B. Introduction.

B.1. Historical events.

In 1972, a new form of dominantly inherited ataxia was described among Portuguese immigrants living in Massachusetts U.S.A. and descendants of an Azorean family called " Machado " (Nakano, *et al.* 1972). In the same year another family of Portuguese descents, who manifested a dominantly inherited nigro-spino-dental degeneration, was reported by Woods, *et al.* (1972). Later on, Rosenberg, *et al.* (1976) reported a third Azorean family, the " Joseph " family, who lived in California, with an autosomal dominant striato-nigral degeneration.

Romanul, *et al.* in 1977 suggested that a mutation in a specific gene could be responsible for all these diseases. They explained the clinicopathological differences among and within the three families as phenotypic variations and they called this new disorder " Machado-Joseph " disease. Later, other families without Portuguese ancestry were found showing the characteristic Machado Joseph disease symptoms in North Carolina (Healton, *et al.* 1980), in Japan (Sakai, *et al.* 1983; Ishino, *et al.* 1971), in North Australia as well as in a Jewish family in Yemen (Goldberg-Stern, *et al.* 1994).

B.2. Pathological aspects.

MJD patients show a moderate atrophy of spinal cord, pons and a shrinkage of the substantia nigra, subthalamic nucleus and globus pallidus. Neuronal loss and gliosis are found in regions of cerebellum, dentate nucleus, pontine nuclei, subthalamic nucleus as well as in pallidum. A more moderate degeneration is present in substantia nigra, brainstem nuclei, the anterior horn of the spinal cord and anterior nerve roots (Takiyama, *et al.* 1994; Schmidt, *et al.* 1998).

B.3. Clinical manifestation.

The MJD disease phenotypes are classified into four types (I to IV) depending on the age of onset and the clinical manifestations (Coutinho, *et al*. 1978; Rosenberg, *et al*. 1983; Barbeau, *et al*. 1984; Rosenberg, *et al*. 1989; Rosenberg, *et al*. 1990; Conner, *et al*. 1993).

Type I: The symptoms are pyramidal and extrapyramidal signs and the disease begins between the teens and the 30s.

Type II: The onset is between the late teens and the 40s and the pyramidal and cerebellar signs are more evident than the extrapyramidal signs.

Type III: The disease starts during the 40s and 60s and the signs are cerebellar.

Type IV: This is a very rare form of MJD and manifests it self as Parkinsonism combined with peripheral neurophaty.

Since the age of onset and the degree of symptoms correlate with the number of CAG repeats, a classification of the disease based on the number of CAG repeats is possible (Nakamoto, *et al.* 1998).

B.4. Molecular mechanisms for inherited neurodegenerative disorders.

Although the inherited neurodegenerative disorders caused by CAG repeats (HD, SCA1, SCA2, SCA3/MJD, SCA6, SCA7, SBMA, DRPLA) manifest diverse symptoms, they share common features: (i) the inheritance of most of these disorders is autosomal dominant; (ii) the age of onset is generally after the 3rd or 4th decade of life and (iii) in succeeding generation the symptoms become more evident and the ages of onset are earlier. This phenomenon is called " anticipation ". (iiii) Finally, there is a typical neuronal cell loss, although in different regions of the central nervous system. All these features suggest the possibility of a common molecular mechanism which could be responsible for the pathology in all these disorders.

Between 1991 and 1993 the first two genes of these neurodegenerative disorders have been identified. La Spada, *et al.* (1992) reported that the

androgen receptor (AR) gene, located on chromosome Xq11-q12, is responsible for spinobulbar muscular atrophy (SBMA) and The Huntington`s Disease Collaborative Research Group (1993) identified the Huntington`s disease gene and mapped it to chromosome 4p16.3. The two genes showed no homology except for the presence of polyglutamine-coding CAG repeats. The repeats were found to be expanded in all affected individuals and they seemed to be unstably transmitted to the next generations with a tendency for further elongation (Duyao, *et al.* 1993). Therefore, it was proposed that this mutation could be also the cause of other autosomal dominant neurodegenerative disorders. This hypothesis was confirmed in late 1993 and early 1994 when the disease genes for spinocerebellar ataxia 1 (SCA1) and dentatorubralpallidoluysian atrophy (DRPLA) were cloned (Banfi, *et al.* 1993; Koide, *et al.* 1994; Nagafuchi, *et al.* 1994). Both genes contained an elongated CAG expansion encoding an elongated glutamine repeat.

B.5. The Machado-Joseph disease gene (*MJD1*).

In order to find the Machado-Joseph disease gene a human brain cDNA library was screened using an oligonucleotide with 13 CTG repeats as probe (Kawaguchi, *et al.* 1994). Several cDNA fragments containing CAG repeats were isolated. One of them consisted of 1776 base pairs with an open reading frame (ORF) of 1169 bases. The CAG repeat was located close to the 3‘ end of the ORF and was predicted to be translated into a glutamine tract. The amino acid sequence of this cDNA showed no apparent homology to any other protein sequences, except for the glutamine stretch. The mRNA, detected by Northern blot analysis, revealed that the *MJD1* gene was expressed in all human tissues and predominantly in testis. RT-PCR of human brain RNA showed the presence of two cDNAs with different CAG repeat lengths, demonstrating the presence of the two alleles which were polymorphic.

Furthermore, *MJD1* cDNAs splice variants were found creating a different C-terminus of the MJD protein also called ataxin-3 (Goto *et al.* 1997; Schmidt, *et*

al. 1998). In addition, Goto, *et al.* (1997) found that 60% of the *MJD1* alleles in the Japanese population had a polymorphism in the stop codon: The TAA sequence was changed to a tyrosine-coding TAC sequence, resulting in an elongated ORF.

The screening of a human genomic library using the *MJD1* cDNA as probe, resulted in the isolation of the *MJD1* gene fragments together with three other *MJD1* related genes. They were named *MJD2, MJD3* and *MJD4.* Using fluorescence *in situ* hybridisation (FISH) the genomic clones were mapped at position 14q32.1, 8q23, 14q21 and Xp22.1, respectively (Kawaguchi, *et al.* 1994). At the same time, using a genetic linkage analysis on Japanese MJD families**,** the MJD1 gene was finally localised on chromosome 14q24.3-32.1 (Takiyama, *et al.* 1993 and 1994).

B.6. CAG expansions and gene dosage-dependent effects in MJD patients.

PCR amplification of the CAG repeats within the *MJD1* gene resulted in products which correspond to both the normal and the expanded alleles. In normal individuals the CAG repeat is 12-40 in length, whereas in affected individuals it is expanded to 55-84 repeat units in at least one of the two alleles (Lunkes, *et al.* 1997; Kakizuka, *et al.* 1998, Chai, et al. 1999). It has been shown that there is an inverse correlation between the age of onset of symptoms and the number of repeats: The longer the CAG repeat, the earlier the onset of the disease (Nakamoto, *et al.* 1998). Moreover, patients who have homozygous mutations develop an earlier onset disease and show different clinical manifestations indicating a gene dosage effect (Kawakami, *et al.* 1995; Sobue, *et al.* 1996).

B.7. MJD molecular genetics.

MJD is the most frequently observed hereditary ataxia (Ranum, *et al.* 1995). Although it was initially proposed that a common founder chromosome

carrying the mutation is responsible of the disease (Takiyama, *et al.* 1995), further haplotype analysis on affected families from different ethnic origins showed that the *MJD1* mutation could also arise from a *de novo* mutation (Giunti, *et al*. 1995).

B.8. Gender effects in MJD.

A study of a group of Japanese MJD patients showed a gender effect in the enhanced earliness of the ages of onset of MJD in males compared to the females (Kawakami, *et al.* 1995). However the same studies in Europe and North America did not give the same statistical difference. These discrepancies might be due to a different ethnic origin of the two MJD patient groups (Dürr, *et al.* 1995).

B.9. CAG Repeat Instabilities in MJD.

The CAG repeats in the pathological range (Q55-84) show intergenerational instability with expansion of CAGs more evident during the paternal transmission than during maternal transmission (Duyao, *et al.* 1993; Koide, *et al*. 1994; Snell, *et al*. 1993; Orr, *et al*. 1993; Jodice, *et al*. 1994; Warner, *et al*. 1995; Yvert, *et al.* 1999). This is believed to be the genetic explanation of the phenomenon called " anticipation ". In contrast, the CAG repeats in the normal range (Q12-40) are stable in both maternal and paternal transmission. The elongation of the CAG repeats in the paternal transmission occurs during spermatogenesis (Takiyama, *et al.* 1995) and the expanded allele is more frequently transmitted to the next generation (Ikeuchi, *et al.* 1996). It has been shown that the CAG repeat length can differ from tissue to tissue (Cancel, *et al*. 1995). This phenomenon is called " somatic mosaicism " and is observed in MJD patients in cerebellum where the repeat length is smaller than in the other brain and peripheral tissues (Lopes-Cendes, *et al.* 1996).

B.10. Experimental models for MJD.

Different experiments have been performed to study the expression of the *MJD1* gene in eucaryotic cell systems. Ikeda, *et al.* (1996) for example expressed a range of different sizes of ataxin-3 protein in cultured COS-7 cells. Immunocytochemical analysis of cells, transfected with constructs expressing full-length ataxin-3 protein with 26 (MJD26), 35 (MJD35) and 79 (MJD79) glutamines as well as a truncated construct containing a C-terminal part of the protein with 35 (Q35C) glutamines, showed a diffuse cytoplasmatic staining. In contrast, in cells expressing the C-terminal fragment of the ataxin-3 protein with 79 glutamines (Q79C), a punctate cytoplasmic staining was detected. Most of the cells with the punctate staining also showed DNA fragmentation as well as condensed nuclei. Furthermore, the cells started to detach from the dishes 24 hours after transfection. Taken together, this indicates that the expression of a truncated ataxin-3 protein with 79 glutamines induces apoptosis.

To examine whether the polyglutamine stretch itself or the RNA is responsible of the cell death, a fragment containing 79 CAG repeats was cloned into the 3' noncoding region of an eucaryotic expression vector containing 35 CAGs [pCMXHA-Q35-$(CAG)_{79}$]. Cells transfected with the plasmid and expressing the RNA did not show any evidence of apoptosis and cell death. Thus, Ikeda, *et al.* (1996) demonstrated that the glutamines, but not the RNA containing a CAG repeat, is toxic for the cells.

Paulson, *et al.* (1997) showed that expression of truncated fragments of ataxin-3 with a polyglutamine repeat in the pathological range resulted in the formation of nuclear ubiquitinated inclusions. Similar inclusions were also detected in some areas of the MJD disease brain such as pons. In another experiment they demonstrated that the expression of the construct with the C-terminal part of ataxin-3 containing 78 CAGs (Q78C), expressed transiently in 293T cells, resulted in the formation of microaggregates in the intra- and perinuclear space and induced cell death, whereas the full- length ataxin-3 protein with 78 (MJD78) and with 27 glutamines (MJD27) only showed a diffuse cytoplasmic

staining. In other studies performed by Perez, *et al.* (1998) a construct expressing a C-terminal fragment of ataxin-3 protein with 27 glutamines (Q27C) formed intranuclear aggregates when targeted to the nucleus by a nuclear localisation signal (NLS). The authors proposed that the nuclear environment could promote the formation of these aggregates because of its less capacity to metabolise misfolded proteins. Moreover, they could also show that the polyglutamine repeat fusion proteins, HA-Q78C and GFP-Q35, were sufficient to recruit other polyglutamine repeat fusion proteins into aggregates. Perez *et al.* (1998) also observed that the wild-type ataxin-3 protein was sequestrated in aggregates when co-expressed with normal and expanded ataxin-1 protein. They proposed that the recruitment of glutamine-containing proteins into nuclear inclusions with ataxin-3 could be an important aspect of the pathological mechanism in MJD. Very recently, co-localization of the 26S proteasome with polyglutamine aggregates was demonstrated in MJD tissue and in transiently transfected cell lines. Inhibitors of the proteasome caused a repeat length-dependent increase in aggregate formation, providing evidence for involvement of the ubiquitin-proteasome pathway in suppressing polyglutamine aggregation in MJD pathogenesis (Chai, *et al.* 1999). In an other work, Evert *et al.* (1999) using a rat mesencephalic dopaminergic cell line (CSM14.1), stably expressing the full-length *MJD1* cDNA with 70 glutamines (SCA3-Q70), demonstrated the formation of nuclear inclusion bodies and an increased non-apoptotic cell death which was not correlated with the formation of the intranuclear inclusions.

B.11. A transgenic mouse model.

In order to study the expression of ataxin-3 protein *in vivo*, a mouse model has been developed by Ikeda, *et al.* (1996). Two different constructs expressing the C-terminal fragment of ataxin-3 protein with 35 and 79 glutamines (Q79C, Q35C) as well as the full-length ataxin-3 protein with 79 glutamines (MJD79) and a fragment containing only 79 glutamines (Q79), were placed in the second

exon of the L-7 gene and transgenic techniques were used to create transgenic mice. For each construct a F1 mouse generation was obtained. It was found that the size of the CAG repeats was maintained stably during the generations. Neurological analysis of the transgenic mice showed that all the Q79C offspring and part of the Q79, had ataxic signs as well as gait disturbance and postures problems after the 3rd week of age. In contrast, the transgenic mice expressing the truncated ataxin-3 protein (Q35C) and the full-length protein (MJD79) did not show any ataxic symptoms. These results clearly demonstrated that the cleaved protein is much more potent in inducing ataxic phenotypes than the full-length protein.

The ataxic phenotypes were stronger when the transgene Q79C was overexpressed from more copies in the genome. Brain sections obtained from mice overexpressing the Q79C construct showed an atrophic cerebellum with one-eighth of the volume of a normal cerebellum. Histological analysis revealed that all three layers of the cerebellum were affected. The molecular layer was thin whereas the Purkinje cells exhibited scarce dendrite formation and the granule cells seemed to have shrunk.

B.12. An invertebrate transgenic model.

In order to see whether the polyglutamine-caused pathomechanism is conserved in invertebrates, a transgenic *Drosophila* model has been constructed by Warrick, *et al.* (1998) expressing the C-terminal part of ataxin-3 protein with a normal (Q27C) and expanded (Q78C) polyglutamine repeat. The expression of the construct with 78 glutamines, analysed in different tissues, led to formation of nuclear inclusions (NI) and late onset cell degeneration whereas flies expressing the construct with the polyglutamine in the normal range (Q27C) did not show a phenotypic effect. When the expression of the Q78C construct was directed to developing eye by a specific gene promoter, the morphology of the eyes was disrupted and the pigmentation was less than normal. Histological analysis of eye tissues demonstrated cell retina

degeneration with late-stage loss of photoreceptor integrity. In brain, neurones of flies showed particular susceptibilities to the expression of the elongated polyglutamine repeats construct. Other cells appeared to be more resistant to polyglutamine proteins (epithelial cells of the legs and wings) and despite the evidence of nuclear inclusions (NI) these inclusions were not sufficient to cause cell degeneration. Using the viral anti-apoptotic gene *P35*, it has been possible to mitigate *in vivo* the induction of the degeneration, demonstrating also the role of apoptosis in this kind of pathologies.

B.13. Molecular model of the polyglutamine aggregates.

In 1994 Max Perutz proposed the theory of the " polar zipper " to explain how polyglutamine could form aggregates (Perutz, *et al.* 1994; Stott, *et al.* 1995). According to his theory a polyglutamine stretch forms β-strands which are linked together by hydrogen bonds between both their main chain and side chain amides. He proposed that the elongation of the polyglutamine stretch beyond ~41 results in a phase change in conformation from a random coil to a β-pleated secondary structure similar to those found in Aβ protein in Alzheimer (AD) and PrP^{sc} in prion disease (Perutz, *et al.* 1996). Electron microscopic analysis of aggregates in Huntington`s disease and in DRPLA showed that the polyglutamine inclusions were heterogeneous in composition and contained a mixture of filament-, granule-, fibril-like structures (Igarashi, *et al.* 1998; DiFiglia, *et al.* 1997), but they were somehow reminiscent of amyloid-like structures observed in AD and in prions. Scherzinger, *et al.* (1997) demonstrated that fusion proteins between the exon 1 of huntingtin with a polyglutamine stretch in the pathological range (Q51) and glutathione S-transferase (GST) formed insoluble amyloid-like birefringent aggregates *in vitro.* This aggregation was enhanced when the GST was cleaved by factor Xa demonstrating that the processing of the polyglutamine-containing protein by proteinases is a crucial step for the formation of aggregates *in vivo* and *in vitro.*

In another theory, Green (1993) proposed that the polyglutamine stretches may aggregate following covalent linkage by cellular transglutaminases (TGase). This theory was supported by *in vitro* experiments (Kahlem, *et al.* 1998; Igarashi, *et al.* 1998). Furthermore, in a recent article Karpuj *et al.* (1999) demonstrated the presence of increased TGase activity in cell nuclei of Huntington brain.

B.14. The rat and chicken ataxin-3 homolog.

Using the human *MJD1* cDNA Schmitt, *et al.* (1997) screened five rat cDNA libraries constructed from total forebrain, hippocampus, substantia nigra and cerebellum as well as testis RNA. A single rat cDNA clone was identified and named *rsca3*. Sequence comparison of the rat (*rsca3*) and human Machado-Joseph disease gene (*MJD1*) revealed an identity of ~88% at DNA level. The two genes differed strongly in CAG repeat number. Whereas in normal human *MJD1* gene the CAG repeat stretch consisted of 12-40 trinucleotide units (Lunkes, *et al.* 1997; Kakizuka, *et al.* 1998, Chai. *et al.* 1999), the *rsca3* gene harboured only three interrupted CAG triplets (caaCAGcatCAGcaaCAG→QQHQQQ), demonstrating that a longer polyglutamine stretch is not important for the normal function of the ataxin-3 protein in rodents. Furthermore, the C-terminal end of the putative protein differed strongly from the published human ataxin-3 protein sequence. By Northern blot analysis a *rsca3* gene transcript of ~6 kb was detected in most tissues including brain. In testis an additional transcript of ~1.3 kb due to a shorter 3´-untranslated region (UTR) was also identified.

Screening a chicken erythroblast cDNA expression library, using antibody raised against an other protein, Linhartova´*et al.* (1999) found by accident the chicken homolog of ataxin-3 protein. The overall amino acid comparison with the human ataxin-3 resulted in 79% identity. The amino-terminal half of the protein (1-198 aa) showed an higher conservation of 94% identity, whereas the carboxy-terminal half (199-359aa) was found to be only 60% identical.

Sequencing 15 alleles from eight chickens belonging to two different strains, they found no evidence for polyglutamine length polymorphism or genetic instability of the CAG repeat in the chicken ataxin-3. The polyglutamine stretch consisted of seven glutamines encoded by CAAs and CAGs and interrupted by a leucine encoded by a CTA (CAGcaacaacaacta CAGCAGCAG→QQQQKQQQ). The glutamine codon preceding the CTA leucine codon showed a CAA/CAG polymorphism. The authors suggested that the stability of the chicken CAG repeat may be due to the presence of interspersing variant codons (CAA and CTA). Also in chicken a short polyglutamine stretch of seven glutamines seems to be sufficient for the normal function of the ataxin-3 protein.