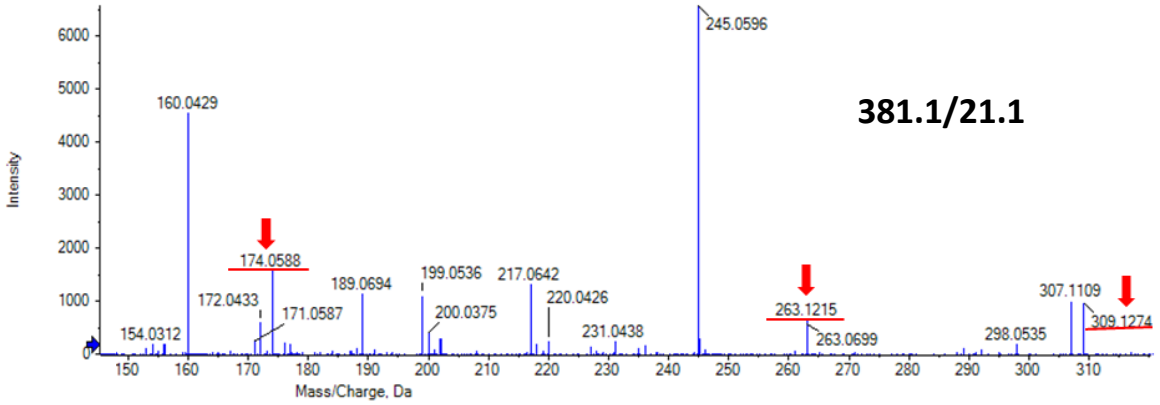
Supplementary Figure 8.3 - Fragmentation spectrum of the precursor ion with m/z value of 398.2 and RT of 35.5 minutes. For this ion there was no fragment to be monitored.



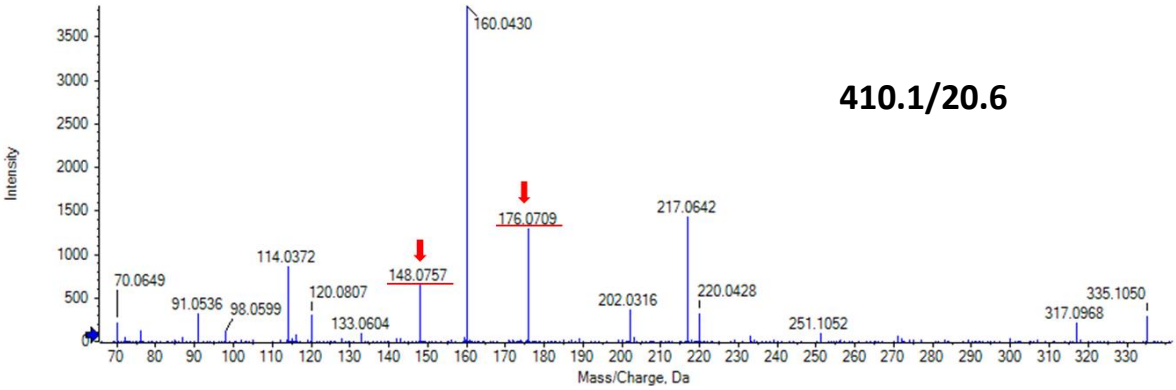
Supplementary Figure 8.4 - Fragmentation spectrum of the precursor ion with m/z value of 335.1 and RT of 21.2 minutes. Red arrows represent the fragments selected for quantification.



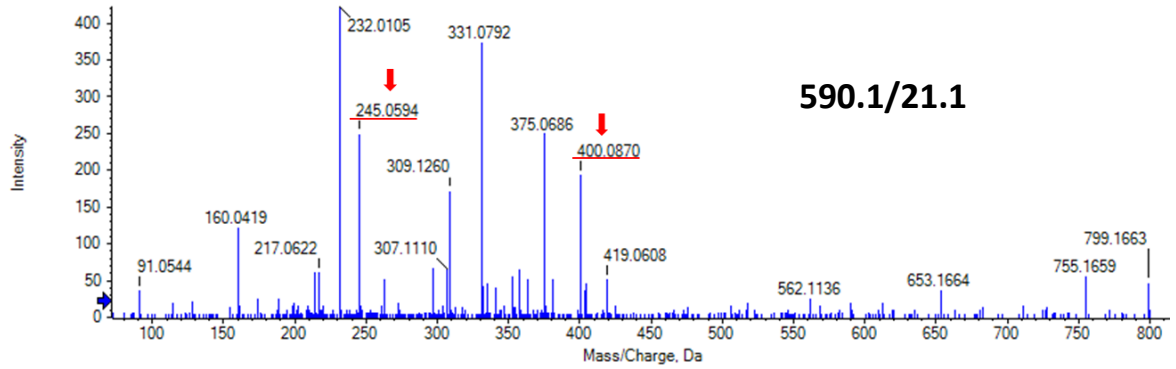
Supplementary Figure 8.5 - Fragmentation spectrum of the precursor ion with m/z value of 363.1 and RT of 21.1 minutes. Red arrows represent the fragments selected for quantification.



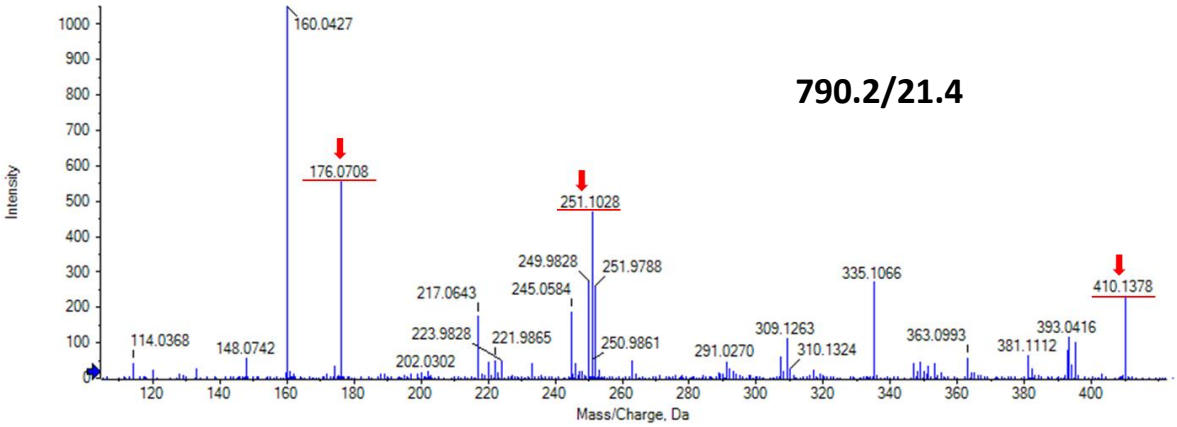
Supplementary Figure 8.6 - Fragmentation spectrum of the precursor ion with m/z value of 381.1 and RT of 21.1 minutes. Red arrows represent the fragments selected for quantification.



Supplementary Figure 8.7 - Fragmentation spectrum of the precursor ion with m/z value of 410.1 and RT of 20.6 minutes. Red arrows represent the fragments selected for quantification.



Supplementary Figure 8.8 - Fragmentation spectrum of the precursor ion with m/z value of 590.1 and RT of 21.1 minutes. Red arrows represent the fragments selected for quantification.



Supplementary Figure 8.9 - Fragmentation spectrum of the precursor ion with m/z value of 790.2 and RT of 21.4 minutes. Red arrows represent the fragments selected for quantification.