SEMI-SYNTHESIS OF UBIQUITINATED ANDROGEN RECEPTOR PEPTIDES USING AN EXPRESSED PROTEIN LIGATION SYSTEM

Shaza Mokhtar

Department of Human Genetics
Master of Science

McGill University

Montreal, Quebec, Canada

November, 2013

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Acknowledgements

First of all, I would like to thank Dr. Mark Trifiro and Dr. Lenore Beitel, my supervisors, who taught me all that is required to be a researcher with great guidance and patience. Special thanks to Dr. Trifiro for the wonderful support while I was sick with UC and hospitalized. Also, it is my pleasure to thank all my present and past lab members, including Dr. Shafinaz Chowdhury, Dr. Milt Paliouras, Rose Lumbroso, Carlos Alvarado and Paresa Giannopoulos for their great help and support. Very special thanks to Rose Lumbroso and Carlos Alvarado for their assistance with cloning methods, PCR, tissue culture work, and western-blot methods. Dr. Beitel was the greatest resource for my thesis. This project would not have been possible without her guidance, patience and unmatched resolve.

Finally, the most important support that I got was from my family to move from Saudi Arabia to Canada to achieve my dream and do my degree at McGill. With the help of my parents and my brothers, especially my mother and Hassan, I was able to finish my Master's and I am very grateful for their encouragement. Special thanks to my brother Fares who has accompanied me all the time while I was doing my Master's.

Table of Contents

List of Abbreviations

AA: Amino Acid
AF-1: Activating Function 1
AF-2: Activating Function 2
AIS: Androgen insensitivity syndrome
ALS: Amyotrophic lateral sclerosis
Amp: Ampicillin
AR: Androgen receptor
ARE: Androgen response element
CAG: cytosine-adenine-guanine
CBD: Chitin-binding domain
DBD: DNA-binding domain
DHT: Dihydrotestosterone
DTT: 1, 4-Dithiothreitol
DUB: Deubiquitinase
EPL: Expressed protein ligation
E1: Ubiquitin activating enzyme
E2: Ubiquitin conjugating enzyme
E3: Ubiquitin ligase
FSA: Fluorigenic substrate assay
GFP: Green fluorescent protein
GSH: Glutathione
GST: Glutathione S-transferase
HAT: Histone acetyltransferase
HA: Synthetic peptide (HA epitope)
H2B: Synthetic peptide (H2B epitope)
HD: Huntington's disease
HDAC: Histone deacetylase
HEK: Human Embryonic Kidney
HUB: Human ubiquitin
IMPACT-TWIN: Intein Mediated Purification with an Affinity Chitin-binding Tag-Two Intein
IPTG: Isopropyl Beta-D-1-thiogalactopyranoside
kDa: Kilodalton
LBD: Ligand-binding domain
LBP: Ligand-binding pocket
Luc: Luciferase
MB: Mibolerone
MG132: Z-Leu-Leu-Leu-aldehyde

MESNA: 2-mercapto-ethanesulfonic acid
NLS: Nuclear localization sequence
NR: Nuclear Receptor
OD: Optical density
PIP: Proteasome-interacting protein
PolyQ: Polyglutamine
SBMA: Spinal and Bulbar Muscular Atrophy
SCA: Spinocerebellar ataxia
T: Testosterone
TAD: Transactivation domain
TAF4: TBP-associated factor
TBP: TATA-binding protein
Ub: Ubiquitin
UBA: Ubiquitin-associating domain
UBL: Ubiquitin-like domain
Ub-H: Ubiquitin aldehyde
UBX: Ubiquitin regulatory X
UIM-1: Ubiquitin interacting motif 1
UIM-2: Ubiquitin interacting motif 2
UPS: Ubiquitin proteasome system for protein degradation
YFP: Yellow fluorescent protein

List of Tables and Figures

Abstract

The 90-kb androgen receptor (AR) gene is located at locus Xq11-12 and has eight exons that encode a protein of about 919 amino acids. Exon 1 encodes the transactivational domain, exons 2 and 3 encode the DNA-binding domain, and exons 4 through 8 encode the ligand-binding domain. The AR protein, a nuclear receptor that belongs to the steroid/thyroid hormone receptor family, functions primarily as a DNA-binding transcription factor to regulate gene expression. The AR also has a role in the development of the male phenotype: in fact, cytoplasmic AR is activated by binding testosterone or dihydrotestosterone hormone before translocating to the nucleus.

Spinal and Bulbar Muscular Atrophy (SBMA), also known as Kennedy's disease, is an X-linked recessive neurodegenerative disease. The cause of SBMA is expansion of the CAG trinucleotide repeat, which codes for glutamine (Gln, Q), in Exon 1 of the AR gene. Therefore SBMA is considered to be a polyglutamine (polyGln) expansion disorder. In unaffected men, 9-33 CAG repeats coding for Gln are found, while males who inherit more than 37 CAG repeats will develop SBMA. The toxic gain of AR function that is characteristic of SBMA results in a slow progressive muscle weakness, atrophy of the tongue, fasciculations of bulbar, facial and limb muscles, and muscle cramping. A loss of AR function is related to breast enlargement and reduced fertility.

The ubiquitin-proteasome system (UPS), which functions in protein degradation, plays an important role in neural cells from SBMA patients. UPS components are found in

polyGln-expanded protein aggregates seen in these cells. Therefore previous studies have suggested that the UPS is impaired in SBMA. Previous work suggests that 0QAR and 20QAR can be degraded via the UPS, however, the 50QAR appeared to inhibit the degradation of specific proteins by the UPS.

It is also hypothesized that these SBMA-inducing mutant ARs can accumulate in the cell and cause disease by evading UPS degradation. The primary goal of this research was to develop a system for ubiquitinating AR by using an expressed protein ligation (EPL) system. To prove this hypothesis, several intermediate goals have been achieved. PCR-amplified fragments of AR (0Q, 20Q, and 50Q) were cloned into the pTWIN2 vector and contructs were analyzed using restriction enzyme digests and gel electrophoresis. Furthermore, 0QAR, 20QAR, 50QAR and human ubiquitin (HUB) fusion proteins were purified using chitin beads and cleaved using 2-mercapto-ethanesulfonic acid (MESNA). Finally, the 0QAR, 20QAR, 50QAR peptides were ligated with an HA-tagged H2B peptide using EPL. Thus, we successfully developed tools which can be used to explore ubiquitinatination of the AR using EPL systems in vitro.

Future work aims to ligate the HUB protein to a peptide to using the EPL method, and then ligating the ubiquitinated peptide to the three different AR peptides. These will be added to active proteasomes to explore the involvement of the UPS in the pathogenesis of SBMA.

Abrégé

Le gène de 90-kb du récepteur androgène (AR) est localisé au locus Xq11-12 et possède huit exons codant pour une protéine d'environ 919 acides aminés. L'exon 1 code pour le domaine transactivationnel, les exons 2 et 3 codent pour le domaine de liaison à l'ADN, et les exons 4 à 8 codent pour le domaine de liaison au ligand. La protéine du AR, un récepteur nucléaire appartenant à la famille des récepteurs hormonaux stéroïdiens/thyroïdiens, agit primordialement en tant que facteur transcriptionnel se liant à l'ADN pour réguler l'expression génique. Le AR possède également un rôle dans le développement du phénotype masculin: en effet, le AR cytoplasmique est activé en se liant aux hormones testostérone ou dihydrotestostérone avant de se translocaliser au noyau.

L'amyotrophie bulbo-spinale (SBMA-Spinal and Bulbar Muscular Atrophy), aussi connue sous le nom de maladie de Kennedy, est une maladie neurodégénérative liée au chromosome X récessif. La cause de la SBMA est une expansion du triplet CAG répété, lequel code pour une glutamine (Gln,Q), dans l'exon 1 du gène AR. Ainsi, la SBMA est considérée comme étant un désordre d'expansion polyglutaminique (poly-Gln). Chez l'homme non-affecté, on retrouve 9-33 répétitions CAG codant pour la Gln, alors que l'homme qui hérite plus de 37 répétitions CAG développera la SBMA. Le gain de fonction toxique du AR caractérisant la SBMA se traduit par: une faiblesse musculaire à progression lente, une atrophie de la langue, des fasciculations des muscles bulbaires,

faciaux et limbiques et enfin par des crampes musculaires. Une perte de fonction du AR est liée à l'élargissement de la poitrine et à la fertilité réduite.

Le système ubiquitine-protéasome (UPS), responsable dans la dégradation protéique, joue un rôle important dans les cellules neurales chez les patients atteints de la SBMA. Les composantes de l'UPS se trouvent dans les aggrégats protéiques de l'expansion poly-Gln dans ces cellules. C'est la raison pour laquelle des études précédentes ont suggéré que l'UPS est affaibli dans la SBMA. D'autres études antérieures suggèrent que les peptides 0QAR et 20QAR peuvent être dégradés par l'UPS, cependant le peptide 50QAR semble inhiber la dégradation par lUPS de certaines protéines spécifiques.

L'hypothèse veut que ces mutants AR responsables de la SBMA peuvent s'accumuler dans la cellule et causer la maladie en s'évadant de la dégradation UPS. Le but premier de ce projet était de développer un système d'ubiquitination du AR en appliquant le système de ligation des protéines exprimées (expressed protein ligation-EPL). Afin de prouver cette hypothèse, quelques buts intermédiaires ont été atteints. Les fragments du AR (0Q, 20Q, et 50Q), amplifiés par PCR, ont été clonés dans le vecteur pTWIN2 et les clones ont été analysés sur gels d'électrophorèse suite aux digestions par enzymes de restriction. De plus, les fragments 0QAR, 20QAR et 50QAR ainsi que les protéines de fusion de l'ubiquitine humaine (HUB) ont été purifiés à l'aide de billes de chitine et clivés par l'acide mercapto-2-éthanesulfonique (MESNA).

Les peptides 0QAR, 20QAR et 50QAR ont été unis par ligation au peptide de fusion H2B doté d'un tag HA en se servant du système EPL. Ainsi, nous avons développé avec succès des outils permettant d'explorer l'ubiquitination du AR en utilisant les systèmes EPL in vitro.

Des expériences à l'avenir viseront à lier par ligation la protéine HUB à un peptide, en utilisant la méthode EPL et ensuite faire la ligation du peptide ubiquitiné aux trois différents peptides AR. Ces derniers seront ajoutés aux protéasomes actifs pour explorer l'implication de l'UPS dans la pathogénèse de la SBMA.

1. General Introduction

The androgen receptor (AR), a nuclear receptor belonging to the steroid/thyroid hormone receptor family, functions primarily as a DNA-binding transcription factor regulating gene expression. The AR is involved in a myriad of cellular functions that have a role in the development and growth of the male reproductive and nonreproductive systems. Cytoplasmic AR is activated by the binding of the testosterone (T) or dihydrotestosterone (DHT) hormone before translocation to the nucleus (Beitel et al., 2005; La Spada et al., 1992; Mhatre et al., 1993). A number of mutations that have been reported in the AR are associated with the inactivation of normal androgen-binding activity. In these instances, a loss-of-function in AR manifests itself in a disorder known as Androgen Insensitivity Syndrome (AIS) (Morais Fonseca et al., 2010), with major effects development of the male phenotype. There is another class of mutations in AR, an expansion of the polymorphic cytosine-adenine-guanine (CAG) (Rusmini et al., 2011) repeat in the coding region of exon 1, known as a trinucleotide repeat expansion (Beitel et al., 2005; La Spada et al., 1992; Mhatre et al., 1993). Such expansions result in spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease, an X-linked recessive neurodegenerative disease. The CAG codon codes for the amino acid glutamine and SBMA is thus considered a polyglutamine (polyQ) expansion disorder.

The 90-kb AR gene located at locus Xq11-12, and is transcribed to form an mRNA transcript 10.6 kb in length. The AR contains eight exons encoding a protein of about 919 amino acids, which has three conserved nuclear receptor (NR) superfamily domains: (1) a transactivation domain (TAD) in exon 1 (Paradas et al. 2008); (2) a DNA-binding domain in exons 2 and 3, and; (3) a ligand-binding domain encoded by exons 4 through 8 (Figure 1).

SBMA affects adult males exclusively; females may show mild neurodegenerative symptoms, as they are heterozygous for the AR mutation. This may be related to the T and DHT level, as the T levels in males are about 10 times more than in females. In unaffected individuals, 9-33 CAG repeats coding for Gln are found in exon 1, while males who inherit more than 37 CAG repeats will develop SBMA (Katsuno et al., 2006; Ranganathan et al., 2010; Katsuno et al., 2012). The toxic gain of function in the AR protein that is characteristic of SBMA results in slowly progressive muscle weakness, atrophy of the tongue, fasciculations of bulbar, facial, and limb muscles, and muscle cramping, while a loss of function in the AR gene is related to breast enlargement and reduced fertility (Katsuno et al., 2006; Ranganathan et al., 2010; Katsuno et al., 2012).

The ubiquitin-proteasome system (UPS), which functions in protein degradation, plays an important role in neuronal cells from SBMA patients. Components of the UPS are found in polyGln-expanded protein aggregates suggesting that the UPS is impaired in SBMA (Beitel et al., 2005; La Spada et al., 1992; Mhatre et al., 1993).

Later in the thesis I will be discussing a method called the IMPACT-TWIN (Intein Mediated Purification with an Affinity Chitin-binding Tag-Two Intein) system. It is a system that has the ability to isolate native recombinant proteins possessing a reactive C-terminal thioester in a single chromatographic step, without the use of exogenous proteases (New England BioLabs, 2006). These reactive groups can be used in intein-mediated protein ligation to attach peptides or proteins to the C-terminus or N-terminus of a target protein. IPL is also known as expressed protein ligation (EPL). One of the advantages in EPL is the use of a thiol reagent that easily removed (Vila-Perello et al., 2010). In fact, the EPL method was specifically used in this project to synthesis AR peptides and ligate test peptides to the AR peptides, in order to develop the tools to generate ubiquitinated AR.

2. Literature Review

2.1 Nuclear Receptor Superfamily

One of the largest families of transcription factors that are activated by ligands is the nuclear receptor (NR) superfamily. It is only found in metazoans (animals). The number of NR genes in different animal species varies widely; for example *C. elegans* has 270 NR genes, but humans, rats and mice have 48, 47 and 49, respectively. Drosophilia can count only 21 NR genes (Robinson-Rechavi et al., 2003).

NRs in general play a vital role in the regulation of many cellular activities like maintaining homeostasis, differentiation, reproduction and particularly development, since they affect gene expression and growth (Robinson-Rechavi et al., 2003). NRs play a direct role in linking signaling molecules and transcription. But they also provide a role in the integration of many signals from diverse signal transduction cascades, since their activity is modulated by post-translational modifications. Most signaling molecules that serve as NR ligand activators are small and hydrophobic, such as the steroid hormones (androgens, estrogens, glucocorticoids), retinoic acid, fatty acids, leukotrienes and prostaglandins. Also it should be noted that a large number of genes with significant homology to classical NR DBD and LBD have been identified, but have no known natural ligands. These are the so-called "orphan receptors". The NR superfamily is of great interest to the pharmacology industry, mainly because of the nature of NR natural ligands (small, hydrophobic, cell permeable, high binding affinity),

and the confirmed involvement in high-profile disease processes like cancer and diabetes. This probably accelerated the current and ongoing understanding of the NR superfamily (Dahlman-Wright et al., 1995).

2.1.1 Domain Structure of the Nuclear Receptor Superfamily

The structural organization of NRs is highly conserved throughout the superfamily (Figure 1). Generally, there are five conserved domains of NRs (from N- to C-terminal): the transactivational domain (TAD), DNA-binding domain (DBD), a flexible linker domain (the hinge region), the ligand-binding domain (LBD), and the C-terminal.

The N-terminal (A/B), DNA-binding (C), and ligand-binding (E) domains function independently, can fold and are structurally stable, whereas the hinge region (D) and the C-terminal (F) domains may have flexible conformations. Overall, this conservation in general domain structure suggests a common mode of ligand-dependent activation across subfamily members (Dahlman-Wright et al., 1995).

The DBD contains a P-box element, a short motif responsible for DNA-binding specificity and is involved in dimerization of NR including the formation of both heterodimers and homodimers. The hinge region (linker) or domain D is located between the DBD (domain C) and the ligand-binding domain (domain E). The region is named “hinge” because it functions as a flexible hinge and contains the nuclear localization signal (NLS), enabling the ligand bound NR to translocate to the nucleus (Tanner et al., 2004). Transcriptional activation is usually associated with

phosphorylation of the hinge region. The ligand-binding domain (located in the E domain) is the region is responsible for binding of the cognate ligand or hormone. This domain also harbors what is called the ligand-regulated transcriptional activation function 2 or AF-2, which is crucial for recruiting transcriptional co-activators, and interacting with the transcriptional activation machinery in general. Another transcriptional activation function known as AF-1 is located N-terminal to the DBD of most NR (Bourguet et al., 1995). Contrary to the relatively conserved AF-2, the AF-1 is weakly conserved across the NR super-family and could have a role in mediating the differential promoter regulation in vivo. The AF-1 sequence also has a role in the ligand-independent transcriptional activation. It should be noted that some receptors contain a carboxyl-terminal region (F domain) of still unknown function.

2.1.2 Mode of Action of NR

In general, nuclear receptors are thought to act in three steps: (1) repression, (2) derepression and (3) transcriptional activation (reviewed by Laudet and Gronemeyer, 2002). Repression is initiated by the apo-nuclear receptor by recruiting a corepressor complex with histone deacetylase activity. Derepression follows, occurring after ligand binding, which in turn dissociates this complex bringing the first coactivator complex with histone acetyltransferase (HAT) activity resulting in chromatin decondensation. This is believed to be necessary, but not sufficient, for activation of the target gene. In the last step, the histone acetyltransferase complex dissociates so that a second coactivator complex is assembled (TRAP/DRIP/ARC), a transcriptional coactivator complex essential in the expression of most genes. Once this is established, contact is initiated with the basal transcription machinery, resulting in the transcriptional activation of the target gene. It should be noted that this is a schematic view of events, but the precise order is still under debate. Also, some exceptions might exist. Some NR could act as activators without the need for a ligand, while others are unable to interact with the target gene promoter without a ligand (repression).

Throughout the NR superfamily the TAD domain is the least conserved domain. Its average amino acid identity compared to the consensus NR TAD sequence is about 29% (Robinson-Rechavi et al., 2003). This large variability in the TAD structure could potentially contribute to the specificity of unique NR response. In the meantime there is

no 3D structure for any NR TAD, although it is unlikely that the solution of one NR TAD would be relevant for homology modeling of other NRs. Structural (NMR) studies of the glucocorticoid receptor (GR) suggest that the AF-1 is natively unfolded (Dahlman-Wright et al., 1995), and that by an "induced fit" mechanism it can be forced into an α-helical conformation (Tjian et al., 1994), creating a stabilized structure competent for transactivation. This suggests that conditional folding and α-helical formation are an important requirement for the interaction with coregulators to achieve transcriptional regulation (Guarente, 1995; Onate et al., 1995).

Even with this large NR TAD variability, there are many conserved regions, like the presence of one constitutively active transactivation region (AF-1) and several other autonomous transactivational subdomains. These domains are constitutively active even in the absence of the DBD and LBD, but are repressed in the apo (unliganded) state of the receptor. The recruitment of the basal transcription machinery is the main reported function of AF-1, and its transactivational potential depends on promoter context, cell type and post-translational modifications, such as phosphorylation and small ubiquitin-like modification (SUMO-ylation) (Cenni et al., 1999; Tzukerman et al.,1994).

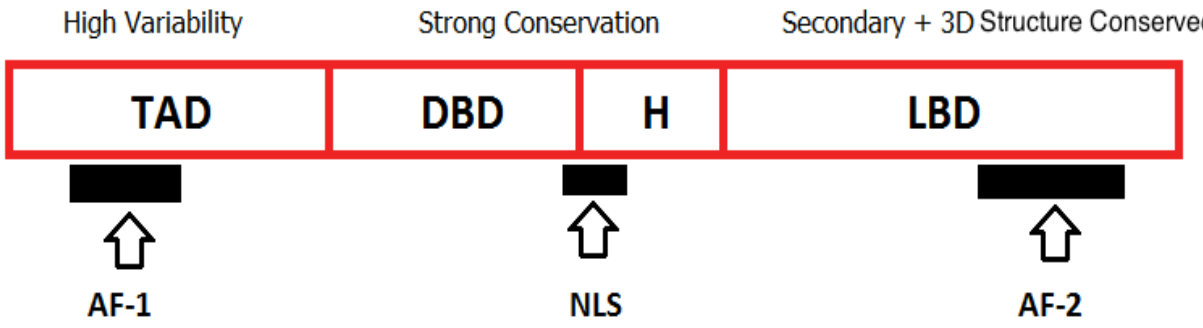

Figure 1: The general organization of domains of the nuclear receptor superfamily

Shown is the transactivation domain (TAD) (Paradas et al., 2008), which has low homology between family members and where the activation function-1 (AF-l) resides, which has a primary role in the interaction with coregulatory proteins. The DNA-binding domain (DBD) is responsible for recognizing the DNA response elements and for dimerization. The hinge region (H) contains the nuclear localization sequence (NLS). Finally, the ligand-binding domain or (LBD), which has a highly conserved structure involved in both ligand-binding and the transactivation domain (TAD) functioning via the activation function-2 (AF-2) subdomain, and by dimerizing.

The DBD is considered well conserved among the NR superfamily, in terms of both primary sequence and 3-D conformation. The main features of the NR DBD structure are the two zinc-finger motifs that make site-specific contact with the DNA response element hexamers formed by three protruding helices. These two zinc-finger motifs are interdependent subdomains and have slight structural and functional differences, but still fold as a unified globular domain whose structure is modestly changed upon DNA binding (Baumann et al., 1993; Hard et al., 1990). This is different from the classical zinc-finger motifs that function as structurally and conformationally stable units in the presence or absence of DNA contributing DNA binding independently (Klevit et al., 1990). In this case, the first NR zinc-finger subdomain aids in the discrimination of receptor-specific DNA response elements (Luisi et al., 1991; Lee et al., 1993). On the other hand, a helix in the second NR zinc-finger has nonspecific affinity for DNA, with a loop in the same subdomain that forms a homodimerization interface called the D box.

The third region of interest is the hinge region, it is small, but has multiple functions. It is the domain that links the DBD and LBD, it harbors the primary NLS, and has roles in DNA binding and modulation of transactivation (Tanner et al., 2004). The NLS is a cluster of basic residues which function to enable a ligand-bound NR to be translocated to the nucleus. It should also be noted that the hinge region has been shown to harbor several lysine residues acting as targets for post-translational modifications such as acetylation and ubiquitination (McEwan, 2004).

Finally, the NR LBD provides the receptor interface for binding of class-specific ligands and is involved in homo- and heterodimerization. It also mediates transcriptional repression and it harbors the ligand-dependent transcriptional activation function (AF-2). When it comes to LBDs, the crystal structures have been solved for more than 20 of the 48 human NR LBDs identified bioinformatically. One interesting finding was that both secondary and tertiary structure of the LBD is highly conserved across the superfamily members.

2.1.3 Nuclear Receptor Coregulators

It is the combined action of coregulatory proteins that makes the transactivational potential of NRs possible. This transcriptional activity of NRs is dependent on the chromatin environment and on the recruitment of RNA polymerase II to target promoter regions. A recent paper on the number of NRC shows that there are more than 350 reported NR coregulatory proteins (Lonard et al., 2012). The two major sites of coregulator recruitment are AF-1 and AF-2, though they have very different modes of action. The AF-1-interacting coregulators are characterized by having no conserved domains and do not bind to a defined surface on the TAD. On the other hand the AF-2 coregulators have a highly conserved motif which fits into the hydrophobic AF-2 location (Lonard et al., 2007).

Coactivators are recruited to the promotoras of performed complexes, so that controlling the rate and identity of proteins in the assembly of various promoter-specific complexes

is a stategy employed by the NR to affect transcriptional activity and integrate information from many signal transduction cascades. Another important role of coactivators is the stabilization of the interaction between the NR and the basal transcription machinery (Darimont et al., 1998). Also, coactivator complexes usually contain enzymes with HAT activity, resulting in the covalent modification of histone proteins in an ATP-dependent manner, finally leading to chromatin remodeling and resulting in increased transcriptional potential (Roth et al., 2001).

NR function can also be repressed by corepressors. They can be divided into two forms, the Nuclear Receptor Corepressor (NCoR) and the Silencing Mediator of Retinoic Acid and Thyroid Hormone Receptors (SMRT). These corepressors reduce the transcriptional activity of NR when there are no ligands or in the presence of antagonists. There are several mechanisms that have been discovered for repression of transcription of NR by corepressors. One example is passive repression, which occurs when corepressive proteins block promoter binding, preventing dimerization or binding of coactivators (Hudson et al., 1990). On the other hand, when the corepressive protein blocks transcriptional initiation by using factors that facilitate blockage there is what is called active silencing (Baniahmad et al., 1993). Another method where NR corepressors can function is by active recruitment of histone deacetylase (HDAC), so that they function in opposition of HATs, in order to catalyze the production of inactive

chromatin structures which are incompatible with high rates of transcriptional activation (Baniahmad, 2005).

In conclusion, the high number and types of NR coregulatory proteins adds an extra level of complexity to the entire process of NR signaling, expanding the ways in which NR-regulated cellular processes have the ability to integrate signals from diverse signal transduction cascades.

2.2 Androgen Receptor

The development of the male phenotype is dependent on steroid hormones in general and androgens in particular. The role of androgens is essential during male sexual differentiation and in both development and maintenance of external genitelia and other sex characteristics including the initiation and maintenance of spermatogenesis (Schweikert et al., 1974). In addition, androgens have roles in other male patterns like hair and in general male baldness (Schweikert et al., 1974). Naturally occurring androgens that are synthesized by humans can be divided into two types; testosterone (T) and 5α-dihydrotestosterone (DHT). The first type is synthesized in the Leydig cells of the testes from cholesterol. Testosterone is a steroid hormone that is lipophilic and it is a rather small molecule, which allows it to passively diffuse across the membranes of target cells. It then either binds directly to the AR, or becomes a substrate for the 5α-reductase enzyme (EC 1.3.99.5) which catalyzes its conversion to a more avid AR-

binding molecule, DHT, which has a higher AR activation potential than T. The two androgens have specific roles in male sexual development, but there is only one AR protein, demonstrating that subtle differences in T-AR and DHT-AR complexes cause distinct modes of activation of the androgen-regulated genes.

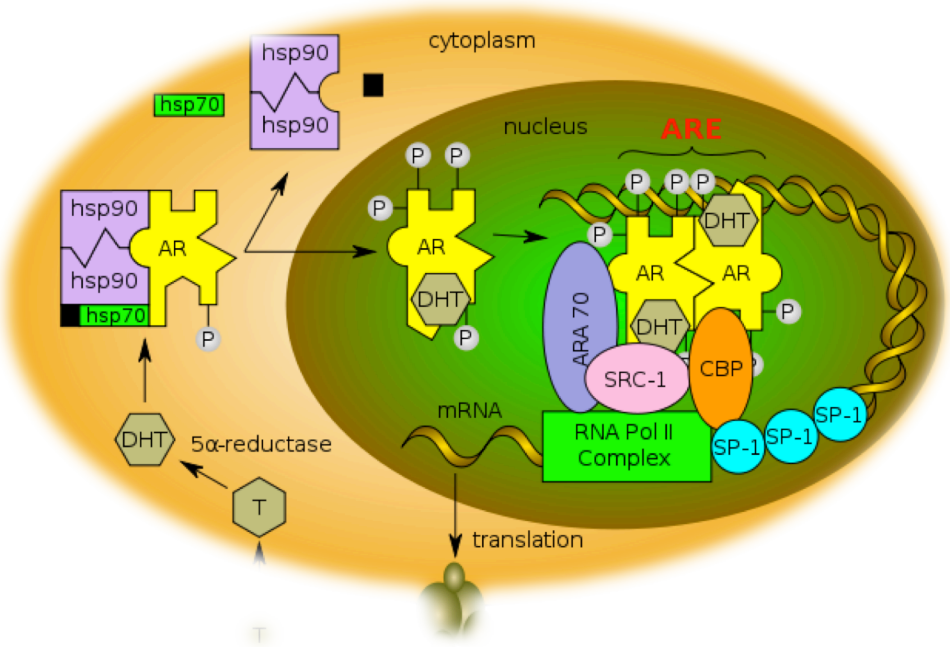

Figure 2: Functional and structural mechanism of androgen receptor

While androgen receptor (AR) is in the cytoplasm it binds to the heat-shock proteins (Hsps). When testosterone (T) enters the cell it is converted into dihydrotestone (DHT). AR releases the heat-shock proteins when T or DHT binds to AR in the cytoplasm and AR translocates to the nucleus where dimerization on the androgen receptor element (ARE) occurs (Galani et al., 2008)

2.3 Androgen Receptor Gene

Our gene of interest is the AR gene. It is located on the X chromosome at Xq 11-12 and spans 90 kb of DNA. After transcription it forms a 10.6 kb mRNA encoding a protein that is approximately 919 amino acids in length (Lubahn et al., 1988; Chang et al., 1988). This genetic organization of the AR is a reflection of the conserved modular domain structure of the NR superfamily mentioned already. Exon 1 encodes the entire TAD (amino acids 1-532), then exons 2 and 3 encode the DBD, and exons 4 to 8 encode the LBD (Figure 3). It should be noted that this organization of the AR exon structure is conserved throughout mammalian evolution from rodents to humans. This conservation of the AR genetic locus across many species of animals shows that there is a significant developmental association of other genes in the vicinity or what is referred to as syntenic genes (Spencer et al., 1991).

The AR is expressed in all human tissues except the spleen, but there is great variability in the expression levels and timing depending on the cell type (Song et al., 1993). The 5' promoter region of the AR notably lacks the traditional TATA or CAAT transcription initiation sites, but instead is rich in GC regions required for recruiting the Specificity Protein 1 (SP1) general transcription factor (Song et al., 1993). There are two promoter binding sites with confirmed variable timing, activity and cell type specificity (Grossmann et al., 1994). Transcriptional control of AR expression is contributed by androgens and other steroid hormones. For example, the rat AR promoter contains several palindromic

hormone response elements trophic for the AR, GR and PR (Baarends et al., 1990), so that they function mainly as a negative feedback loop for downregulating AR signalling.

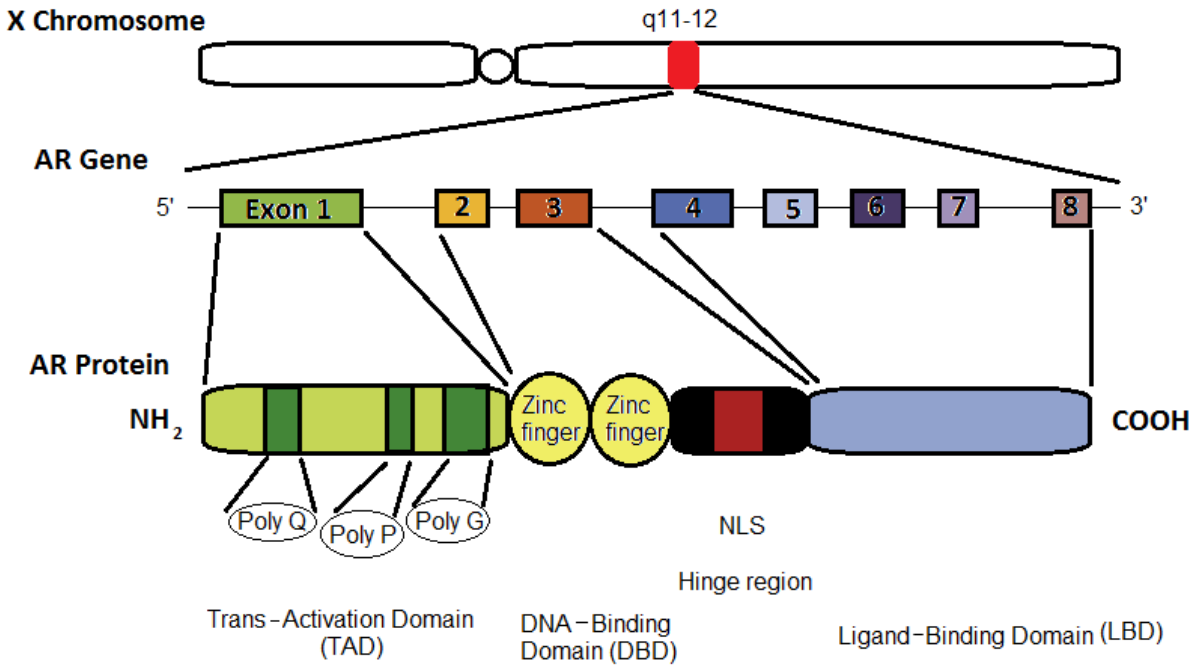

Figure 3: The genomic and protein domain structure of the androgen receptor

The AR gene is located at Xq11-12 on the long arm of the X chromosome and is approximately 90 kb in length. It is organized into eight exons that form the three functional domains of the AR protein as shown. The polymorphic glutamine tract or poly Q is located in the transactivation domain (TAD) (Paradas et al., 2008). Similarly, repeat sequences encoding the CCT/GCG (polyproline) and GGT/GGC (polyglycine) repeats are displayed.

2.4 Genetic Features of Androgen Receptor

Since the AR is located on the X chromosome, males will express only one allele and so they are sensitive to mutations without the balance of a codominant allele. Also, the AR has been shown to have a large record of spontaneous mutations (Gottlieb et al., 2004). These mutations have been used to understand the structure/function relationships of the AR. Thus mutations in the AR have a broad range of functional consequences that are clinically manifested as phenotypic variabilities of the Androgen Insensitivity Syndrome (AIS) (Morais Fonseca et al., 2010). This syndrome is further classified according to the clinical severity, where there are at least three categories: Mild Androgen Insensitivity, Partial Androgen Insensitivity and Complete Androgen Insensitivity. This severity of AIS is dependent on the degree of impairment of the AR function caused by defects in certain molecular mechanisms like binding affinity, rates of ligand dissociation (off-rates), AR N/C-terminal interactions and coactivator recruitment.

The AR gene has an interesting feature, which is the presence of several regions of repetitive DNA sequences. The first exon contains three such regions: a noninterrupted CAG repeat which codes for polyglutamine, beginning at codon 58, that is polymorphic in length. This polyglutamine (polyQ) tract has proven functional consequences both in ligand-dependent AR transcriptional activation (Mhatre et al., 1993), and as a toxic element causing SBMA (La Spada et al., 1991). The CCT/GCG repeat encoding

polyproline is from codons 371-379, and the GGT/GGC repeat coding for polyglycine is from codons 449-472. Unfortunately, the functional consequences of the polyproline and polyglycine tracts are not yet well understood.

An important feature of the CAG tract is its polymorphic length, which could be caused by slippage of the DNA polymerase during replication. The average length of the CAG repeat in the human population is around 21, yet the normal range is anything between 14 and 35, varying with ethnicity and race (La Spada et al., 1991). There appears to be an inverse correlation with the transcriptional activation potential of the AR with respect to the polyQ tract length. For example, men with very short tracts may have high risks for prostate cancer, indicating hypersensitivity of the AR (Giovannucci et al., 1997), while on the other hand men with rather long tracts have reduced AR activity and so they develop some symptoms of AIS (La Spada et al., 1991). Elongation of the CAG repeat to over 37 repeats will, in addition to the functional consequences and effects on transcriptional activation, result in a late-onset neurodegenerative disorder known as Kennedy's disease or SBMA (La Spada et al., 1991).

2.5 Trinucleotide Repeat Disorders

Trinucleotide repeat disorders (sometimes known as trinucleotide repeat expansion disorders or triplet repeat expansion disorders) are a type of genetic disorder caused by an increase beyond the normal, stable, threshold in the number of trinucleotide repeats in certain genes. Every gene that is affected by trinucleotide repeat expansion has a different range of repeat number that constitutes the normal threshold and the number that results in manifestation of disease, with an age of onset of the disease that is dependent on that number of repeats (Meyers, October 2005, Wiley-Blackwell).

The trinucleotide repeat disorders are divided into two main categories, depending on the type of repeat. The most common repeat and the one of most interest to us is the triplet CAG, which codes for the amino acid glutamine (Q) when it is in the coding region of a gene. So, such disorders are referred to as the polyglutamine (polyQ) disorders and I will describe them in some detail. The remaining disorders have the CAG triplet in a noncoding region of the gene or they do not involve the CAG triplet repeat, therefore, they are referred to as the non-polyglutamine disorders and will only be discussed briefly (Meyers, October 2005, Wiley-Blackwell).

2.5.1 Non-Coding Trinucleotide Repeat Disorders

The non-coding trinucleotide repeat diseases or non-polyglutamine disorders are characterized typically by multiple tissue dysfunction or degeneration that is a result of large and variable repeat expansions. Phenotypic effects of a single disease are variable, perhaps due to the pronounced degree of somatic heterogeneity, since it is often the case that the larger mutations originate from a small set of clinically silent intermediate size expansions, termed pre-mutations. Despite having many such similarities, the trinucleotide repeat sequences themselves are different (CGG, GCC, GAA, CTG and CAG), and it is clear that the particular trinucleotide repeat sequence, along with its location relative to a gene, are important defining factors in developing the unique mechanism of pathogenesis for each disease (Meyers, October 2005, Wiley-Blackwell). The pathogenic mechanism will vary from disease to disease, depending on the effects of the lost and/or gained function of the respective (toxic transcript) proteins (Meyers, October 2005, Wiley-Blackwell).

2.5.2 Polyglutamine Disorders

1) Dentatorubropallidoluysian Atrophy (DRPLA)

All polyglutamine repeat disorders affect both the brain and the body, and dentatorubropallidoluysian atrophy (DRPLA) is no exception (Yamada et al., 2008). The main symptoms are abrupt muscle jerking, involuntary movements, and dementia. Some cases might suffer from progressive intellectual decline (Yamada et al., 2008).

The Gene and Protein:

The gene involved in DRPLA is actually named “DRPLA” and it lies on chromosome 12 (Yamada et al., 2008). Normal individuals have between 6 and 35 copies of CAG in the DRPLA allele, however, disease occurs when a person has between 49 and 88 copies (Yamada et al., 2008). At present, not enough data exist to fully understand the effect that alleles with 35-49 copies of CAG will have on individuals. The protein product of the DRPLA gene is called Atrophin-1. It is thought to be involved in the insulin pathway (Yamada et al., 2008).

2) Huntington's disease (HD)

Huntington's disease (often abbreviated "HD") first occurred in the medical literature in 1872, reported by 22 year old Dr. George Huntington (Walker, 2007). The disease affects both sexes alike, and its rate of occurrence is about one in every 10,000 in most Western countries. The age of onset of Huntington's disease is mostly between 30 and 50 years old. HD is a neurodegenerative disorder with a very distinct phenotype; it is characterized by chorea, dystonia, incoordination, cognitive decline, and behavioral difficulties. Patients show progressive neural cell loss and atrophy in the caudate and putamen. There is also a prodromal phase of mild psychotic and behavioral symptoms which precedes frank chorea by up to 10 years (Rangone et al., 2004). Another type of HD is Juvenile-onset Huntington's disease when the onset is before the age of 20 (Rangone et al., 2004). Roughly less than 10% of all HD cases are of early onset. Juvenile-onset Huntington disease is usually transmitted from an affected father and is associated with very large CAG repeat sizes (60 or more) in the HTT gene (Walker, 2007). It should be noted that in normal individuals, the range of repeat numbers is 9 to 36, while in HD patients the repeat number is over 37. The repeat is located in the gene encoding *Huntingtin* (HTT; 613004) on chromosome 4p16.3 (Walker, 2007).

Early signs of Huntington's disease might appear while the brain shows no signs of neurodegeneration (Myers, 2004). Decreased staining of nerve fibres in the cortical neurons is usually the first sign, along with decreased staining in the neurofilaments,

tubulin, and microtubule-associated protein along with diminished complexin concentrations (Walker, 2007). All these elements have been shown to be associated with synaptic function and axonal transport, suggesting an important role for cortical dysfunction in the pathogenesis of the disorder (Myers, 2004).

One of the main pathological characteristics of Huntington's disease is the appearance of nuclear and cytoplasmic inclusions that contain mutant Huntingtin and polyglutamine aggregates (Rangone et al., 2004). Although they indicate the presence of pathological polyglutamine processing and are apparent in affected individuals long before symptoms show, some suggest that these inclusions are not predictors of disease activity and instead they seem to be related more with intermediate stages of polyglutamine aggregates. In some transgenic mouse models inclusions only develop after the symptoms are already evident (Walker, 2007) and cells with inclusions survive longer than those without. It should be noted that little correlation is observed between the various cellular and animal models of the disorder and human Huntington's disease in the case of appearance of inclusions in histopathological samples and the onset of neurological symptoms (Myers, 2004).

3) Spinal and Bulbar Muscular Atrophy (SBMA)

In 1968 SBMA was known as an X-linked progressive neurodegenerative disorder, however, this was before the polyQ tract expansion in the AR gene was determined to be the mutation responsible for the disease (La Spada et al., 1991). SBMA was the first identified polyQ expansion disease. SBMA affects adult males, whereas females do not show visible neurological symptoms, as they are heterozygous for the AR mutation (La Spada et al., 1991). The clinical symptoms in affected individuals usually occur between 30 and 50 years of age and the prevalence of SBMA is 1-2 in 100,000 individuals (Katsuno et al., 2006). In unaffected individuals, 9-33 CAG repeats coding for Gln are found in AR exon 1, while males who inherit more than 37 CAG repeats go on to develop SBMA. The toxic gain of AR function characteristic of SBMA results in slowly progressive muscle weakness, atrophy of the tongue, fasciculations of bulbar, facial, and limb muscles, and muscle cramping, while a loss of AR function is related to breast enlargement and reduced fertility (Katsuno et al., 2006; Ranganathan et al., 2010; Katsuno et al., 2012). SBMA is a progressive disease with patients in the latest stages developing dysphagia and dysarthia and also patients in their sixties will require a wheelchair for mobility (Katsuno et al., 2006).

SBMA is progressively neurodegenerative, which is characterized by loss of motor neurons in the spinal cord and dorsal root ganglia, and thus degeneration of the motor neurons in the brainstem (Sobue et al., 1989).

4) Spinocerebellar Ataxia Type 1 (SCA1)

Spinocerebellar ataxia type 1 (SCA1) is one of the spinocerebellar ataxias (SCAs), a family of many closely related disorders. All are characterized by atrophy of the cerebellum, which is partly responsible for the major symptoms of the disorder such as a loss of coordination, difficulty in articulating speech, and decreased sensation in the limbs (Spears et al., 2010).

The Gene and Protein:

The gene involved in SCA1 is located on Chromosome 6 and is also called SCA1. Typically, in normal individuals there are between 6 and 44 copies of CAG in the SCA1 allele. In a person with the disease, the allele has anywhere between 39 and 81 copies of CAG (Spears et al., 2010).

The protein product of the SCA1 gene is called a Ataxin-1. It is thought that there is an interaction between ataxin-1 and another protein called LANP. LANP appears to have a major effect on cell communication, which is vital for nerve cell survival. When the ataxin-1 protein structure is changed, it will have an altered interaction with LANP and then this could cause degeneration of the nerve cell (Spears et al., 2010).

5) Spinocerebellar Ataxia Type 2 (SCA2)

Spinocerebellar ataxia type 2 (SCA2) causes a general slowing of some of the body's normal functions (Lastres-Becker et al., 2008). In addition to the loss of coordination that is typical to all SCAs, people with SCA2 also tend to shift the focus of their eyes from one point to another in a very deliberate manner, to the point where partial paralysis of the eyes has been described in some cases (Lastres-Becker et al., 2008).

The Gene and Protein:

The gene involved in SCA2 lies on chromosome 12 and is also named SCA2. Typically, in normal individuals there are between 14 and 31 copies of CAG in the SCA2 allele. In a person affected with the disease the allele has between 36 and 64 copies. Individuals with 31-36 copies of CAG may or may not develop the symptoms of the disease (like most cases of mid-range number of repeats). The protein product of SCA2 is called ataxin-2 (Lastres-Becker et al., 2008).

6) Spinocerebellar Ataxia Type 3 or Machado-Joseph Disease (SCA3)

Spinocerebellar ataxia type 3 (SCA3) is also known as Machado-Joseph Disease. Other than the loss of coordination that is common to all SCAs, the most common symptoms of SCA3 are bulging eyes, small contractions of the facial muscles, and general rigidity along with other non-unique symptoms (Alves et al., 2008).

The Gene and Protein:

The gene involved in SCA3 lies on chromosome 14 and is also named SCA3 (although sometimes it is named “MJD1″). Typically, in normal individuals there are between 12 and 43 copies of CAG in the SCA3 allele. In a person with the disease the allele will contain anywhere between 56 and 86 copies (Alves et al., 2008). The protein product of SCA3 is called ataxin-3. Its function is not known, but it normally resides in the cytoplasm of the cell in unaffected individuals. However, in affected individuals SCA3 aggregates in the nucleus (Alves et al., 2008).

7) Spinocerebellar Ataxia Type 6 (SCA6)

People with SCA6 mostly experience slowly progressing ataxia with random episodes (Jacobi et al., 2013).

The Gene and Protein:

The gene involved in SCA6 lies on chromosome 19 and is also named SCA6 (Jacobi et al., 2013). Typically, in normal individuals there are between 4 and 18 copies of CAG in the SCA6 allele (Jacobi et al., 2013). In affected individuals the allele contains between 21 and 33 copies. It is interesting that this is the smallest number of trinucleotide repeats known to cause disease. Instead of its own separate protein product, the SCA6 gene codes for a subunit of the calcium channels that exist in all nerve cells (Jacobi et al., 2013).

8) Spinocerebellar Ataxia Type 7 (SCA7)

Similar to the other SCAs, the most common symptom of SCA7 is loss of coordination and some people with SCA7 have difficulties with vision (Wolf et al., 2013).

The Gene and Protein:

The gene involved in SCA7 lies on chromosome 3 and is also named SCA7 (Wolf et al., 2013). Typically, in normal individuals there are between 4 and 19 copies of CAG in the SCA7 allele (Wolf et al., 2013). In individuals with the disease, the allele has between 37 and 306 copies. The protein product of the SCA7 gene is called ataxin-7 (Wolf et al., 2013). It is thought that when ataxin-7 proteins are changed they clump together in the nucleus producing neuronal inclusions (NIs), which are also found in certain nerve cells of individuals affected with SBMA, HD, and some other SCAs (Wolf et al., 2013).

9) Spinocerebellar Ataxia Type 12 (SCA12)

Spinocerebellar ataxia type 12 has been identified quite recently and because of the small number of individuals known to have SCA12, many clinical manifestations have been not recognized yet (O'Hearn et al., 2012). For now it is characterized by onset of tremor of the upper extremities that is slowly progressing, leading to ataxia and other cerebellar and cortical signs (O'Hearn et al., 2012).

The Gene and Protein:

The gene involved in SCA12 lies on chromosome 5 and is named PPP2R2B (O'Hearn et al., 2012). Typically, in normal individuals there are between 4 and 32 copies of CAG in the PPP2R2B allele. In a person with the disease the allele has between 41 and 65 copies. The protein product of the PPP2R2B gene is also called PPP2R2B and it is implicated in the negative control of cell growth and division (O'Hearn et al., 2012).

10) Spinocerebellar Ataxia Type 17 (SCA17)

Spinocerebellar ataxia type 17 is the last of the SCAs to fall under the category of polyglutamine diseases. It is sometimes known as Huntington Disease-Like 4 or HDL4 because its clinical phenotype and inheritance pattern are similar to Huntington's disease (Musova et al., 2013). As in SCA12, SCA17 has been identified quite recently and because of the small number of individuals known to have SCA17 many clinical manifestations have been not recognized yet (Musova et al., 2013). SCA17 is characterized by characterized by ataxia, pyramidal and extrapyramidal signs, cognitive impairments, psychosis, and seizures (Musova et al., 2013).

The Gene and Protein:

SCA17 is caused by an expansion in the TATA box-binding protein. Typically, in normal individuals there are between 25 and 44 copies of CAG in the TBP. In a person with the disease the allele carries between 47 and 63 copies (Musova et al., 2013).

2.6 PolyGln Tract

More than 20 years ago Dr. Kurt Fischbeck and his colleagues (La Spada et al., 1991) identified a novel mutational mechanism that alters the sequence of a protein: the addition of glutamines to a polyglutamine tract within the androgen receptor (AR). Shortly afterwards the cloning of the genes affected in HD (Huntington's Disease Collaborative Research Group 1993), and SCA1 (Orr et al., 1993) occurred, and subsequently the basis for many other neurodegenerative diseases were identified (Table 1). It has therefore become clear that an expansion of a polyglutamine tract is the mutational mechanism causing a number of neurodegenerative diseases.

The evidence that a polyglutamine tract is itself toxic stems from studies with both transgenic mice and transfected cell lines. Transgenic mice expressing a small fragment of the huntingtin protein including an expanded polyglutamine tract have been shown to develop abnormal neurological signs and neuropathology (Mangiarini et al., 1996). Similarly, a peptide containing essentially only the expanded polyglutamine tract from the SCA3 gene product has also been shown to be toxic to cerebellar Purkinje cells in transgenic mice (Ikeda et al., 1996). Also it should be mentioned that inserting a polyglutamine tract into the HPRT (hypoxanthinephosphoribosyl transferase) gene resulted in mice with a neurological phenotype very similar to that observed in mice expressing the mutant HD truncated protein (Ordway et al., 1997). In general there are a number of studies that have demonstrated that expanded polyglutamine tracts within

truncated proteins are more toxic to transfected cells than expanded tracts embedded in full-length proteins (Hackam et al., 1998). Such results imply that proteolytic cleavage is a necessary step in the initiation of polyglutamine disease.

The class of polyQ expansion diseases includes SBMA, Huntington's Disease (HD), dentatorubal-pallidoluysian atrophy (DRPLA), and spinocerebellar ataxias (SCA) types 1, 2, 3, 6, 7, and 17 (Table 1). While these genes do not share any homology outside of the CAG tract and do not seem to be related in terms of the biochemical function, interestingly, the disorders themselves do share several clinical, pathological, and molecular features. PolyQ expansion diseases are all considered late-onset neurological disorders sharing a dominant mode of inheritance, with the exception of an X-linked recessive mode for SBMA. This genetically dominant mode of inheritance suggests a toxic "gain-of-function" change is altering the disease protein by the polyQ expansion itself, and could explain why proteins with such unrelated biochemical or metabolic function might have a common pathology (La Spada et al., 1991). Another conserved feature is that polyQ diseased patients have both somatic and germ-line instability of triplet repeats (Chong et al., 1995), which results in expansion of repeat number in subsequent generations. This clinical attribute, termed genetic anticipation has been shown to be correlated with both age of onset and disease severity (Trottier et al., 1994). Also, all reported polyQ expansion diseases were shown to have an inverse

correlation between polyQ tract length and age of onset, but a direct correlation between tract length and disease severity.

Another common feature among polyQ expansion diseases is the histopathological observation of insoluble protein deposits also called "protein aggregates" or "inclusion bodies" containing the polyQ-expanded protein. One additional similarity between these diseases is the minimum glutamine tract length limit required for disease initiation (Trottier et al., 1994). Normal alleles of these disease-causing genes have widely variable average polyQ tract lengths, but the number of glutamines required to acquire the disease is a relatively constant 36-39Q across almost all polyQ expansion diseases (Table 1). Unfortunately, the reason behind this “pathological threshold” for toxicity remains unknown.

One unexplained phenomenon in polyQ expansion diseases are their late-onset. This neuronal specific toxicity in non-overlapping brain regions was observed in all these disorders, even though it is confirmed that the expression of the polyQ expansion is in almost every tissue type. All of this suggests a highly complex mechanism of toxicity that so far has proven to be difficult to unravel.

Table 1: Summary of Polyglutamine Diseases (Orr, 2001)

Disease	Gene name	Chromosomal location	Pattern of inheritance	Protein	Normal repeat length	Disease repeat length
Spinobulbar muscular atrophy (Kennedy disease)	*AR*	Xq13–21	X-linked recessive	androgen receptor (AR)	9–36	38–62
Huntington disease	*HD*	4p16.3	autosomal dominant	huntingtin	6–35	36–121
Dentatorubral-pallidoluysian atrophy (Haw River syndrome)	*DRPLA*	12p13.31	autosomal dominant	atrophin-1	6–35	49–88
Spinocerebellar ataxia type 1	*SCA1*	6p23	autosomal dominant	ataxin-1	6–44	39–82
Spinocerebellar ataxia type 2	*SCA2*	12q24.1	autosomal dominant	ataxin-2	15–31	36–63
Spinocerebellar ataxia type 3 (Machado-Joseph disease)	*SCA3*	14q32.1	autosomal dominant	ataxin-3	12–40	55–84
Spinocerebellar ataxia type 6	*SCA6*	19p13	autosomal dominant	$\alpha 1_A$-voltage-dependent calcium channel subunit	4–18	21–33
Spinocerebellar ataxia type 7	*SCA7*	3p12–13	autosomal dominant	ataxin-7	4–35	37–306
Spinocerebellar ataxia type 17	*SCA17*	6q27	autosomal dominant	TATA binding protein	25–42	45–63

2.6.1 Molecular Pathology of PolyQ Expansion Disease

One might be inclined to suggest from the observation of this common, gain-of-function mechanism for polyQ expansion mediated pathology that there is a single conserved molecular etiology of toxicity in every polyQ expansion disease (Shao et al., 2007). This is probably is an oversimplification of the actual mechanism, but conserved mechanistic features of toxicity surely exist (Shao et al., 2007). There are several good hypotheses for the expansion-mediated toxicity of polyQ. These include toxicity of protein aggregates, aberrant transcriptional properties of polyQ-expanded proteins, and inhibition of the UPS. These hypotheses have been extensively investigated and tested in cell culture experiments, animal models and in human tissue isolated from patients, but unfortunately have yielded confusing, often conflicting results. Even though some progress has been made in understanding the polyQ expansion toxicity, the pathogenic mechanisms of the polyQ expansion have remained elusive.

2.6.2 Role of Aggregation of the PolyQ-Expanded Protein

As mentioned previously, a unifying pathological phenomenon in polyQ expansion diseases is the presence of discrete deposits of cellular material. These can be easily seen in patient tissue samples by microscopy, in cultured cells and animal models (Michalik et al., 2003). Immunohistochemical staining of these deposits universally stain positive for the polyQ-expanded protein. These deposits are usually referred to as "aggregates" or "inclusion bodies" (Michalik et al., 2003). Protein "aggregation" is

defined as a biochemical phenomenon that occurs when a self-associated novel structure (likely a ß-sheet) forms, distinct from the native structure of the polyQ tract, between two or more polyQ protein molecules (likely a random coil) (Michalik et al., 2003).

Second, the term "inclusion" is historically used to describe abnormal intracellular structures observed histologically (Taylor et al., 2003), and will be referred to here as a "cellular response" resulting in relocalization of polyQ proteins to a small area. It is highly likely that the activity of heat shock proteins and the aggresome machinery are involved in the process of inclusion body formation, but may or may not be dependent on prior polyQ tract self-association. This means inclusion bodies are to be considered a distinct process from aggregate formation. One should note that there is no requirement for the existence of polyQ expansion in either definition, since normal length polyQ tracts have been demonstrated to aggregate in vitro, and are also sequestered into inclusion bodies.

2.6.3 Aggregate Formation

X-ray diffraction analysis of synthetic peptides consisting of short glutamine tracts was used in the first attempts to describe the propensity of the polyQ expansion to self-associate. It showed that consecutive glutamines favor a ß-pleated sheet conformation, which in turn would explain the self-association of polyQ proteins as the formation of "polar zippers" (Perutz et al., 1994). On the other hand, short glutamine tracts considered native proteins were nonetheless shown to be in a random coil conformation by nuclear magnetic resonance (NMR). However, based on entropic factors it is predicted that elongated polyQ tracts should form a more stable intramolecular structure, likely a ß-hairpin. This is an inherently different structure than the polyQ-expanded protein and would in turn facilitate in the formation of the polar zippers, resulting in the formation of a nucleus of aggregated protein that might rapidly recruit other polyQ protein monomers (Perutz, 1996). This may explain the pathogenic threshold observed across polyQ expansion disorders, but recent studies have shown this hypothesis to be flawed. Current models of soluble polyQ proteins derived by Circular Dichroism (CD) spectra (Tzukerman et al. 1994) and NMR spectra demonstrate a random coil structure of both normal and expanded polyQ tracts, and that the formation of a ß-sheet conformation occurs concomitantly with the aggregate formation in polyQ tracts of any length (Chen et al., 2002; Masino et al., 2002).

This data invalidates the hypothesis that the structure of expanded polyQ tracts is implicitly novel and different from the normal length polyQ tract, suggesting instead that a rare folding event could be the culprit.

In order to reconcile the structural equivalence of both the native forms of normal and expanded polyQ tract with the observation that aggregate formation or disease causing aggregates are only found in patients expressing polyQ tracts that are longer than the pathological threshold (Table 1), one can hypothesis the following; it could be that a very critical parameter is the timing of "nucleation" of aggregate formation, which could in turn be proportional to tract length. Indeed, it has been shown that short polyQ tracts aggregate in vitro, but with delayed kinetics compared to proteins with longer tracts (Chen et al., 2002; Chen et al., 2001). Therefore, polyQ tracts of subpathological length can still be expected and should form aggregates in the cell, but at such slow rates as to be inconsequential and harmless with respect to the normal half-life of the soluble polyQ protein. Taken together, these findings suggest aggregate formation in cells could be dependent on the integration of several factors, probably not limited to tract length, timing and level of expression, and also the ability to sequester or degrade soluble polyQ proteins. These points are common to all proposed mechanisms of polyQ mediated toxicity theories (Chen et al., 2002; Chen et al., 2001).

2.6.4 Toxicity of Aggregates / Inclusion Bodies

The common paradigm used to explain the mechanism of toxicity is the formation of inclusion bodies or aggregates in cells expressing polyQ-expanded proteins. Yet, there is still considerable debate regarding whether aggregates and inclusion bodies are actually toxic (similar to the debate with prions in the mid eighties and nineties). An alternate hypothesis is that inclusion body formation is actually a cellular response that protects the cell from toxic soluble polyQ-expanded protein (Cummings et al., 1999). The fact that aggregate size and presence have been shown to correlate to affected/unaffected regions of postmortem tissue and disease severity was used to both prove and disprove the toxic aggregate hypothesis. Unfortunately, it is difficult to reach a conclusion due to largely conflicting evidence. One camp, for example, showed that aggregate presence and density positively correlate with disease severity in several polyQ expansion diseases including SBMA (Li et al., 1998), SCA1 (Skinner et al., 1997), and SCA3 (Paulson et al., 1997). The other camp argues that aggregates form in both affected and unaffected neuronal regions of the SCA7 brain (Holmberg et al., 1998). Also in HD, polyQ expansion-containing aggregates occur only in less than 1% of striatal neurons, which is the most affected neuronal region, yet they are quite common in the cortex, where aggregate size positively correlates with advanced grade tissue (Gutekunst et al., 1999). A number of questions remain unanswered including: Are the diffuse staining of polyQ-expanded proteins an indication of soluble protein or

microaggregates? Are neurons with the highest load of aggregates eliminated prior to postmortem analysis? Is there a bias in comparing analyses performed with a benchtop fluorescent microscope to those done with electron microscopy? Is it possible that signaling from "healthy" neurons with aggregates impact adjacent affected neurons that do not have aggregates?

To shed light on these issues, there must be a strict definition of how aggregates might cause toxicity. A large number of cell culture experiments have been performed, contributing to the explanation of this "toxic aggregate" hypothesis. On the other hand, several cell culture models have been used to illustrate the conflicting nature of the evidence for the "toxic aggregate" hypothesis. Treatment of an immortalized striatal cell line expressing the polyQ-expanded Huntingtin (Htt) with a number of caspase inhibitors prevented nuclear and cytoplasmic inclusion formation significantly, however, it did not alter the polyQ-mediated toxicity (Kim et al., 1999). Also, another study showed that a conditional-expression system for expression of the polyQ expanded protein demonstrated that aggregate formation is a dynamically reversible process with no link to neuronal death (Martin-Aparicio et al., 2001). As mentioned before, several studies suggest that sequestration of aggregates into inclusion bodies could be a protective mechanism for affected cells (Kopito, 2000; Taylor et al., 2003). A study showed that inclusion body formation reduced the risk of neuronal death in striatal neurons expressing pathogenic polyQ tracts by using automated live-cell imaging (Arrasate et

al., 2004). The study showed that when pathogenic length polyQ tracts of N-terminal Htt fragments were imaged over time, a positive correlation between the diffusion of Htt and cell death was established, and the formation of inclusion bodies showed a reduction of Htt diffusion levels and cell death (Arrasate et al., 2004). This strongly supports the idea that inclusion bodies actually form as a protective cellular response to a toxic protein. Such studies seem to remove the association between aggregation and the cellular toxicity of polyQ-expanded proteins. But one must remember that many cell culture studies still suggest that polyQ aggregates play a role as toxic particles. For example, a number of transfection studies with mutant Htt showed that rapid inclusion body formation resulted in nuclear envelope disruption and eventually cell death (Waelter et al., 2001). However, it should be noted that many early studies trying to establish a link between the formation of aggregates to toxicity were inherently flawed, since the expression of soluble monomelic polyQ-expanded protein was required for aggregate formation; but the presence of soluble polyQ expanded protein might have also contributed to observed pathology, which creates a bias. A solution was proposed by Yang and colleagues (Yang et al., 2002). In their study, aggregates of synthetic polyQ peptides of both normal and expanded forms were allowed to form, with and without NLS sequences. These were introduced into cells by liposomes and they observed that cell viability was reduced compared to soluble monomer polyQ peptides of matched length after the preformed aggregates targeted the nucleus, including peptides with

subpathogenic polyQ tract lengths. Such studies demonstrate a solid link between aggregation and toxicity.

3. Sequestration of Proteins and Transcriptional Dysregulation

Finding a correlation between protein aggregation/inclusion bodies and cellular toxicity is not sufficient for explaining the mechanism of toxicity of the polyQ-mediated neurodegeneration. There are a number of proposed mechanisms by which aggregate formation could induce the observed neurotoxicity of polyQ expansion disease. Depletion of polyQ protein that is caused by aggregation could cause haploinsufficiency, yet this is unlikely to be the primary mechanism for toxicity, since males with a complete loss of function of the AR develop AIS, but lack any discernable neurological phenotype similar to what is observed in patients with SBMA disease. The same is observed with individuals that are hemizygous for SCA1 or HD genes.

Another proposed mechanism for aggregate-mediated toxicity is the sequestration of proteins essential for viability. It was proposed after immunohistochemical analysis from patients (de Pril et al., 2004) and mass spectrometry analysis of aggregates purified from cell culture (Mitsui et al., 2006) showed that a large number of cellular proteins are recruited to protein aggregates formed in polyQ expansion disease. This mislocalization into protein aggregates prevents the proteins from carrying out their normal function, which leads to toxicity. The likely candidates are transcription factors.

The fact that in all polyQ expansion diseases there is a correlation between nuclear localization of polyQ-expansions and toxicity indicates that polyQ-expanded proteins might disrupt transcriptional events and hence cause toxicity. First, it is well known that

global regulation of transcription is a delicate and vulnerable synchronization of many separate events, including regulation of a complex network of protein-protein interactions and protein trafficking. Second, an important feature of many transcription factors is that they tend to have glutamine-rich sequence that has a confirmed role in promoting protein-protein interactions (Tanese et al., 1993). There are many examples of transcription factors that have glutamine-rich sequences confirmed to localize to intracellular polyQ aggregates including the cAMP response element binding protein binding protein (CBP) (McCampbell et al., 2000), the TATA-binding protein (TBP, i.e. SCA17) (Perez et al., 1998), the TBP-associated factor (TAF4) (Shimohata et al., 2000) and SP1 (Shimohata et al., 2000).

At least two reports have shown aberrant transcriptional activity caused by direct interactions of a transcription factor and a polyQ-expansion protein. For example the polyQ-expanded protein in SCA7, ataxin-7, has been shown to be a component of a histone acetyltransferase complex called the TATA-binding protein-free TBP-associated factor-containing complex (TFTC). In a mouse model of SCA7, polyQ-expanded ataxin-7-containing TFTC complex was highly recruited to specific promoters in a subset of genes expressed in rod photoreceptors, resulting in hyperacetylation of histone H3 in the promoter region and so a detectable change in transcription of these genes (Helmlinger et al., 2006). Another example is that polyQ-expanded Htt has shown the capacity to repress in vitro transcription by TFIID via direct interactions with SP1, TAF4

and RNA polymerase II-associated protein 30 kDa (RAP30) (Zhai et al., 2005). This raises another big question, which is how sequestration or inactivation of a general transcription factor by proteins expressed in myriad tissue types results in the selective degeneration of neurons only, characteristic of every polyQ expansion disease. This is a major topic of debate.

Altered chromatin acetylation is another potential mechanism of transcriptional dysregulation that is induced by the polyQ expansion and as a possible cause of neurotoxicity. The expression of polyQ protein in the nucleus of Saccharomyces cerevisiae resulted in repression that is normally caused by a specific set of genes regulated by the Spt/Aga/Gen5 acetyltransferase complex (SAGA). Also, treatment with a histone deacetylase inhibitor reversed this transcriptional repression (Hughes et al., 2001). It should be noted that the SAGA complex is a histone acetyltransferase with high homology to human CBP, which is a confirmed polyQ-interacting protein. This encouraged several groups to investigate the role of CBP's in polyQ-expansion toxicity. Their reports showed that interaction of CBP with polyQ was dependent on the acetyltransferase domain, and that a reduction in enzymatic activity was observed when mutant Htt is expressed (Steffan et al., 2001). Furthermore, CBP overexpression caused the amelioration of the polyQ-expansion toxicity (Nucifora et al., 2001). This hypothesis that polyQ-expansion toxicity could be directly caused or exacerbated by aberrant acetylation is very interesting because this phenomenon could be regulated by

an existing class of drugs, the histone deacetylase inhibitors. Therefore, it might be the most fruitful to further investigate.

4. Axonal Transport in Spinal and Bulbar Muscular Atrophy

Abnormalities in axonal transport have often been implicated in neurodegenerative diseases, which might explain the high vulnerability of long projection neurons (motoneurons for example) (Perlson et al., 2010; Morfini et al., 2009). Axonal transport is a process that occurs throughout the lifetime of a neuron and is essential for the growth and survival of neurons. It transports axonal proteins between the neuronal cell body and the axon tip (Curtis et al., 1998). Also, it is well accepted that the AR mediates most of the actions of androgens at the genomic level by acting as a transcription factor. However, it should be noted that recent evidence has shown that AR might have effects in the nervous system in a nongenomic manner, where androgens act independently of the AR (Foradori et al., 2008; Losel et al., 2003). This, combined with the fact that AR is located in high concentrations in lower motorneurons of the spinal cord and brainstem (Matsumoto, 1997; Yu et al., 2001; Tetzlaff et al., 2007) and that androgens regulate the growth of the very long axons, would highlight, if we used the right experimental techniques, whether SBMA is caused by a mechanism that affects axonal transport or not.

It has been suggested that SBMA may prevent the normal pattern of connectivity of motorneurons by interfering with the control of their neurite outgrowth. Motoneurons that express mutant SBMA AR exhibit neurites that are typically short and dystrophic with abnormal branching patterns (Poletti, 1999); These changes in morphology raise the

possibility that axonal transport may initiate motoneuron pathogenesis (Avila et al., 2003; Pozzi et al., 2003; Szebenyi et al., 2003). Eventually, the loss of connectivity with the target musculature could cause the death of the motoneuron (Fargo et al., 2008). This possibility is further reinforced because SBMA mutations lead to both a gain-of-function and a loss of function problem. The former is suspected in forming aggregates that would interfere with trafficking and connectivity in the cell. It is possible that the very long axons are susceptible to such interruptions in axonal trafficking (Ellerby et al., 1999). The latter, would interrupt the normal role of androgens in upregulating neuritin, which otherwise might have mitigated the effects of the aggregates, and maintain or recover normal axonal functionality (Fargo et al., 2008). Another study that supports this view found abnormalities in the endosomal trafficking in sciatic nerve axons using live imaging. It showed disease-induced deficits in the flux and run length of endosomes movement toward the cell body in both knock-in (KI) and myogenic transgenic (TG) male mouse models of SBMA (Kemp et al., 2011).

Some argue against the SBMA - axonal transport relation. A study carried out in 2004, where the first 20 and 100 CAG repeats were introduced into mouse embryonic stem cells to produce transgenic mice that show symptoms of SBMA motor neuronopathy, reported that induced death of motor neurons could be rescued by VEGF. This suggests that altered expression of VEGF could be involved in SBMA motor neuronopathy (Sopher et al., 2004). A more recent study examining the expression

levels of motor and microtubule-associated protein tau in the spinal cord and sciatic nerve of both WT and SBMA mouse found no difference at various stages of disease progression (Malik et al., 2011). There were no differences in the binding properties of motor proteins and tau to microtubules (Malik et al., 2011). Most importantly, axonal transport rates of in vitro and in vivo adult WT and mutant SBMA mice were analyzed and demonstrated no significant axonal transport deficits. Such results indicate that impairment of axonal transport does not correlate with the pathogenesis of SBMA (Malik et al., 2011).

5. Ubiquitin Proteasome System for Protein Degradation

The ubiquitin proteasome system (UPS) functions as the primary non-lysosmal mechanism for protein degradation in eukaryotic cells. The analysis of polyQ-expansion-induced protein aggregates in cells shows the presence of several components of the UPS, including UPS adaptor proteins, ubiquitin and proteasome subunits (Cummings et al., 1998; DiFiglia et al., 1997). In fact, this has been shown in cell culture, animal models and in patient tissues in all polyQ expansion disease types (Cummings et al., 1998; DiFiglia et al., 1997). Most reports on the toxicity of the polyQ involve the UPS in the clearance of soluble or aggregated polyQ-expanded protein (Cummings et al., 1998; DiFiglia et al., 1997). There are several mechanisms that show that the UPS is inhibited in a general manner such as the direct inhibition of proteasomes by the polyQ-expanded protein as a result of failure to degrade and release the protein. This has been shown by determining the activity of the UPS required for proteasome recognition, catalytic proteolysis, ATP-dependent unfolding and deubiquitination, and post-translational targeting of substrates (Cummings et al., 1998; DiFiglia et al., 1997). The enzymatic activity of the proteasome was determined by using small peptide fluorigenic substrates and the fluorigenic substrate assays used to prove the proteasome impairment in the expression of polyQ-expanded protein (Cummings et al., 1998; DiFiglia et al., 1997). Functional studies in vivo confirmed the association between the UPS and aggregates or inclusion bodies (Cummings et al., 1998; DiFiglia

et al., 1997). Previous studies mentioned that the proteasome plays a role in the degradation of aggregates containing polyQ-expanded protein that were found in a mouse model of HD (Martin-Aparicio et al., 2001). Previous work suggests that the AR without a polyQ tract (0QAR) and the normal AR (20QAR) can be degraded via the UPS. However, the longer polyQ expanded AR (50QAR) appears to inhibit degradation of particular proteins by UPS (Mandrusiak et al., 2003). It is hypothesized that these SBMA-inducing mutant ARs can accumulate in the cell and cause disease by evading UPS degradation.

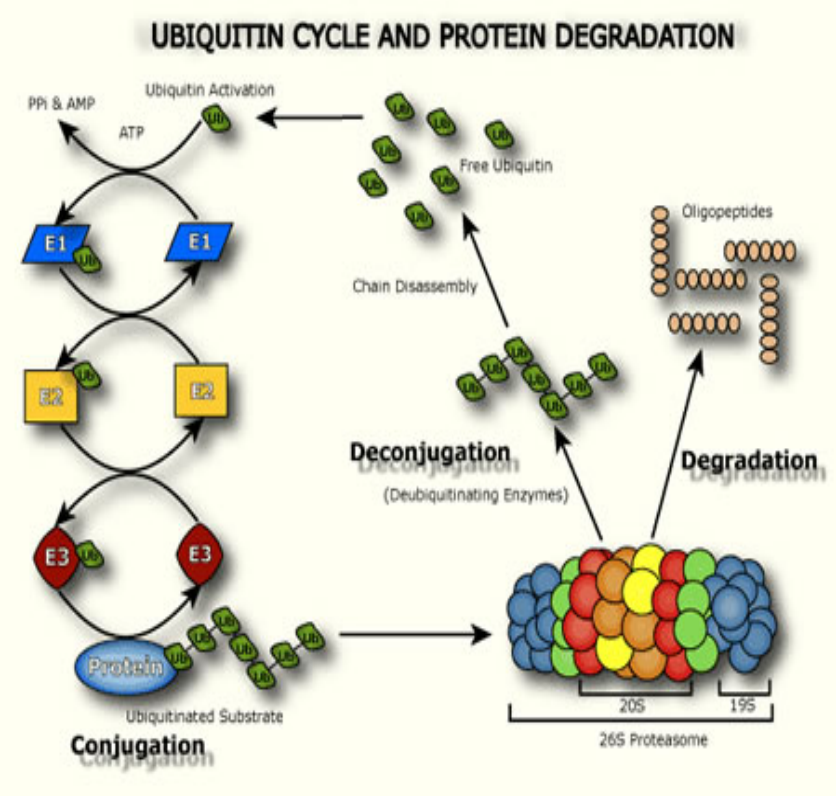

Figure 4: The process of ubiquitination and protein degradation

E1, E2 and E3 enzymes add ubiquitin monomers to protein substrates by the process of ubiquitin conjugation. Degradation by the 26S proteasome requires the attachment of at least 4 ubiquitin moieties to signal targeting protein substrates. Small peptides exit the proteasome by passive diffusion after their release by proteolysis in the central core of the proteasome. Deubiquitinating enzymes recycle ubiquitin chains (BOSTON BIOCHEM, INC).

6. Expressed Protein Ligation System

Expressed protein ligation (EPL) is a semisynthetic technique in which a recombinant protein thioester generated by thiolysis of an intein fusion protein is reacted with a synthetic or recombinant peptide with an N-terminal cysteine to produce a native peptide bond (Ghosh et al., 2011). This technique is also known as intein-mediated protein ligation or IPL. It is named this way because the technique uses an intein to make a protein possessing a C-terminal thioester, which is then ligated to a protein or peptide with an amino-terminal cysteine (Ghosh et al., 2011).

In the late eighties, chemical ligation emerged as an approach for the total synthesis of proteins. It is a process which allows unprotected peptides to be chemoselectively joined together. However, it was not until the early 1990s that the use of chemoselective coupling reaction with fully unprotected synthetic peptide building blocks was introduced as an idea in the Kent laboratory (Muir, 2003). In a ground breaking paper, the 99-residue HIV-1 protease was assembled from two 50-residue unprotected peptides using a thioester bond–forming chemical reaction which was performed in an aqueous solution at pH 7 (Ghosh et al., 2011). A large amount of research effort has gone into expanding and refining the chemical ligation approach, partially because of the elegance and practicality of the approach. Such efforts made a lot of changes to the original thioester ligation strategy. It has now been joined by ligation approaches based on thioether, oxime/hydrazone, thiazolidine, and amide derivatives (Ghosh et al., 2011).

E. coli is usually used for heterologous protein production because of its fast growth, easy fermentation and general simplicity of its nutritional and sterility requirements (Ghosh et al., 2011). A variety of intein-enabled approches have solved the problem of size. Unlike most chemical labeling methods, EPL uses a thiol reagent that can be easily removed. The ligated peptide or protein need only contain an N-terminal cysteine. It should be noted that a peptide as small as two amino acids can be added to a protein of interest, which reduces perturbations to the native structure of the protein (Ghosh et al., 2011).

To do EPL, the first step is to clone the gene encoding the protein of interest. The target protein is fused to the N-terminus of a modified intein and a small chitin-binding domain (CBD, 6 kDa) (Figure5). Expression and affinity purification of the three-part fusion protein from E. coli is then carried out, and the intein cleavage is initiated with a thiol reagent like 2-mercaptoethanesulfonic acid to release the target protein (Ghosh et al., 2011; McGinty et al., 2009). Then the resulting thioester-tagged protein is combined with a peptide or protein possessing an amino-terminal cysteine so that ligation occurs spontaneously to produce a native peptide bond between the proteins or protein peptide complex (Figure 6). In most cases protein purification is done by using 1,4-dithiothreitol (DTT) as the intein cleavage reagent (New England BioLabs, 2006).

It should be noted that higher ligation efficiencies are associated with the use of thiophenol or 2-mercaptoethanesulfonic acid (MESNA) rather than DTT to induce cleavage.

In vivo and thiol-induced cleavage at the N-terminus of an intein is affected by the target protein sequence, where the C-terminal amino acid residue of the target sequence has an important role on the effect. It is thought that every intein has evolved in its own host protein context and so may display different reaction rates when responding to non-native amino acid residues adjacent to the peptide bond. The many choices of inteins allows for the optimization of cleavage and ligation reactions depending on the protein of interest to isolate (Ghosh et al., 2011; McGinty et al., 2009).

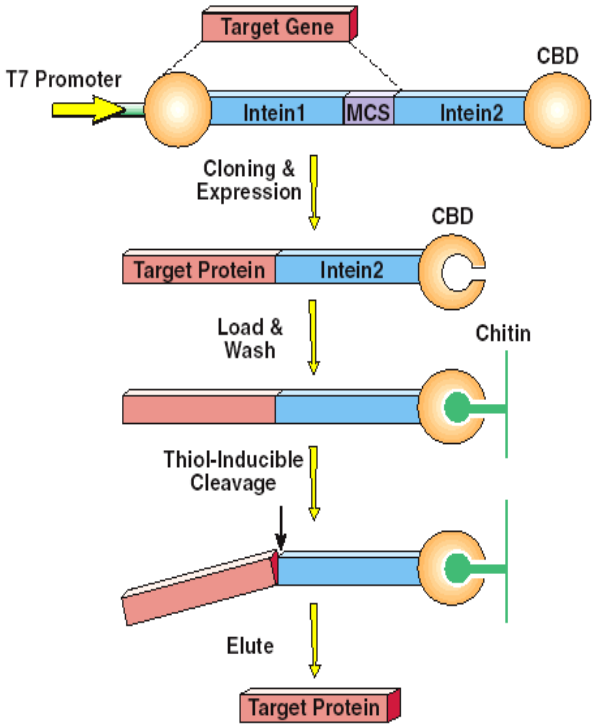

Figure 5: Expressing and purifying the cloned proteins (0Q,20Q,50Q) and human ubiquitin by using pTWIN2 expression system (New England BioLabs, 2006).

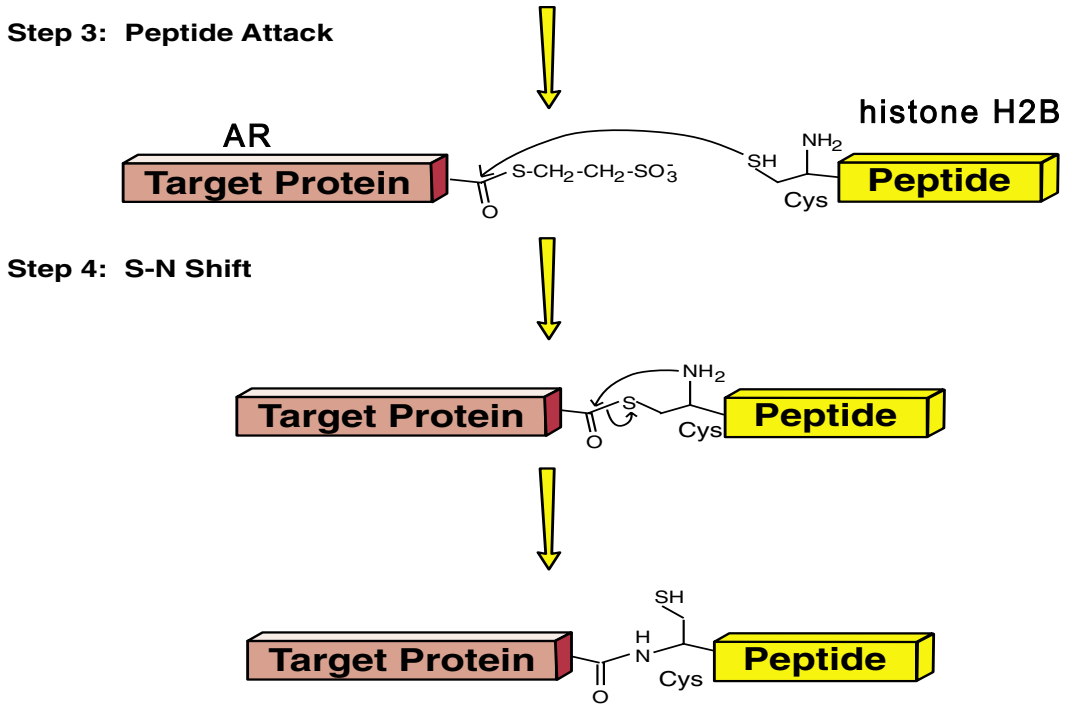

Figure 6: Mechanism showing the ligation of the histone H2B peptide to AR protein

Schematic of expressed protein ligation (EPL). Expression of the target protein (AR) is at the N-terminus of an intein–chitin binding domain (CBD) protein fusion. After binding of the expressed fusion protein to the chitin resin, cleavage between the AR and the intein is done by adding the thiol reagent 2-mercaptoethanesulfonic acid (MESNA) to release the AR with a C-terminal thioester. Then, histone H2B (peptide made in house) is ligated to the thioester-tagged protein by a peptide bond to finally produce a AR protein tagged with the histone H2B peptide (New England BioLabs, 2006).

In general the EPL procedure is very easy to perform. The product can be used in a variety of biochemical studies because it contains a native peptide bond between the reactants at the point of ligation. For example, enhanced sensitivity in protein array and ELISA analysis can be easily achived by using EPL for ligation of peptides to carrier proteins (Ghosh et al., 2011). Generation of phosphorylated substrates to study phosphatase activity by Western blot analysis can be easily done when using EPL (Ghosh et al., 2011).

In conclusion, EPL is a semisynthetic method that has been developed for the creation and joining of small or large polypeptides. This approach has been used in various applications including the production of toxic proteins; introduction of non-coded amino acids, fatty acids, or chemically-modified amino acid residues; segmental protein labeling; generation of positive controls for western blot analysis; and production of substrates for enzymatic assays and epitope mapping (Ghosh et al., 2011).

Moreover, in the ubiquitin (Ub) field, a major diffculty was producing homogenously ubiquitylated histones in vitro. EPL overcomes this as a useful method to ligate the human ubiquitin (HUB) with the synthetic peptide (H2B) (McGinty et al., 2009). The molecular weight of HUB is ~8 KDa (76 amino acids). In this project I was successfully able to purify HUB and ligate the H2B peptide to the HUB using the EPL method (McGinty et al., 2009).

7. Work Accomplished

The EPL method was used in this project to synthesis AR peptides and ligate test peptides to the AR peptides. In order to achieve this goal the following steps were involved as follows:

1. Fragments of the AR containing different polyGln sizes were amplified by PCR and cloned into the pTWIN2 vector.
2. The cloned proteins (AR0Q, AR20Q, AR50Q) were then expressed and purified using the pTWIN2 system.
3. Cleaved proteins (AR0Q, AR20Q, AR50Q) were produced by using thiol-induced cleavage with 2-mercapto-ethanesulfonic acid (MESNA) and 1, 4-Dithiothreitol (DTT).
4. Cleaved proteins (AR0Q, AR20Q, AR50Q) were ligated to different synthetic peptides (HA epitope) using EPL.
5. Human proteasomes were purified and their activity was tested by using succinly-Leu-Leu-Val-Tyr-7-amino-4-methylcoumarin (Suc-LLVY-AMC), a fluorogenic substrate.
6. Contributed to a review on “Mechanisms Mediating Spinal and Bulbar Muscular Atrophy: Investigations into Polyglutamine-Expanded Androgen Receptor Function and Dysfunction” by Lenore K. Beitel, Carlos Alvarado, **Shaza**

Mokhtar, Miltiadis Paliouras, and Mark Trifiro. Front Neurol, May 2013, volume 4, pages 53.

8. Methods & Results

Polymerase chain reaction (PCR) method

Fragments of the (5 µL) AR (0Q, 20Q, 50Q) were amplified by PCR using primer A 5`...GGTGGTCATATGCACCACCATCACCATCACGACGATGACGATAAG...3` as the forward primer (NdeI) and primer B 3`...CTAGAAGGCGACTCGGATCCGTCGCCTTCTCGTTGGTGG...5` as the reverse primer (SapI). The primers were purchased from Invitrogen. Moreover, for making these fragments of the AR (0Q, 20Q, 50Q) (part of the N-terminal of exon 1), a master mix was used. The master mix had 6 µL 5x PCR buffer (QIAGEN), 0.6 µL 10 mM dNTP, 21.8 µL ddH_2O, 0.15 µL primer A (NdeI), 0.18 µL primer B (SapI) and 0.3 µL DNA polymerase (HotStarTaq Plus) and 17 µL DNA template (pcDNA 0Q, pcDNA 20Q, and pcDNA 50Q) (Abdullah et al., 1998; Paliouras et al., 2011). The conditions of the PCR were 98°C for 30 seconds on the initial cycle, and then 10 seconds at 98°C and 74°C for 45 seconds for each of 30 cycles. Also, I did a maxi prep (as suggested by the manufacturer QIAGEN Plasmid Purification Handbook) for the pTWIN2 plasmid (NEB Impact kit) and calculated the concentration of the DNA to be 745.9 ng/µL (373 µg). The PCR products were cut by NdeI at the 5' end and by Sap I at the 3' end. The digestion conditions were 3 µL buffer 4 (New England Biolabs), 1 µL NdeI, 1 µL SapI, 5 µL ddH_2O, 20 µL DNA, and incubated overnight at 37 °C, and then samples (15 µL) were run on 1% agarose gel (Figure 8). The pTWIN2 vector was cut by the same restriction

enzymes for the ligation the AR gene into the pTWIN2 vector. The following map (Figure 7) shows the location of the restriction enzymes including (Nde I /Sap I) (New England Biolabs) on the pTWIN2 vector. All AR fragments and the digested pTWIN2 vector were cut out of the agarose gel and purified using gel extraction kit (Qiagen).

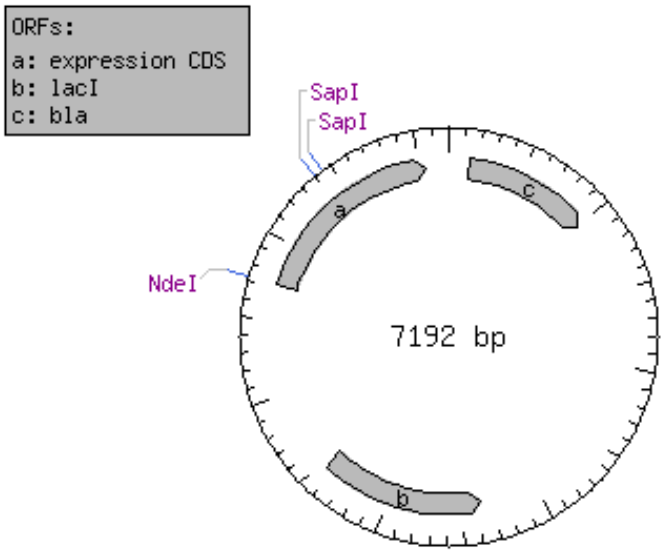

Figure 7: Map showing the unique Nde I and Sap I sites in pTWIN2

NdeI/SapI digest

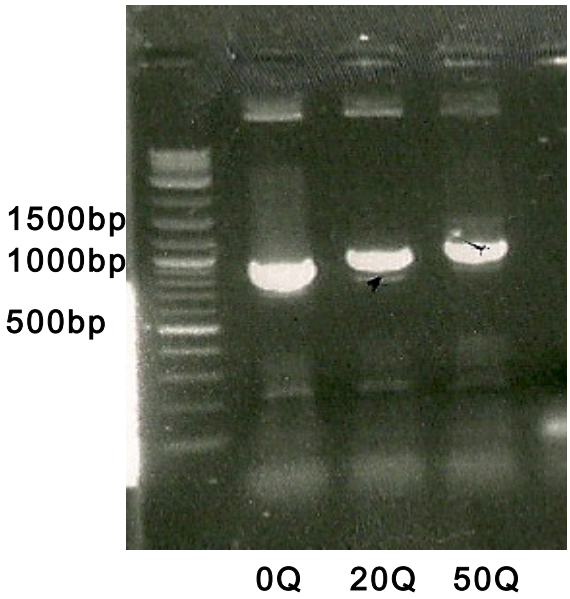

Figure 8: Cloning the polyGln tract of the human androgen receptor by PCR

Agarose gel (1%) showing the differences in the size of the inserts for the androgen receptor (AR) . These AR fragments (0Q,20Q,50Q) were made by using the restriction enzyme digestion by NdeI and SapI after polymerase chain reaction (PCR) amplification. Tris-Borate-EDTA (TBE) buffer was used to run the gel for an hour at 100 volts. The observed sizes of the AR fragments were 1000 bp, 1142 bp, 1218 bp for the AR 0Q, 20Q, and 50Q fragments, respectively. The Quick-Load 2-log DNA Ladder was used as a marker (Qiagen).

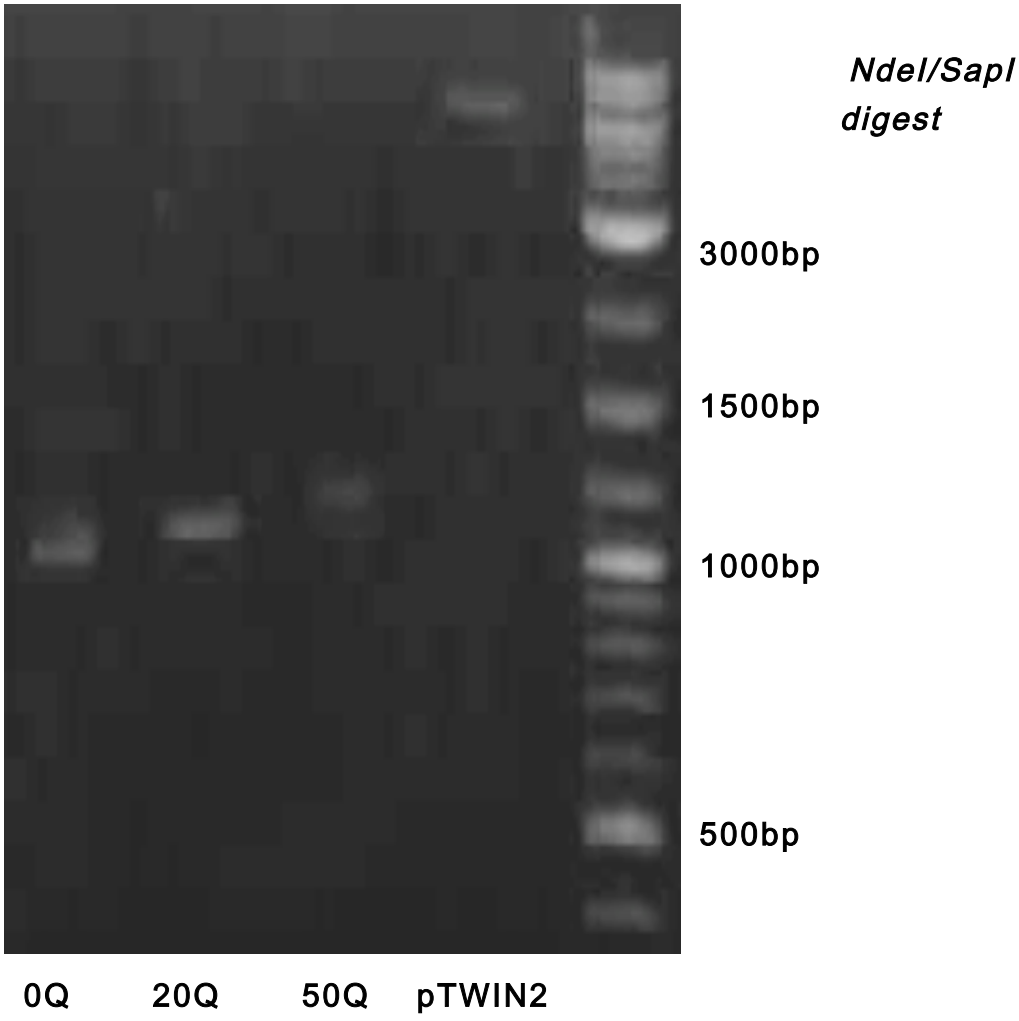

Figure 9: Cloning the polyGln tract of the human androgen receptor by PCR

Agarose gel showing inserts for the androgen receptor (AR) and the vector (pTWIN2). AR (0Q,20Q,50Q) fragments were made by restriction enzyme digests with NdeI and SapI after polymerase chain reaction (PCR) amplification. The pTWIN2 vector was digested with the same restriction enzymes and expected to produce a DNA fragment of 6471 bp. The Quick-Load 2-log DNA Ladder was used (Qiagen).

Cloning method

Purified PCR amplified fragments of AR (0Q, 20Q, and 50Q) were cloned into the NdeI to Sap I sites in pTWIN2 (6471 bp). pTWIN2 contains a SapI site that allows the AR gene to be cloned before an intein tag without the addition of vector-derived sequences at either terminus of the AR gene. Ligations were done by adding ligation mix (10x T4 DNA ligase buffer, T4 DNA ligase) (New England Biolabs), 1 µL insert (0Q, 20Q, 50Q), 1 µL vector (pTWIN2), and ddH_2O and incubated overnight at room temperature. Then, I did transformation for the pTWIN2 vectors containing the inserts. The conditions for the BL21 (DE3) bacterial transformation were heat shock for 30 seconds and 45 mins incubation time for the sample in the shaker as suggested by the manufacturer (New England Biolabs). The LB ampicillin plates were incubated overnight at 37°C. Plasmid DNA was extracted from bacteria incubated overnight with shaking in 5ml LB Broth media with 100 µg/ml Ampicillin using Miniprep kit from Qiagen. The phUB(1-75) plasmid was obtained from Dr. Tom Muir (Muir, 2003).

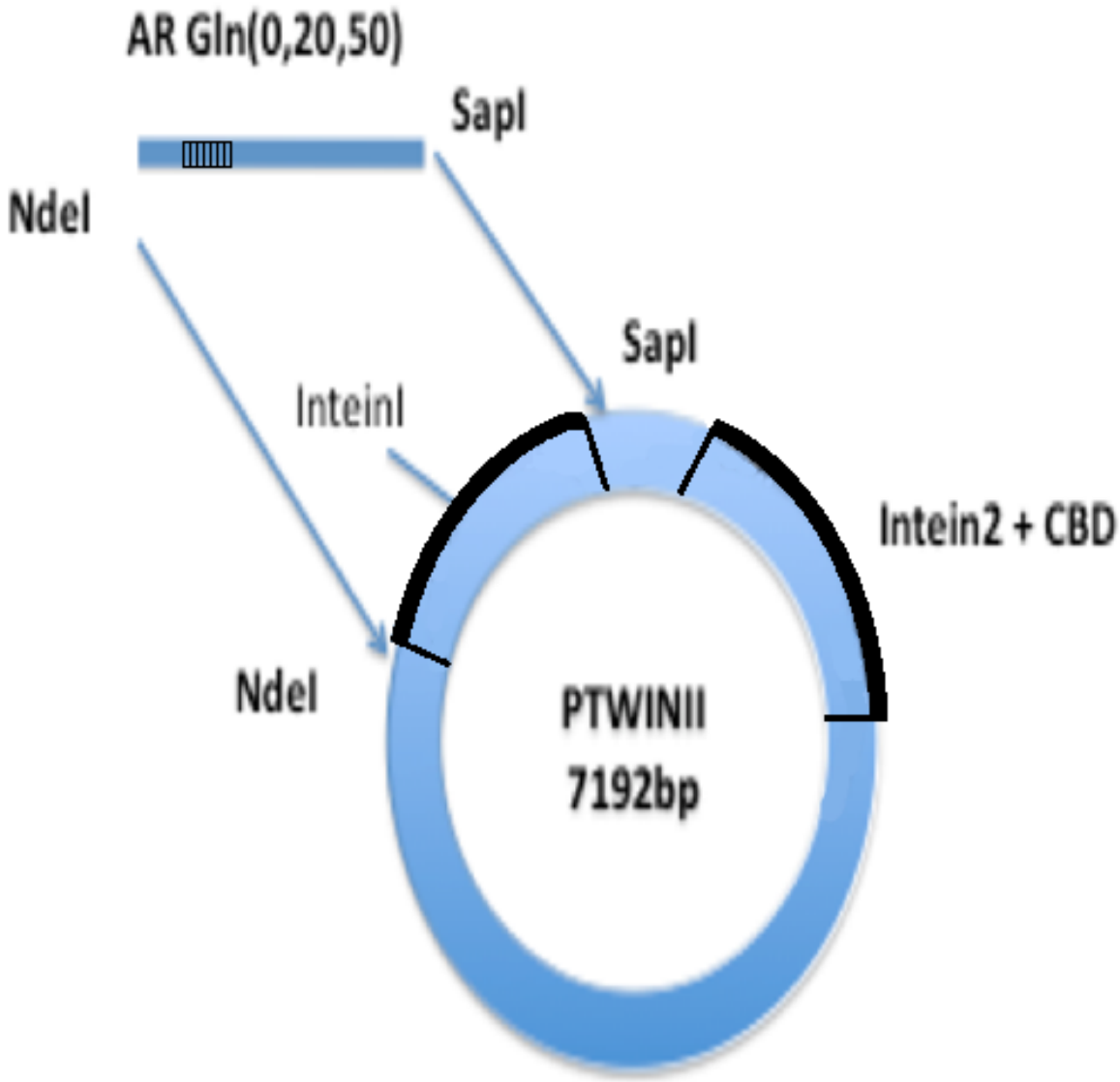

Figure 10: Diagram showing the cloning of AR Gln (0, 20,50) into the NdeI to SapI sites in pTWIN2

Next, I did double digests of the miniprep DNA by using different restriction enzymes (REs) such as NdeI/AvrII (see Figures 11,12,13), and NdeI/AflII (results not shown) to check if the inserts were the expected size. The samples were run on 1% agarose gels to check the fragment sizes. Nine positive clones were found for the different AR sizes (i.e. 0Q, 20Q, 50Q). Detected sizes were compared with predicted size, obtained by digesting the AR-pTWIN2 using the Enzyme Cutter tool provided by New England Biolabs (see Table 2). The positive clones were digested with the REs, which I ran each time, however, undigested AR, and undigested vector, AR in pcDNA with REs was run sometimes. The table below shows the expected sizes of the band for each digest.

Table 2: Predicted sizes for double digests

Restriction Enzymes	Cut size for AR 0Q, 20Q, 50Q (bp) in pTWIN2	Uncut size for AR 0Q, 20Q, 50Q (bp) in pTWIN2
NdeI/AvrII	0Q: 6,477+ 943 20Q: 6,477+ 1003 50Q: 6,477+1093	7,420 7,480 7,570
NdeI/AflII	0Q: 6,930+ 490 20Q: 6,930+550 50Q: 6,930+640	7,420 7,480 7,570

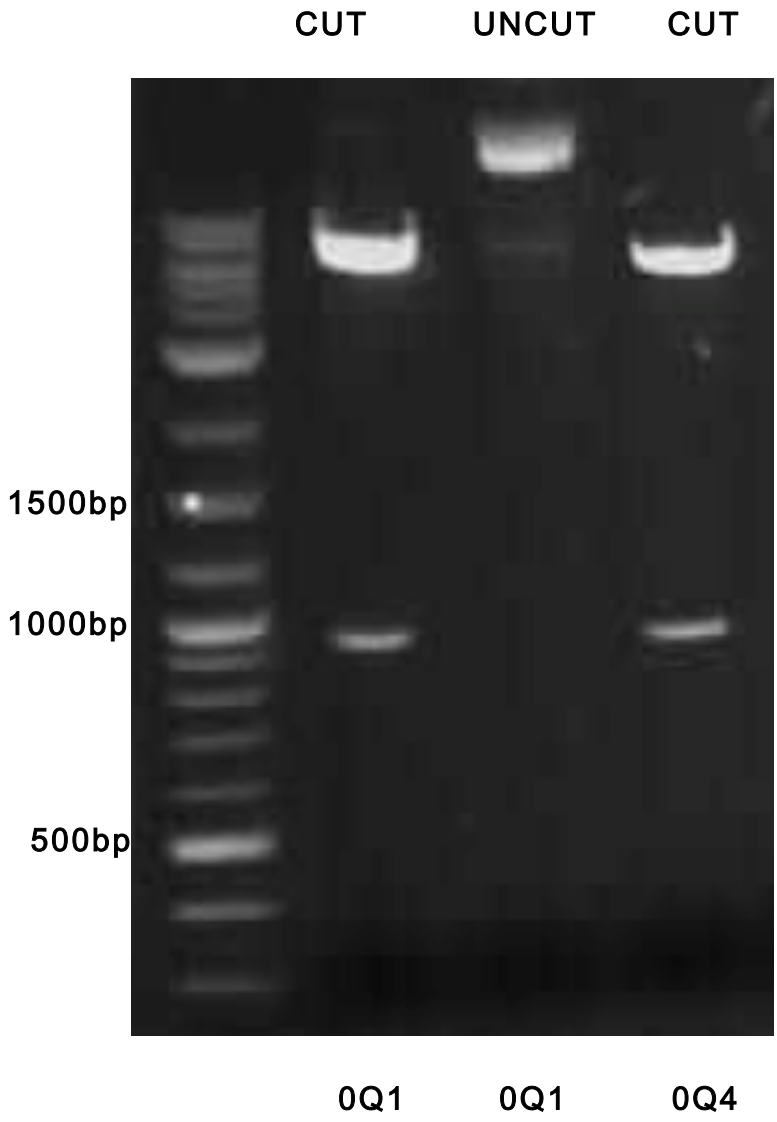

Figure 11: Double digest of potential 0QAR human androgen receptor clones by NdeI/AvrII

This figure shows the agarose gel with cut and uncut AR 0Q clones. Clones 0Q1 and 0Q4 AR plasmid DNAs were double digested by using the restriction enzymes NdeI and AvrII. The observed sizes for the lower fragments were 900 bp and 940 bp for the AR 0Q1 and 0Q4 fragments, respectively. The Quick-Load 2-log DNA Ladder was used.

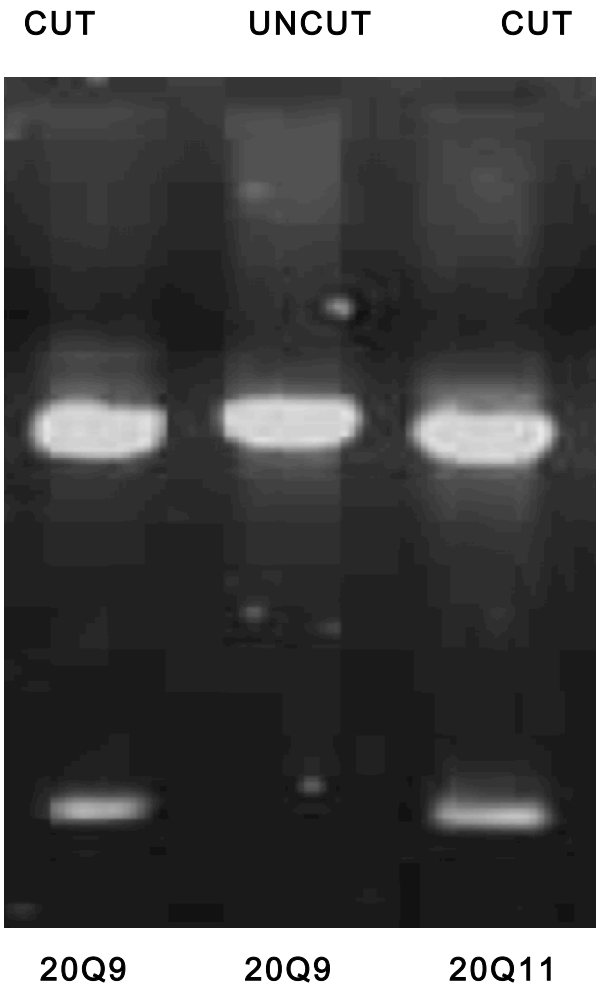

Figure 12: Double digest of potential 20QAR human androgen receptor clones by NdeI/AvrII

This figure shows agarose gel with cut and uncut AR 20Q clones. Clones (20Q9 and 20Q11) AR plasmid DNAs were double digested by using the restriction enzymes (NdeI and AvrII). The observed size for the lower fragments were 1022 bp for the AR 20Q9 and 20Q11 fragments, respectively. The Quick-Load 2-log DNA Ladder was used.

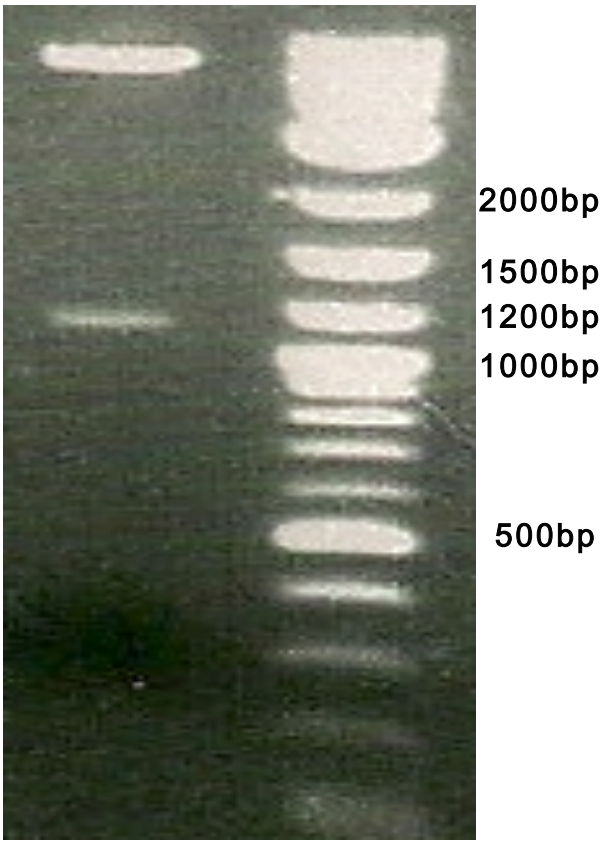

Figure 13: Double digest of potential 50QAR human androgen receptor clone by NdeI/AvrII

This figure shows agarose gel with AR 50Q clone. Clone 50Q AR plasmid DNAs was double digested by using the restriction enzymes (NdeI and AvrII). The observed size for the lower fragment was 1210 bp for the AR 50Q2 fragment. The Quick-Load 2-log DNA Ladder was used.

Sequencing

Plasmid DNA samples were sent to the McGill University and Genome Quebec Innovation Centre to confirm the sequence. Sequencing, completed by Genome Quebec, confirmed that the nine clones AR the correct AR sequence and cloning sites. The primers used were T7 (TAATACGACTCACTATAGGG), 1.2A (TCCTGGATGAGGAACAGCAA), 1.3A (CCACTTTTGGGAGTTCCACC), and 1.3A°(GCTCCCACTTCCTCCAAGGACAATTAC). CLUSTALW and other web sites helped me align the sequences to confirm there were 0, 20, and 50 CAG/CAA codons in clones 0Q, 20Q, 50Q, respectively. Also, I calculated the protein molecular weight to compare them to the AR peptides produced fragments (ClustalW2).

Table 3: The following sequence shows the human AR 20Q region from NdeI and Sap I sites as amplified with primers A and B.

The 5' NdeI site and the reverse complement of the 3' SapI site are underlined. The ATG start codon is in bold.

GGTGGTCAT**ATG**CACCACCATCACCATCACGACGATGACGATAAGATGGAAGTGC
AGTTAGGGCTGGGAAGGGTCTACCCTCGGCCGCCGTCCAAGACCTACCGAGGAG
CTTTCCAGAATCTGTTCCAGAGCGTGCGCGAAGTGATCCAGAACCCGGGCCCCAG
GCACCCAGAGGCCGCGAGCGCAGCACCTCCCGGCGCCAGTTTGCTGCTGCTGCA
GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC
AGCAGCAAGAGACTAGCCCCAGGCAGCAGCAGCAGCAGCAGGGTGAGGATGGTT
CTCCCCAAGCCCATCGTAGAGGCCCCACAGGCTACCTGGTCCTGGATGAGGAACA
GCAACCTTCACAGCCGCAGTCGGCCCTGGAGTGCCACCCCGAGAGAGGTTGCGT
CCCAGAGCCTGGAGCCGCCGTGGCCGCCAGCAAGGGGCTGCCGCAGCAGCTGC
CAGCACCTCCGGACGAGGATGACTCAGCTGCCCCATCCACGTTGTCCCTGCTGGG
CCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTTAAAGACATCCTGAGC
GAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCAGTATCCGAAG
GCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCCAAG
GACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAA
GGCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGAGTCC
AGGGGAACAGCTTCGGGGGGATTGCATGTACGCCCCACTTTTGGGAGTTCCACCC
GCTGTGCGTCCCACTCCTTGTGCCCCATTGGCCGAATGCAAAGGTTCTCTGCTAG
ACGACAGCGCAGGCAAGAGCACTGAAGATACTGCTGAGTATTCCCCTTTCAAGGG
AGGTTACACCAAAGGGCTAGAAGGCGAGAGCCTAGGCTGCGGAAGAGCAACCAC
C

Tabel 4: Sequence of AR OQ1 clone aligned with predicted pTWIN2-AR0Q sequence.

The AR 0Q1 clone was sequenced using primer (1.2A). The sequence was aligned with pTWIN2 sequence using ClustalW2.

```
pTWIN2          GCCCTGGAGTGCCACCCCGAGAGAGGTTGCGTCCCAGAGCCTGGAGCCGCCGTGGCCGCC 6120
0Q Seq          -----------------CNGNNNNNNTTGCGTCCCAGAGCCTGGAGCCGCCGTGGCCGCC 43
                                 *       **********************************

pTWIN2          AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 6180
0Q Seq          AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 103
                ************************************************************

pTWIN2          TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 6240
0Q Seq          TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 163
                ************************************************************

pTWIN2          AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 6300
0Q Seq          AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 223
                ************************************************************

pTWIN2          GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 6360
0Q Seq          GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 283
                ************************************************************

pTWIN2          AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 6420
0Q Seq          AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 343
                ************************************************************

pTWIN2          GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 6480
0Q Seq          GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 403
                **************************************************
```

Table 5: Sequence of AR 2OQ1 clone aligned with predicted pTWIN2-AR20Q sequence.

The AR 20Q1 clone was sequenced using primer (1.2A). The sequence was aligned with pTWIN2 sequence using ClustalW2.

```
20Q1            ----------------CCGAGNGNNNTTGCGTNNN-GAGCCTGGAGCCGCCGTGGCCGCC 43
pTWIN2          GCCCTGGAGTGCCACCCCGAGAGAGGTTGCGTCCCAGAGCCTGGAGCCGCCGTGGCCGCC 6120
                                 ***** *   ******    ************************

20Q1            AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 103
pTWIN2          AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 6180
                ************************************************************

20Q1            TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 163
pTWIN2          TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 6240
                ************************************************************

20Q1            AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 223
pTWIN2          AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 6300
                ************************************************************

20Q1            GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 283
pTWIN2          GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 6360
                ************************************************************

20Q1            AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 343
pTWIN2          AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 6420
                ************************************************************

20Q1            GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 403
pTWIN2          GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 6480
                **************************************************
```

Table 6: Sequence of AR 5OQ2 clone aligned with predicted pTWIN2-AR50Q2 sequence

The AR 50Q2 clone was sequenced using primer (1.2A). The sequence was aligned with pTWIN2 sequence using ClustalW2.

```
pTWIN2          GCCCTGGAGTGCCACCCCGAGAGAGGTTGCGTCCCAGAGCCTGGAGCCGCCGTGGCCGCC 6120
50Q             ------------------------------GTCCCAGAGCCTGGAGCCGCCGTGGCCGCC 30
                                               ******************************

pTWIN2          AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 6180
50Q             AGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA 90
                ************************************************************

pTWIN2          TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 6240
50Q             TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTT 150
                ************************************************************

pTWIN2          AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 6300
50Q             AAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA 210
                ************************************************************

pTWIN2          GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 6360
50Q             GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCC 270
                ************************************************************

pTWIN2          AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 6420
50Q             AAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG 330
                ************************************************************

pTWIN2          GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 6480
50Q             GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGA 390
                **************************************************
```

Table 7: Sequence shows the glutamine tract regions of pTWIN-AR 0Q, 20Q, 50Q clones obtained using T7 primer.

The sequencing confirmed there were 0, 20, and 49 CAG/CAA codons in clones 0Q, 20Q, and 50Q, respectively. The AR ATG start codon is underlined.

```
20Q1_T7_        TCCCTNNNANNNTTTTGTTNNNTTNNGAAGGAGATATACATATGCACCACCATCACCATC 60
50Q2_T7_        ------------TTTTGTTTACTTTAGAAGGAGATATACATATGCACCACCATCACCATC 48
0Q1_T7_         ------------TTTTGTTTACTNNNGAAGGAGATATACATATGCACCACCATCACCATC 48
                           *******   *   ************************************

20Q1_T7_        ACGACGATGACGATAAGATGGAAGTGCAGTTAGGGCTGGGAAGGGTCTACCCTCGGCCGC 120
50Q2_T7_        ACGACGATGACGATAAGATGGAAGTGCAGTTAGGGCTGGGAAGGGTCTACCCTCGGCCGC 108
0Q1_T7_         ACGACGATGACGATAAGATGGAAGTGCAGTTAGGGCTGGGAAGGGTCTACCCTCGGCCGC 108
                ************************************************************

20Q1_T7_        CGTCCAAGACCTACCGAGGAGCTTTCCAGAATCTGTTCCAGAGCGTGCGCGAAGTGATCC 180
50Q2_T7_        CGTCCAAGACCTACCGAGGAGCTTTCCAGAATCTGTTCCAGAGCGTGCGCGAAGTGATCC 168
0Q1_T7_         CGTCCAAGACCTACCGAGGAGCTTTCCAGAATCTGTTCCAGAGCGTGCGCGAAGTGATCC 168
                ************************************************************

20Q1_T7_        AGAACCCGGGCCCCAGGCACCCAGAGGCCGCGAGCGCAGCACCTCCCGGCGCCAGTTTGC 240
50Q2_T7_        AGAACCCGGGCCCCAGGCACCCAGAGGCCGCGAGCGCAGCACCTCCCGGCGCCAGTTTGC 228
0Q1_T7_         AGAACCCGGGCCCCAGGCACCCAGAGGCCGCGAGCGCAGCACCTCCCGGCGCCAGTTTGC 228
                ************************************************************

20Q1_T7_        TGCTGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300
50Q2_T7_        TGCTGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 288
0Q1_T7_         TGCTGCTG-------------------------------------------------------- 236
                ********

20Q1_T7_        AANAG----------------------------------------------------------- 305
50Q2_T7_        AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 348
0Q1_T7_         ------------------------------------------------------------


20Q1_T7_        --------------------------------ACTAANNNNNGGCCCCAG-CAGCAGCAGC 333
50Q2_T7_        AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGAGACTAGCCCCAGGCAGCAGCAGC 408
0Q1_T7_         -----------------------------------GAGACTAGCCCCAGGCAGCAGCAGC 261
                                                   ******* **********
```

Table 8: Predicted 0Q AR protein sequence after purification and cleavage using pTWIN2 system

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence: 0Q SEQ

```
       10         20         30         40         50         60
MHHHHHHDDD DKMEVQLGLG RVYPRPPSKT YRGAFQNLFQ SVREVIQNPG PRHPEAASAA

        70         80         90        100        110        120
PPGASLLLLQ ETSPRQQQQQ QGEDGSPQAH RRGPTGYLVL DEEQQPSQPQ SALECHPERG

       130        140        150        160        170        180
CVPEPGAAVA ASKGLPQQLP APPDEDDSAA PSTLSLLGPT FPGLSSCSAD LKDILSEAST

       190        200        210        220        230        240
MQLLQQQQQE AVSEGSSSGR AREASGAPTS SKDNYLGGTS TISDNAKELC KAVSVSMGLG

       250        260        270        280        290        300
VEALEHLSPG EQLRGDCMYA PLLGVPPAVR PTPCAPLAEC KGSLLDDSAG KSTEDTAEYS

       310
PFKGGYTKGL EGESLG
```

Theoretical pI/Mw: 4.94 / 33299.81

Table 9: Predicted 20Q AR protein sequence after purification and cleavage using pTWIN2 system

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence: 20Q SEQ

```
   10         20         30         40         50         60
MHHHHHHDDD DKMEVQLGLG RVYPRPPSKT YRGAFQNLFQ SVREVIQNPG PRHPEAASAA

        70         80         90        100        110        120
PPGASLLLL QQQQQQQQQQ QQQQQQQQQQ ETSPRQQQQQ QGEDGSPQAH RRGPTGYLVL

       130        140        150        160        170        180
DEEQQPSQPQ SALECHPERG CVPEPGAAVA ASKGLPQQLP APPDEDDSAA PSTLSLLGPT

       190        200        210        220        230        240
FPGLSSCSAD LKDILSEAST MQLLQQQQQE AVSEGSSSGR AREASGAPTS SKDNYLGGTS

       250        260        270        280        290        300
TISDNAKELC KAVSVSMGLG VEALEHLSPG EQLRGDCMYA PLLGVPPAVR PTPCAPLAEC

       310        320        330
KGSLLDDSAG KSTEDTAEYS PFKGGYTKGL EGESLG
```

Theoretical pI/Mw: 4.94 / 35734.29

Table 10: Predicted 50Q AR protein sequence after purification and cleavage using pTWIN2 system

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence: 50Q SEQ

```
  10         20         30         40         50         60
MHHHHHHDDD DKMEVQLGLG RVYPRPPSKT YRGAFQNLFQ SVREVIQNPG PRHPEAASAA

         70         80         90        100        110        120
PPGASLLLL QQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ

        130        140        150        160        170        180
ETSPRQQQQQ QGEDGSPQAH RRGPTGYLVL DEEQQPSQPQ SALECHPERG CVPEPGAAVA

        190        200        210        220        230        240
ASKGLPQQLP APPDEDDSAA PSTLSLLGPT FPGLSSCSAD LKDILSEAST MQLLQQQQQE

        250        260        270        280        290        300
AVSEGSSSGR AREASGAPTS SKDNYLGGTS TISDNAKELC KAVSVSMGLG VEALEHLSPG

        310        320        330        340        350        360
EQLRGDCMYA PLLGVPPAVR PTPCAPLAEC KGSLLDDSAG KSTEDTAEYS PFKGGYTKGL


EGESLG
```

Theoretical pI/Mw: 4.94 / 39450.08

Protein purification from E. coli bacteria

The plasmids bearing the AR gene were transformed into E. coli bacteria using BL21 (DE3) competent cells for protein expression. A 25 ml bacteria culture was grown in LB Amp overnight and then diluted in 250 ml LB Amp. Cells were cultured at 37 °C for 3 hours until an OD600 of 0.5-0.7 was obtained, at which point protein expression was induced by adding 1 mM IPTG and incubation for 3 hours in the shaker. AR proteins expressed from 0QAR, 20QAR, and 50QAR pTWIN vectors and HUB protein from the PHUB (1-75) plasmid were purified from the bacterial pellet as recommended (New England BioLabs, 2006). Bacteria were suspended with 8 ml of lysis buffer B2 that consisted of 20 mM Tris pH = 7.0, 500 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, and a Complete Protease Inhibitor Cocktail pill (Roche), and sonication was used to break the cells at intensity 2.5 for 10 times for 20 second. The lysate was centrifuged for 15 mins in the cold room (1.5 ml tube) at 13,000 rpm and the clear lysate bound to 2.5 ml chitin beads for one hour in the cold room with rocking. The chitin beads were washed three times with 5 ml buffer B2 before and after the binding. The protein sizes after purification, and before cleavage of the AR + (intein + CBD) for AR 0Q, 20Q, and 50Q were predicted to be ~33.2 KDa + ~28 kDa = ~61.2 kDa, ~35.7 KDa+ ~28 KDa = ~63.7 kDa & ~39.4 KDa + ~ 28 KDa = ~67.4 kDa, respectively. However, as shown in the Figure 14 and 15, there were different in the observed sizes because in Figure 14 samples were prepared in SDS sample buffer with DTT and boiling in DTT caused

cleavage of the AR-intein-CBD fusion proteins. Also, 10% more SDS sample buffer was added to 0Q1, and 20Q9 samples as shown in Figure 15.

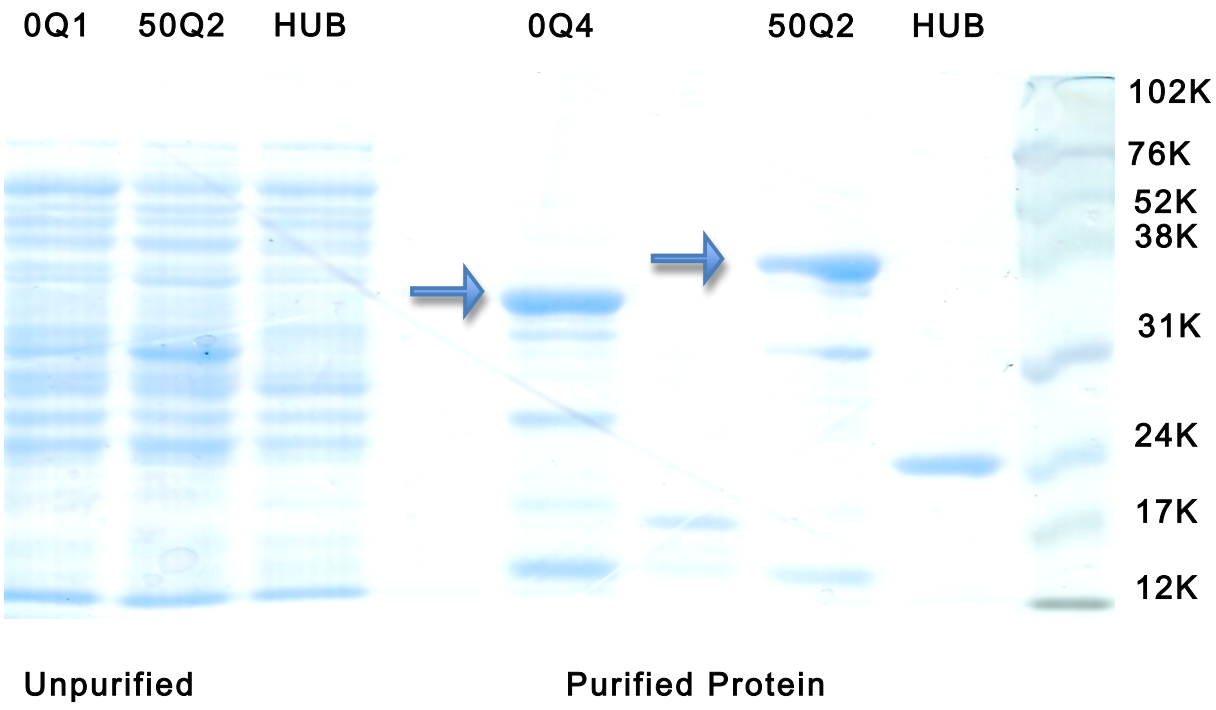

Figure 14: Expressing and purifying the cloned proteins

This figure shows the Coomassie Blue stained 8% SDS-PAGE gel of AR 0Q and 50Q and the human ubiquitin (HUB) purified using the Chitin-binding domain (CBD). The size for AR 0Q and 50Q were lower than what was expected because samples were prepared in SDS sample buffer with DTT and boiling in DTT caused cleavage of the AR-intein-CBD fusion proteins. The observed sizes of purified proteins were 37.5 KDa, and 42.8 KDa for AR 0Q4-intein-CBD, 50Q2-intein-CBD fusion proteins respectively. The Full-Range Rainbow Molecular Weight Markers (RPN800E) (GE Healthcare Life Sciences) was used.

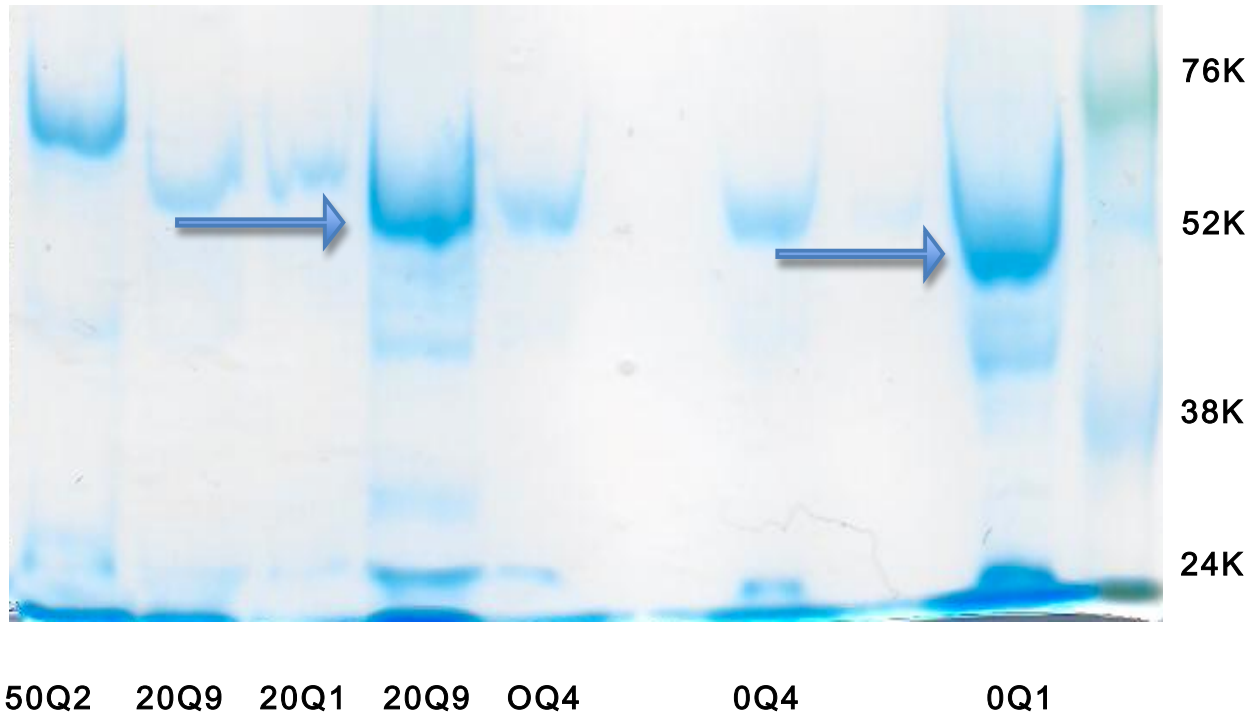

Figure 15: Expressing and purifying the cloned proteins

This figure shows the Coomassie Blue stained 8% SDS-PAGE gel of AR 0Q, 20Q, and 50Q purified using Chitin-binding domain (CBD) beads. The observed sizes of purified proteins were ~50 kDa, ~54.7 kDa, ~56.4 kDa, ~59.7 kDa, & ~70 kDa for AR 0Q1, 0Q4, 20Q9, 20Q1 and 50Q2 respectively. Note that 0Q1 and 20Q9 showed lower bands than what was expected, because perhaps samples were prepared in 10% more SDS sample buffer and different amounts of protein were added. The Full-Range Rainbow Molecular Weight Markers (RPN800E) (GE Healthcare Life Sciences) was used.

Cleavage of AR protein from chitin beads

The AR 0QAR, 20QAR, 50QAR and HUB polypeptides were cleaved from the AR + Intein + CBD fusion proteins using thiol-induced cleavage (DTT & MESNA). The cleavage efficiencies were significantly higher when MESNA was used instead of DTT for inducing cleavage. The results showing lower efficiency cleavage using DTT are not shown. The length of time allowed to bind the fusion proteins to the beads, the temperature, the number of chitin beads, and the concentration of the buffer were optimized for cleavage as suggested (Ghosh et al., 2011). Beads with fusion protein were spun down for 5 mins, the supernatant was removed and 1 ml cleavage buffer with fresh MESNA added to each sample. Cleavage buffer B4 consisted of 20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA. MESNA was added fresh for each cleavage and each cleavage was done for 24 hours with rocking at 4 °C. The AR protein sizes of 0Q, 20Q, and 50Q after cleavage were predicted to be ~33.2 kDa, ~35.7 kDa, and ~39.4 kDa, respectively (Tables 8-10), analyzed on 8% & 12% SDS-PAGE gels.

The conditions of western blots analysis were 8% and 12% SDS-PAGE gels for the separation, and after that proteins were transferred to a nitrocellulose membrane that was blocked for 1 hour at room temperature or O/N at 4 °C in Tris buffered saline (TBS) pH 7.5, 500 mM NaCl with 0.5% Tween and 5% skim milk powder. After blocking and washing, the Western blots were incubated with a 1:1000 dilution of purified Anti-HA

Tag monoclonal antibody CLH104AP (Cedarlane) as primary antibody to detect the HA peptide. Anti-AR441 (Neo Marker) was used as primary antibody to detect the AR. After washing, the blots were incubated with a 1:10,000 dilution of anti-mouse HRP-conjugated secondary antibody for 1 hour. Also, I used a monoclonal antibody, which is known as polyubiquitin-C (Covance), as a primary antibody to detect the HUB (1:10,000) and anti-mouse HRP-conjugated secondary antibody. The optimal working dilution was 1:10,000. The ECL Prime Western Blotting Detection Reagent kit (GE Healthcare Life Sciences) was used for detection of the secondary antibodies.

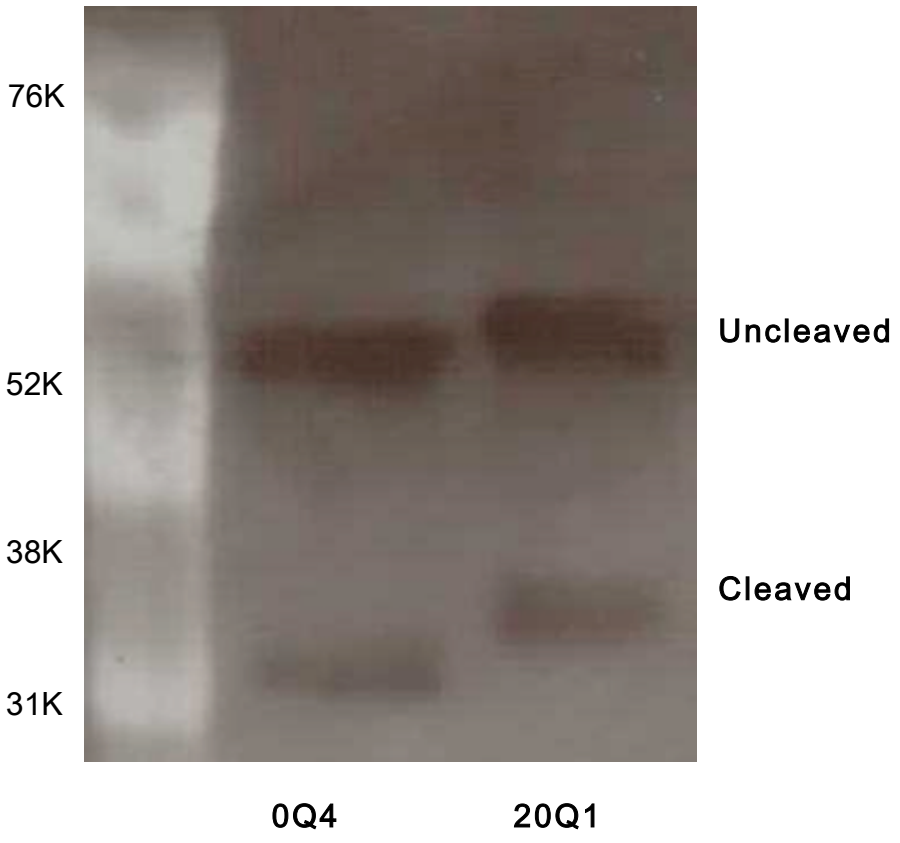

Figure 16: Cleavage of AR 0Q and 20Q using 2-mercaptoethanesulfonic acid (MESNA)

Western blot showing that the AR protein sizes for 0Q, and 20Q after cleavage were ~33 kDa, and ~37 kDa, respectively. The higher molecular weight bands represent uncleaved fusion proteins. Proteins were separated on 8% SDS-PAGE after cleavage using MESNA. An AR antibody was used as the primary antibody. In this gel, the cleavage efficiencies for 0Q4, 20Q1 were 45%, and 39%.

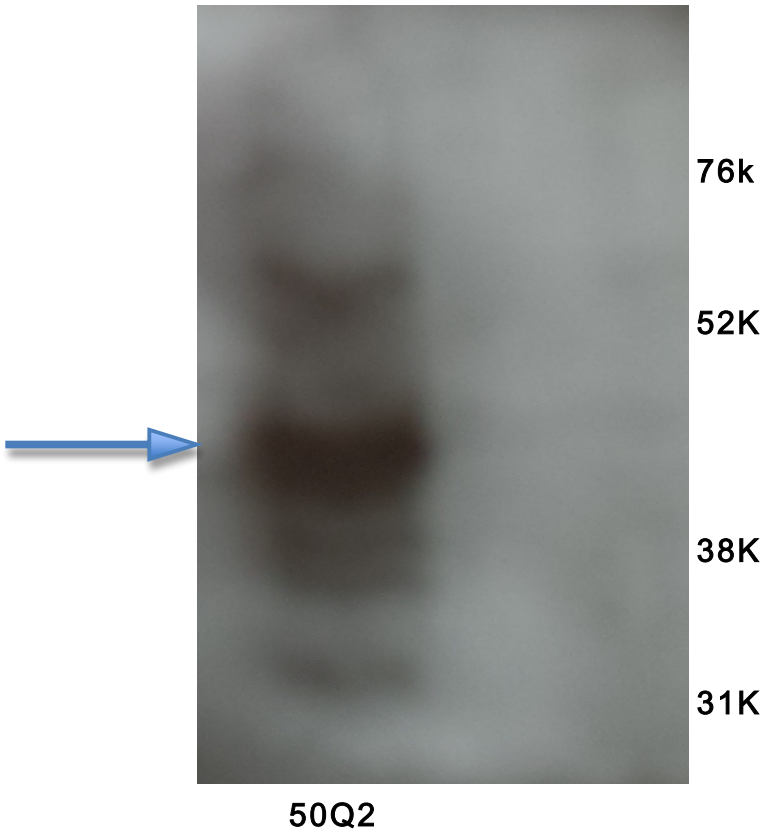

Figure 17: Cleavage of AR 50Q using 2-mercaptoethanesulfonic acid (MESNA)

Western blot showing that the AR protein size for 50Q2 after cleavage was ~43.4 kDa. Proteins were separated on 8% SDS-PAGE after cleavage using MESNA and an AR antibody was used as the primary antibody.

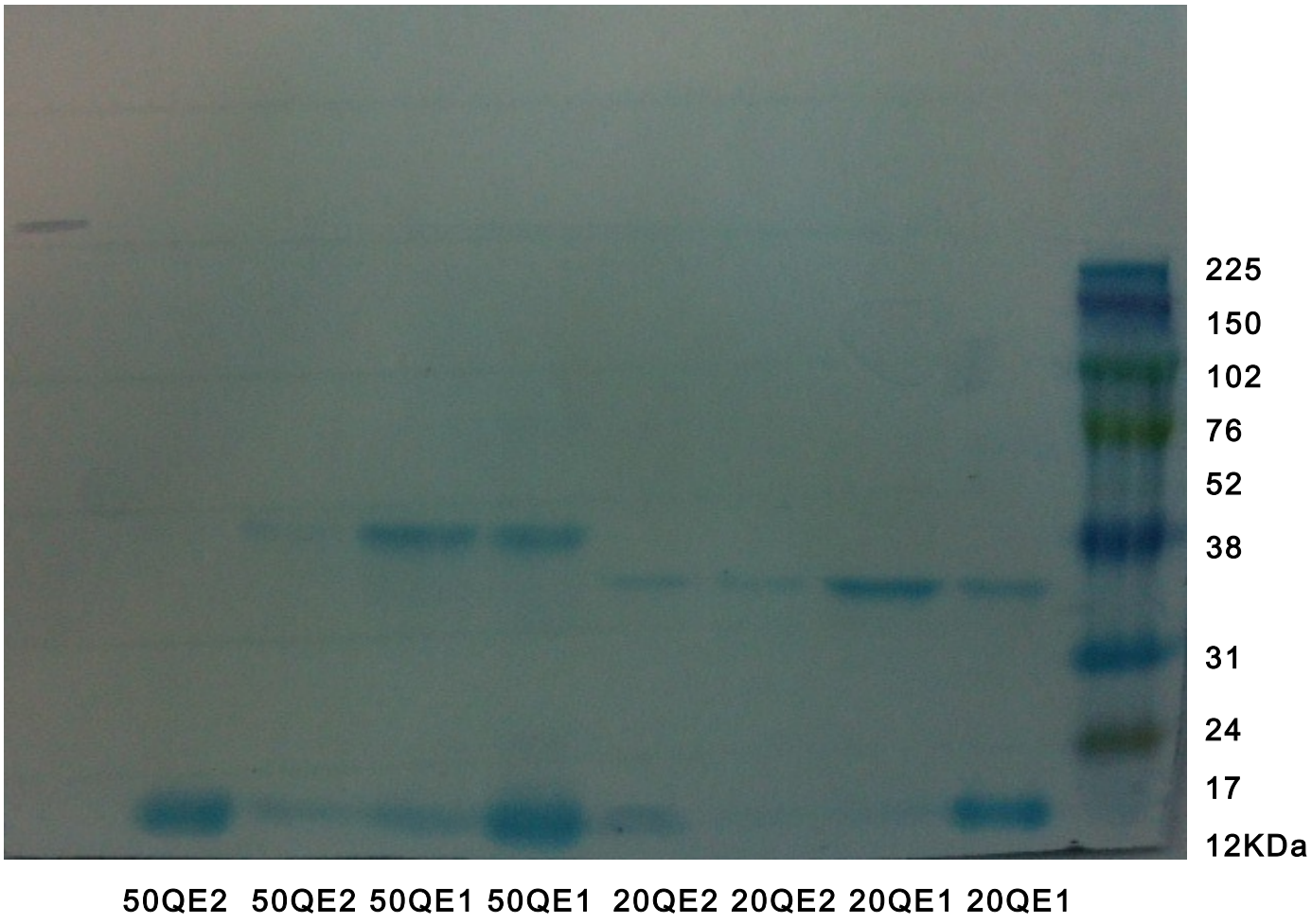

Figure 18: Cleavage of AR 20Q, and 50Q using MESNA

An 8% SDS-PAGE gel was stained with Coommassie Blue to detect the AR protein for the 20Q9 clone from the first elution (E1) and second elution (E2) with MENSA, and the 502Q clone (E1 and E2). The observed sizes after cleavage were ~32 kDa, and ~38 kDa for AR 20Q9 and AR 50Q2. Note that this result shows the same proteins as Figure 16 and 17, but they were detected using Coomassie Blue instead of western bloting. The Full-Range Rainbow Molecular Weight Markers (RPN800E) (GE Healthcare Life Sciences) was used.

Ligation method and western blots analysis

Successfully cleaved AR 0Q, 20Q, and 50Q were ligated to different synthetic peptides (HA epitope, H2B peptide) using EPL. Ligation time was 0 time, 4 hours at room temperature, and O/N at 4 °C, and the buffer concentration and the peptide concentration were optimized. Ten times ligation buffer that consisted of 100 mM Tris pH 8.85, 250 mM NaCl, and 10mM MESNA was used in the ligation and the peptide concentration used was 1-2 mM. Western blot method was used to check the ligation of the 0QAR, 20QAR, and 50QAR with HA epitope. Western blots detected that ligation successfully yielded the correct AR lengths. The efficiency of the ligation of each length AR to the HA peptide is currently being evaluated.

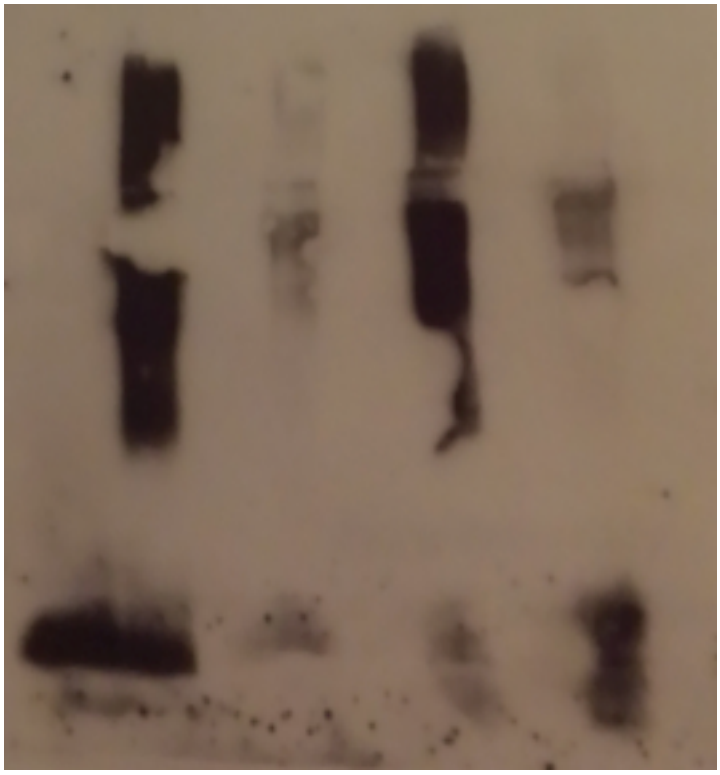

Figure 19: Ligation of 0QAR with H2B-HA peptide by EPL

This figure shows Western blot that detected the ligation of the 0QAR protein from first elution (E1) and the second elution (E2) with H2B-HA peptide in different conditions (0 time, 4 hours, and O/N). The size of the 0QAR+ H2B-HA peptide was ~38.5kDa (~36.5 kDa + ~2kDa). Proteins were separated on a 12% SDS-PAGE and detected by using anti-HA as the primary antibody and anti-mouse-HRP as the secondary antibody. High molecular weight bands around 38.6 kDa show EPL was successful. Note that lane 8 shows a better pattern than lane 6 or 4 with increasing ligation time.

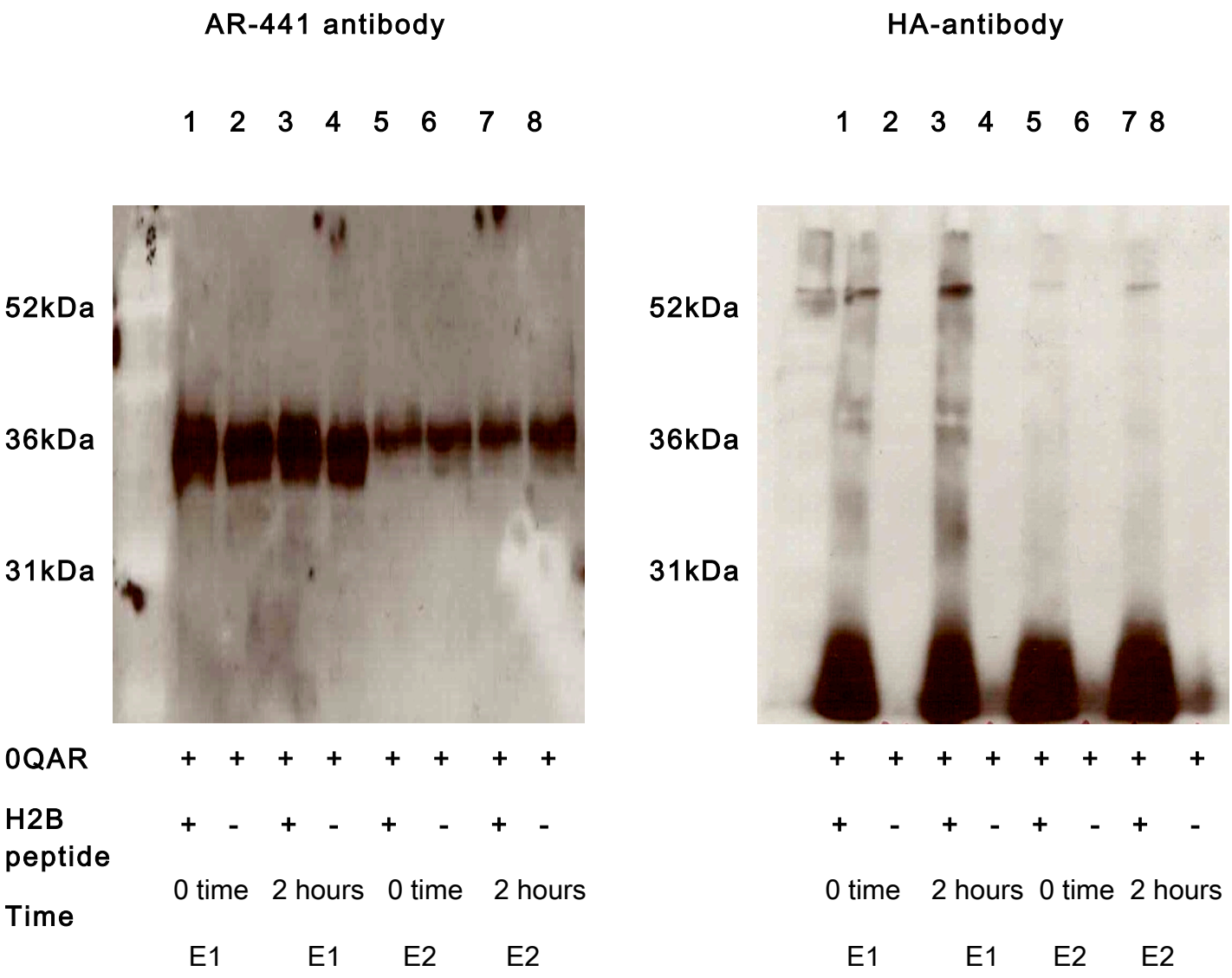

Figure 20: Ligation of 0QAR with H2B-HA peptide by EPL

This figure shows Western blot that detected the ligation of the 0QAR protein with H2B-HA peptide under different conditions (0 time and 2 hours). Proteins were separated on a 12% SDS-PAGE gel and detected by using HA-antibody or AR-441 as the primary antibody and anti-mouse-HRP as secondary antibody, on the same blot. The observed size of the 0QAR + H2B-HA peptide was ~38.5kDa (~36.5 kDa + ~2kDa).

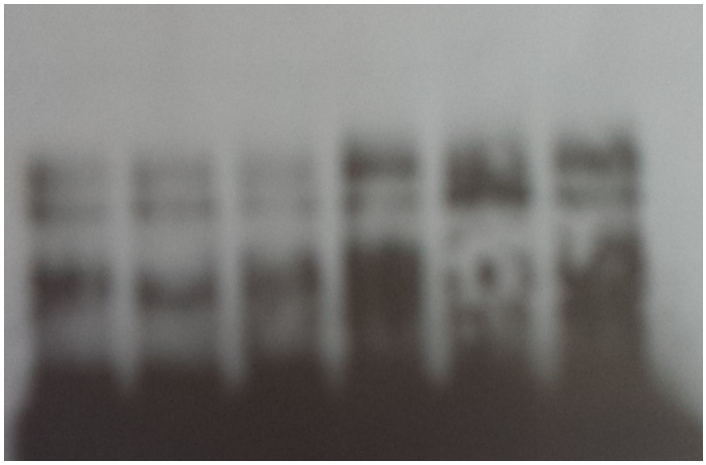

Figure 21: Ligation of 20QAR with H2B-HA peptide by EPL

This figure shows Western-blot that detected the 20QAR protein with H2B-HA peptide under different conditions (0 time, 4 hours, and O/N). The size of the 20QAR+ H2B-HA peptide was ~40.2kDa (~38.2 kDa + ~2kDa). Proteins were separated on a 12% SDS-PAGE gel and detected using AR-441 as the primary antibody and anti-mouse-HRP as secondary antibody. Note that there may be protein degradation in the sample producing lower molecular weight bands.

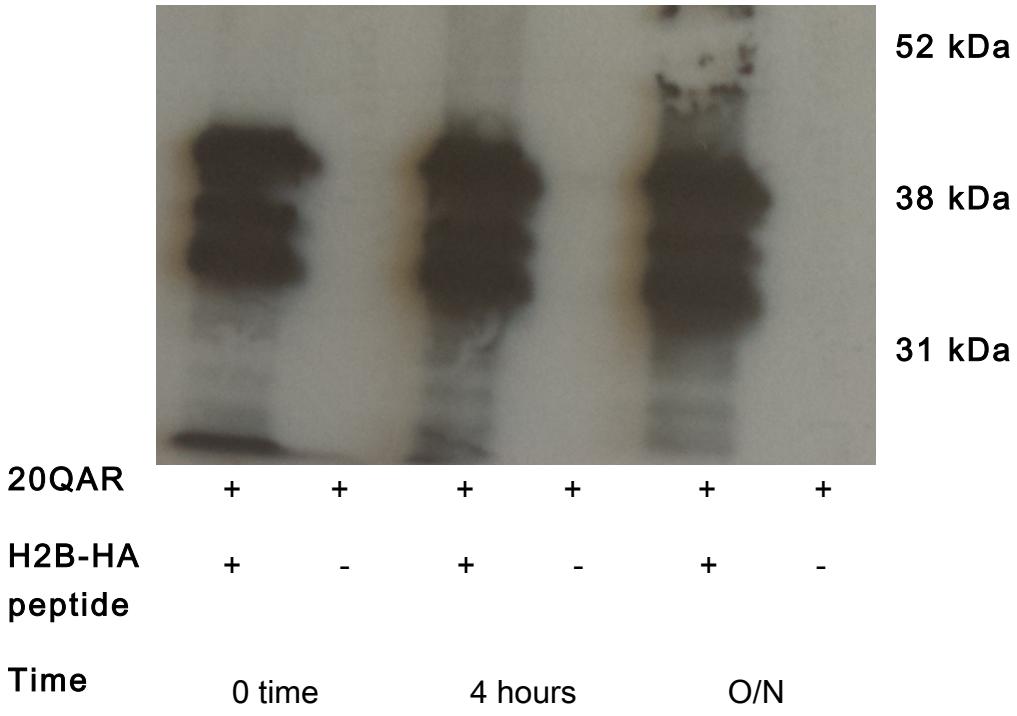

Figure 22: Ligation of 20QAR with H2B-HA peptide by EPL

This figure shows the Western blot that detected the ligation of the 20QAR protein with H2B-HA peptide under different conditions (0 time, 4 hours, and O/N). The size of the 20QAR+ H2B-HA peptide was ~40.2kDa (~38.2 kDa + ~2kDa). Proteins were separated on a 12% SDS-PAGE gel and detected by using HA-antibody as the primary antibody and anti-mouse-HRP secondary antibody. High molecular weight bands around 40.2 kDa show EPL was successful. Note that this protein preparation was the same one as in Figure 22.

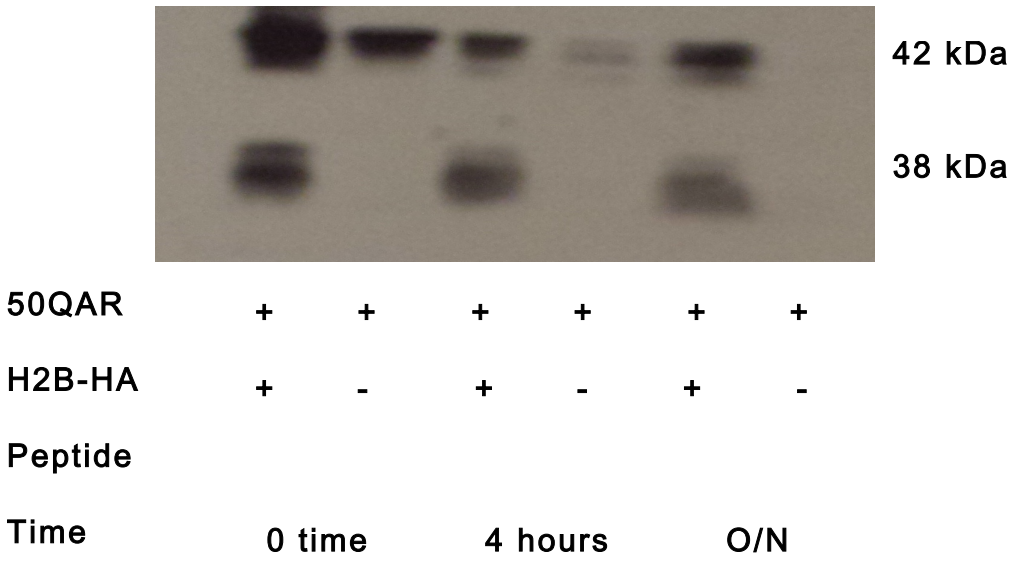

Figure 23: Ligation of 50QAR with H2B-HA peptide by EPL

This figure shows Western blot that detected ligation of the 50QAR protein with the H2B-HA peptide under different conditions (0 time, 4 hours, and O/N). The size of the 50QAR+ H2B-HA peptide was ~44.8 kDa (~42.8 kDa + ~2 kDa). Proteins were separated on a 12% SDS-PAGE and detected by using AR-441 as the primary antibody and anti-mouse-HRP secondary antibody. Note that it is possible that the last band was not transfered.

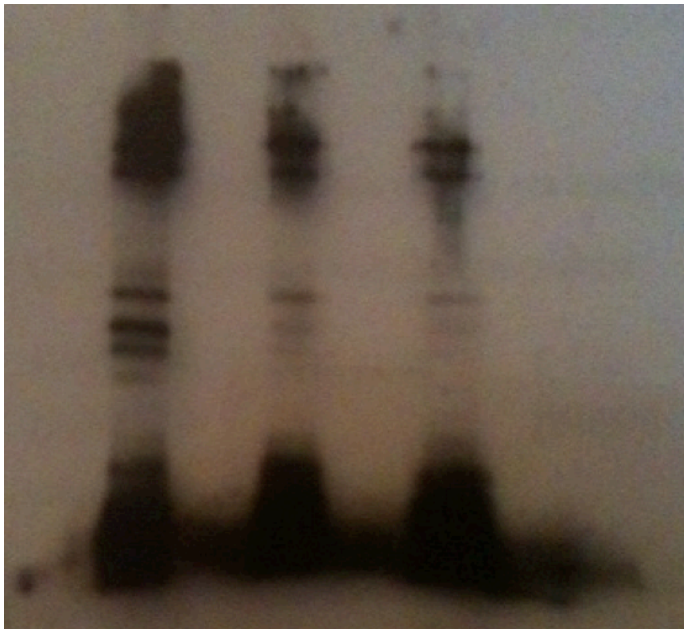

Figure 24: Ligation of 50QAR with H2B-HA peptide by EPL

This figure shows the Western blot that detected the ligation of the 50QAR protein with the H2B-HA peptide under different conditions (0 time, 4 hours, and O/N). The size of the 50QAR + H2B-HA peptide was ~44.8 kDa (~42.8 kDa + ~2 kDa). Proteins were separated on a 12% SDS-PAGE and detected by using HA-antibody as the primary antibody and anti-mouse-HRP secondary antibody. Note that the >52 kDa bands may be related to the aggregation of the 50QAR.

Tissue Culture and Proteasome Activity Assay

Proteasomes were successfully purified from Human Embryonic Kidney (HEK 293) cells that were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 5% serum, 10 U/ml penicillin and 10 μL/ml streptomycin using the method developed in our laboratory (Scanlon et al., 2009). Glutathione S-Transferase (GST) fused to ubiquitin-like domain (UBL) was used to isolate proteasomes from HEK293 cell lysates. To prepare the GST-UBL, I grew the BL21 bacteria containing the plasmid in 25 ml LB Broth, 100 μg/ml ampicillin and incubated overnight in the shaker. The overnight culture was centrifuged two times for ten mins at 13,000 rpm and then the pellets were suspended with 2 ml LB. After that I added the pellets to 250 ml LB with 100 μg/ml ampicillin and cells were cultured at 37 °C for 3 hours until an OD600 of 0.5-0.7 was obtained, at which point protein expression was induced by adding 1 mM IPTG and incubation for 3 hours in the shaker. The lysate was centrifuged for ten mins at 13,000 rpm and pellets were suspended with 10 ml buffer that consisted of 20 mM Tris, 500 mM NaCl, and 1 mM EDTA. Next, I added 5 ml buffer that consisted of 20 mM Tris, 500 mM NaCl, and 1 mM EDTA with rocking for 5 mins at 4°C and then centrifuged for 10 mins at high speed. The GST-UBL then was rocked for one hour at 4°C with 150 μL glutathione-agarose beads and centrifuged for 10 mins at 13,000 rpm. GST-UBL on beads was washed by using 3 ml buffer that consisted of 20 mM Tris, 1 mM EDTA, 500 mM NaCl, and 10% glycerol. The Z-Leu-Leu-Leu-aldehyde (MG132) was used as a proteasome inhibitor to

compare proteasome activity with and without MG132 (Scanlon et al., 2009). Purified fractions were tested for proteasomal activity by the fluorigenic substrate assay (FSA) (Scanlon et al., 2009). Samples were incubated in 10 μM Suc-LLVY-AMC + 5 μM MG132 with buffer B that consisted of 25 mM HEPES-buffered saline pH 7.5, 50 mM NaCl, and 2 mM ATP for one hour at 37 °C and then 800 ml 1% SDS added. The free AMC emission fluorescence was measured with a LS55 Luminescence spectophotmeter. The fluorometer settings were excitation at 380 nm, and emission at 460 nm. An increase in the relative fluoresce was observed as the amount of GST-Ubl-purified lysate in the reaction mix was increased from 5μl to 25μl, and this activity inhibited by the addition of MG132 (result not shown).

9. Future Directions

The purified AR 0Q, 20Q and 50Q proteins will be added to active proteasomes to compare the rates at which poly Gln expanded AR and wild-type AR are degraded. A ubiquitinated peptide will be attached to the AR with each of the variable polyGln tracts (AR0Q, AR20Q, AR50Q) by EPL technology. The proteasome degradation of the ubiquitinated expanded polyGln AR and the normal AR will also be compared. Finally, the capacity of the ubiquitinated expanded AR to prevent normal proteasome function will be evaluated.

10. Conclusion

Previous work suggests that 0QAR and 20QAR can be degraded via the UPS. However, 50QAR appeared to inhibit the degradation of particular proteins by UPS (Mandrusiak et al., 2003). It is thus hypothesized that these SBMA-inducing mutant ARs can accumulate in the cell and cause disease by evading UPS degradation.

In this thesis the main goal was to develop a system for ubiquitinating AR by using an expressed protein ligation (EPL) system. In order to attain this goal, several tools have successfully been developed. As I have shown, PCR-amplified fragments of AR (0Q, 20Q, 50Q) were cloned into the pTWIN2 vector and contructs were analyzed using restriction enzyme digests and gel electrophoresis. Then, 0QAR, 20QAR, 50QAR and human ubiquitin (HUB) fusion proteins were purified using chitin beads and cleaved using 2-mercapto-ethanesulfonic acid (MESNA). Finally, the 0QAR, 20QAR, 50QAR peptides were ligated with an H2B peptide using EPL. Also, the activity of purified proteasomes was tested in this thesis. With all these tools, the proposed future work can be carried out which aims to ligate a human ubiquitin (HUB) to a peptide by the EPL method, and then ligating the ubiquitinated peptide to the three different sizes of AR. The ubiquitinated AR proteins will then be added to active proteasomes to explore the involvement of the UPS in the pathogenesis of spinal and bulbar muscular atrophy.

Bibliography

Abdullah, A., Trifiro, M. A., Panet-Raymond, V., Alvarado, C., de Tourreil, S., Frankel, D., Pinsky, L. (1998). Spinobulbar muscular atrophy: polyglutamine-expanded androgen receptor is proteolytically resistant in vitro and processed abnormally in transfected cells. *Hum Mol Genet, 7*(3), 379-384.

Alves, S., Regulier, E., Nascimento-Ferreira, I., Hassig, R., Dufour, N., Koeppen, A., Carvalho, A. L., Simoes, S., de Lima, M. C., Brouillet, E., Gould, V. C., Deglon, N., & de Almeida, L. P. (2008). Striatal and nigral pathology in a lentiviral rat model of Machado-Joseph disease. *Hum Mol Genet, 17*(14), 2071-2083. doi: 10.1093/hmg/ddn106

Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., & Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature, 431*(7010), 805-810. doi: 10.1038/nature02998

Avila, D. M., Allman, D. R., Gallo, J. M., & McPhaul, M. J. (2003). Androgen receptors containing expanded polyglutamine tracts exhibit progressive toxicity when stably expressed in the neuroblastoma cell line, SH-SY 5Y. *Experimental Biology and Medicine, 228*(8), 982-990.

Baarends, W. M., Themmen, A. P., Blok, L. J., Mackenbach, P., Brinkmann, A. O., Meijer, D., Faber, P. W., Trapman, J., & Grootegoed, J. A. (1990). The rat androgen receptor gene promoter. *Mol Cell Endocrinol, 74*(1), 75-84.

Baniahmad, A. (2005). Nuclear hormone receptor co-repressors. *J Steroid Biochem Mol Biol, 93*(2-5), 89-97. doi: 10.1016/j.jsbmb.2004.12.012

Baniahmad, A., Ha, I., Reinberg, D., Tsai, S., Tsai, M. J., & O'Malley, B. W. (1993). Interaction of human thyroid hormone receptor beta with transcription factor TFIIB may mediate target gene derepression and activation by thyroid hormone. *Proc Natl Acad Sci U S A, 90*(19), 8832-8836.

Baumann, H., Paulsen, K., Kovacs, H., Berglund, H., Wright, A. P., Gustafsson, J. A., & Hard, T. (1993). Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry, 32*(49), 13463-13471.

Beitel, L. K., Scanlon, T., Gottlieb, B., & Trifiro, M. A. (2005). Progress in Spinobulbar muscular atrophy research: insights into neuronal dysfunction caused by the polyglutamine-expanded androgen receptor. *Neurotox Res, 7*(3), 219-230.

Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., & Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature, 375*(6530), 377-382. doi: 10.1038/375377a0

Cenni, B., & Picard, D. (1999). Ligand-independent activation of steroid receptors: new roles for old players. *Trends Endocrinol Metab, 10*(2), 41-46.

Chang, C. S., Kokontis, J., & Liao, S. T. (1988). Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science, 240*(4850), 324-326.

Chen, S., Berthelier, V., Hamilton, J. B., O'Nuallain, B., & Wetzel, R. (2002). Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry, 41*(23), 7391-7399.

Chen, S., Berthelier, V., Yang, W., & Wetzel, R. (2001). Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol, 311*(1), 173-182. doi: 10.1006/jmbi.2001.4850

Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R., & Zoghbi, H. Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet, 10*(3), 344-350. doi: 10.1038/ng0795-344

Cummings, C. J., Mancini, M. A., Antalffy, B., DeFranco, D. B., Orr, H. T., & Zoghbi, H. Y. (1998). Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet, 19*(2), 148-154. doi: 10.1038/502

Cummings, C. J., Reinstein, E., Sun, Y., Antalffy, B., Jiang, Y., Ciechanover, A., Orr, H. T., Beaudet, A. L., & Zoghbi, H. Y. (1999). Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. *Neuron, 24*(4), 879-892.

Curtis, R., Tonra, J. R., Stark, J. L., Adryan, K. M., Park, J. S., Cliffer, K. D., Lindsay, R. M., & DiStefano, P. S. (1998). Neuronal injury increases retrograde axonal transport of the neurotrophins to spinal sensory neurons and motor neurons via multiple receptor mechanisms. *Mol Cell Neurosci, 12*(3), 105-118. doi: 10.1006/mcne.1998.0704

Dahlman-Wright, K., Baumann, H., McEwan, I. J., Almlof, T., Wright, A. P., Gustafsson, J. A., & Hard, T. (1995). Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. *Proc Natl Acad Sci U S A, 92*(5), 1699-1703.

Darimont, B. D., Wagner, R. L., Apriletti, J. W., Stallcup, M. R., Kushner, P. J., Baxter, J. D., Fletterick, R. J., & Yamamoto, K. R. (1998). Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev, 12*(21), 3343-3356.

de Pril, R., Fischer, D. F., Maat-Schieman, M. L., Hobo, B., de Vos, R. A., Brunt, E. R., Hol, E. M., Ross, R. A., & van Leeuwen, F. W. (2004). Accumulation of aberrant ubiquitin induces aggregate formation and cell death in polyglutamine diseases. *Human Molecular Genetics, 13*(16), 1803-1813. doi: 10.1093/hmg/ddh188

DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P., & Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science, 277*(5334), 1990-1993.

Ellerby, L. M., Hackam, A. S., Propp, S. S., Ellerby, H. M., Rabizadeh, S., Cashman, N. R., Trifiro, M. A., Pinsky, L., Wellington, C. L., Salvesen, G. S., Hayden, M. R., & Bredesen, D. E. (1999). Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *Journal of Neurochemistry, 72*(1), 185-195.

Fargo, K. N., Galbiati, M., Foecking, E. M., Poletti, A., & Jones, K. J. (2008). Androgen regulation of axon growth and neurite extension in motoneurons. *Hormones and Behavior, 53*(5), 716-728. doi: 10.1016/j.yhbeh.2008.01.014

Foradori, C. D., Weiser, M. J., & Handa, R. J. (2008). Non-genomic actions of androgens. *Front Neuroendocrinol, 29*(2), 169-181. doi: 10.1016/j.yfrne.2007.10.005

Galani, A., Kitsiou-Tzeli, S., Sofokleous, C., Kanavakis, E., & Kalpini-Mavrou, A. (2008). Androgen insensitivity syndrome: clinical features and molecular defects. *Hormones (Athens), 7*(3), 217-229.

Ghosh, I., Considine, N., Maunus, E., Sun, L., Zhang, A., Buswell, J., Evans, T. C., Jr., & Xu, M. Q. (2011). Site-specific protein labeling by intein-mediated protein ligation. *Methods Mol Biol, 705*, 87-107. doi: 10.1007/978-1-61737-967-3_6

Giovannucci, E., Stampfer, M. J., Krithivas, K., Brown, M., Dahl, D., Brufsky, A., Talcott, J., Hennekens, C. H., & Kantoff, P. W. (1997). The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci U S A, 94*(7), 3320-3323.

Gottlieb, B., Beitel, L. K., Wu, J. H., & Trifiro, M. (2004). The androgen receptor gene mutations database (ARDB): 2004 update. *Hum Mutat, 23*(6), 527-533. doi: 10.1002/humu.20044

Gottlieb, B., Lumbroso, R., Beitel, L. K., & Trifiro, M. A. (2005). Molecular pathology of the androgen receptor in male (in)fertility. *Reprod Biomed Online, 10*(1), 42-48.

Grossmann, M. E., Lindzey, J., Blok, L., Perry, J. E., Kumar, M. V., & Tindall, D. J. (1994). The mouse androgen receptor gene contains a second functional promoter which is regulated by dihydrotestosterone. *Biochemistry, 33*(48), 14594-14600.

Guarente, L. (1995). Transcriptional coactivators in yeast and beyond. *Trends Biochem Sci, 20*(12), 517-521.

Gutekunst, C. A., Li, S. H., Yi, H., Mulroy, J. S., Kuemmerle, S., Jones, R., Jones, R., Rye, D., Ferrante, R. J., Hersch, S. M., & Li, X. J. (1999). Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J Neurosci, 19*(7), 2522-2534.

Hackam, A. S., Wellington, C. L., & Hayden, M. R. (1998). The fatal attraction of polyglutamine-containing proteins. *Clin Genet, 53*(4), 233-242.

Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J. A., & Kaptein, R. (1990).

Solution structure of the glucocorticoid receptor DNA-binding domain. *Science, 249*(4965), 157-160.

Helmlinger, D., Hardy, S., Abou-Sleymane, G., Eberlin, A., Bowman, A. B., Gansmuller, A., Picaud, S., Zoghbi, H. Y., Trottier, Y., Tora, L., & Devys, D. (2006). Glutamine-expanded ataxin-7 alters TFTC/STAGA recruitment and chromatin structure leading to photoreceptor dysfunction. *PLoS Biol, 4*(3), e67. doi: 10.1371/journal.pbio.0040067

Holmberg, M., Duyckaerts, C., Durr, A., Cancel, G., Gourfinkel-An, I., Damier, P., Faucheux, B., Trottier, Y., Hirsch, E. C., Agid, Y., & Brice, A. (1998). Spinocerebellar ataxia type 7 (SCA7): a neurodegenerative disorder with neuronal intranuclear inclusions. *Hum Mol Genet, 7*(5), 913-918.

Hudson, L. G., Santon, J. B., Glass, C. K., & Gill, G. N. (1990). Ligand-activated thyroid hormone and retinoic acid receptors inhibit growth factor receptor promoter expression. *Cell, 62*(6), 1165-1175.

Hughes, R. E., Lo, R. S., Davis, C., Strand, A. D., Neal, C. L., Olson, J. M., & Fields, S. (2001). Altered transcription in yeast expressing expanded polyglutamine. *Proc Natl Acad Sci U S A, 98*(23), 13201-13206. doi: 10.1073/pnas.191498198

Ikeda, H., Yamaguchi, M., Sugai, S., Aze, Y., Narumiya, S., & Kakizuka, A. (1996). Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. *Nat Genet, 13*(2), 196-202. doi: 10.1038/ng0696-196

Jacobi, H., Reetz, K., du Montcel, S. T., Bauer, P., Mariotti, C., Nanetti, L., Rakowicz, M., Sulek, A., Durr, A., Charles, P., Filla, A., Antenora, A., Schols., L., Schicks, J., Infante, J., Kang, J. S., Timmann, D., Fabio, R. D., Masciullo, M., Baliko, L., Bela, M., Boesch, S., Burk, K., Peltz, A., Schulz, J. B., Dufaure-Gare, I., & Klockgether, T. (2013). Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study: analysis of baseline data. *Lancet Neurol,* 12(7), 650-658. doi: 10.1016/S1474-4422(13)70104-2

Katsuno, M., Adachi, H., Waza, M., Banno, H., Suzuki, K., Tanaka, F., Doyu, M., & Sobue, G. (2006). Pathogenesis, animal models and therapeutics in spinal and bulbar muscular atrophy (SBMA). *Exp Neurol, 200*(1), 8-18. doi: 10.1016/j.expneurol.2006.01.021

Katsuno, M., Tanaka, F., Adachi, H., Banno, H., Suzuki, K., Watanabe, H., & Sobue, G. (2012). Pathogenesis and therapy of spinal and bulbar muscular atrophy (SBMA). *Prog Neurobiol, 99*(3), 246-256. doi: 10.1016/j.pneurobio.2012.05.007

Kemp, M. Q., Poort, J. L., Baqri, R. M., Lieberman, A. P., Breedlove, S. M., Miller, K. E., & Jordan, C. L. (2011). Impaired motoneuronal retrograde transport in two models of SBMA implicates two sites of androgen action. *Human Molecular Genetics, 20*(22), 4475-4490. doi: 10.1093/hmg/ddr380

Kim, M., Lee, H. S., LaForet, G., McIntyre, C., Martin, E. J., Chang, P., Kim, T. W., Williams, M., Reddy, P. H., Tagle, D., Boyce, F. M., Won, L., Heller, A., Aronin, N., & DiFiglia, M. (1999). Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J Neurosci, 19*(3), 964-973.

Klevit, R. E., Herriott, J. R., & Horvath, S. J. (1990). Solution structure of a zinc finger domain of yeast ADR1. *Proteins, 7*(3), 215-226. doi: 10.1002/prot.340070303

Kopito, R. R. (2000). Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol, 10*(12), 524-530.

La Spada, A. R., Roling, D. B., Harding, A. E., Warner, C. L., Spiegel, R., Hausmanowa-Petrusewicz, I., Yee, W. C., & Fischbeck, K. H. (1992). Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nat Genet, 2*(4), 301-304. doi: 10.1038/ng1292-301

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature, 352*(6330), 77-79. doi: 10.1038/352077a0

Lastres-Becker, I., Rub, U., & Auburger, G. (2008). Spinocerebellar ataxia 2 (SCA2). *Cerebellum, 7*(2), 115-124. doi: 10.1007/s12311-008-0019-y

Lee, M. S., Kliewer, S. A., Provencal, J., Wright, P. E., & Evans, R. M. (1993). Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science, 260*(5111), 1117-1121.

Li, M., Miwa, S., Kobayashi, Y., Merry, D. E., Yamamoto, M., Tanaka, F., Doyu, M., Hashizume, Y., Fischbeck, K. H., & Sobue, G. (1998). Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann Neurol, 44*(2), 249-254. doi: 10.1002/ana.410440216

Lonard, D. M., Lanz, R. B., & O'Malley, B. W. (2007). Nuclear receptor coregulators and human disease. *Endocr Rev, 28*(5), 575-587. doi: 10.1210/er.2007-0012

Lonard, D. M., & O'Malley, B. W. (2012). Nuclear receptor coregulators: modulators of pathology and therapeutic targets. *Nat Rev Endocrinol, 8*(10), 598-604. doi: 10.1038/nrendo.2012.100

Losel, R. M., Falkenstein, E., Feuring, M., Schultz, A., Tillmann, H. C., Rossol-Haseroth, K., & Wehling, M. (2003). Nongenomic steroid action: controversies, questions, and answers. *Physiol Rev, 83*(3), 965-1016. doi: 10.1152/physrev.00003.2003

Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S., & Wilson, E. M. (1988). Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science, 240*(4850), 327-330.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature, 352*(6335), 497-505. doi: 10.1038/352497a0

Malik, B., Nirmalananthan, N., Bilsland, L. G., La Spada, A. R., Hanna, M. G., Schiavo, G., Gallo, J. M., & Greensmith, L. (2011). Absence of disturbed axonal transport in spinal and bulbar muscular atrophy. *Hum Mol Genet, 20*(9), 1776-1786. doi: 10.1093/hmg/ddr061

Mandrusiak, L. M., Beitel, L. K., Wang, X., Scanlon, T. C., Chevalier-Larsen, E., Merry, D. E., & Trifiro, M. A. (2003). Transglutaminase potentiates ligand-dependent proteasome dysfunction induced by polyglutamine-expanded androgen receptor. *Hum Mol Genet, 12*(13), 1497-1506.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S. W., & Bates, G. P. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell, 87*(3), 493-506.

Martin-Aparicio, E., Yamamoto, A., Hernandez, F., Hen, R., Avila, J., & Lucas, J. J. (2001). Proteasomal-dependent aggregate reversal and absence of cell death in a conditional mouse model of Huntington's disease. *J Neurosci, 21*(22), 8772-8781.

Masino, L., Kelly, G., Leonard, K., Trottier, Y., & Pastore, A. (2002). Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett, 513*(2-3), 267-272.

Matsumoto, A. (1997). Hormonally induced neuronal plasticity in the adult motoneurons. *Brain Res Bull, 44*(4), 539-547.

McCampbell, A., Taylor, J. P., Taye, A. A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., & Fischbeck, K. H. (2000). CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet, 9*(14), 2197-2202.

McEwan, I. J. (2004). Sex, drugs and gene expression: signalling by members of the nuclear receptor superfamily. *Essays Biochem, 40*, 1-10.

McGinty, R. K., Chatterjee, C., & Muir, T. W. (2009). Semisynthesis of ubiquitylated proteins. *Methods Enzymol, 462*, 225-243. doi: 10.1016/S0076-6879(09)62011-5

Meyers, Robert A. (October 2005, Wiley-Blackwell). *Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd Edition* (Vol. 15).

Mhatre, A. N., Trifiro, M. A., Kaufman, M., Kazemi-Esfarjani, P., Figlewicz, D., Rouleau, G., & Pinsky, L. (1993). Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nat Genet, 5*(2), 184-188. doi: 10.1038/ng1093-184

Michalik, A., & Van Broeckhoven, C. (2003). Pathogenesis of polyglutamine disorders: aggregation revisited. *Human Molecular Genetics, 12 Spec No 2*, R173-186. doi: 10.1093/hmg/ddg295

Mitsui, K., Doi, H., & Nukina, N. (2006). Proteomics of polyglutamine aggregates. *Methods Enzymol, 412*, 63-76. doi: 10.1016/S0076-6879(06)12005-4

Morais Fonseca, D., Rosada, R. S., Paula, M. O., Wowk, P. F., Franco, L. H., Soares, E. G., Silva, C. L., & Deperon Bonato, V. L. (2010). Experimental tuberculosis: designing a better model to test vaccines against tuberculosis. *Tuberculosis (Edinb), 90*(2), 135-142. doi: 10.1016/j.tube.2010.01.005

Morfini, G. A., Burns, M., Binder, L. I., Kanaan, N. M., LaPointe, N., Bosco, D. A., Brown, R. H., Jr., Tiwari, A., Hayward, L., Edgar, J., Nave., K. A., Garberrn, J., Atagi, Y., Song, Y., Pigino, G., & Brady, S. T. (2009). Axonal transport defects in neurodegenerative diseases. *J Neurosci, 29*(41), 12776-12786. doi: 10.1523/JNEUROSCI.3463-09.2009

Musova, Z., Sedlacek, Z., Mazanec, R., Klempir, J., Roth, J., Plevova, P., Vyhnalek, M., Kopeckova, M., Apltova, L., Krepelova, A., & Zumrova, A. (2013). Spinocerebellar ataxias type 8, 12, and 17 and dentatorubro-pallidoluysian atrophy in Czech ataxic patients. *Cerebellum, 12*(2), 155-161. doi: 10.1007/s12311-012-0403-5

Muir, T. W. (2003). Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem, 72*, 249-289. doi: 10.1146/annurev.biochem.72.121801.161900

Myers, R. H. (2004). Huntington's disease genetics. *NeuroRx, 1*(2), 255-262. doi: 10.1602/neurorx.1.2.255

New England BioLabs. (2006). IMPACT-TWIN Purification, Ligation and Cyclization of Recombinant Proteins Using Self-Cleavable Affinity Tags 1-60.

Nucifora, F. C., Jr., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., Dawson, T. M., & Ross, C. A. (2001). Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science, 291*(5512), 2423-2428. doi: 10.1126/science.1056784

Nuclear Receptors Nomenclature Committee. (1999). A unified nomenclature system for the nuclear receptor superfamily. *Cell, 97*(2), 161-163.

O'Hearn, E., Holmes, S. E., & Margolis, R. L. (2012). Spinocerebellar ataxia type 12. *Handb Clin Neurol, 103*, 535-547. doi: 10.1016/B978-0-444-51892-7.00034-6

Onate, S. A., Tsai, S. Y., Tsai, M. J., & O'Malley, B. W. (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science, 270*(5240), 1354-1357.

Ordway, J. M., Tallaksen-Greene, S., Gutekunst, C. A., Bernstein, E. M., Cearley, J. A., Wiener, H. W., Dure, L. S., Lindsey, R., Hersch, S. M., Jope, R. S., Albin, R. L., & Detloff, P. J. (1997). Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell, 91*(6), 753-763.

Orr, H. T. (2001). Beyond the Qs in the polyglutamine diseases. *Genes Dev, 15*(8), 925-932. doi: 10.1101/gad.888401

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P., & Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet, 4*(3), 221-226. doi: 10.1038/ng0793-221

Paliouras, M., Zaman, N., Lumbroso, R., Kapogeorgakis, L., Beitel, L. K., Wang, E., & Trifiro, M. (2011). Dynamic rewiring of the androgen receptor protein interaction network correlates with prostate cancer clinical outcomes. *Integr Biol (Camb), 3*(10), 1020-1032. doi: 10.1039/c1ib00038a

Paradas, C., Solano, F., Carrillo, F., Fernandez, C., Bautista, J., Pintado, E., & Lucas, M. (2008). Highly skewed inactivation of the wild-type X-chromosome in asymptomatic female carriers of spinal and bulbar muscular atrophy (Kennedy's disease). *Journal of Neurology, 255*(6), 853-857. doi: 10.1007/s00415-008-0766-1

Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H., & Pittman, R. N. (1997). Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron, 19*(2), 333-344.

Perez, M. K., Paulson, H. L., Pendse, S. J., Saionz, S. J., Bonini, N. M., & Pittman, R. N. (1998). Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol, 143*(6), 1457-1470.

Perlson, E., Maday, S., Fu, M. M., Moughamian, A. J., & Holzbaur, E. L. (2010). Retrograde axonal transport: pathways to cell death? *Trends Neurosci, 33*(7), 335-344. doi: 10.1016/j.tins.2010.03.006

Perutz, M. F. (1996). Glutamine repeats and inherited neurodegenerative diseases: molecular aspects. *Curr Opin Struct Biol, 6*(6), 848-858.

Perutz, M. F., Johnson, T., Suzuki, M., & Finch, J. T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A, 91*(12), 5355-5358.

Poletti, A. (1999). CAG expansion in androgen receptor gene and neuronal cell death. *Recent Research Developments in Neurochemistry, 2*, 507-515.

Pozzi, P., Bendotti, C., Simeoni, S., Piccioni, F., Guerini, V., Marron, T. U., Martini, L., & Poletti, A. (2003). Androgen 5-alpha-reductase type 2 is highly expressed and active in rat spinal cord motor neurones. *J Neuroendocrinol, 15*(9), 882-887.

Ranganathan, S., & Fischbeck, K. H. (2010). Therapeutic approaches to spinal and bulbar muscular atrophy. *Trends Pharmacol Sci, 31*(11), 523-527. doi: 10.1016/j.tips.2010.08.005

Rangone, H., Humbert, S., & Saudou, F. (2004). Huntington's disease: how does huntingtin, an anti-apoptotic protein, become toxic? *Pathol Biol (Paris), 52*(6), 338-342. doi: 10.1016/j.patbio.2003.06.004

Robinson-Rechavi, M., & Laudet, V. (2003). Bioinformatics of nuclear receptors. *Methods Enzymol, 364*, 95-118.

Roth, S. Y., Denu, J. M., & Allis, C. D. (2001). Histone acetyltransferases. *Annu Rev Biochem, 70*, 81-120. doi: 10.1146/annurev.biochem.70.1.81

Rusmini, P., Simonini, F., Crippa, V., Bolzoni, E., Onesto, E., Cagnin, M., Sau, D., Ferri, N., & Poletti, A. (2011). 17-AAG increases autophagic removal of mutant androgen receptor in spinal and bulbar muscular atrophy. *Neurobiology of Disease, 41*(1), 83-95. doi: 10.1016/j.nbd.2010.08.023

Sartor, O., Zheng, Q., & Eastham, J. A. (1999). Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. *Urology, 53*(2), 378-380.

Scanlon, T. C., Gottlieb, B., Durcan, T. M., Fon, E. A., Beitel, L. K., & Trifiro, M. A. (2009). Isolation of human proteasomes and putative proteasome-interacting proteins using a novel affinity chromatography method. *Exp Cell Res, 315*(2), 176-189. doi: S0014-4827(08)00460-6 [pii]10.1016/j.yexcr.2008.10.027

Schweikert, H. U., & Wilson, J. D. (1974). Regulation of human hair growth by steroid hormones. I. Testerone metabolism in isolated hairs. *J Clin Endocrinol Metab, 38*(5), 811-819.

Shao, J., & Diamond, M. I. (2007). Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Human Molecular Genetics, 16 Spec No. 2*, R115-123. doi: 10.1093/hmg/ddm213

Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y., Ishiguro, H., Sakoe, K., Ooshima, T., Sato, A., Ikeuchi, T., Oyake, M., Sato, T., Aoyagi, Y., Hozumi, I., Nagatsu, T., Takiyama, Y., Nishizawa, M., Goto, J., Kanazawa, I., Davidson, I., Tanese, N., Takahashi, H., & Tsuji, S. (2000). Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. *Nat Genet, 26*(1), 29-36. doi: 10.1038/79139

Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servadio, A., Zoghbi, H. Y., & Orr, H. T. (1997). Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures. *Nature, 389*(6654), 971-974. doi: 10.1038/40153

Sobue, G., Hashizume, Y., Mukai, E., Hirayama, M., Mitsuma, T., & Takahashi, A. (1989). X-linked recessive bulbospinal neuronopathy. A clinicopathological study. *Brain, 112 (Pt 1)*, 209-232.

Song, C. S., Her, S., Slomczynska, M., Choi, S. J., Jung, M. H., Roy, A. K., & Chatterjee, B. (1993). A distal activation domain is critical in the regulation of the rat androgen receptor gene promoter. *Biochem J, 294 (Pt 3)*, 779-784.

Sopher, B. L., Thomas, P. S., Jr., LaFevre-Bernt, M. A., Holm, I. E., Wilke, S. A., Ware, C. B., Jin, L. W., Libby, R. T., Ellerby, L. M., & La Spada, A. R. (2004). Androgen

receptor YAC transgenic mice recapitulate SBMA motor neuronopathy and implicate VEGF164 in the motor neuron degeneration. *Neuron, 41*(5), 687-699.

Spears, M. D., Melton, S., Mao, Q. W., Payne, D., Rakheja, D., Hatanpaa, K. J., Burns, D. K., Sequeiros, J., & Alonso, I. (2010). Ataxia and progressive encephalopathy in a 4-year-old girl. *Labmedicine, 41*(1), 5-9. doi: Doi 10.1309/Lmsg0l113u5qtwrs

Spencer, J. A., Watson, J. M., Lubahn, D. B., Joseph, D. R., French, F. S., Wilson, E. M., & Graves, J. A. (1991). The androgen receptor gene is located on a highly conserved region of the X chromosomes of marsupial and monotreme as well as eutherian mammals. *J Hered, 82*(2), 134-139.

Steffan, J. S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B. L., Kazantsev, A., Schmidt, E., Zhu, Y. Z., Greenwald, M., Kurokawa, R., Housman, D. E., Jackson, G. R., Marsh, J. L., & Thompson, L. M. (2001). Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila. *Nature, 413*(6857), 739-743. doi: 10.1038/35099568

Szebenyi, G., Morfini, G. A., Babcock, A., Gould, M., Selkoe, K., Stenoien, D. L., Young, M., Faber, P. W., MacDonald, M. E., McPhaul, M. J., & Brady, S. T. (2003). Neuropathogenic forms of huntingtin and androgen receptor inhibit fast axonal transport. *Neuron, 40*(1), 41-52.

Tanese, N., & Tjian, R. (1993). Coactivators and TAFs: a new class of eukaryotic transcription factors that connect activators to the basal machinery. *Cold Spring Harb Symp Quant Biol, 58*, 179-185.

Tanner, T., Claessens, F., & Haelens, A. (2004). The hinge region of the androgen receptor plays a role in proteasome-mediated transcriptional activation. *Ann N Y Acad Sci, 1030*, 587-592. doi: 10.1196/annals.1329.068

Taylor, J. P., Tanaka, F., Robitschek, J., Sandoval, C. M., Taye, A., Markovic-Plese, S., & Fischbeck, K. H. (2003). Aggresomes protect cells by enhancing the degradation of toxic polyglutamine-containing protein. *Hum Mol Genet, 12*(7), 749-757.

Tetzlaff, J., Tanzer, L., & Jones, K. J. (2007). Cellular localization of androgen and estrogen receptors in mouse-derived motoneuron hybrid cells and mouse facial motoneurons. *Dev Neurobiol, 67*(10), 1362-1370. doi: 10.1002/dneu.20505

Tjian, R., & Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. *Cell, 77*(1), 5-8.

Trottier, Y., Biancalana, V., & Mandel, J. L. (1994). Instability of CAG repeats in Huntington's disease: relation to parental transmission and age of onset. *J Med Genet, 31*(5), 377-382.

Tzukerman, M. T., Esty, A., Santiso-Mere, D., Danielian, P., Parker, M. G., Stein, R. B., Pike, J. W., & McDonnell, D. P. (1994). Human estrogen receptor transactivational capacity is determined by both cellular and promoter context

and mediated by two functionally distinct intramolecular regions. *Mol Endocrinol, 8*(1), 21-30.

Vila-Perello, M., & Muir, T. W. (2010). Biological applications of protein splicing. *Cell, 143*(2), 191-200. doi: 10.1016/j.cell.2010.09.031

Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H., & Wanker, E. E. (2001). Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. *Mol Biol Cell, 12*(5), 1393-1407.

Walker, F. O. (2007). Huntington's disease. *Lancet, 369*(9557), 218-228. doi: 10.1016/S0140-6736(07)60111-1

Wolf, N. I., & Koenig, M. (2013). Progressive cerebellar atrophy: hereditary ataxias and disorders with spinocerebellar degeneration. *Handb Clin Neurol, 113*, 1869-1878. doi: 10.1016/B978-0-444-59565-2.00057-5

Yamada, M., Sato, T., Tsuji, S., & Takahashi, H. (2008). CAG repeat disorder models and human neuropathology: similarities and differences. *Acta Neuropathol, 115*(1), 71-86. doi: 10.1007/s00401-007-0287-5

Yang, W., Dunlap, J. R., Andrews, R. B., & Wetzel, R. (2002). Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells. *Hum Mol Genet , 11*(23), 2905-2917.

Yu, W. H., & McGinnis, M. Y. (2001). Androgen receptors in cranial nerve motor nuclei of male and female rats. *J Neurobiol, 46*(1), 1-10.

Zhai, W., Jeong, H., Cui, L., Krainc, D., & Tjian, R. (2005). In vitro analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell, 123*(7), 1241-1253. doi: 10.1016/j.cell.2005.10.030

http://www.bostonbiochem.com/products/ubiquitin/ubiquitin-mutants.). Boston Biochem.

http://www.ebi.ac.uk/Tools/msa/clustalw2/.). ClustalW2.